A Preliminary Investigation of the Psychometric Properties of the WISC-V and KABC-II for South African Children Experiencing Barriers to Learning and Development

Adriana Catherina Vorster

702844

University of the Witwatersrand, Johannesburg

Supervised by: Professor Zaytoon Amod

A thesis submitted to the Discipline of Psychology in the School of Human and Community Development in fulfilment of the requirements for the degree of Doctor of Philosophy by thesis in the Faculty of Humanities, University of the Witwatersrand.

Johannesburg, South Africa March 30, 2025

Declaration

I, Adriana Catherina Vorster, hereby declare that the work conducted on this study is my own work. I have referenced all work utilised and cited in this report by other authors in this field of study. This thesis is being submitted for the Degree of Doctor of Philosophy at the University of the Witwatersrand, Johannesburg. It has not been submitted before for any degree or examination at any other university.

AVorster

Adriana Catherina Vorster

Signed on this 30th day of March 2025, at Horsham, West Sussex, United Kingdom.

Research Contribution

In collaboration with former colleagues at the University of the Witwatersrand, some results from the current study contributed to a book chapter in a forthcoming text on *Psychological Assessment in South Africa*. These results include findings on the reliability of the WISC-V and KABC-II subtests, as well as the outcome of confirmatory factor analyses conducted.

Acknowledgements

A project such as this can only be completed with the support of many people. I would like to express my deepest gratitude to:

- My supervisor, Prof Zaytoon Amod. You have been both a colleague and a mentor and remain a dear friend. I truly appreciate your valuable insight and guidance, not only in completing this study, but also in my development as an Educational Psychologist.
- The participants of this study and their three schools. This study would not have been possible without your contribution and willingness to participate.
- The professionals who assisted me at various stages with data collection and test administration.
- Dr Petra Gaylard, who supported me in analysing my data.
- My family – Liam, Alban, Carla, Louis, Paul, Nathan, Kyle, and Sanet. I could never have done this without the love, laughter and support of my family. You are my joy, and I love you.

Dedication

This study is dedicated to **my parents**.

Without your love, continuous encouragement and support, I would never have reached this stage of my academic journey. I miss you every day and love you always. You raised me on a foundation of love and faith and taught me that, “With man this is impossible, but with God all things are possible.”

(Matthew 19:26)

Thank you for always loving your children and grandchildren the most.

And to my son, **Liam**.

You are the most beautiful soul. The world is your oyster, but always remain as kind as you are.

I love you 3000 x infinity + 1.

Table of Contents

List of Tables

List of Figures

Abbreviations and Acronyms

2PLM:	Two-Parameter Logistic Model
ADHD:	Attention-Deficit/Hyperactivity Disorder
AERA:	American Educational Research Association
AIDS:	Acquired Immune Deficiency Syndrome
ANCOVA:	Analysis of Covariance
APA:	American Psychiatric Association
APIL:	(Conceptual) Ability, Processing of Information, and Learning Battery
CAPS:	South African Curriculum Assessment Policy Statements
CFA:	Confirmatory Factor Analysis
CFI:	Comparative Fit Index
CHC:	Cattell-Horn-Carroll (CHC) Theory
CNE:	Christian and National Education Schools
CRC:	Categorical Response Curves
DIF:	Differential Item Functioning
DoE:	[South African] Department of Education
DoBE:	[South African] Department of Basic Education
DSM-5-TR:	Diagnostic and Statistical Manual of Mental Disorders – Fifth Edition, Text Revision
EAL:	English Additional Language
EFA:	Exploratory Factor Analysis
EFL:	English First Language
FCI:	Fluid-Crystallised Index
FSIQ:	Full-Scale Intelligence Quotient
GDE:	Gauteng Department of Education
GFI:	Goodness of Fit Index
GLM:	General Linear Model
GRM:	Graded Response Model
HIV:	The Human Immunodeficiency Virus Infection
HPCSA:	Health Professions Council of South Africa
ICC:	Item Characteristic Curve

ICD-11:	International Statistical Classification of Diseases and Related Health Problems – Eleventh Revision
IDEIA:	Individuals with Disabilities Education Improvement Act
IFI:	Bollen Fit Index (Incremental Fit Index)
IRT:	Item Response Theory
IQ:	Intelligence Quotient
K-ABC:	Kaufman Assessment Battery for Children
KABC-II:	Kaufman Assessment Battery for Children – Second Edition
KAIT:	Kaufman Adolescent and Adult Intelligence Test
KTEA-II:	Kaufman Test of Educational Achievement – Second Edition
LDA:	Learning Disabilities Association of America
LPCAT:	Learning Potential Computerised Adaptive Test
LSM:	Living Standards Measure
MPI:	Mental Processing Index
NCESS:	National Committee on Education Support Services
NCSNET:	National Commission on Special Needs in Education and Training
NFI:	Normed Fit Index
PASS:	Planning, Attention, Simultaneous and Successive (PASS) Cognitive Processing Theory
PMT:	Panga Munthu Test
RMSEA:	Root Mean Square Error of Approximation
SIAS:	Policy on Screening, Identification, Assessment, and Support
SLD:	Specific Learning Disorder
SLD-M:	Specific Learning Disorder in Mathematics
SLD-R:	Specific Learning Disorder in Reading
SLD-RW:	Specific Learning Disorder in Reading and Written Expression
SSAIS-R:	Senior South African Individual Scale – Revised
SSARF:	South African Audience Research Foundation
TRAM:	Transfer, Automatisation and Memory Batteries
UK:	United Kingdom
UNESCO:	United Nations Educational, Scientific and Cultural Organisation
UNICEF:	The United Nations Children's Fund
US:	United States

WAIS: Wechsler Adult Intelligence Scale

WAIS-III: Wechsler Adult Intelligence Scale – Third Edition

WAIS-IV: Wechsler Adult Intelligence Scale – Fourth Edition

WAIS-IV-SA: Wechsler Adult Intelligence Scale – Fourth South African Edition

WAIS-5: Wechsler Adult Intelligence Scale – Fifth Edition

WHO: World Health Organisation

WISC: Wechsler Intelligence Scale for Children

WISC-R: Wechsler Intelligence Scale for Children – Revised

WISC-III: Wechsler Intelligence Scale for Children – Third Edition

WISC-IV: Wechsler Intelligence Scale for Children – Fourth Edition

WISC-V: Wechsler Intelligence Scale for Children – Fifth Edition (Original US Version)

WISC-V^{CDN}: Canadian Version of the Wechsler Intelligence Scale for Children – Fifth Edition

WISC-V^{FR}: Canadian Version of the Wechsler Intelligence Scale for Children – Fifth Edition

WISC-V A&NZ: Australian and New Zealand Version of the Wechsler Intelligence Scale for Children - Fifth Edition

WJ: Woodcock-Johnson III Tests of Cognitive Abilities

WPPSI: Wechsler Preschool and Primary Scale of Intelligence

WPPSI-IV: Wechsler Preschool and Primary Scale of Intelligence – Fourth Edition

ZamCAT: Zambia Child Assessment Tool

1. Introduction

This introductory chapter provides the background and rationale for the study, outlining its aims and objectives. For each of the four objectives, corresponding research questions are presented. Thereafter, important concepts used in this study, including psychometric properties, barriers to learning and development, essential academic skills, intelligence, and intelligence tests, are clarified. Moreover, as the terms *sex*/*gender* and *race*/*ethnicity* are reported in the study, their usage in South Africa will be explained. The chapter concludes by providing an overview of the thesis structure, summarising the content of each subsequent chapter.

1.1. Background and Rationale

Following the establishment of a democratic government in South Africa, Nell (1994) acknowledged the ongoing debate about the use of intelligence tests. These tests were, however, deeply embedded in various sectors of society and would remain in demand. International and local researchers have reiterated this view. For example, Nisbett et al. (2012) acknowledge the controversy that surrounds intelligence testing; however, they still consider it an important achievement in the discipline of psychology. Other international authors, such as Underwood (2007), Gilmore and Campbell (2009), and McKown (2010), not only recognise the administration of intelligence tests as an essential component of the work that school and educational psychologists perform, but also as one of the most essential capabilities of psychologists. South African authors such as Foxcroft et al. (2004) and Kumar (2017) emphasise the considerable amount of time that psychologists in the country devote to intelligence testing. Van der Merwe et al. (2024) recognise the potential of valid and reliable intelligence tests for the purpose of identifying cognitive strengths and weaknesses of children. This can aid both the development and the implementation of suitable interventions and guide appropriate support and school placement.

Despite the evident importance of intelligence testing within the field of psychology, it is essential to acknowledge the complexities and challenges associated with its use, particularly in South Africa. Psychologists must confront the reality that psychological assessments, including intelligence tests, are inextricably linked to South Africa's history of segregation (Laher, 2024; Laher & Cockcroft, 2013; Van der Merwe et al., 2022). In this regard, Foxcroft (1997) and Claassen (1997) noted that the history of psychology in South Africa has long been closely tied to the prevailing socio-political system. The practices and politics of racial segregation subsequently had a significant impact on the development and use of intelligence tests in South Africa. Prior to changes in South Africa's socio-political landscape, psychological assessments were predominantly developed, standardised, and normed for use with White populations. Further, intelligence tests standardised and normed exclusively for White populations were administered to other racial groups without due consideration for fairness, validity, and reliability, and the results were manipulated to justify segregation policies by claiming the intellectual superiority of one group over another (Foxcroft et al., 2018; Laher, 2024; Laher & Cockcroft, 2014). As the socio-political landscape began to shift during the 1980s and 1990s, these factors led to a widespread mistrust of psychological assessments, including intelligence tests, particularly regarding their ability to produce fair, valid, and reliable results within the diverse South African context (Foxcroft et al., 2004; Laher, 2024; Laher & Cockcroft, 2013, 2014).

As awareness of bias, equivalence, and fairness in psychological testing grew, researchers increasingly recognised the critical need to develop assessment measures that would be appropriate and applicable across South Africa's diverse population (Laher & Cockcroft, 2014). This resulted in the transformation and growth of the discipline of psychology in South Africa, including the development and use of psychological assessments (Cooper, 2014; Laher, 2024). Although now considered outdated, one of the first intelligence tests to be standardised for White, Indian, and Coloured South African children, with additional norms for children from under-resourced backgrounds, was the Senior South

African Individual Scale – Revised (SSAIS-R) (Van Eeden, 1991). Since then, the field of cognitive and intelligence testing saw the publication of several assessments of learning potential, including the Learning Potential Computerised Adaptive Test (LPCAT) (De Beer, 2013), the (Conceptual) Ability, Processing of Information, and Learning Battery (APIL), and the Transfer, Automatisation and Memory Batteries (TRAM) (Taylor, 2013). In 2014, JvR Psychometrics adapted, normed, and published the Wechsler Adult Intelligence Scale – Fourth South African Edition (WAIS-IV-SA) (Wechsler, 2014a). In 2023, a book entitled *Cross-Cultural Cognitive Test Norms: An Advanced Collation from Africa*, which contains a collection of normative data for commonly used cognitive assessment measures in Africa, including intelligence tests such as the Wechsler Intelligence Scale for Children – Fifth Edition (WISC-V) and the Kaufman Assessment Battery for Children – Second Edition (KABC-II), was published by two South African authors (Shuttleworth-Edwards & Truter, 2023). However, the development of fair, valid, and reliable intelligence tests for the diverse South African context remains a challenge, and there is a shortage of culturally applicable intelligence tests (Foxcroft et al., 2004; Laher, 2024; Mitchell et al., 2018; Van der Merwe et al., 2024).

With limitations in the availability of locally developed intelligence tests, referred to as emic assessments (Laher, 2024) and concerns about the psychometric properties and outdated norms of existing emic assessments such as the SSAIS-R (Abdool Gafoor et al., 2021; Cockcroft, 2013; Te Nijenhuis et al., 2011; Van Eeden & De Beer, 2018), researchers and practitioners often use internationally developed or etic intelligence tests (Foxcroft et al., 2004; Kumar, 2017). However, the use of etic intelligence tests increases the risk of “culturally biased errors” (Shuttleworth-Edwards, 2016, p. 977), which compromises the reliability and validity of the test results and, consequently, the conclusions drawn, and the interventions recommended. Culturally biased test errors occur when there is a discrepancy between a person’s actual intellectual ability and their ability as measured on an intelligence test due to socio-cultural factors such as English language proficiency, quality of education, and socio-economic status. In South Africa, these socio-cultural factors are complex and inter-related, contributing to the country's diversity, and can have a significant impact on a person's performance on intelligence tests (Cockcroft et al., 2015; Seabi & Amod, 2009; Skuy & Skuy, 2005; Shuttleworth-Edwards, 2016; Shuttleworth-Edwards et al., 2004; Shuttleworth-Edwards, Van der Merwe, et al., 2013; Van der Merwe et al., 2022).

In terms of English language proficiency and socio-economic status, a recent census (Statistics South Africa, 2023) indicated that only 8.7% of South African households speak English as a first language, while the World Bank (2023) reported that 55.5% of South Africans were living in poverty in 2014. More recent statistics by Hall (2024) indicate that in 2022, 11.1% of South African children lived in poverty, and 7.9 million did not have access to appropriate nutrition as they lived below the food poverty line. In addition to poor nutrition, limited access to appropriate healthcare, and lower levels of parental education, poverty also regularly impacts the quality of education a child has access to (Mitchell et al., 2018). Schirmer and Visser (2023), Shuttleworth-Edwards, Van der Merwe, et al. (2013), and Spaull (2015) emphasise the poor learning outcomes of many South African children as a result of low-quality education linked to poverty and related factors such as high teacher-pupil ratios, a lack of essential resources such as books and desks, poorly trained teachers, and high rates of absenteeism for both teachers and children. These socio-cultural factors not only impact a children's performance on intelligence tests, but given their contribution to diversity, complicates the development, standardisation, and norming of intelligence tests that can be considered culturally applicable across different subgroups within the broader South African society (Shuttleworth-Edwards, 2016; Van der Merwe et al., 2022).

Regardless, with South Africa remaining "at the bottom of all international tables on learning outcomes: reading, maths, science" (Schirmer & Visser, 2023, The Silent Crisis section, para. 1), the development of culturally applicable intelligence tests that can inform appropriate and targeted interventions to support children's learning, should be a priority. Laher (2024) stresses the importance of developing emic tests that incorporate materials and methods more familiar to South Africans, as was done in the development of the Panga Munthu Test (PMT) and the Zambia Child Assessment Tool (ZamCAT) (Kuthria & Serpell, 1998; Matafwali & Serpell, 2014). The author, however, also mentions the importance of adapting and norming etic tests for use with South African populations. Van der Merwe et al. (2022) agree but emphasise the importance of developing intelligence tests that are fair, reliable, and valid when applied to specific subgroups within the broader South African society. In this regard, Shuttleworth-Edwards (2016) questions the utility of developing population-based norms for diverse societies such as South Africa and subsequently proposes more within-group research studies. In the context of intelligence testing, population-based norms enable the comparison of an individual's performance with that of age-matched peers from a census-matched population. Within-group research studies investigate the performance of specific subgroups within a population, allowing their scores to be meaningfully compared to the broader population norms and to serve as a reference point for interpreting the performance of individuals within those subgroups (Sunderaraman et al., 2016).

A number of these research studies have been done by South African researchers with South African samples to investigate the psychometric properties of intelligence tests that have not been standardised for the South African context, including the Raven's Coloured Progressive Matrices (August, 2017; Knoetze et al., 2005), the Wechsler Adult Intelligence Scale – Third Edition (WAIS-III) (Cockcroft et al., 2015; Shuttleworth et al., 2004; Shuttleworth-Edwards, Gaylard, et al., 2013), the Kaufman Assessment Battery for Children (K-ABC) (De Sousa et al., 2010; Jansen & Greenop, 2008), the Wechsler Intelligence Scale for Children – Revised (WISC-R) (Skuy et al., 2001), the WISC-R and K-ABC (Skuy et al., 2000), the Wechsler Intelligence Scale for Children – Fourth Edition (WISC-IV)(Bickell, 2015; Ferrett, 2011; Shuttleworth-Edwards, Van der Merwe, et al., 2013), the KABC-II (Mitchell, 2015; Mitchell et al., 2018), and the WISC-V and KABC-II (Cassoojee, 2020). The Cassojee (2020) study was a preliminary study, conducted under the researcher's supervision, focusing on the first 50 participants of this research study. It specifically compared the scores of participants on specific scales and indices of the WISC-V and KABC-II, which are said to assess similar constructs. Nonetheless, these research studies remain scarce, and aside from the studies by Cassojee (2020) and Skuy et al. (2000) in August 2017, the other research studies did not specifically include children who experience barriers to learning and development. As intelligence tests are regularly used in the assessment of children who experience barriers to learning and development to recommend suitable interventions and inform support and placement (Learning Disabilities Association of America [LDA], 2019; Nel & Grosser, 2016; Poletti, 2016), August (2017) recommends more research with samples of children who experience barriers to learning and development, this was deemed important.

Furthermore, while being cognisant of the limited number of South African research studies on the psychometric properties of the WISC-V and the KABC-II, specifically with children who experience barriers to learning and development, and the recommendations by Shuttleworth-Edwards (2016) and Van der Merwe et al. (2022) regarding the importance of within-group research studies for diverse countries such as South Africa, the present study aimed to investigate the psychometric properties of the WISC-V and KABC-II in a sample of South African children who experience barriers to learning and development. In doing so, it endeavoured to contribute to the existing research on the psychometric properties of etic intelligence tests when used with specific subgroups within the broader South African context. In addition, it attempted to contribute to the calls by Mitchell et al. (2018) and Van der Merwe et al. (2022; 2024) for South African research studies that can contribute to an understanding of the validity, reliability, and fairness of recently developed etic intelligence tests when used with South African children to determine appropriate school placement and suitable intervention programmes.

1.2. Aims and Objectives of the Research Study

The primary objective of the research study was to conduct a preliminary investigation into the psychometric properties of the WISC-V and the KABC-II in a sample of South African children in the intermediate (Grades 4 – 6) and early senior phases (Grades 7 and 8) of education, who face barriers to learning and development that negatively impact the acquisition and development of essential academic skills. Aligned to this, two objectives were set:

1. Objective 1: To investigate the internal consistency reliability of the WISC-V and KABC-II subtests using the Spearman-Brown split-half reliability method within the framework of Classical Test Theory.
2. Objective 2: To investigate the construct validity of the WISC-V and KABC-II through the use of correlation coefficients and confirmatory factor analysis (CFA), within the framework of Classical Test Theory.

In addition, the study also aimed to investigate the suitability of these two measures in a South African sample through item analysis and a comparison of the performance of English First Language (EFL) and English Additional Language (EAL) participants. In line with this, two additional objectives were set:

3. Objective 3: To determine whether statistically significant differences in performance exist between EFL participants and EAL participants, using a General Linear Model (GLM) with participants' scores as the dependent variables and English language background as the independent variable.
4. Objective 4: To investigate differential item functioning (DIF) using Item Response Theory (IRT), focusing specifically on the subtest items of the WISC-V and KABC-II.

1.3. Research Questions

Corresponding research questions were posed for each of the four objectives:

1. Objective 1: To investigate the internal consistency reliability of the WISC-V and KABC-II subtests using the Spearman-Brown split-half reliability method within the framework of Classical Test Theory.
    - Is the WISC-V reliable for use with a sample of South African children who experience barriers to learning and development that negatively impact the acquisition and development of essential academic skills?
    - Is the KABC-II reliable for use with a sample of South African children who experience barriers to learning and development that negatively impact the acquisition and development of essential academic skills?
2. Objective 2: To investigate the construct validity of the WISC-V and KABC-II through the use of correlation coefficients and CFA, within the framework of Classical Test Theory.
    - Does the factor structure of the WISC-V in this sample of South African children who experience barriers to learning and development that negatively impact the acquisition and development of essential academic skills align with its theoretical model based on correlation coefficients and CFA?
    - Does the factor structure of the KABC-II in this sample of South African children who experience barriers to learning and development that negatively impact the acquisition and development of essential academic skills align with its theoretical model based on correlation coefficients and CFA?
3. Objective 3: To determine whether statistically significant differences in performance exist between EFL participants and EAL participants, using a GLM with participants' scores as the dependent variables and English language background as the independent variable.

- Do statistically significant differences exist between the scores of EFL and EAL participants in this sample of South African children who experience barriers to learning and development that negatively impact the acquisition and development of essential academic skills?

4. Objective 4: To investigate DIF using IRT, focusing specifically on the subtest items of the WISC-V and KABC-II.
    - Do the subtest items of the WISC-V show DIF when administered to a sample of South African children who experience barriers to learning and development that negatively impact the acquisition and development of essential academic skills?
    - Do the subtest items of the KABC-II show DIF when administered to a sample of South African children who experience barriers to learning and development that negatively impact the acquisition and development of essential academic skills?

1.4. Concept Clarification

In this section, important concepts used consistently throughout this research study are clarified. These include psychometric properties, intelligence, intelligence tests, barriers to learning and development, and essential academic skills. Moreover, given that the terms *ethnicity/race* and *sex/gender* are used, it is also important to clarify how these concepts are understood in the South African context.

1.4.1. Ethnicity and Race

The Population Registration Act of 1950 divided the South African population into four distinct *racial* groups: Black, Coloured, Indian, and White (Clark, 2007). The most recent South African census (Statistics South Africa, 2023) still uses the terms Black African, Coloured, Indian or Asian, and White; however, it refers to these as *population groups* rather than races. Within the context of this research study, similar to the delineations made by Laher (2024) and Shuttleworth-Edwards (2016), Black persons refer to people of African origin, who make up 81.4% of the South African population; Coloured persons refer to people of mixed-race, who make up 8.2% of the South African population; Indian/Asian persons refer to people with an Asian heritage, who make up 2.7% of the South African population; and White persons refer to people with a European ancestry, who make up 7.3% of the South African population (Statistics South Africa, 2023). Also, as is done by Laher (2024), preference will be given to the term *ethnicity*, unless used in the context of South Africa's past and its history of racial segregation. Moreover, when reference is made to the socio-cultural factors that can impact learning, development, and people's performance on intelligence tests, cognisance needs to be taken that these are complex and broader than ethnicity and can include many inter-related factors, such as English language proficiency, quality of education, parent education level, access to resources and healthcare, and socio-economic status (Cockcroft et al., 2015; Seabi & Amod, 2009; Skuy & Skuy, 2005; Shuttleworth-Edwards, 2016; Shuttleworth-Edwards et al., 2004; Shuttleworth-Edwards, Van der Merwe, et al., 2013; Van der Merwe et al., 2022).

1.4.2. Sex and Gender

According to the South African Litigation Centre (2016), the concepts of *sex* and *gender* are often used interchangeably, albeit incorrectly. In South Africa, an understanding of gender is a socially constructed concept that refers to "the set of roles, activities, expectations and behaviours assigned to females and males by society" (South African Litigation Centre, 2016, p. 3). Sex, however, is assigned at birth and distinguishes between females and males based on biological factors. As used in the most recent South African census, the concept of sex will be given preference, which refers to males and females (Statistics South Africa, 2023).

1.4.3. Psychometric Properties

When considering the cultural applicability of etic intelligence tests, such as the WISC-V and KABC-II, when administered to South African children, it is essential to investigate their psychometric properties. This is done to ensure that the results obtained from an intelligence test are valid, reliable, and fair, as are the interventions recommended to address the challenges experienced by children (De Beer, 2004; De Kock & Foxcroft, 2018). Among these psychometric properties are validity, reliability, and DIF. Validity and reliability are both situated in Classical Test Theory. While validity refers to the ability of an intelligence test to assess the constructs it was designed to measure accurately, reliability refers to the consistency with which the test can do so across time and repeated administrations (Roodt & De Kock, 2018a, 2018b). Item analysis, situated in IRT, is another statistical technique that can be used to investigate the cultural applicability of an intelligence test by analysing item bias. Referred to as DIF, item bias can be investigated through item analysis and refers to the way that the items of a particular intelligence test create differentiation or variance in the performance of examinees due to socio-cultural factors rather than ability. If DIF is present, the intelligence test cannot be considered fair for use with individuals from specific socio-cultural contexts (De Beer, 2004; De Kock & Foxcroft, 2018; International Test Commission, 2017; Lai et al., 2005; Taylor, 2009).

1.4.4. Barriers to Learning and Development and Essential Academic Skills

Within the context of this study and as explained by the National Commission on Special Needs in Education and Training (NCSNET) and the National Committee on Education Support Services (NCESS) (Department of Education [DoE], 1997), barriers to learning and development is understood as a broad concept that includes a variety of factors that can negatively impact the acquisition and development of a child's essential academic skills. Some of these factors may be extrinsic to the child, such as contextual factors, while others may be intrinsic, for example, learning impairments or learning disabilities. Thus, barriers to learning and development encompass a multitude of factors, often interrelated, that negatively impact the child's development and learning.

In referring to essential academic skills, the study acknowledges that barriers to learning and development can negatively impact the development of several academic and cognitive skills, including receptive and expressive language, reading, written expression, mathematics, sensory processing, memory, executive functions, and fine and gross motor skills (Nel & Grosser, 2016). From these, the South African Curriculum Assessment Policy Statements (CAPS) for the Foundation Phase emphasise the importance of a child developing listening, speaking, reading, and written expression in their home language, as well as mathematics (Department of Basic Education [DoBE], 2011a, 2011b). Thus, as these are considered foundational to a child's learning and continued development, essential academic skills in the context of this research study refer to listening, speaking, reading, written expression, and mathematics. It further acknowledges that cognitive abilities and processes such as memory, sensory integration and processing, and executive functions, as well as fine and gross motor skills, impact the development of these academic skills.

1.4.5. Intelligence and Intelligence Tests

According to Kaufman et al. (2016a, p. 10), concepts such as "IQ test, cognitive ability test, and intelligence test" are often used interchangeably in the literature; however, they tend to refer to measures that assess the cognitive abilities of individuals. Given its use in international and South African literature (Kaufman et al., 2016a; Lichtenberger et al., 2006; McGill, 2015; Naglieri, 2009; Renner et al., 2023; Shuttleworth-Edwards, Van der Merwe et al., 2013; Skuy et al., 2000; Van der Merwe et al., 2024) to refer to assessment measures, such as the Kaufman Assessment Battery for Children and Wechsler Intelligence Scale for Children, this study will use the concept intelligence tests. It, however, recognises that, based on the theoretical perspectives underlying the development and interpretation of the WISC-V and KABC-II, intelligence is understood from both a psychometric and a structural perspective, as well as a cognitive or information-processing viewpoint. Psychometric and structural theories of intelligence, such as Wechsler's conceptualisation of intelligence and the Cattell-Horn-Carroll (CHC) theory, view intelligence in relation to specific cognitive abilities and employ psychometric techniques, particularly factor analysis, to identify the factors or cognitive abilities that constitute an individual's intellectual functioning. Conversely, information-processing theories, such as Luria's neuropsychological theory and the Planning, Attention, Simultaneous and Successive (PASS) cognitive processing theory, view intelligence in relation to the cognitive processes that underpin intellectual functioning. Thus, in this study, intelligence is understood as the cognitive abilities and cognitive processes that play a critical role in a person's intellectual functioning and their ability to perform a wide variety of tasks.

1.5. Outline of the Research Thesis

Cresswell and Cresswell (2020) suggest that a research thesis adopting a quantitative approach follows a structured sequence, typically beginning with an introductory chapter, followed by a review of relevant literature and theory, a detailed description of the methods used, and concluding with the presentation and discussion of findings. In alignment with this structure, this thesis has been organised as follows:

Chapter 1: This chapter provided the background and rationale for the study. Thereafter, the aims, objectives and research questions were posed, and important concepts used in the study were explained. The chapter ends with this outline of the research thesis.

Chapter 2: This chapter provides an overview of literature related to the core concepts under investigation in this research study, starting with an exploration of the concept of barriers to learning and development. Thereafter, the concept of intelligence is explored in more depth through a review of three theoretical frameworks that underlie the intelligence tests used in this study. The chapter then presents an overview of the WISC-V and the KABC-II, followed by an investigation into available research on the psychometric properties of these intelligence tests and their cross-cultural applicability. It also provides an overview of investigations of DIF in intelligence tests. It concludes with an exploration of African and South African research which utilised the Kaufman Assessment Batteries for Children and the Wechsler Intelligence Scales for Children.

Chapter 3: This chapter discusses the methods used to address the study's aims and objectives. Included are an overview of the research approach, paradigm, methods, and designs, as well as the sample and sampling technique. As discussed in the previous chapter, this section provides information on the measurement instruments used in this research study, along with a description of the parent questionnaire that parents or legal guardians are asked to complete. Thereafter, an overview of the procedure followed in this research study is provided. This chapter concludes with a discussion of the data analysis techniques employed in answering the research questions and the ethical standards considered during the research process.

Chapter 4: This chapter presents the findings of the research study, beginning with descriptive statistics that provide the means, standard deviations, minimum values, and maximum values of all the global, index/scale, and subtest scores calculated for this study. Thereafter, the results of the data analyses as they pertain to the four objectives of the study are included.

Chapter 5: This chapter presents the findings of this study and concludes the thesis with an overview of the study’s strengths and limitations, followed by recommendations for future research.

2. Literature Review

2.1. Introduction

Internationally and in South Africa, intelligence tests, such as the Wechsler Intelligence Scale for Children – Fifth Edition (WISC-V) and the Kaufman Assessment Battery for Children – Second Edition (KABC-II), are routinely used in the assessment of children who experience barriers to learning and development that impact their acquisition of essential academic skills (Learning Disabilities Association of America [LDA], 2019; Nel & Grosser, 2016; Poletti, 2016). Countries like the United States (US) refer to these children as having a *specific learning disability* (Individuals with Disabilities Education Improvement Act [IDEIA], 2004). In the Diagnostic and Statistical Manual of Mental Disorders – Fifth Edition, Text Revision (DSM-5-TR) preference is given to *specific learning disorder* (American Psychiatric Association [APA], 2022), while the International Statistical Classification of Diseases and Related Health Problems – Eleventh Revision (ICD-11) (World Health Organisation [WHO], 2022) refers to *developmental learning disorder*. In explaining these concepts, reference is often made to the role of neurological and biological factors in their origin and development. In the South African context, accurately diagnosing a specific disorder with a distinct cause is frequently challenging due to the presence of multiple barriers that can adversely affect a child's development and acquisition of essential academic skills. Consequently, the term *barriers to learning and development* is widely adopted.

This literature review will begin by exploring these thoughts and elaborating on the history of education, as well as the barriers to learning and development in South Africa. Thereafter, a review of the work of prominent authors in the field, as well as relevant legislation and policy documents, will help define barriers to learning and development, supporting the adoption of this concept in the South African context and this research study. The literature review will then continue by examining the concept of intelligence through the lens of different theoretical perspectives. In this regard, attention will be given to three prominent theories of intellectual functioning that underlie the development and interpretation of the WISC-V and KABC-II: Wechsler's conceptualisation of intelligence and the Cattell-Horn-Carroll (CHC) theory, both psychometric theories, and Luria's neuropsychological theory as conceptualised in the Planning, Attention, Simultaneous and Successive (PASS) cognitive processing model. The literature review concludes with an overview of the WISC-V and KABC-II, as well as an exploration of studies conducted to investigate their psychometric properties. This includes a review of the concepts of validity and reliability, in conjunction with research related to the validity and reliability of the WISC-V and the KABC-II, followed by an exploration of studies that have examined the cross-cultural applicability of these two measures. Item analysis and Item Response Theory (IRT) will also be explored, and the rationale for using item analysis in this research study will be substantiated. The chapter concludes with a review of African and South African research, which included the current and former editions of the Wechsler Intelligence Scale for Children (WISC) and the Kaufman Assessment Battery for Children (K-ABC).

2.2. South African History of Education and Barriers to Learning and Development

Alnaim (2016) emphasises the significance of understanding the historical context of concepts such as barriers to learning and development, as it offers teachers, parents, psychologists, and researchers valuable insights into how children experiencing these challenges have been educated over time and how their right to quality education has been addressed. Clark (2007) provides a review of the history of education in South Africa, dividing it into four time periods: the Early Years (pre-1652), the Colonial Period (1652-1948), the Apartheid Era (1948-1994), and the Democratic Era (1994-present)., The history of barriers to learning and development can be incorporated within these four periods.

2.2.1. The Early Years

Prior to the arrival of the first European settlers in 1652, the Khoikhoi, San, and various other African tribes living in southern Africa employed both informal and formal techniques to support the holistic development of children. During this time, which Clark (2007) labels the Early Years and Seroto

(2011a) refers to as the Pre-Colonial Period, the inhabitants of southern Africa viewed education as a means of transmitting cultural knowledge and values from one generation to the next. Learning was focused on living and adapting successfully to one’s environment, whilst protecting and preserving it and its natural resources. In addition to cultural knowledge, language, and social roles, basic human needs such as food security, nutrition, hydration, safety, and shelter took precedence. Closely linked to the social life and structures of a community, as well as the well-being of the entire community, education was also about character building, instilling societal values and traditional law, conflict management, and developing ethical conduct and moral behaviour in citizens (Clark, 2007; Kaya & Seleti, 2013; Seroto, 2011b). Within this context, with an emphasis on survival and adaptation, where learning primarily occurred through practice, experimentation, observation, physical involvement, and tribal customs and rituals, the concept and identification of barriers to learning and development would not have been a priority (Clark, 2007; Dreyer, 2008).

2.2.2. *The Colonial Period*

With the Dutch being the first European settlers to arrive in the Cape Colony from 1652 onwards, schools were guided by the teachings and beliefs of the Dutch Reformed Church and the Ecclesiastical Council, focusing on instilling Christian beliefs while teaching children reading, written expression, and mathematics. Dutch was the primary language of learning and teaching in schools (Clark, 2007; Le Roux, 2011a). However, from the start of the second period of British rule in 1806, numerous efforts were made to reform and strengthen formal education and schools in the Cape Colony. With the arrival of more British settlers from 1820 onwards, English became the language of learning and teaching. In line with the status quo in Britain, the Anglicisation policy ensured that English was not only the medium of instruction but was also included in the curriculum as a subject. In addition, the curriculum became more secular; yet Anglican views were still promoted. Further, following the appointment of Lord Charles Somerset as Governor of the Cape Colony in 1814, English schools became state-owned enterprises, and appropriately trained English teachers were appointed to teach children reading, written expression, mathematics, music, and the principles of Christianity (Clark, 2007; Le Roux, 2011b). These more secular, English-language schools were not well supported by the Dutch, later Afrikaans-speaking part of the population, who resented English rule in general and sought the influence of their church in the education of their children. These concerns, among other factors, led to a large number of Dutch-Afrikaans-speaking people moving inland from the 1830s onwards. Issues of language and religion were intensified after the Anglo-Boer War (1899-1902), which led to the rise of Christian National Education (CNE) schools (Booyse, 2011; Le Roux, 2011b; Van Eeden & Vermeulen, 2005). CNE schools were private schools for Dutch-Afrikaans-speaking children, where the values and teachings of Calvinism dominated the curriculum, allowing children to be taught in their native language.

Although efforts were made to strengthen education for English- and Dutch-Afrikaans-speaking children, education for enslaved people, Black, and Coloured children was provided mainly by missionary societies (Clark, 2007; Le Roux, 2011b). Their work continued after the emancipation of enslaved people in 1834, focusing not only on teaching the children in their care about the Bible and the Christian faith, but also on reading and writing. However, as noted by Lewis and Steyn (2003, p. 101), "the work of missionaries and mission education was profoundly influenced by their historical backgrounds, culture, understanding of reality, personalities, social positions, ecclesiastical tradition, personal context, motivation, and ideology". Coming from a predominantly Western perspective, missionaries often disregarded the educational needs, culture, backgrounds, beliefs, and traditions of the people whom they educated. Furthermore, in line with the political needs of the ruling government at the time, it was believed that missionaries and the education they provided were intended to prevent uprisings and war between Europeans and the indigenous tribes of South Africa (Msila, 2007).

While missionaries also provided education to some Muslim children living in the Cape Colony, it was Muslim Imams who took primary responsibility for the education of their children. Given the importance of religion in the education of Muslim children, the first Madrasah was established in Cape Town in 1793, and the first mosque in 1799. At the Madrasah, children were not only educated in the Holy Qur'an, but also taught to read and write in Arabic (Le Roux, 2011a). Missionaries, however, continued to provide education to Indian/Asian children in the then Natal. These were often children of labourers, imported from India to assist in the establishment of sugar cane farms by the British Government. The first English-medium school for Indian/Asian children and adults was established in 1869. After the enrolment of Indian and Asian children in White-only schools was prohibited at the end of the 1800s, additional schools for these children were established in Durban in 1899 and Pietermaritzburg in 1903 (Le Roux, 2011b).

Following the establishment of the Union of South Africa in 1910, free education for White children became compulsory, and the education for Dutch-Afrikaans and English-speaking children was prioritised to reconcile the divide left by the Anglo-Boer War (Booyse, 2011; Clark, 2007). The Constitution of the Union of South Africa recognised English and Dutch (later Afrikaans) as mediums of instruction for all White children (De Wet & Wolhuter, 2009). Of further importance is the fact that, as formalised by the School Board Act of 1905 (Cape Province) and the Education Act 25 of 1907 (Transvaal Province), separate education and schools were established for White, Coloured, Black, and Indian/Asian children. As a result of these events and the laws that came into effect, education in South Africa could be split into three sections around the start of the Second World War. The first was English-medium, state-funded schools, which supported the earlier Anglicisation policy and modelled their structure and curriculum on that of British education. Parallel to these were the CNE schools, where Dutch and, later, Afrikaans were the medium of instruction, and Calvinistic beliefs were incorporated into the curriculum to promote the identity and cultural heritage of this part of the population. As noted, these English and Dutch-Afrikaans schools primarily catered to white children. As a result, the third section primarily included missionary primary and secondary schools, as well as colleges, for Black, Coloured, and, to a lesser extent, Indian and Asian children (McKeever, 2017). These later became state-owned schools.

In addition to a racial and linguistic divide in schools, the Colonial Period also witnessed a divide in education for typically developing children and those who faced barriers to learning and development. Oakland (2004) states that countries often concentrate on first providing educational opportunities for children with sensory and physical impairments. This was certainly the case for South Africans during the Colonial Period. Among the schools for children who experienced barriers to learning and development were the Grimley Institute for the Deaf, founded in 1863. This later became the *Dominican Grimley School for the* Deaf in Cape Town, which catered for White children, and the *Dominican School for the Deaf at Wittebome*, which accommodated Coloured children (Clark, 2007). Schools for visually impaired children were also established, including the *Doofstommen en Blinden Instituut*, established in 1881 at Worcester (Vergottini & Weyers, 2020). The *Athlone School for the Blind* catered explicitly for visually impaired Coloured children. It is worth noting that none of these schools were funded by the government of the time. It was only with the passing of the Vocational Education and Special Education Act 29 of 1928 that specialised schools fell under the jurisdiction of the Department of Education (Clark, 2007; Dreyer, 2008; Vergottini & Weyers, 2020). Furthermore, with the inception of the Special Schools Amendment Act in 1937, compulsory education and placement in existing special schools for White children were mandated; however, no provision was made for Black children and once again, the education of Black children who experienced barriers to learning and development lay predominantly in the hands of the church. At the time, the Roman Catholic Church and the Dutch Reformed Church established, amongst others, the *Kutlwanong School for the Deaf* and the *Tshilidzini School* for children with physical disabilities (Clark, 2007; Dreyer, 2008; Vergottini & Weyers, 2020).

2.2.3. The Apartheid Era

With the foundation laid during the Colonial Period, the Apartheid Era was marked by educational inequality, division, and exclusion. In line with the political ideology of the time, the South African education system promoted segregation and exclusion based on factors such as race and language (Naicker, 2000). Clark (2007) mentions the Population Registration Act of 1950, which divided the South African population into four distinct racial groups, Black, Coloured, Indian/Asian, and White. This racial division would support an education system that offers noticeable differences in the quality and quantity of education to different racial groups (McKeever, 2017; Walton & Engelbrecht, 2024). Children from different racial backgrounds were not only forced to attend separate schools, but the differences in funding for White schools, contrasted to other schools, led to a lack of appropriately qualified teachers and physical resources in schools catering for children from other racial groups. These divisions, in turn, lead to a system which saw a much higher education attainment for White children when compared to Black, Indian/Asian, and Coloured children (Fiske & Ladd, 2006; Lu & Treiman, 2011; McKeever, 2017; Naicker, 2000; Walton, 2011).

Whilst promoting inequality and division based on language and race, the Apartheid Era also promoted division based on ability by creating a dual-education system. This dual-education system resulted in children being seen as abled attending mainstream schools and those seen as disabled attending special schools (Naicker, 2000; Walton, 2011). This trend continued throughout the Apartheid Era and is evident in the Education Affairs Act, No. 70 of 1988. This act referred to “handicapped” children who required access to special schools, as they were unable to benefit from the education offered in ordinary schools and classrooms (Republic of South Africa, 1988, p. 5). In this dual-education system, schools that accommodated White children again enjoyed more significant funding and support, and to a lesser extent, schools for Indian/Asian and Coloured children. Black children who experienced barriers to learning and development had limited options and were often forced to either attend their local schools with no additional support or interventions offered, or not attend school altogether (Naicker, 2000; Walton, 2011). In addition, Clark (2007) notes that White special schools not only received better funding but also benefited from education through appropriate training courses offered by universities and colleges, which specially trained and qualified teachers taught. Black, Coloured, and Indian/Asian children with barriers to learning and development were transferred from mission schools to state-funded schools created by the different education departments catering for children from different racial backgrounds. However, the development of special schools amongst these learners was generally slow, and a severe discrepancy existed in the value, quantity, and quality of specialised education offered.

Introduced by the Vocational Education and Special Education Act 29 of 1928 (Clark, 2007; Dreyer, 2008) and strengthened by the Special Schools Act of 1948 (Engelbrecht, 2006; Naicker, 2007), the Apartheid Era also continued to apply a medical-deficit model in understanding disability in South Africa. This was not unique to South Africa. For instance, the concept of learning disabilities was introduced to the US in the 1960s. Previously, parents often sought the help of medical professionals for their children who struggled to learn and acquire essential academic skills. These professionals would use concepts “such as brain damage, minimal brain dysfunction, neurological impairment, or perceptual handicap” (Grünke & Cavendish, 2016, p. 2). The application of these concepts, combined with the assistance of medical professionals, in diagnosing and addressing the challenges faced by children, situates the field within a medical paradigm.

Regardless, Swart and Pettipher (2019) state that the medical-deficit model can be considered a paradigm that shapes not only people’s beliefs and values, but in practice impacts the explanations they provide for the challenges that children experience, as well as their behaviour and the methods employed in assessing and in assisting these children. In essence, it is a paradigm that attributes educational challenges solely to the child; hence, the concept of the within-child model is employed

(Engelbrecht, 2006; Dreyer, 2008; Nel et al., 2016). As a result, the focus of an assessment process often falls on diagnosing these difficulties within the child and then on treating the child (Ainscow & César, 2006; Swart & Pettipher, 2019). Reindall (2008) agrees with this assumption of the medical-deficit model, referring to it as the individual-materialist tradition. This author goes further in saying that within the paradigm, the challenges that children experience are viewed as being biologically, neurologically, and physically based, which is understandable given that the model has its roots firmly established in the medical field.

As a result of these assumptions, the medical-deficit model often views the child who experiences specific challenges as “the unfortunate individual”, underpinned by the “view that disability is a personal tragedy due to reduced function” (Reindall, 2008, p. 141). This premise can lead to stigma, labelling, categorisation, and the perception that children who experience specific challenges are the victims of their circumstances and individual difficulties and disabilities (Dreyer, 2008; Reindall, 2008; Nel et al., 2016; Swart & Pettipher, 2019). In this regard, Dreyer (2008) and Green and Engelbrecht (2007) aptly point out that it is a model which supports the notion that children who experience barriers to learning and development are seen as requiring the charity of others, whilst Clark (2007) speaks to the fact that it disrespects and disempowers individuals and leaves them dependent on others. Additional concerns about the medical-deficit paradigm include a disregard for external factors that may be causing or contributing to the child’s challenges, as well as those within the education system (Ainscow & César, 2006). The medical-deficit model is linear and directly links the child’s biology with the difficulties they experience, without considering additional external factors and their role (Engelbrecht, 2006; Reindall, 2008). Nel et al. (2016) describe it as a discriminatory model, one which supports the view that the compulsory segregation of children based on their ability levels is required, as those who are believed to be *disabled* cannot benefit from the same education as those who are identified as *able*.

2.2.4. **The Democratic Era**

Inheriting a fragmented, unfair, and unequal education system, the Democratic Era focused on addressing past injustices and incorporating broader processes of transformation and unification in South Africa into its education. The Constitution of South Africa (Republic of South Africa, 1996a), inclusive of the Bill of Rights, and the South Africa Schools Act, No. 84 of 1996 (Republic of South Africa, 1996b), were the first documents to lay the foundation for inclusive education as a strategy to assist transformation and contribute to a more just and equitable society (Engelbrecht, 2006). Several important laws and policy documents followed this, for instance the Education White Paper 6: Special Needs Education: Building an Inclusive Education and Training System (Department of Education [DoE], 2001), Conceptual and Operational Guidelines for the Implementation of Inclusive Education: District-Based Support Teams (DoE, 2005), the Guidelines for Full-Service/Inclusive Schools (Department of Basic Education [DoBE], 2010), and the Policy on Screening, Identification, Assessment, and Support (SIAS) (DoBE, 2014). These policies outlined the implementation of inclusive education in South Africa, aiming to enhance and encourage inclusive education by establishing procedures for identifying, assessing, and supporting children with barriers to learning and development in inclusive classrooms.

As stated by Engelbrecht (2006) and Green and Engelbrecht (2007), there is no one single definition of inclusive education. Regardless, important aspects of inclusive education can be identified through an analysis of these South African documents, as well as the international frameworks, statements, and policies that influenced their development. Included would be the Universal Declaration of Human Rights, adopted by the United Nations General Assembly (United Nations, 1948), the Education for All framework (United Nations Educational, Scientific and Cultural Organisation [UNESCO], 1990) and the Salamanca Statement (UNESCO, 1994). The DoBE (2014) also emphasises the importance of the Convention on the Rights of the Child (United Nations, 1989) and the Convention on the Rights of Persons with Disabilities (United Nations, 2006), whilst Nel et al. (2016) includes the African

Charter on the Rights and Welfare of the Child (United Nations Children's Fund [UNICEF], 1999) and the Dakar Framework for Action Education for All: Meeting our Collective Commitments (UNESCO, 2000).

Inclusive education is more than the integration of children who experience barriers to learning and development in the general or mainstream classroom, which many believe it to be (Ainscow & César, 2006; Nel et al., 2016). This is referred to as mainstreaming (Donald et al., 2014). Inclusive education seeks to realise equitable levels of integration and the creation of a unified education system that provides quality education for all (UNESCO, 2001). It is linked to the Education for All movement, as inclusive education strives to provide all persons with quality education in a non-discriminatory and inclusive society (UNESCO, 2005). It is a commitment to human rights, based on the belief that children's access to quality education should be considered a fundamental human right. In addition, it strives to respect this fundamental right through the provision of equitable learning opportunities for all, regardless of factors such as ethnicity, ability, language, sex, religion, class, and socio-economic status (Ainscow & César, 2006; Du Plessis, 2013; Engelbrecht, 2006; UNESCO, 2001, 2005).

According to UNESCO (2005), an education system built on human rights has three core principles: "access to free and compulsory education; equality, inclusion and non-discrimination; the right to quality education, content and processes" (p. 12). Guided by these principles, inclusive education seeks to foster an integrated and just society by addressing how differences and diversity are perceived and treated (Engelbrecht, 2006). This diversity extends not only to children's educational needs and the barriers they encounter, but also to the other factors reported (Ainscow & César, 2006). It also supports the development of an integrated society and aligns itself with the development of a unified education system, far removed from the dual-education system prevalent in South Africa's educational history (Engelbrecht, 2006; UNESCO, 2005). Inclusive education challenges exclusion and segregation, viewing diversity and the barriers to learning and development that children experience as an opportunity to enhance the learning of all children and build the capacity of a country's education services (Ainscow & César, 2006; Du Plessis, 2013; UNESCO, 2001). Tied to the belief that most children will encounter some barriers at some point in their lives, some of which may go unrecognised or undetected, inclusive education is a child-centred pedagogy that values the child and enhances the learning experiences of all children, even those whose barriers remain hidden. Additionally, it enables teachers from diverse backgrounds to combine their knowledge, skills, and experience to provide high-quality education to all children (Ferguson, 2008). Whereas some view inclusive education as *access*, it goes beyond to include *active participation*. Thus, children with barriers to learning and development should not only be given access to schools but also be encouraged to participate actively in the learning process. At the same time, their needs are accommodated and addressed (Ainscow & César, 2006).

UNESCO (2005) raises an important consideration by describing inclusive education as a process. This process continually seeks to meet the distinct learning needs of all children, whilst actively involving all in the learning process, and eradicating any form of exclusion. In doing so, children are not required to change; instead, the responsibility falls on schools and other institutions of learning to adapt teaching methods and the culture of the school, prepare the physical layout of the school so that children with for instance physical disabilities can be accommodated, and provide differentiated learning experiences and content to meet the learning needs of a diverse group of children (Nel et al., 2016). Of further importance is the fact that inclusive education recognises that not all difficulties and disabilities originate within the child (Ainscow & César, 2006; UNESCO, 2005). This is central to the South African conceptualisation of barriers to learning and development.

2.3. Barriers to Learning and Development

A country's history, its political and social context, its legislation, and the definitions proposed by leading authors, researchers, and organisations, determine how a concept such as barriers to learning and development is understood, identified, and addressed (Al-Yagon et al., 2013; Büttner & Hasselhorn, 2011; Dudley-Marling, 2004; Grünke & Cavendish, 2016; Loyd et al., 2007; Sideridis, 2007).

As such, various concepts are used to describe the challenges that negatively impact children's acquisition of essential academic skills. For instance, in the US, the term *specific learning disability* is a legislative definition "used to describe a disorder in one or more of the basic psychological processes involved in understanding or in using language, spoken or written, which disorder may manifest itself in imperfect ability to listen, think, speak, read, write, spell, or do mathematical calculations" (IDEIA, 2004, p. 13). The aetiology of a specific learning disability is ascribed to "perceptual disabilities, brain injury, minimal brain dysfunction, dyslexia, and developmental aphasia" (IDEIA, 2004, p. 13). However, disabilities and difficulties related to visual or auditory acuity, motor functioning, global intellectual development, emotional and/or behavioural factors, and socio-economic, cultural, linguistic, and contextual dynamics are disregarded as causal factors. In Australia, the term *learning difficulties* is more commonly used in primary and secondary schools, referring to children who struggle to make scholastic progress due to difficulties in learning (Van Kraayenoord & Chapman, 2016). Learning Difficulties Australia (n.d.) notes that both internal and external factors must be considered as potential causal or contributing factors in the development of learning difficulties. Internal causal factors include medical, biological, and neurological factors, whilst external factors can include familial, socio-economic, educational, and other contextual factors. Like the United States, the New Zealand Federation of Specific Learning Disabilities (n.d.) and the Learning Disabilities Association of New Zealand (n.d.) both use the term *specific learning disabilities*. In contrast, the New Zealand Ministry of Education (2021) uses the term *students with learning support needs*, which is similar to *students with special educational needs*, a concept used in the New Zealand Education and Training Act (New Zealand Government, 2020). These terms include a wide range of difficulties that can impact a child's learning and academic progress, including acquired neurological damage, medical difficulties, dyslexia, autism spectrum disorder, and English language proficiency.

In Canada, the situation is complicated by the fact that the duties of education, identifying barriers to learning and development, and providing support to identified children are the responsibilities of different territories and provinces, rather than the federal government. This has led to a discrepancy in the way difficulties related to learning and the acquisition of academic skills are identified and ultimately addressed (D'Intino, 2017; Stegemann, 2016). The only formal definition is provided by the Learning Disabilities Association of Canada, which adopted the term *learning disabilities* in 2002 and subsequently re-endorsed it in 2015. This definition refers to several "disorders which may affect the acquisition, organisation, retention, understanding or use of verbal or nonverbal information. These disorders affect learning in individuals who otherwise demonstrate at least average abilities essential for thinking and/or reasoning" (Learning Disabilities Association of Canada, 2015, n.p.). It further highlights the neurobiological and genetic nature of learning disabilities, noting that the difficulties caused cannot be attributed to socio-economic, educational, cultural, or linguistic factors, nor visual or hearing impairments. It is also distinguished from an intellectual developmental disorder. This appears different from the United Kingdom (UK), where the term *learning disabilities* is used to refer to a more general developmental delay or intellectual developmental disorder. In contrast, the term *learning difficulties* refers to children who struggle to progress academically due to challenges that relate specifically to their learning (Gates & Mafuba, 2016). Included within the latter are learning disorders and attention-deficit/hyperactivity disorder (ADHD). Although, like South African education in the Democratic Era, the UK aligns itself more closely with a social model of disability, rather than a medical one, leading to different concepts being used to capture the difficulties children experience in progressing scholastically, as well as a broader view of the factors which contribute to them (Gates & Mafuba, 2016). One such concept, *special educational needs*, is defined in the United Kingdom Children and Families Act of 2014 as "a learning difficulty or disability which calls for special educational provision to be made …" (Children and Families Act, 2014, p. 19).

While these country-specific definitions guide the identification, support, and education of children within these countries, psychologists often refer to two prominent classification systems in the formal diagnosis of children whose challenges impact the acquisition of essential academic skills. These include the DSM-5-TR (APA, 2022), which refers to *specific learning disorder*, and the ICD-11 (WHO, 2022), which refers to *developmental learning disorder*. Although the terms used by these two classification systems differ, several notable similarities are observed. For instance, both include these disorders in the broader category of Neurodevelopmental Disorders, indicating that the child's functioning in one or more academic domains is significantly below what would be expected for their age and intellectual ability. While these academic domains include reading, written expression, and mathematics, the classification systems also reiterate the importance of specifying the impact of the child's challenges on more specific subskills, such as reading comprehension, spelling, and mathematical reasoning. Similar to some of the country-specific definitions stated, both classification systems exclude contextual factors such as low-quality education, high rates of absenteeism, lack of access to education, and socio-economic status, as well as other contributing factors, such as intellectual developmental disorder, neurological damage, motor disorders, and visual and hearing impairments, in the aetiology of these disorders (APA, 2022; WHO, 2022). This has contributed to the debate around the value of classification systems.

While not disregarding the advantages of these classification systems, the Division of Clinical Psychology in the British Psychological Society (2013) report that much of the debate surrounding classification systems centres on the notion that "physical disease models and medical classification" are applied "to the realms of thoughts, feelings and behaviours, as implied by terms such as 'symptoms' and 'mental illness' or 'psychiatric disease'" (p.1). The Division of Clinical Psychology (British Psychological Society, 2013) further states that this medical approach does not adequately consider the role of socio-cultural factors in the aetiology of clinical diagnoses; a statement with which Kinderman et al. (2013)

agree. Vanheule (2012) notes that medical diagnoses often overlook the uniqueness of the individual in presenting specific difficulties, which in turn has important implications for the design of appropriate interventions. It may also unjustly increase the prevalence of specific disorders in specific contexts. While not ignoring biological and neurological factors and their causal role in the development of difficulties with learning and the acquisition of essential academic skills, the Division of Clinical Psychology (British Psychological Society, 2013) emphasises the importance of a multifaceted and contextualised approach in considering the factors that underlie these challenges. This approach, as supported internationally (Adelman & Taylor, 2018; Department for Education & Department of Health and Social Care, 2015; Forlin & Chambers, 2023; Franck, 2022; Hannant, 2024) and in South Africa (Flack, 2005; Kern, 2022; Malahlela & Johnson, 2024; Walton & Engelbrecht, 2024), forms the foundation for the understanding and conceptualisation of barriers to learning and development.

One of the first South African documents to mention the concept of barriers to learning and development, was a report submitted by the National Commission on Special Needs in Education and Training (NCSNET) and the National Committee on Education Support Services (NCESS) (DoE, 1997), entitled *Quality Education for All: Overcoming Barriers to Learning and Development*. This report stated that preference is given to this concept, rather than *special needs*, as the challenges that children experience in education are often multifactorial and contextual. Nel et al. (2016) support this by stating that the term *"special needs"* focuses on the "*specialness* of learners" (p. 20), which still places the challenges a child experiences within that child and thus aligns itself with the medical-deficit model. Conversely, inclusive education aligns itself with a socio-ecological perspective, viewing the challenges children experience and considering a multitude of factors that can hinder a child's development and progress (Green & Engelbrecht, 2007; Nel et al., 2016). The concept of barriers to learning and development was subsequently adopted by White Paper 6 (DoE, 2001) and the SIAS policy (DoBE, 2014) to describe multiple, often interrelated factors, either extrinsic or intrinsic to the child, that can negatively impact their learning and development.

From a socio-ecological perspective, specifically Bronfenbrenner's bio-ecological systems model (Bronfenbrenner, 1979; Donald et al., 2014; Nel et al., 2016; Swart & Pettipher, 2019), the focus is on considering a multitude of factors, both extrinsic and intrinsic to the child, that influence their learning and development. Within the framework of Bronfenbrenner's model, extrinsic factors can be located in any of the systems within which the child learns and develops, including the microsystem, mesosystem, exosystem, or macrosystem. Included can be socio-economic dynamics and poverty, which in turn impact a child's access to other essential services, such as adequate medical and healthcare services, quality education, and access to water and sanitation. In addition, poverty can negatively impact a child's nutrition, sense of safety, living conditions, and shelter, and can be linked to parental unemployment, stress, and poor educational attainment. Other extrinsic factors can be harmful attitudes towards children and, more generally, people who experience specific challenges, which can impact access to education and the quality of that education, as well as broader societal challenges such as substance abuse and violence. Multiple factors within the education system can also negatively impact the development and learning of children. Included can be poorly trained teachers, under-resourced schools, overcrowded classrooms, an uncompromising curriculum, unsafe school buildings, a lack of support services available in schools, the language of learning and teaching, the beliefs and attitudes of teachers and school management, and a failure in school-parent/community collaboration and partnerships. Educational policies that are unclear and/or poorly implemented can also contribute to a breakdown in learning (DoE, 1997; Donald et al., 2014; Donohue & Bornman, 2014; Nel et al., 2014, 2016; Swart & Phasha, 2019; Venter & Venter, 2019; Walton & Engelbrecht, 2024).

While many barriers to learning and development are extrinsic to the child, Bronfenbrenner's model also highlights intrinsic factors, such as person characteristics, which can influence learning and development (Bronfenbrenner & Morris, 2006; Swart & Pettipher, 2019; Tudge et al., 2009). Included in these intrinsic factors can be physical and neurological disabilities, such as epilepsy and cerebral palsy, intellectual developmental disorder, autism spectrum disorder, ADHD, visual and hearing impairments, and chronic diseases such as the human immunodeficiency virus infection (HIV) and acquired immune deficiency syndrome (AIDS) (DoBE, 2014; DoE, 1997; Donald et al., 2014; Nel et al., 2016; Walton & Engelbrecht, 2024). According to Du Plessis and Dednam (2019) and Nel and Grosser (2016), learning disabilities and learning impairments would fall within this category. Although concepts such as learning impairments, learning difficulties, and learning disabilities are often used interchangeably in South Africa, these authors note that learning difficulties are often related to specific subjects, or aspects of specific subjects, and can be addressed through appropriate interventions and subject choices. Conversely, learning disabilities and learning impairments are more pervasive and intrinsic. Furthermore, whereas learning difficulties tend to be more subject-specific, learning impairments and disabilities tend to vary in severity and the number of academic and cognitive skills affected. Included can be reading, written expression, mathematics, listening and speaking, executive functions, memory, sensory processing, and/or fine and gross motor skills. Du Plessis and Dednam (2019) and Nel and Grosser (2016) moreover state that while learning impairments and learning disabilities are often intrinsic, stemming from genetic, teratogenic, organic, or biological factors, the extrinsic factors stated earlier can either increase children's risk of developing these, or worsen the intrinsic challenges they already face.

Taking this into consideration, this research study prioritises the concept of barriers to learning and development, which is broader in scope than learning disabilities or learning impairments, in line with South African legislation. It acknowledges that within the South African context, various external and internal barriers can negatively impact learning and the acquisition and development of essential academic skills. It further recognises that not all children who experience barriers to learning and development will be formally diagnosed, and that various internal and external factors may necessitate further intervention and support (Astle & Fletcher-Watson, 2020). Given the multitude of factors that can negatively impact a child's development and learning, the identification of barriers to learning and development requires a comprehensive assessment, considering multiple risk and protective factors and involving several stakeholders, including teachers, parents, children, and school- and district-based support teams (DoBE, 2014; Du Plessis & Dednam, 2019; Nel & Grosser, 2016). Although the SIAS policy (DoBE, 2014) does not require the administration of intelligence tests as part of the assessment process, these can be included to encourage and guide the development of appropriate interventions to support the child's challenges and ensure that quality education is provided to all children. Additionally, as noted by Nel and Grosser (2016), teachers and other stakeholders often lack confidence in identifying specific barriers to learning and development, and thus rely on other assessment measures, such as intelligence tests, to support the identification and assessment process. However, all standardised tests, including those assessing intelligence, should consider cultural fairness and applicability. The next section of the literature review will examine two intelligence tests commonly used (Flanagan et al., 2018; Kaufman et al., 2005a; Saklofske et al., 2019) in this assessment process: the WISC-V and KABC-II. It will start by elaborating on the concept of intelligence and the theories which underlie these two tests. Thereafter, these two intelligence tests will be reviewed.

2.4. Intelligence

According to international and South African authors, the psychological and scientific communities have yet to reach a consensus on a definition of intelligence (Goldstein, 2013; Moerdyk, 2022; Van Eeden & De Beer, 2018; Wasserman, 2018). In 1996, Neisser and colleagues addressed this issue by stating that human beings differ in their capacity to comprehend complex and abstract concepts, successfully adapt to their environments, learn from experience, reason, and problem-solve. While these differences can be substantial, a person's intellectual functioning can fluctuate depending on the situation, the area of focus, and the criteria or measures used for assessment. Definitions of intelligence aim to clarify and organise these complex ideas, yet no definition has been successful in doing so or gained universal agreement (Neisser et al., 1996). More recent publications highlight this continued lack of consensus (Croker, 2012; Goldstein, 2013; Van Eeden & De Beer, 2018). Wechsler (2003) contributes to the debate by asserting that intelligence tests, as well as the theoretical frameworks that inform their design and interpretation, are inherently situated within and influenced by specific social contexts from which they cannot be separated. Grigorenko et al. (2001), Holden and Tanenbaum (2023), Serpell (2011), Sternberg and Grigorenko (2006), and Yang and Sternberg (1997) corroborate this perspective, emphasising that conceptualisations of intelligence are subject to cultural shaping and variability. Croker (2012) and Greenfield (2020) conversely note that definitions of intelligence tend to be vague and change over time, as do the tests that are developed to assess it. This becomes evident when one explores the history of intelligence testing.

At the start of the twentieth century, the dominant theory of intelligence that guided the development of intelligence tests "emphasised a single, underlying construct of intelligence as largely responsible for an individual's performance on all mental tasks" (Wechsler, 2014b, p. 1). The concept of the *g*-factor was introduced by Spearman in 1904 (Croker, 2012; Kaufman et al., 2005b; Holden & Tanenbaum, 2023). Based on this theory and the social context in which it was originally developed, early intelligence tests focused on assessing this single construct and distinguishing individuals based on their score differences (Croker, 2012; Wechsler, 2003). Simon and Binet developed the first classic intelligence test, known as the Binet-Simon Intelligence Scale, within this framework. The Binet-Simon Intelligence Scale contained specific test instructions and clear guidelines for test administration. In later revisions the concept of mental age versus chronological age and the comparison of inter-individual's scores were introduced (Boake, 2002; Holden & Tanenbaum, 2023). This encouraged William Stern to develop the intelligence quotient (IQ). A person's IQ score was calculated by dividing their mental age by their chronological age and multiplying it by 100 (Croker, 2012; Kovacs & Pléh, 2023).

In 1916, Terman and his colleagues published the Stanford-Binet, an adapted version of the Binet-Simon Intelligence Scale, standardised for use in America. The Stanford-Binet still primarily focuses on classifying individuals with severe cognitive delays (Boake, 2002; Wechsler, 2003). Regardless, it has been adapted numerous times, with the latest version known as the Stanford-Binet Intelligence Scales - Fifth Edition, published in 2003 (Flanagan & Kaufman, 2004). As the US entered World War I, there was a need to develop a test to assess the intellectual abilities of military recruits. This led to the creation of the Army Alpha Test (Boake, 2002; Wechsler, 2003). However, this test focussed on verbal skills which posed a challenge for recruits with limited verbal abilities. A nonverbal intelligence test, the Army Beta Test, was developed in response to this challenge (Warne et al., 2019; Wechsler, 2003). As a result of ongoing criticism, several nonverbal intelligence tests were developed. This inspired Wechsler to develop the Wechsler-Bellevue Intelligence Scale (Wechsler-Bellevue) in the late 1930's, which combined both verbal and nonverbal components (Boake, 2002; Croker, 2012; Flanagan & Kaufman, 2004; Wechsler, 1981, 2003). Unlike earlier intelligence tests, that provided only an IQ score, the Wechsler-Bellevue Scale offered a more detailed analysis of intellectual abilities. The test assessed areas such as verbal comprehension, abstract reasoning, perceptual organisation, quantitative reasoning, memory, and processing speed (Niileksela & Reynolds, 2019; Wechsler, 2003). In 1942, the Wechsler-Bellevue was revised for use with the American Armed Forces and came to be known as the Army-Wechsler. With further extensions and revisions, it was then published in 1946 as the Wechsler-Bellevue Form II (Wechsler, 1981).

Following the publication of the 1946 Wechsler-Bellevue Form II, the first WISC was published in 1949. This was followed by the Wechsler Preschool and Primary Scale of Intelligence (WPPSI) in 1967. Since 1949, the WISC has undergone four revisions (1974, 1991, 2003, 2014), with the latest version, WISC-V, published in 2014. Similarly, the WPPSI has been revised three times (1989, 2002, 2012), with the fourth edition (WPPSI-IV) published in 2012. Finally, with roots dating back to the Wechsler-Bellevue Scale of 1939, the Wechsler Adult Intelligence Scale (WAIS) has undergone five revisions, culminating in the publication of the fifth version, the WAIS-5, in 2024 (Pearson Assessments US, 2024).

With the rise of the cognitive revolution, developments in several disciplines encouraged psychologists to re-examine the concept of intelligence (Amod, 2013; Bolhuis et al., 2023; Das et al., 1994; Fagan & Holland, 2007; Mandal, 2021; Miller, 2003; Vauclair & Perret, 2003). Educational reform was also encouraged by the publication of the Education for All Handicapped Children Act of 1975, which discouraged division between children based on IQ scores (Flanagan & Kaufman, 2004; Wechsler, 2003). This increasingly resulted in criticism against traditional intelligence tests which struggled to appropriately assess the complexity of human intelligence and were considered to be culturally biased and unfair. In addition, these tests failed to provide guidance to teachers in supporting children who experience learning difficulties (Amod, 2013; Amod et al., 2018; Amod & Seabi, 2013; Nisbett et al., 2012; Seabi, 2012a). This stimulated the development of intelligence tests underpinned by theories of information or cognitive processing. A shift was observed from traditional views of intelligence to viewing intellectual functioning in terms of cognitive abilities and processes (Frischkorn et al., 2022; Naglieri & Das, 1990; Naglieri et al., 1991).

Theories which view intelligence in terms of specific cognitive abilities are often psychometric or structural. Structural theories of intelligence employ psychometric testing and statistical techniques, specifically factor analysis, to identify the factors which constitute a person's intelligence. These factors relate to the cognitive abilities assessed by specific intelligence tests. While some structural and psychometric theories place great emphasis on the general intelligence factor, others consider intelligence to consist of multiple independent factors, such as Thurstone's multiple-factor theory. Others, such as Wechsler's conceptualisation of intelligence and the CHC theory, identify multiple cognitive abilities and structure these to demonstrate their relationship to a general intelligence factor (Flanagan & Dixon, 2014; McGrew, 2009, 2023; Sternberg et al., 2008; Van Eeden & De Beer, 2018; Wasserman, 2018). Although psychometric and structural theories of intelligence have been consistently applied in understanding the concept of intelligence since Spearman identified the *g*-factor in the early 1900s (Goldstein, 2013), the cognitive revolution allowed theories of information processing or cognitive psychology to take prominence. Whereas the focus of structural and psychometric theories of intelligence falls on understanding a person's intelligence in terms of their cognitive abilities, cognitive processing theories aim to define intelligence in terms of the cognitive processes that underpin intellectual functioning (Amod, 2013; Amod et al., 2018; Das, 2002; Frischkorn & Schubert, 2018; Frischkorn et al., 2022; Naglieri, 2015; Naglieri & Das, 1990; Naglieri et al., 1991; Sternberg et al., 2008). As captured by J. P. Das, cognitive processing theories, such as the PASS theory, define intelligence as "the sum total of all cognitive processes. It entails planning, coding of information and attention, as well as arousal" (Plucker & Esping, 2014, p.19). The following section will define intelligence by exploring the theoretical perspectives that underlie the development and interpretation of the intelligence tests included in this study. These include two structural and psychometric theories: Wechsler's conceptualisation of intelligence and CHC theory, and one cognitive processing theory, PASS theory, based on Luria's neuropsychological theory.

2.4.1. Wechsler's Conceptualisation of Intelligence

Wechsler's traditional definition of intelligence, which has remained unchanged since it was first applied to the Wechsler-Bellevue Scale, reads as follows:

> "Intelligence is the aggregate or global capacity of an individual to act purposefully, reason, and deal effectively with their environment. It is global because it characterises the individual's behaviour as a whole; it is an aggregate because it is composed of elements or abilities which, though not entirely

independent, are qualitatively differentiable. By measurement of these abilities, we ultimately evaluate intelligence." (Wechsler, 1939, p.3)

Although published several decades before CHC theory gained prominence, traces of it can be found in Wechsler's conceptualisation of intelligence as an amalgamation of intellectual abilities that, in combination, represent a person's general intellectual functioning (Cockcroft, 2013; Kaufman et al., 2016a). In a time where there was much debate between Spearman's general intelligence factor (*g*-factor) and Thurstone's belief in divergent forms of intelligence (Van Eeden & De Beer, 2018), Wechsler proposed that a person's intelligence is represented by a general intelligence factor which is composed of more specific cognitive abilities (Kaufman et al., 2016a). Regarding his earlier tests, the general intelligence factor was represented by the Full-Scale Intelligence Quotient (FSIQ), while two additional scales measured more specific intellectual capabilities. These were initially referred to as the Verbal Scale, also known as Verbal IQ, and the Nonverbal Performance Scale, also known as Performance IQ.

Given Wechsler's experiences in working for the American Armed Forces where the Army Alpha and Beta tests were regularly used, the initial Wechsler-Bellevue Scale, included four Verbal Scale subtests from the Army Alpha test, whereas the Army Beta test contributed two subtests to the Performance Scale (Boake, 2002; Grégoire, 2013). While Wechsler initially explained the inclusion of the two scales based on their clinical utility, his experience in working in the field of intelligence testing, and their appearance in other intelligence tests, specifically the Army Alpha and Beta tests and the Performance Ability Scale published by Cornell and Coxe in 1934 (Boake, 2002), he later found both theoretical and practical support in the research done by Alexander in 1935. Alexander was one of the first researchers to apply factor-analytic studies in investigating intelligence tests and found that the scales proposed by Wechsler were correlated, unsurprisingly, since both are said to contribute to the general intelligence factor. Moreover, Alexander also found that these scales represented a person's utilisation of different intellectual capabilities in completing specific tasks, some related to the Verbal

Scale and others related to the Performance Scale (Grégoire, 2013). These two factors persisted for decades, until after Wechsler died in 1981 (Keith et al., 2006). In what would probably have been one of his last personal publications, Wechsler (1981), however, wrote that he was aware of research indicating more than two factors comprising a person's general intellectual functioning.

Such research, published by, for instance, Cohen in the 1960s and then Kaufman in the 1970s, led to modifications and the introduction of a three-factor model in the publication of the Wechsler Intelligence Scale for Children – Third Edition (WISC-III) in 1991. In this test, the Verbal and Perceptual Organisation Scales were accompanied by a Freedom from Distractibility Scale, which included the Symbol Search subtest. However, with additional factor-analytic studies indicating that this subtest, together with Coding, loaded on a fourth factor — processing speed — support for a fourth factor was starting to emerge. As such the WISC-III manual subsequently indicated that four factors should be applied in the analysis of the results from this test, with the Symbol Search and Coding subtests loading on Processing Speed and the Arithmetic and Digit Span subtests loading on Freedom from Distractibility (Grégoire, 2013). These four factors, however, did not gain widespread popularity, with many preferring the traditional Verbal and Performance Scale structure. Furthermore, as the WISC-III still allowed for the calculation of these, there was little practical need to consider four factors.

Fundamental change in the factor structure of the Wechsler scales was consequently only introduced in the publication of the Wechsler Intelligence Scale for Children – Fourth Edition (WISC-IV) in 2003, which made the calculation of only a Performance IQ and Verbal IQ impossible, and formally introduced four indices - Verbal Comprehension, Perceptual Reasoning, Processing Speed, and Working Memory (Keith et al., 2006). These four indices, along with their accompanying subtests, all contributed to the calculation of the FSIQ and corresponded to CHC theory, which was also gaining prominence at the time. Further research and factor analytic studies (see, for instance, Weiss et al., 2013) were conducted. With the publication of the WISC-V, a five-factor model was introduced with the five factors represented in the five primary indices, Verbal Comprehension, Processing Speed, Working Memory, Fluid Reasoning, and Visual-Spatial (Wechsler, 2014b; Weiss et al., 2016). These again correlated with five of the CHC broad abilities; however, as noted by Weiss et al. (2013), the WISC-V five-factor model has been guided more by ongoing research on the part of Wechsler test development team, as well as ongoing neuropsychological and clinical research, rather than by the CHC model and its conceptualisation of intelligence. Grégoire (2013) also notes that, out of the 16 broad CHC abilities, only five are represented in the WISC-V primary indices. Furthermore, as noted earlier, the inclusion of Verbal IQ and Performance IQ in the initial Wechsler-Bellevue Scale was based on Wechsler’s clinical experience. The inclusion of new subtests and additional indices or factors in later Wechsler scales resulted from further research and further factor-analytic studies, rather than a specific and dominant theoretical model.

However, considering the initial definition of intelligence as described by Wechsler above, it is evident that it is closely aligned with CHC theory; thus, the Wechsler scales have often been interpreted from this perspective (Caemmerer et al., 2020; Kaufman et al., 2016a; Keith et al., 2006). Also, subtests from the initial Wechsler scales, which have stood the test of time, including Arithmetic, Digit Span, Comprehension, Block Design, and Information, have been grouped according to this theory. Similarly, newer or revised subtests, including Matrix Reasoning and Visual Puzzles, have been developed whilst considering the measurement of specific CHC narrow abilities (Kaufman et al., 2016a). Thus, although the Wechsler scales have often been labelled old-fashioned or lacking in theoretical grounding, evidence suggests that their development has been guided by practical experience, research, and advances in the fields of intelligence, cognitive psychology, child development, and neuropsychology, as well as by prominent theories of intelligence. Further, in the interpretation of these scales, advancements based on prominent theories of intelligence have also been observed, despite fewer changes to the internal structure of the Wechsler scales (Boake, 2002; Keith et al., 2006). CHC theory is considered one of the theories that have informed the development of the Wechsler scales and guided their interpretation; however, it cannot be labelled as the sole theoretical perspective underlying the WISC-V (Kaufman et al., 2016a; Weiss et al., 2016). To shift test interpretation from structure to function, PASS theory has, for instance, also been applied in understanding the Wechsler scales (Kaufman et al., 2016a).

2.4.2. Cattell-Horn-Carroll Theory

According to Schneider and McGrew (2018), the CHC theory of intelligence categorises cognitive abilities; however, is not a mere list. It is, rather, grounded in theory that aim to deepen an understanding of human cognition and individual differences based on cognitive abilities. Early theories, like Spearman's two-factor theory, identified a general intelligence factor (*g*) alongside specific abilities (*s*) (Kaufman et al., 2005b; Schneider & McGrew, 2018; Sternberg et al., 2008). Cattell and Horn expanded this theory by viewing crystallised (*Gc*) and fluid intelligence (*Gf*) as two distinct components (Kaufman et al., 2005b; Keith & Reynolds, 2010; Schneider & McGrew, 2018; Sternberg et al., 2008; Wechsler, 2003). *Gf* pertains to problem-solving and both inductive and deductive reasoning, influenced by biological and neurological factors. *Gc* pertains to acquired knowledge gathered through cultural, formal and informal educational experiences (Alfonso et al., 2005; Flanagan & Dixon, 2014; Kaufman et al., 2005b; Sternberg et al., 2008). Horn increased the number of cognitive abilities, which led to an extended *Gf*-*Gc* theory. Added abilities included visual processing (*Gv*), short term memory (*Gsm*), long term storage and retrieval (*Glr*), processing speed (*Gs*), auditory processing (*Ga*), reaction time and decision speed (*Gt*), quantitative reasoning (*Gq*) and broad reading-writing factors (*Grw*) (Flanagan & Dixon, 2014; Keith & Reynolds, 2010; Schneider & McGrew, 2018; Wechsler, 2003). Although Horn viewed these abilities as equal, Carroll presented a hierarchical structure incorporating three levels. This three-stratum theory (Keith & Reynolds, 2010) included Spearman's *g*-factor, broad abilities, and narrow or more specialised abilities (Flanagan & Dixon, 2014). Similarities between Carroll's theory and the one proposed by Cattell and Horn, warranted their amalgamation by McGrew into CHC theory (Caemmerer et al., 2020; Floyd et al., 2003; Kaufman et al., 2005b). CHC is a valid and inclusive theory of intelligence, underpinned by research and factor analysis (Evans et al., 2002; Floyd et al., 2003; Flanagan & Dixon, 2014). It defines intelligence as a hierarchical framework of cognitive abilities that are both multidimensional and functionally integrated (McGrew et al., 2023; Schneider & McGrew, 2018). It is multidimensional in the sense that, in addition to a general intelligence factor, presented by Stratum III, it also views intelligence in terms of 16 broad abilities on Stratum II and more than 80 narrow abilities placed on Stratum I (Floyd et al., 2003; Flanagan & Dixon, 2014). Additionally, it is functionally integrated, as the execution of cognitive tasks in real-world contexts relies on several cognitive abilities. For instance, research by Cormier et al. (2017) and Evans et al. (2002) highlighted the correlation between several CHC broad and narrow abilities and reading, while McGill (2017) emphasised the importance of a general intelligence factor in predicting reading proficiency.

Within the scope of this research study, CHC theory is deemed suitable in defining the cognitive abilities that constitute intelligence. It is further evident in the most recent publication of the WISC-V (Keith & Reynolds, 2010; Sternberg et al., 2008) and is one of the two theories that underpin the KABC-II (Kaufman et al., 2005b). Regarding the WISC-V, global cognitive ability is represented in the FSIQ, the Nonverbal Index, or the General Ability Index, while several broad, Stratum II, abilities are represented in the primary, ancillary, and complementary indices. Stratum I narrow abilities are represented in cognitive skills required by different subtests (Kaufman et al., 2016a; Wechsler, 2014b). Kaufman et al. (2005b, p.9) note that the *g*-level is not considered a "theoretical construct but as a practical one to provide a summary score." In the KABC-II, these summary scores are represented by two global indices, the Fluid-Crystallised Index (FCI) and the Mental Processing Index (MPI). In addition, five of the CHC Stratum II abilities are represented in the five KABC-II scales, whereas the subtests reflect more narrowly defined cognitive abilities (Kaufman et al., 2005b; Keith & Reynolds, 2010).

2.4.3. Luria's Neuropsychological Theory

Luria proposed that the brain contained three functional units or systems that can influence a person's performance of a specific task (Kaufman & Kaufman, 2004; Kaufman et al., 2005b; Naglieri & Das, 2005). These three units are responsible for a number of cognitive processes, including attention, information processing and storage, and executive functions such as planning (Kaufman & Kaufman, 2004; Kaufman et al., 2005b). Luria stressed that the three systems are not only integrated and interdependent, but also continuously interact with a person's acquired knowledge or crystallised intelligence (Amod, 2013; Naglieri, 1999). In conjunction with the work of others, including Broadbent in 1958 and Hunt in 1980, Luria's neuropsychological theory laid the foundation for PASS theory (Das et al., 1994).

Das (1973), Das et al. (1975, 1994), Kirby and Das (1978), Naglieri (1999), Naglieri and Das (1990, 2005) and Naglieri et al. (1989, 1991) propose that intelligence should be defined in terms of the cognitive processes that underpin intellectual functioning and task performance. PASS theory emphasise four cognitive processes, including planning, attention, successive and simultaneous processing (Naglieri, 1999). Strongly linked to a person's executive functions, planning allows for problem-solving through the selection, application and evaluation of suitable solutions (Naglieri, 1999). Attention incorporates focused attention, sustained attention, and selective attention (Das et al., 1994; Naglieri, 1999). Successive processing involves "arranging input in sequential or serial order to solve a problem, where each idea is linearly and temporally related to the preceding one" (Kaufman & Kaufman, 2004, p. 16). Finally, simultaneous processing involves the "grouping of stimuli or recognition of a common characteristic or interrelationship amongst stimuli" (Amod, 2013, p. 106). Applied to the WISC-V and KABC-II, Luria's neuropsychological theory, as encapsulated in PASS theory, is the second of the two theories that underpin the KABC-II (Kaufman et al., 2005b). Further, with the WISC-V development team incorporating more recent theories of information and cognitive processing in the design of this intelligence test, Luria’s neuropsychological theory can be applied to interpret children’s performance on specific subtests (Flanagan & Kaufman, 2004; Kaufman et al., 2005b; Wechsler, 2014b).

Earlier, it was reported that standardised intelligence tests, such as the KABC-II and WISC-V, are often used in the assessment of children with barriers to learning and development; however, their cultural applicability should be considered. When determining the cultural applicability of an intelligence test, it is crucial to assess its psychometric properties when administered to specific populations. The section that follows will provide an overview of the WISC-V and the KABC-II. Thereafter, pertinent psychometric concepts, including validity, reliability, and differential item functioning (DIF), will be explained. Research conducted to investigate the psychometric properties of the WISC-V and KABC-II, as well as studies examining the cultural applicability of these two measures, will be reviewed. The chapter will conclude with an overview of investigations into the DIF of intelligence tests and research related to the use of the Kaufman Assessment Batteries for Children and the Wechsler Intelligence Scales for Children in African and South African research studies.

2.5. Wechsler Intelligence Scale for Children – Fifth Edition

The WISC-V is an intelligence test that assesses the cognitive abilities and intellectual functioning of children from 6 years to 16 years (Wechsler, 2014c). Whilst “incorporating new research on intelligence, cognitive development, neurodevelopment, cognitive neuroscience, and processes important to learning” (Wechsler, 2014c, p. 1), the WISC-V assesses a child’s general cognitive ability, reflected in one of three global composite scores, as well as more specific broad and narrow cognitive abilities, reflected in the indices and subtests. While Kaufman et al. (2016b) assert that the FSIQ constitutes the most reliable of the composite scores, as it provides the most accurate representation of an individual’s overall cognitive ability, the Nonverbal Index and the General Ability Index can also be considered as reflective of global cognitive ability. Whereas the Nonverbal Index is often used as an indication of global cognitive ability for children who experience expressive language difficulties, the

General Ability Index provides an estimation of global cognitive ability that is less influenced by processing speed and working memory. It is consequently often used when assessing children who experience neurodevelopmental disorders that impact these skills, including ADHD and autism spectrum disorder (Kaufman et al., 2016b).

In addition to these global composite scores, the WISC-V also yields several specific composite scores, as reflected in the primary, ancillary, and complementary indices. These provide pertinent information regarding a child's functioning in specific cognitive domains, assist in assessing cognitive strengths and weaknesses, and in recommending interventions to support challenges (Wechsler, 2014b). Included are the five primary indices, Verbal Comprehension, Visual-Spatial, Fluid Reasoning, Working Memory, and Processing Speed. Added to these are the five ancillary indices, including Quantitative Reasoning, Auditory Working Memory, Nonverbal, General Ability, and Cognitive Processing Indices. The three complementary indices include Naming Speed, Symbol Translation, and Storage and Retrieval.

Contributing to these indices are 21 subtests, which have been summarised in Table 1. A child's FSIQ can be calculated through the administration of only seven subtests; however, Wechsler (2014c) recommends the administration of the 10 primary subtests for a more comprehensive assessment of a child's intellectual functioning and cognitive abilities. Together with the administration of two of the six secondary subtests, Arithmetic and Letter Number Sequencing, the administration of the 10 primary subtests allows for the calculation of all the primary and ancillary indices. The calculation of all three complementary indices necessitates the inclusion of an additional five subtests (Kaufman et al., 2016b; Schneider et al., 2017; Wechsler, 2014b; Weiss et al., 2016).

Table 1

WISC-V Indices and Subtests

Index	Subtest	Category
Verbal Comprehension Index	Similarities	Primary (FSIQ)
	Vocabulary	Primary (FSIQ)
	Information	Secondary
	Comprehension	Secondary
Visual-Spatial Index	Block Design	Primary (FSIQ)
	Visual Puzzles	Primary
Fluid Reasoning Index	Matrix Reasoning	Primary (FSIQ)
	Figure Weights	Primary (FSIQ)
	Picture Concepts	Secondary
	Arithmetic	Secondary
Working Memory Index	Digit Span	Primary (FSIQ)
	Picture Span	Primary
	Letter-Number Sequencing	Secondary
Processing Speed Index	Coding	Primary (FSIQ)
	Symbol Search	Primary
	Cancellation	Secondary
Quantitative Reasoning Index	Figure Weights	Primary
	Arithmetic	Secondary
Auditory Working Memory Index	Digit Span	Primary
	Letter Number Sequencing	Secondary
Nonverbal Index	Block Design	Primary
	Visual Puzzles	Primary
	Matrix Reasoning	Primary
	Figure Weights	Primary
	Picture Span	Primary
	Coding	Primary
General Ability Index	Similarities	Primary
	Vocabulary	Primary
	Block Design	Primary
	Matrix Reasoning	Primary
	Figure Weights	Primary
Cognitive Proficiency Index	Digit Span	Primary
	Picture Span	Primary
	Coding	Primary
	Symbol Search	Primary
Naming Speed Index	Naming Speed Literacy	Complementary
	Naming Speech Quantity	Complementary
Symbol Translation Index	Immediate Symbol Translation	Complementary
	Delayed Symbol Translation	Complementary
	Recognition Symbol Translation	Complementary
Storage and Retrieval Index	Naming Speech Index + Symbol Translation Index	

Note. Adapted from *WISC-V: Wechsler Intelligence Scale for Children – Fifth Edition, Technical and Interpretive Manual* (pp. 6-16), by D. Wechsler, 2014, Pearson Clinical Assessment. Copyright 2014 by Pearson Clinical Assessment.

2.5.1. Full-Scale Intelligence Quotient

As per Wechsler's conceptualisation of intelligence, as well as CHC theory, the WISC-V FSIQ indicates a child's general underlying cognitive ability, also referred to as general intelligence, represented by the *g*-factor. Within the WISC-V, the FSIQ is calculated based on a child's performance on seven primary subtests, being Block Design, Similarities, Matrix Reasoning, Digit Span, Coding, Vocabulary, and Figure Weights; however, as noted in Table 1, the test is made up of 10 primary subtests, six secondary subtests, and five complementary subtests. The five primary indices relate to five of the broad abilities of CHC theory, but also to Wechsler's more specific intellectual capabilities.

2.5.2. Verbal Comprehension Index

The Verbal Comprehension Index consists of four subtests, of which two are primary, Similarities and Vocabulary, and two are secondary, Comprehension and Information. The Verbal Comprehension Index helps assess a child's English language skills, specifically listening comprehension and verbal expression, as well as their general knowledge, verbal reasoning, and verbal concept formation skills (Kaufman et al., 2016c). In essence, it is an assessment of crystallised intelligence (Schneider et al., 2017). The Vocabulary subtest, comprising 29 items, assesses both receptive and expressive vocabulary, as well as the child's ability to verbally express their understanding of English words. Individual words are read aloud to the child, and they must verbally express the meaning of each word. The Similarities subtest, with its 23 items, also relies on lexical knowledge, as well as the ability to form concepts and reason on a more abstract verbal level. Here, children are required to express the similarity between two verbal concepts, for example, red and green. The Information subtest, consisting of 31 items, assesses a child's general knowledge by asking the child to respond verbally to a series of general information questions that cover a wide array of topics. The Comprehension subtest, comprising 19 items, requires more general and social knowledge to logically reason and provide answers to questions that involve everyday situations (Kaufman et al., 2016c; Schneider et al., 2017; Wechsler, 2014c; Weiss et al., 2016). Given its focus on verbal comprehension and crystallised intelligence, the Verbal Comprehension Index is the index most influenced by language and formal and informal learning experiences. The cognitive skills assessed by the Verbal Comprehension Index and its subtests are summarised in Appendix A, Table A1.

2.5.3. Visual-Spatial Index

As indicated in the name, the focus of the Visual-Spatial Index falls on visual processing, specifically visual-spatial processing, and visualisation. Other important skills include visual and nonverbal reasoning, visual attention to detail, visual-motor integration, and the ability to perceive the relationship between the parts and the whole of a geometric design. Within the Visual-Spatial Index, there are two subtests, Block Design and Visual Puzzles. Block Design, with its 13 items, primarily assesses a child's visual-spatial and visualisation skills by asking the child to use red and white coloured blocks to replicate designs presented individually on cards. Visual Puzzles, comprising 29 items, requires a child to select three pieces that, when combined, form a geometric puzzle. The puzzle is presented as a whole, and the child then needs to choose the three puzzle pieces that, together, would form the completed puzzle (Kaufman et al., 2016c; Schneider et al., 2017; Wechsler et al., 2014b; Weiss et al., 2016). The cognitive skills and processes assessed by the Visual-Spatial Index and its subtests are summarised in Appendix A, Table A2).

2.5.4. Fluid Reasoning Index

As implied by the name, the Fluid Reasoning Index focuses on assessing fluid and abstract reasoning. Also included are elements of simultaneous processing, quantitative reasoning and quantitative knowledge, inductive reasoning, and broad visual intelligence and problem-solving abilities. Although the Fluid Reasoning Index includes four subtests, only two are considered primary subtests: Matrix Reasoning and Figure Weights. Whilst Matrix Reasoning and its 32 items require the application of inductive reasoning skills to detect the underlying principle that guides a pattern, Figure Weights relies more on quantitative reasoning, with its 34 items asking the child to look at a weight missing a piece, and then choosing a piece from a row of possibilities that would balance the weight. The first of the two secondary subtests is Picture Concepts, which was not administered as part of this study. For more information, readers are referred to Kaufman et al. (2016c) and Wechsler (2014c). The second secondary subtest is Arithmetic, comprising 34 math problems (word problems) that the child must solve mentally. Presented one at a time, it not only relies heavily on quantitative knowledge and numerical reasoning skills, but also auditory working memory (Kaufman et al., 2016c; Schneider et al., 2017; Wechsler, 2014b; Weiss et al., 2016). The cognitive skills and processes assessed by the Fluid Reasoning Index and its subtests are summarised in Appendix A, Table A3.

2.5.5. Working Memory Index

The name of this primary index implies that its focus is on working memory, encompassing both auditory and visual aspects. While one primary, Digit Span, and one secondary subtest, Letter-Number Sequencing, assist in assessing auditory working memory, Picture Span assists in assessing visual working memory. Starting with the primary auditory working memory subtest, Digit Span consists of three components: Digits Forward, Digits Backwards and Digits Sequencing, each consisting of nine items. Digits Forward requires the child to repeat a row of numbers read to them and depends significantly on memory span as a component of short-term memory. Digits Backwards and Digits Sequencing rely more heavily on auditory working memory and working memory capacity, as the former requires children to listen to a row of numbers and then repeat them in reverse order. At the same time, the latter requires them to repeat the row of numbers in the correct sequential or numerical order. Letter-Number Sequencing, a secondary auditory working memory subtest consisting of 10 items, requires children to listen to a row of both numbers and letters and then repeat it, first in numerical order and then in alphabetical order. As reported, Picture Span is a visual working memory subtest, included as a primary subtest in the Working Memory Index. The child is shown a row of pictures, which become progressively longer, for 3-5 seconds, and then needs to identify the pictures shown on subsequent pages in the correct order. In total, there are 26 items in the Picture Span subtest (Kaufman & Kaufman, 2004; Kaufman et al., 2016c; Schneider et al., 2017; Wechsler, 2014b; Weiss et al., 2016). The cognitive skills and processes assessed by the Working Memory Index and its subtests are summarised in Appendix A, Table A4.

2.5.6. Processing Speed Index

The two primary and the one secondary subtest of the Processing Speed Index focusses on assessing processing speed of visually presented information and thus includes the speed and accuracy with which these visual stimuli can be processed, as well as short-term memory, visual discrimination, visual-motor integration, perceptual scanning, attention and concentration, cognitive flexibility, and rate-of-test taking. The first of the Processing Speed Index subtests is Coding, which requires children to copy relatively simple designs associated with specific numbers. Symbol Search, the second primary subtest, requires the child to scan a row of visual stimuli and mark target stimuli (Kaufman et al., 2016c; Schneider et al., 2017; Wechsler, 2014b; Weiss et al., 2016). Cancellation was not administered as part of this research study. For more information, readers are referred to Kaufman et al. (2016c) and Wechsler (2014b, 2014c). The cognitive skills assessed by the Processing Speed Index and its subtests are summarised in Appendix A, Table A5.

2.5.7. Ancillary Indices

Whilst the secondary subtests are not routinely administered as part of a psycho-educational assessment, these assist in calculating five ancillary indices. The first of these is the Quantitative Reasoning Index, which incorporates the Figure Weights and Arithmetic subtests. As such, the focus of this ancillary index is on assessing quantitative reasoning, quantitative knowledge, and maths achievement. The second of the ancillary indices is the Auditory Working Memory Index, which includes the Letter Number Sequencing and Digit Span subtests, and, as the name implies, focuses on assessing a child's auditory working memory skills. The third index is the Nonverbal Index, which includes a number of primary subtests focussed on assessing a child's nonverbal and visual reasoning and problem-solving skills. Included in the Nonverbal Index are two of the Visual-Spatial Index subtests, Block Design and Visual Puzzles, two of the Fluid Reasoning Index subtests, Matrix Reasoning and Figure Weights and then two subtests from the Working Memory and Processing Speed Indices respectively, Coding and Picture Span. With the focus falling on nonverbal reasoning, none of the subtests from the Verbal Comprehension Index are included. As stated, the General Ability Index also renders a global composite score; however, the effects of working memory and processing speed on a child's intellectual functioning is minimised through the exclusion of all the Working Memory and Processing Speed subtests. Included in the General Ability Index are thus five of the seven subtests included in the calculation of the FSIQ: Similarities, Vocabulary, Matrix Reasoning, Figure Weights, and Block Design. Finally, in contrast to the General Ability Index, the Cognitive Proficiency Index focusses on assessing processing speed and working memory and thus includes the four primary subtests from the Working Memory and the Processing Speed Indices (Kaufman et al., 2016c; Schneider et al., 2017; Wechsler, 2014b; Weiss, Saklofske, et al., 2019).

2.5.8. Complementary Indices

The Complementary Index consists of three indices, the Naming Speed Index and the Symbol Translation Index which, in combination, form the Storage and Retrieval Index. The subtests included in these indices do not contribute to the FSIQ or the primary and ancillary indices but have been included because of their importance in a child's scholastic development and progress. Starting with the Naming Speed Index, two subtests are included: Naming Speed Literacy and Naming Speed Quantity. Both of these focus on naming speed and automaticity; however, where the Naming Speed Literacy subtest focusses on identifying letters and numbers as fast as possible, the Naming Speed Quantity subtest focusses on identifying and naming quantities. The CHC broad skill assessed by these subtests is long-term storage and retrieval; described in Lurian terms as learning. However, the narrow CHC skill assessed by Naming Speed Literacy is naming facility, whereas Naming Speed Quantity assesses the narrow CHC skills of naming facility and number facility. The Naming Speed Quantity subtest also involves processing speed as another broad ability. Additional cognitive abilities assessed by the Naming Speed Literacy subtest include perceptual speed, rate-of-test taking, attention, scanning, and automaticity in visual-verbal associations. Similarly, the Naming Speed Quantity subtest also assesses attention, rate-of-test taking and scanning ability; however, it also involves visual processing and discrimination and number sense (Schneider et al., 2017; Wechsler, 2014b; Weiss et al., 2016). Like the Naming Speed Index, the subtests of the Storage and Retrieval Index also assess long-term storage and retrieval as a broad CHC skill, termed learning by Luria. As none of these subtests were included in this study, they will not be elaborated on. For additional information, readers are referred to Kaufman et al. (2016c), Schneider et al. (2017), Wechsler (2014b), and Weiss et al. (2016).

2.6. Kaufman Assessment Battery for Children – Second Edition

Published in 2004, the KABC-II is based on two theoretical models, Luria's neuropsychological model and CHC theory. As such it renders two global scores, including the MPI from a Lurian perspective and the FCI, situated in CHC theory (Kaufman & Kaufman, 2004). In the assessment of children, only one of the two global scores is measured, and must be decided at the outset. This decision is primarily based on the reason for referral; however, as noted by Kaufman et al. (2005b), there are several factors to consider in deciding whether the MPI or the FCI will be administered. Included are the child's linguistic skills and socio-cultural background, evidence of neurodevelopmental disorders such as autism spectrum disorder, specific learning disorder, ADHD, language disorder, or intellectual developmental disorder, hearing impairments, emotional or behavioural difficulties, and the potential of a child being regarded as gifted.

Developed for children from 3 to 18 years, the KABC-II assesses a child's cognitive abilities and processes on five scales. The number of subtests administered, and the scales included, depend on the age of the child, the reason for referral and the theoretical stance taken. While very young children (age 3 years), can only be assessed on a global measure of intellectual functioning, children from 4-6 years can be assessed on three or four scales, depending on the theoretical stance. Older children between the ages of 7 and 18 years can be assessed on either four or five scales. Table 2 summarised the KABC-II scales and subtests (Kaufman & Kaufman, 2004; Kaufman et al., 2005b).

Table 2

KABC-II Scales and Subtests

Scales	Subtests
Sequential Scale: Luria Term: Sequential Processing CHC Term: Short-Term Memory	Word Order Number Recall Hand Movements
Simultaneous Scale: Luria Term: Simultaneous Processing CHC Term: Visual Processing	Rover Triangles Block Counting Conceptual Thinking Face Recognition Gestalt Closure
Learning Scale: Luria Term: Learning Ability CHC Term: Long-Term Storage and Retrieval	Atlantis Atlantis Delayed Rebus Rebus Delayed
Planning Scale: Luria Term: Planning Ability CHC Term: Fluid Reasoning/Intelligence	Pattern Reasoning Story Completion
Knowledge Scale: CHC Term: Crystallised Intelligence	Riddles Verbal Knowledge Expressive Vocabulary

Note. Adapted from "Overview," by A.S. Kaufman, E.O. Lichtenberger, E. Fletzer-Jansen, and N.L. Kaufman, *Essentials of KABC-II Assessment* (pp. 14-19), by A.S. Kaufman, E.O. Lichtenberger, E. Fletzer-Jansen, and N.L. Kaufman, 2005, John Wiley & Sons, Inc. Copyright 2005 by John Wiley & Sons, Inc.

The KABC-II further allows for the calculation of a Nonverbal Scale, often used for children with significant hearing difficulties, speech and language challenges, or limited English language proficiency. The subtests included in the Nonverbal Scale are Hand Movements, Triangles, Conceptual Thinking, Face Recognition, Pattern Reasoning, Story Completion, and Block Counting (Kaufman et al., 2005b). As is the case with the Wechsler scales, the KABC-II has core and supplementary subtests. These will be elaborated on below.

2.6.1. *Fluid-Crystallised Index and Mental Processing Index*

As stated, the KABC-II renders two global scores. The first of the global scores is the FCI, which is primarily based on CHC theory and thus incorporates a measure of crystallised intelligence in terms of the Knowledge Scale. This scale is accompanied by the other four scales, all focussed on assessing different broad abilities related to CHC theory (Kaufman et al., 2005c). The second global score is the MPI, which is based on Luria's neuropsychological theory and thus excludes a measure of crystallised intelligence. However, the other four scales are all included as these provide valuable information related to a child's functioning in terms of sequential and simultaneous processing, planning, and learning (Kaufman & Kaufman, 2004; Kaufman et al., 2005b).

2.6.2. Sequential Scale

The Sequential Scale consists of three subtests, of which Number Recall is a core subtest for children between the ages of 4 – 18 years. Word Order is a core subtest for children between the ages of 3 – 18 years, and finally, Hand Movements which is a supplementary subtest. The focus of the Sequential Scale is on assessing a child's ability to organise information in a serial or successive order to solve a problem or perform a certain task. The information is linearly and temporally related. Short-term memory also plays a role (Kaufman et al., 2005b). Number Recall, consisting of 22 items, is similar to the WISC-V Digit Span Forward subtest because a row of numbers, increasing in capacity, are read to the child and they then need to repeat it in the correct order. In the Word Order subtest, the assessor names a series of objects, following which the child needs to indicate the objects mentioned, in the correct order, on a card containing silhouettes of the mentioned objects. More difficult items include an interference task where the child needs to name a number of colours between stimulus and response. In total there are 27 items, with the last 9 including the interference task (Kaufman & Kaufman, 2004).

For further information on the Hand Movement subtest, which was not included in this research study, the reader is referred to Kaufman and Kaufman (2004) and Kaufman et al. (2005b). The cognitive abilities and processes assessed by the Sequential Scale and its subtests, are summarised in Appendix A, Table A6.

2.6.3. Simultaneous Scale

From a Lurian perspective the Simultaneous Scale assesses simultaneous processing; the ability to process information by synthesising information simultaneously and then interpreting it as a whole. The focus here is on the ability to integrate and make meaning of the information presented. From a CHC perspective the focus falls on visual processing, specifically visual-spatial processing, and visualisation. Other important skills included are nonverbal reasoning, visual attention to detail, visual-motor integration, and the ability to see the relationship between the parts and the whole of a geometric design (Drozdick et al., 2018; Kaufman & Kaufman, 2004; Kaufman et al., 2005b; Shneider et al., 2017).

The Simultaneous Scale contains a number of subtests, with Conceptual Thinking and Face Recognition generally administered to younger children and thus not included in this study. Gestalt Closure is another supplementary subtest that can be administered to all age groups; however, it was also not included in this study as it is not necessary in calculating any of the scales included in this study. For further information on these, readers are referred to Kaufman and Kaufman (2004) and Kaufman et al. (2005b). Triangles are administered to children between the ages of 3 – 12 years, Block Counting for children between the ages of 13 – 18 years, and Rover for all the KABC-II age groups. Triangles and its 27 items contain easier and more difficult items. The easier items require children to represent a model made by the assessor with their own set of plastic shapes. More difficult items require the child to replicate a design with a number of foam triangles that are yellow on the one side and blue on the other. Block Counting includes 35 items that require the child to identify the number of blocks in a picture. These blocks differ in size from one picture to the next and the child has to understand that some of the blocks in the picture are hidden from view. Rover consists of 22 items that require children to move a toy dog to its bone in the fewest number of moves possible. The toy dog moves on a grid (checkerboard-like) and sometimes must move either over or around obstacles (Drozdick et al., 2018; Kaufman & Kaufman, 2004; Kaufman et al., 2005b). The cognitive skills and processes assessed by the Simultaneous Scale and its subtests, are summarised in Appendix A, Table A7.

2.6.4. Planning Scale

Within a Lurian framework, the focus of the Planning Scale falls on planning, decision making, problem-solving, organisation and impulse control, whereas CHC theory sees it as a measure of fluid reasoning and intelligence. This scale and its two core subtests are only administered to children from the age of 7 years onwards. The first of the two core subtests is Story Completion, where the child views a row of pictures that tell a story; however, some of the pictures are missing. The child is then required to complete the 18 items by choosing a picture or pictures that would best fit the story, from a row of stimulus cards. Pattern Reasoning is like Matrix Reasoning in that the child completes the 36 items by choosing, from an array of pictures, the one that would complete a pattern presented in a logical and linear row (Drozdick et al., 2018; Kaufman & Kaufman, 2004; Kaufman et al., 2005b). The cognitive skills and processes assessed by the Simultaneous Scale and its subtests, are summarised in Appendix A, Table A8.

2.6.5. Learning Scale

The Lurian perspective would describe the Learning Scale as the integration of cognitive abilities from all three of Luria’s functional units. To learn and retain information, attention is required (Unit 1) as well as strategies to organise, retain, and later retrieve the new information (Unit 3). In addition, the learned information can be processed either sequentially or simultaneously to derive meaning from it and as required by the KABC-II subtests, pair stimuli to be retrieved later (Unit 2). From a CHC perspective, this scale requires long-term storage and retrieval, with associative memory being the narrow skills (Drozdick et al., 2018). The Atlantis subtest and its 54 items is considered a core subtest for ages 3 – 18 years. The child is tasked with associating nonsense names to pictures of shells, fish, and plants. They must recall these names and point to the correct visual stimulus when asked later. Rebus, and its 21 items, is also a core subtest between the ages of 4 – 18 years. Like Atlantis, it requires the child to learn, retain, and retrieve the associations between words and visual stimuli or drawings to read phrases and sentences (Drozdick et al., 2018; Kaufman & Kaufman, 2004; Kaufman et al., 2005b).Two additional subtests that can be added and administered approximately 15-25 minutes after Atlantis and Rebus respectively, is Atlantis Delayed and Rebus Delayed; however, these were not administered as they do not directly contribute to the calculation of the Learning Scale. For further information on these, readers are referred to Kaufman and Kaufman (2004) and Kaufman et al. (2005b). The cognitive skills and processes assessed by the Simultaneous Scale and its subtests, are summarised in Appendix A, Table A9.

2.6.6. Knowledge Scale

As stated, the Knowledge Scale is not included in calculating the MPI; however, it plays an important role in the calculation of the FCI from a CHC perspective. Like the WISC-V Verbal Comprehension Index, the Knowledge Scale assesses skills and knowledge related to crystallised intelligence, including language development, verbal comprehension, lexical knowledge, and general information or knowledge. This scale includes three subtests that are considered core subtests across the different age groups. The first is Expressive Vocabulary, a core subtest for children between the ages of 3 – 6 years. Given its relevance to this age group, it was not included in this research study. For further information, readers are referred to Kaufman and Kaufman (2004) and Kaufman et al. (2005b).

Verbal Knowledge, the longest subtest, contains 90 items and is considered a core subtest for children between the ages of 7 – 18 years. The child is presented with a vocabulary word or a question and must choose from a selection of six pictures, the picture that would either define the word best or answer the question posed. Finally, the Riddles subtest, consisting of 51 items and a core subtest for children between the ages of 3 – 18 years, requires the child to identify and name the concrete or abstract concept associated with the several characteristics read by the assessor (Drozdick et al., 2018; Kaufman & Kaufman, 2004; Kaufman et al., 2005b). The cognitive skills assessed by the Knowledge Scale and its subtests, are summarised in Appendix A, Table 10.

2.7. Psychometric Properties of Intelligence Tests

There are two factors, related to the psychometric properties of an assessment measure, which need to be considered in the use of intelligence tests (Van Eeden & De Beer, 2018). The first relates to the reliability and validity of a measure when administered to people in a specific socio-cultural context. The second relates to the cultural applicability of a particular measure. With reference to the latter, De Kock and Foxcroft (2018) draw attention to the concepts of bias and equivalence. Equivalence refers to the fact that when a specific construct is assessed, for example fluid reasoning, and two equally capable examinees, regardless of their socio-cultural backgrounds, are assessed, they should obtain comparable results. However, if there is disparity in their results, bias is present and the assessment measure and its items, can be considered non-equivalent.

Van de Vijver (2002), supported by De Kock and Foxcroft (2018), discuss different types of bias to consider in understanding the cultural applicability of intelligence tests. Included in these are construct bias, method bias, and item bias, also referred to as DIF. Construct bias occurs when an assessment measure does not assess the same constructs equally across people from different socio-cultural contexts. Within the context of intelligence tests, these constructs refer to the cognitive abilities and processes assessed. Method bias refers to the bias that is introduced in the administration of a particular assessment measure within a specific context. For instance, bias can be introduced when an intelligence test is administered in English to an individual who speaks English as an additional language and subsequently may not fully grasp the meaning of the instructions and/or the test items because of a language barrier (De Kock & Foxcroft, 2018). Finally, item bias, which Taylor (2009) considers a form of method bias, arises when people from diverse socio-cultural contexts respond to items on an assessment measure contrarily because of the items in a measure, a potential familiarity with the item content, the way in which items have been phrased and constructed and the potential irrelevance of certain items for people in specific socio-cultural contexts (De Kock & Foxcroft, 2018; Meiring et al., 2005; Van de Vijver, 2002). De Kock and Foxcroft (2018) for instance mention that some socio-cultural groups value speed, while others place more emphasis on thoughtfulness and careful consideration of the response given. This can impact an individual's results on for instance subtests of the Processing Speed Index that are timed. When these factors are inherent to some items, certain people might be advantaged whilst others might be disadvantaged; thus, leading to bias.

Whether etic intelligence tests are adopted for use in South Africa, or whether emic measures are developed, DIF, as well as validity and reliability are important psychometric properties to investigate in ensuring that the inferences made, and conclusions drawn from an intelligence test are well-founded (De Beer, 2004; De Kock & Foxcroft, 2018). Boone (2016) stipulates that there are a variety of statistical techniques that can be employed in examining said psychometric properties. Amongst those regularly employed, are the calculation of reliability coefficients and factor analyses, which provide insight into the validity and reliability of a specific assessment measure (Boone, 2016). However, Boone (2016), De Beer (2004), and Van Eeden and de Beer (2018) also recommend item analysis; a statical technique aimed at investigating DIF. The following section will extrapolate on these thoughts by first exploring the concepts of validity, reliability and DIF in more depth. Following this, research related to the psychometric properties of the WISC-V and the KABC-II will be explored, as well as research investigating the cultural applicability of these measures. The chapter concludes with an exploration of African and South African research involving the Kaufman batteries and Wechsler scales.

2.7.1. Reliability

Reliability refers to the *consistency* with which an assessment measure, such as an intelligence test, evaluates what it is designed to assess across time and repeated administrations (Aldridge et al., 2017; Revelle & Condon, 2019; Roodt & De Kock, 2018a; Vetter & Cubbin, 2019). Consistency is a key concept in understanding reliability, because reliability pertains to the ability of an intelligence test to measure its intended constructs consistently across time and repeated administrations. The American Educational Research Association (AERA) et al. (2014) and Revelle and Condon (2019) refer to this as precision, which is related to consistency in test scores across repeated administrations of the intelligence test, irrespective of the method used to measure consistency. To further understand reliability, it is essential to consider the role of measurement error. Furr (2021), Roodt and De Kock (2018a), Vetter and Cubbin (2019), Wechsler (2014b), and Westen and Rosenthal (2003) observe that most assessment tools contain some degree of error. Consider for instance the number of factors that can impact a person's performance on an intelligence test. Wechsler, for example, acknowledged these across his career and termed them "non-intellective factors of intelligence" (Wechsler, 1943, p. 103), which included incentive, drive, goal-orientation, temperament, self-regulation, motivation, and persistence (Wechsler, 1939, 1950, 1975, 1981). These factors can introduce errors and subsequently affect the consistency with which an intelligence test is able to measure its intended constructs (AERA et al., 2014; Roodt & De Kock, 2018a).

It is important to be cognisant that some of these errors might be random and subsequently occur by change, for example, fatigue experienced by the individual on the day of the administration. Other errors are more systematic in nature and thus lead to more constant variations in test scores (Roodt & De Kock, 2018b). For instance, Shuttleworth-Edwards (2016) refers to the potential impact of test-wiseness, which can continually lead to differences in test scores among people from different social-cultural backgrounds. While some people may be more familiar with test-taking procedures and expectations because of social-cultural factors such as quality of education, others may be disadvantaged because of a lack of exposure to standardised assessments. Similarly, other socio-cultural factors, such as the language in which the instructions are presented, or the familiarity of the individual with the constructs measured, may also lead to systematic errors (Roodt & De Kock, 2018b; Shuttleworth-Edwards, 2016). Regardless, as a result of measurement error, the score a person obtains on an intelligence test is a theoretical estimate rather than a precise reflection of innate intelligence (Shuttleworth-Edwards, 2016). Reliability is assessed through a reliability coefficient, which quantifies the consistency of the test. This coefficient essentially represents the correlation between a person's true intellectual ability and their measured test score (Revelle & Condon, 2019; Sattler, 2014). Reliable psychological assessment measures show a small amount of measurement error; thus, there is little variance between the individual's true intellectual ability and their measured score.

Various methods can be employed to investigate the reliability of an intelligence test (Furr, 2021; Revelle & Condon, 2019; Roodt & De Kock, 2018a; Sattler, 2014; Vetter & Cubbin, 2019), including internal consistency reliability. Internal consistency reliability refers to the consistency with which the items of a subtest assess the constructs it was designed to measure by investigating whether these items are correlated and contribute to the subtest's score (AERA et al., 2014; Kaufman & Kaufman, 2004; Murphy & Davidshofer, 2005; Roodt & De Kock, 2018a; Sattler, 2014; Taber, 2018; Wechsler, 2014b). Split-half reliability, which is a form of internal consistency reliability, is often recommended for investigating the internal consistency reliability of intelligence tests as the items of a subtest not only increase in difficulty, but start and discontinuation criteria further ensure that not all test takers complete all the items of a subtest (Kaufman & Kaufman, 2004). This is done by splitting subtests into odd and even numbers, computing the overall scores for each of these halves, and then correlating these (Murphy & Davidshofer, 2005; Revelle & Condon, 2019; Wechsler, 2014b). However, as split-half reliability is considered an underestimation of reliability as the coefficient is only based on half the subtest (Roodt & De Kock, 2018a), the Spearman-Brown formula is applied to address this concern and provide an estimate of the full subtest's reliability (Gravetter & Forzano, 2018). As this method was employed in investigating the reliability of subtests in the standardisation of the WISC-V (Wechsler, 2014b) and the KABC-II (Kaufman & Kaufman, 2004), it was also included in this study. Reliability coefficients are typically reported with the letter *r* to indicate the strength of the relationship between the measured and true scores (Sattler, 2014). Based on information obtained from George and Mallery (2003), Murphy and Davidshofer (2005) and Sattler (2014), Table 3 provides guidelines on interpreting reliability coefficients, as well as levels of reliability coefficients typically reported for certain assessment measures.

Table 3

Interpretation of Reliability Coefficients/Reliability Coefficients Typically Reported for Different Assessment Measures

<table>
<tr><th>Reliability Estimate</th><th>Type of Assessment Measure</th><th>Interpretation of Reliability</th></tr>
<tr><td>.95
.90</td><td>Standardised Intelligence Tests</td><td>High or Excellent Reliability/Measurement Errors has Virtually No Effect</td></tr>
<tr><td>.85
.80</td><td>Standardised Tests of Achievement</td><td>Moderately High or Good Reliability</td></tr>
<tr><td>.75
.70</td><td rowspan="2">Classroom Multiple-Choice Tests and Rating Scales</td><td>Moderate or Fair Reliability</td></tr>
<tr><td>.65</td><td rowspan="2">Low or Poor Reliability</td></tr>
<tr><td>.60</td><td rowspan="2">Some Projective Measures</td></tr>
<tr><td>< .59</td><td>Very Low or Very Poor Reliability</td></tr>
</table>

Note. Adapted from *SPSS for Windows step by step: A simple guide and reference. 11.0 update* (4th ed., p. 231), by D. George and P. Mallery, 2003, Allyn & Bacon Publishers. Copyright 2003 by Allyn & Bacon Publishers. *Psychological Testing: Principles and Applications* (6th int. ed., p. 150), by K.R. Murphy and C.O. Davidshofer, 2005, Pearson Education International. Copyright 2005 by Pearson Education International. *Foundations of Behavioural, Social and Clinical Assessment of Children* (6th ed., p. 65), by J.M. Sattler, 2014, Jerome M Sattler, Publisher, Inc. Copyright 2014 by Jerome M Sattler, Publisher, Inc.

2.7.2. Validity

Whilst the AERA et al. (2014) and Wechsler (2014b) consider validity as the most important factor to consider in the development of psychological assessment measures, Sattler (2014) adds that it is the validity of an assessment measure that substantiates the conclusions drawn from test results.

Validity is conceptualised as the ability of an assessment to *accurately* measure what it is designed to assess (Furr, 2021; Roodt & De Kock, 2018c; Sattler, 2014; Truijens et al., 2019; Vetter & Cubbin, 2019; Wechsler, 2014b; Westen & Rosenthal, 2003). Sattler (2014), Vetter and Cubbin (2019) and Wechsler (2014b) mention three forms of validity, including content validity, construct validity, and criterion-related validity. These authors further add that the different forms of validity can be referred to as validation procedures or facets of validity seen on a continuum. Roodt and De Kock (2018c) and Sattler (2014) also caution that validity is not a specific property and does not have a universally accepted definition in the same manner that reliability has. Rather, psychological assessment measures can be considered more, or less, valid for various reasons. This is because validity is dependent on the assessment measure and its ability to assess appropriately, and sufficiently what it is designed to assess, but also on the person being assessed. Within the context of this study, the KABC-II and WISC-V might thus be considered valid when administered to a group of participants within a specific socio-cultural context; however, if the purpose of the assessment is to investigate the intellectual functioning of a group of participants from a different socio-cultural context, it may no longer proof to be sufficiently valid. It is for this reason that Sattler (2014) advises practitioners that, while tests developers investigate the validity of a specific assessment measure during the development of it, it is the responsibility of practitioners to understand the concept, to ensure that assessment measures are administered in such a manner that validity is not compromised and, to interpret and use test results in an acceptable and ethical manner. Furthermore, it is important that continuous research be conducted on the validity of published measures when applied in particular socio-cultural contexts and to specific samples within that context.

While the concepts content validity, construct validity, and criterion-related validity are evident in literature (Sattler, 2014; Truijens et al., 2019; Vetter & Cubbin, 2019; Wechsler, 2014b), the AERA et al. (2014), Anunciacao (2018), and Colliver et al. (2012) also support the view that different kinds of evidence are required to support the degree to which an assessment measure can be considered more, or less valid. As such, this research study, investigated evidence of the validity of the KABC-II and the WISC-V in a sample of South African children who experience barriers to learning and development through investigations into the internal structures of these intelligence tests, also referred to as construct validity.

This type of investigation is often conducted through intercorrelation studies and factor analysis (Anunciacao, 2018). AERA et al. (2014) clarify that an investigation into the internal structure of an assessment measure analyses the intercorrelations between test components to determine the degree to which the results of the assessment can be used in making valid deductions about a participant's performance in relation to specific constructs. With reference to the intelligence tests used in this study, these test components may include test items, subtests, and indices/scales, all of which assess specific constructs related to a child's intellectual functioning. Through statistical analysis of the intercorrelations between these subtests, indices, and global scores, researchers can gain insight into the ability of an intelligence test to measure the constructs it was designed to assess.

In investigating the internal structure of an intelligence test, Furr (2021), Roodt and De Kock (2018c) and Sattler (2014) distinguish between convergent and divergent validity, also referred to as discriminant validity. Whilst convergent and discriminant validity can be investigated to establish correlations between measures focused on the same constructs, for instance, the KABC-II and the WISC-V, Wechsler (2014b) notes that these concepts can also be applied to subtests that measure specific cognitive abilities or processes. For instance, convergent validity looks at the extent to which subtests focussed on specific cognitive abilities (for example visual-spatial processing assessed by Block Design and Visual Puzzles), correlate with each other, as well as the extent to which subtests within a specific index/scale correlate with that index/scale (for example Block Design and Visual Puzzles with the Visual-Spatial Index). Discriminant validity on the other hand will look at the extent to which subtests that measure different cognitive abilities and processes, show weaker correlations and, the extent to which the subtests within an index (for example Similarities and Vocabulary with the Verbal Comprehension Index) show weaker correlations with another index (for example the Visual-Spatial Index).

In investigating these correlations between subtests, between indices, between subtests and indices and between subtests, indices and global scores, Wechsler (2014b) draws on a classic article published by Campbell and Fiske (1959). In their article, Campbell and Fiske presented a multitrait-multimethod matrix that can be used to study the said intercorrelations and, as such, provide evidence of convergent, as well as discriminant validity (Campbell & Fiske, 1959). In essence the aim of the multitrait-multimethod matrix is on providing a framework where discriminant and convergent validity can be established by investigating various traits through various methods. In doing so, researchers and practitioners can establish whether score variances is because of the assessment methods employed, or the result of participants' capabilities in relation to the traits being assessed. In the context of an intelligence test, the traits investigated in the multitrait-multimethod matrix, are represented by the different cognitive abilities and processes that need to be assessed, whilst the methods are represented by the different subtests within specific indices/scales that are used to assess the traits through different means (Campbell & Fiske, 1959). By relying on this method in investigating the validity of the WISC-V, Wechsler (2014b) writes that through a multitrait-multimethod matrix, the intercorrelation patterns between subtests and index or global scores can be investigated "where relatively high correlations (i.e., convergent validity) are predicted for some pairs of variables and relatively low correlations (i.e., discriminant validity) are predicted for other pairs" (Wechsler, 2014b, p. 72). If the data obtained from the normative sample supports the intercorrelation-hypotheses made at the outset, the assessment measure would be considered more valid.

Thus, through intercorrelation studies based on a multitrait-multimethod matrix, correlations between subtests, subtests and indices/scales, different indices/scales, and subtests, indices/scales, and global scores, can provide evidence to support a measure's validity as well as its ability to measure the constructs it set out to measure. Hypotheses about these interrelationships are generally made on theoretical and empirical knowledge and when these hypotheses can be accepted based on data obtained from intercorrelation studies, the assessment is considered more valid, and the results can for instance provide guidance on a participant's functioning in relation to specific constructs (AERA et al., 2014; Wechsler, 2014b).

Once convergent and discriminant validity has been investigated, factorial validity can be employed to further analyse and explain the noted correlations (Anunciacao, 2018; Roodt & De Kock, 2018c). Factorial validity is determined through factor analysis, which aims to investigate and explain the relationships between specific latent variables or latent factors and measured variables that can be directly observed. According to Anunciacao (2018), Byrne (2005), MacCallum (2012), and Schreiber et al. (2006), factor analysis rests on the premise that some variables assessed by, for instance, intelligence tests, cannot be directly observed. These are considered latent variables or latent factors. Although these latent factors cannot be directly observed, information about them can be gathered by noting their relationship with observable or measured variables. For instance, although fluid reasoning cannot be directly observed, the impact of a child's fluid reasoning skills can be observed in their performance on the Matrix Reasoning subtest and the score they obtain. In this example, fluid reasoning would be considered a latent factor, whereas the Matrix Reasoning subtest and the score the child obtains would be considered measured variables. Through factor analysis, the theoretical relationship between latent factors and measured variables can be investigated.

Anunciacao (2018) and Byrne (2005) further develops the explanation of factor analysis by distinguishing between two types: exploratory factor analysis (EFA) and confirmatory factor analysis (CFA). The latter, which was used to analyse the results of this study, is deductive, theory-driven, and based on empirical research (Byrne, 2005; Wechsler, 2014b; Westen & Rosenthal, 2003). It focuses on confirming or investigating a particular prespecified or hypothesised factor model, which can be theoretically labelled and understood. Theories of intelligence thus play a critical role in understanding and confirming a particular factor model, for without it, the factors cannot be fittingly labelled, nor can their relationship with measured variables be appropriately explained (Byrne, 2005).

In intelligence tests, CFA enables researchers to examine the relationships between subtests, indices, or scales and overarching factors, such as global scores. Typically, subtests assess specific cognitive abilities, which are organised into indices or scales (first-order factors). Global scores represent a higher-order factor. Importantly, CFA enables researchers to examine whether the factor structure holds across different socio-cultural contexts, thereby analysing whether the factor model applies similarly in diverse populations (Kaufman et al., 2005b; Roodt & De Kock, 2018c). Kaufman and Kaufman (2004) further emphasise that CFA provides insight into the strength of correlations between latent factors and their observed variables, using statistical measures of fit. According to Wechsler (2014b), goodness-of-fit in CFA is commonly assessed through the chi-square test, which evaluates if the variance-covariance matrix of the hypothesised model aligns with the observed data matrix. However, in smaller samples such as the one in the present study, goodness-of-fit metrics, such as the Comparative Fit Index (CFI) and Root Mean Square Error of Approximation (RMSEA), are commonly applied (Chen, 2007).

2.7.3. Item Response Theory and Differential Item Functioning

Thus far, this thesis has reviewed statistical techniques situated solely within Classical Test Theory (Taylor, 2009; Van Zyl & Taylor, 2012). While these techniques remain valuable in investigating the validity and reliability of psychological assessment measures, recent years have seen a rise in the use of item analysis, including IRT-based item analysis (Akinboboye et al., 2024; Nguyen et al., 2014; Moses, 2017; Sellbom & Tellegen, 2019; Thomas, 2019). In this study, item analysis was added to the other psychometric investigations discussed to specifically investigate whether the items of the WISC-V and

KABC-II subtests exhibited DIF regarding the participants' English language proficiency. When item bias is present, the results obtained from psychological assessment measures administered to people developing and living in specific socio-cultural contexts, cannot be considered accurate and a true reflection of an examinee's abilities or traits (De Beer, 2004).

Situated within IRT, item analysis can be used to investigate item bias and, where needed adjust or eliminate specific test items (An & Yung, 2014). Many authors, including Boone (2016), De Beer (2004) and Foxcroft and De Kock (2018) agree and recommend item analysis as a statistical technique that can assist in examining test items and as such, the potential bias of psychological assessment measure when administered to a specific sample, such as the one in this study. In her article on item and bias analysis, De Beer (2004) notes that the concept of item bias has largely been replaced by DIF, which can be investigated through item analysis. DIF refers to the way that the items of a particular assessment measure create differentiation or variance in the performance of examinees because of socio-cultural factors, such as English language proficiency, and not ability (De Beer, 2004; De Kock & Foxcroft, 2018; International Test Commission, 2017; Lai et al., 2005; Taylor, 2009). For instance, the language used to frame test items may be interpreted differently because of contextual and cultural differences; thus, creating bias when for instance an intelligence test is administered to children growing up and living in a specific socio-cultural context (Taylor, 2009).

Considering an examination into DIF the most appropriate method to analyse variability in test performance, De Beer (2004) adds that it is a crucial step in the design of an assessment measure. Ariffin et al. (2010) and Toland (2014) agree and add that item analysis assists in developing assessment measures and in examining their cross-cultural applicability for use in a specific context, because the statistical analyses conducted happens not only at the level of the overall scale, but also at item level. Moreover, as reported by Boone (2016), Glas (2008) and Lai et al. (2005), item analysis can also assist in monitoring and improving the quality and precision of an existing assessment measure, and in highlighting potentially challenging and biased test items when an assessment measure is administered to people in a socio-cultural context different from that of the normative sample and the context in which the measure was developed and standardised.

The origins of IRT can be found in the work of several psychometricians between the 1950s and 1970s. The work of Lord (1953) and other psychometricians of the era, including Rasch (1960), Birnbaum (1968), Samejima (1969), and Bock (1972) (Boone, 2016; Embretson, 2004; Glas, 2008; Matteucci & Stracqualursi, 2006; Zanon et al., 2016) advanced into different IRT approaches, also referred to as IRT models. The model used in this study is referred to as the Graded Response Model (GRM). The GRM was deemed appropriate for this study as not all the WISC-V and KABC-II subtests' items are dichotomous in nature (An & Yung, 2014). Some subtests, including Block Design, Similarities, Digit Span, Vocabulary, Picture Span, Letter-Number Sequencing, Comprehension, Atlantis, Story Completion, Rover, Rebus, Triangles, Block Counting, Word Order, and Pattern Reasoning allow examinees to obtain scores between 1 –7 (depending on the subtest). Although the small sample size is acknowledged, especially given the more complex nature of the GRM in comparison to for instance a Rasch Model, Nguyen et al. (2014) state that a small sample can be used in a preliminary study such as this.

The purpose of the GRM is to investigate specific test items by mathematically predicting and estimating the relationship between the latent constructs that subtests and their items are designed to measure, and the abilities of the participants about these latent constructs (Matteucci & Stracqualursi, 2006; Nguyen et al., 2014; Taylor, 2009). This prediction is based on the premise that the probability of providing a correct response is higher for participants with higher ability levels in the latent construct being measured than for those who are less proficient in the same construct. Thus, the more proficient a participant is in the construct measured, the higher the probability of providing a correct response (Reid et al., 2007). When this probability of providing a correct response is plotted against ability levels, the resulting probability curve generally takes the shape of an *S* (De Beer, 2004). In the interpretation of the

GRM, it is important to understand that whilst some IRT models, such as the Rasch model, only include one curve per item that is referred to as an Item Characteristic Curve (ICC), others such as the GRM, can include multiple curves per item depending on the nature of the item and the number of response options available (Ariffin et al., 2010; Lai et al., 2005; Van Zyl & Taylor, 2012). These are referred to as Categorical Response Curves (CRCs).

With a child's score on a subtest reflecting their underlying ability about the latent construct measured, any variance in performance should be ascribed to differences in ability levels. However, given that the probability curves represented by specific IRT models are based on the stated assumption, rather than on data obtained from a specific sample, one would expect to see the same curve regardless of the sample under investigation. Thus, if variance is observed between the estimated model and the data obtained in a particular sample, one needs to further investigate the characteristics of the subtests' items for bias (De Beer, 2004; De Kock & Foxcroft, 2018; Lai et al., 2005; Reise & Waller, 2009; Taylor, 2009). While techniques situated within Classical Test Theory express the reliability and validity of an assessment measure in single numbers, such as a Spearman-Brown split-half reliability score, IRT aims to measure model fit, or the extent to which the estimated relationship reflected by the IRT model aligns with the actual data obtained. When a misfit is observed, it is essential to be aware that the specific characteristics of certain items may have impacted model fit (Cody & Smith, 2014).

In item analysis, it is also important to consider that all items have specific parameters. As an extension of a two-parameter logistic model (2PLM), the GRM focuses on two specific parameters, discrimination and difficulty (Krabbe, 2017). The first parameter, or *a*-value, takes into consideration the discriminatory ability of items. Consequently, the *a*-value indicates the precision with which an item at a particular level of difficulty can distinguish the ability levels of examinees about the latent constructs measured by specific subtests (Cody & Smith, 2014; SAS Institute, 2018). Thus, higher *a*-values indicate that subtest items are more sensitive in discriminating between different participants based on their ability levels related to the latent constructs measured, whilst lower *a*-values indicate less sensitivity in discriminating between participants based on their ability levels related to the latent constructs measured (Matteucci & Stracqualursi, 2006). As the purpose of an intelligence test is to provide information on children's intellectual functioning about cognitive abilities and processes, subtest items should ideally be able to distinguish between children who have higher ability levels and those who are less proficient in the cognitive abilities or processes assessed. When they are unable to do so, they are less able to provide the assessor with information on a specific child's ability levels and cognitive strengths and weaknesses (SAS Institute, 2018). The second parameter, or *b*-value, represents the difficulty level of items. In a GRM, these are referred to as thresholds, described by Cody and Smith (2014) as the point at which the probability of a participant providing a correct response to an item is 0.50 (50%). As stated, the items within the intelligence tests used in this study increase in difficulty across the subtests, and it would be expected for the *b*-values or thresholds to reflect this.

While some researchers in the humanities and education employ IRT to develop, examine, and, where necessary, adjust existing assessment measures, many continue to use techniques grounded in Classical Test Theory (Akinboboye et al., 2024; Boone, 2006; Moses, 2017; Sellbom & Tellegen, 2019; Thomas, 2019). Taylor (2009) argues that it is beneficial to employ statistical techniques that are grounded in both Classical Test Theory and IRT. As such, this study investigated the reliability and validity of the KABC-II and WISC-V in a sample of South African children who experience barriers to learning and development, using the Spearman-Brown split-half reliability method, intercorrelation investigations, and factor analyses. In addition, it investigated DIF between the two language groups, English as a First Language (EFL) and English as an Additional Language (EAL), through item analysis as an additional measure to examine the cross-cultural applicability of the intelligence tests administered. The next section of the literature review will examine research conducted to investigate the psychometric properties of the WISC-V and the KABC-II, with a focus on studies that employed similar statistical analyses to those used in this research study.

2.7.4. Psychometric Properties of the WISC-V

2.7.4.1. Reliability of the WISC-V.

During the standardisation of the WISC-V, investigations into the internal consistency reliability of the subtests were conducted through split-half comparisons and the Spearman-Brown method (Wechsler, 2014b). As not all subtests contain individual items, these analyses could not be conducted for Coding, Symbol Search, Cancellation, Naming Speed (Literacy and Quantity), Immediate Symbol Translation, and Delayed Symbol Translation. In these instances, test-retest coefficients were used to determine reliability. Before discussing the results of these analyses, it is essential to be aware that the WISC-V was standardised on data collected from 2,200 English-speaking children between the ages of 6 years and 16 years, 11 months. The normative sample included an equal number of male and female children across 11 age groups, except for the 9-year-old group, which had two more females (n = 101) than males (n = 99). Each age group included 200 participants. In line with data obtained from the 2012 US census, the sample was stratified to represent the broader US population in terms of sex, ethnicity, parent education level, and geographical region (Wechsler, 2014b). As such, it included 1,228 White participants, 312 African American participants, 458 Hispanic participants, 89 Asian participants, and 113 participants from diverse ethnic backgrounds (Kaufman et al., 2016e). Reliability coefficients for the WISC-V subtests are reported in Flanagan and Alfonso (2017), Reynolds and Hadorn (2016), and Wechsler (2014b). In the overall sample, subtest reliability coefficients ranged from good to excellent, with values ranging from .81 for Symbol Search to .94 for Figure Weights. Within age groups, estimates ranged from 0.67 for Symbol Search in the 12- and 13-year age groups, and 0.96 for Figure Weights in the 10-year-old age group. As far as the complementary indices and subtests used in this study are concerned, Wechsler (2014b) reports that the Naming Speed Literacy and Naming Speed Quantity subtests obtained good internal reliability coefficients of .86 and .83, respectively.

In addition to the normative sample, Wechsler (2014b) also discusses administering the test to smaller samples of children who experienced barriers to learning and development. Included in these were 30 children who had been diagnosed with Specific Learning Disorder in Reading (SLD-R), 22 children who had been diagnosed with Specific Learning Disorder in Reading and Written Expression (SLD-RW), 27 children who had been diagnosed with Specific Learning Disorder in Mathematics (SLD-M), and 48 children who had been diagnosed with ADHD. Within the SLD-R group, there were more females (56.7%) than males (43.34%). The majority of participants were White children (63.3%), aged between 7 and 16 years. The SLD-RW sample also included mainly White children (50%), aged between 6 and 14 years. The sample included more males (68.2%) than females (31.8%). The SLD-M sample included an equal number of males and females between the ages of 9 and 16 years. Most participants were White (46.4%). Finally, the ADHD sample included more males (62.5%) than females (37.5%), and was predominantly White children (77.1%) between the ages of 6 and 16 years.

Internal consistency reliability coefficients were also calculated for the subtests administered to these samples and are presented in Wechsler (2014b). In the SLD-R sample, internal reliability coefficients ranged from .81 for Letter-Number Sequencing to .95 for Figure Weights. In the SLD-RW sample, a low internal reliability coefficient of .67 was obtained for the Matric Reasoning subtest, while the highest internal reliability coefficient was .95 for the Similarities, Figure Weights, Arithmetic, and Letter-Number Sequencing subtests. Similar to the SLD-R sample, the majority of subtests administered to the SLD-M sample obtained good to excellent internal reliability coefficients, ranging from .86 for Similarities to .97 for Figure Weights. The only exception is Letter-Number Sequencing, which obtained an internal reliability coefficient of .78, which is fair. Finally, internal reliability coefficients calculated for subtests administered to the ADHD sample ranged from good (.89) on the Picture Concepts and Picture

Span subtests to excellent on the remaining subtests. The highest internal consistency reliability coefficient (.97) was calculated for the Figure Weights subtest.

2.7.4.2. WISC-V Intercorrelation Studies.

During the standardisation of the WISC-V, intercorrelation studies were conducted to investigate convergent and discriminant validity (Reynolds & Hadorn, 2016; Wechsler, 2014b). This included intercorrelation studies focused on subtest and index scores, based on the assumption that all subtests and primary indices measure the *g*-factor to a greater or lesser extent, and thus, all scores should at least reflect some weak to moderate correlation. In addition, it was also assumed that, as the subtests within a particular index intend to measure a particular construct, these should display numerically higher correlations than they would when compared to the subtests of other indices. While some findings will be highlighted here, readers are referred to the WISC-V Technical and Interpretive Manual (Wechsler, 2014b) for results of these intercorrelation studies.

As expected, investigations into the WISC-V primary indices and their subtests confirmed that there is at least some correlation between all the subtests and the subtests within specific indices demonstrate stronger correlations than when compared to subtests from the other indices. Results further showed moderate correlations between the Verbal Comprehension subtests and those found in the Visual-Spatial, Fluid Reasoning and Working Memory Indices. The latter was expected as the subtests in these indices have higher *g*-loadings, indicating a stronger relationship between them and the *g*-factor than subtests from the Processing Speed Index. Another interesting finding indicated that the subtests of the Verbal Comprehension Index showed stronger correlations with the subtests of the Auditory Working Memory Index than with the Picture Span subtest. However, a correlation between the Picture Span and Verbal Comprehension subtests suggested that verbal mediation may have been employed in the Picture Span subtest (Schneider et al., 2017; Wechsler, 2014b).

Correlations between the subtests of the Visual-Spatial Index and those from the Processing Speed Index ranged from low to moderate. Wechsler (2014b) suggests that this may be attributed to the fact that less emphasis was placed on speed in the Visual-Spatial subtests during the revision and standardisation of the WISC-V. Also noteworthy is that the developers found higher correlations between the Visual-Spatial subtests and two of the three Fluid Reasoning subtests, specifically Matrix Reasoning and Figure Weights, indicating that these subtests rely to some degree on visual-spatial processing. In addition, Wechsler (2014b) highlights the moderate correlations between the Fluid Reasoning and Working Memory subtests, which reiterates the relationship between working memory and fluid reasoning. Subtests of the Working Memory Index showed moderate to high correlations. As expected, given their emphasis on auditory working memory, Digit Span and Letter-Number Sequencing showed stronger correlations than each of these had with Picture Span. As far as the Processing Speed subtests are concerned, moderate correlations were found between these; however, the correlation between the Coding and Symbol Search subtests was higher than the correlations between these and the Cancellation subtest. Finally, regarding the primary indices, given their factorial complexity, Arithmetic showed moderate correlations with most of the other subtests, except for Picture Concepts and the Processing Speed subtests (Wechsler, 2014b).

Regarding the WISC-V ancillary indices, higher correlations were found between these indices and the primary indices with which they share constructs and subtests (Wechsler, 2014b). For instance, given that the Auditory Working Memory Index comprises two subtests of the Working Memory Index —Letter-Number Sequencing and Digit Span —these indices showed a high correlation. In addition, the Quantitative Reasoning Index, which comprises Figure Weights and Arithmetic, as well as two Fluid Reasoning subtests, showed a high correlation with the Fluid Reasoning Index. The Nonverbal and General Ability Indices correlated highly with the Full-Scale; however, whereas the Nonverbal Index demonstrated a stronger correlation with its nonverbal counterpart, the Visual-Spatial Index, the

General Ability Index correlated strongly with the Verbal Comprehension and Visual-Spatial Indices. Given that several of the Visual-Spatial subtests are included in the calculation of the Nonverbal Index, as are subtests from the Verbal Comprehension and Visual-Spatial Indices in determining the General Ability Index, this can be expected. Another expected finding is the high correlation between the subtests of the Processing Speed and Working Memory Indices and the Cognitive Proficiency Index, once again considering that the latter is calculated by including the subtests of the first two indices (Wechsler, 2014b). Given their inclusion in this study, it is also important to note that the Naming Speed subtests —Naming Speed Literacy and Naming Speed Quantity —showed some correlation with the Working Memory and Processing Speed subtests. Given their importance in the diagnosis of learning difficulties and their focus on automaticity, the numerically lower correlation between these and the other ability subtests that contribute more to the *g*-factor is understandable. Here, it is essential to note that the Naming Speed subtests were not included in the calculation of the *g*-factor, but rather to contribute to an understanding of children's scholastic functioning. Also, given their focus on automaticity, their lower correlations with the Processing Speed Indices, indicate that these indices indeed measure different cognitive processes (Wechsler, 2014b).

2.7.4.3. WISC-V Factor Analysis.

During the standardisation of the WISC-V, factor analyses were conducted in investigating the internal structure of the test, specifically CFA. These factor analyses were guided by recent developments in neuropsychology, information processing, psychometry, child cognitive development, and an understanding of intellectual functioning and its composition (Wechsler, 2014b; Weiss et al., 2016). Here, Weiss et al. (2016) warn that whilst there is considerable overlap between CHC theory and the Wechsler understanding of intelligence, the five-factor model supported in the WISC-V has been guided more by ongoing research on the part of the Wechsler test development team, as well as ongoing neuropsychological and clinical research, rather than by the CHC model and its conceptualisation of intelligence. However, Weiss et al. (2013) also mention that given the overlap between the Wechsler and CHC conceptualisations of intelligence, the latter is regularly used in understanding the factor model of the Wechsler scales and thus concepts acknowledged by both are used in interpreting the results of CFA. Keith et al. (2006) add that CHC theory is often utilised in understanding intelligence tests and in investigating their underlying factors as it is both comprehensive and contemporary.

Wechsler (2014b, p. 77) confirms that “the purpose of a factor analysis is to identify a factor model that is, a set of factors and the structure of their relations with one another and with subtests – that does an optimal job of accounting for the observed correlations among subtests.” In naming these factors, a certain theoretical perspective is employed and thus, a theoretical framework in interpreting factor-analytic studies is critical (Golstein, 2013). Should the results of this investigation confirm the predefined factors and the expected interrelationship amongst subtests based on a theoretical framework, the factor model can be accepted as it shows good construct validity. However, in some instances, adjustments might be necessary, but these can only be considered in a theoretical context (Wechsler, 2014b). As noted earlier, in the publication of earlier versions of the Wechsler scales, a two-factor model, as proposed by Wechsler’s conceptualisation of intelligence, was supported. In these assessment measures, the FSIQ, was supplemented by the Verbal IQ and the Performance IQ (Wechsler, 2014b; Weiss et al., 2016). Over the years, this model was extended to include four intelligence factors, as seen in the WISC-IV and the Wechsler Adult Intelligence Scale – Fourth Edition (WAIS-IV), and more recently to a five-factor model with the publication of the WPPSI-IV in 2012 and then the WISC-V in 2014. Prior to these publications, there was much debate over the clinical and practical value of the four- versus the five-factor models (Wechsler, 2014b; Weiss et al., 2016). This culminated in the publication of several papers in the *Journal of Psychoeducational Assessment* in 2013 (no. 31, issue 2).

In this special issue, the practical and theoretical value of the four- versus the five-factor model in the WAIS-IV and WISC-V was considered (Tobin, 2013). Weiss et al. (2013) specifically investigated the factor-structure of the WISC-IV by examining data from its normative sample. Results indicated support for both the CHC resultant four- and five-factor models and indicated that both could be used in analysing a child's results when assessed on this intelligence test. However, when cross-loadings were allowed, based on the premise that some subtests (for example, Matrix Reasoning, Arithmetic and Symbol Search) do not measure only one of the CHC abilities, the five-factor model was a slightly better fit. Interestingly though, these authors also investigated whether the WISC-IV measured the same constructs across normative and clinical samples by noting that if the assessment measure is used for diagnostic and clinical purposes, results can only be considered meaningful if the same constructs are measured across these samples (Weiss et al., 2013). The results again indicated that both the four- and five-factor models could be applied in understanding and interpreting the results of the clinical sample, thus providing support for the clinical utility of these factor structures. Schwartz (2013) however raised some concern with this finding, stating that the heterogeneity of the clinical sample elicits doubt about the clinical utility of the four- and five-factor models. In this regard, it is noted that the clinical sample consisted of a combination of children diagnosed with specific learning disorder, mild to moderate intellectual developmental disorder, ADHD, combined ADHD and specific learning disorder, autism spectrum disorder, motor impairment, traumatic brain injury, and language disorder (Weiss et al., 2013).

In addition to the study by Weiss et al. (2013), factor analytic studies of the WISC-IV were also conducted by Bodin et al. (2009), Chen and Zhu (2012), and Keith et al. (2006). Whereas the results of the Bodin et al. (2009) and Chen and Zhu (2012) studies showed support for the WISC-IV four-factor model, Keith et al. (2006) employed CHC theory in their investigation which showed support for a five-factor model. Based on these studies, as well as the inclusion of additional subtests to support the assessment of specific factors, a five-factor model was investigated in the standardisation of the WISC-V.

In the lead-up to the five-factor model, a series of simpler models were examined, grounded in Wechsler's conceptualization of intelligence, where the general ability factor influences a person's performance on different indices and subtests. For further information on these, readers are referred to Wechsler (2014b). Regardless, substantiating the inclusion of five primary indices in the WISC-V, all five-factor models investigated showed a significant improvement in fit compared to the four-factor models. The last of these five-factor models was used to guide the factor structure of the WISC-V, as it showed an excellent fit when the Arithmetic subtest was cross-loaded onto the verbal comprehension, fluid reasoning, and working memory factors. Once this model was accepted as valid across all age groups, a CFA was conducted, including only the 10 primary subtests. The analysis examined how well the factor model accounted for the established intercorrelations among the 10 primary subtests. Goodness-of-fit statistics indicated an excellent fit, supporting a higher-order five-factor model as an adequate explanation of the intercorrelations among the 10 primary subtests. Interestingly, a standardised path coefficient of 0.99 was observed between the second-order *g*-factor and the fluid reasoning factor (Wechsler, 2014b). These findings were replicated in a study by Chen et al. (2015), where the five-factor model was tested on a US normative sample with the explicit goal of investigating any possible sex differences. The outcome of the Chen et al. (2015) study revealed that the higher-order five-factor model proposed by Wechsler (2014b) was not only suitable for interpreting test results but also demonstrated sex invariance between female and male children. For both female and male children, the results indicated that the WISC-V primary indices and subtests measure the same underlying latent factors. In addition, the cross-loading of the Arithmetic subtest on the Fluid Reasoning Index, the Working Memory Index, and the Verbal Comprehension Index was substantiated and again found to have similar cross-loadings for male and female children. Whilst concern regarding the 1.00 standardised path coefficient between the second-order *g*-factor and the fluid reasoning factor was acknowledged, Chen et al. (2015) considered this to be evidence of the belief that fluid reasoning is critical in a person's intellectual functioning.

In addition to the work by Chen et al. (2015), Reynolds and Keith (2017) again used data from the US normative sample and also found support for the five-factor model, where five factors or cognitive abilities are represented by the five first-order factors, and the general intelligence factor by the second-order factor. It was further found that the first-order factors related in an anticipated manner to the second-order *g*-factor, and that the fluid reasoning factor had the highest g-loading, while the processing speed factor had the lowest. However, in their research, these authors employed both bifactor and higher-order models. Although support was still found for the five-factor model described in the WISC-V Technical and Interpretive Manual (Wechsler, 2014b), some minor differences were noted. First, when compared to the original WISC-V five-factor model, the research by Reynolds and Keith (2017) indicated that while there was a strong relationship between the *g*-factor and the fluid reasoning factor, it was not quite as perfect as it was in the test manual model. Second, correlations were also found between the visual-spatial and fluid reasoning factors, even after the *g*-factor had been partialled out to consider the relationship between these first-order factors. These authors explained this by indicating the possibility of a nonverbal factor which links the two. Another interesting finding in this study is the direct relationship between the Arithmetic subtest and the *g*-factor, with the subtest loading on the working memory factor to a lesser degree. When one considers the strong correlation between the *g*-factor and fluid reasoning, this finding is understandable.

In addition to the above, Reynolds and Keith (2017) also conducted EFAs, in contrast to the WISC-V test development team's focus on CFAs. Regardless, the outcome again supported the five-factor model. Unlike other EFAs, it also supported the separation of the visual-spatial and fluid reasoning factors, indicating an additional higher-order relationship between these indices. This might be attributed to an emerging perceptual reasoning factor, with the authors supporting the consideration of an additional nonverbal higher reasoning score, where subtests from the Fluid Reasoning and Visual-Spatial Indices can be combined, but not at the expense of the individual visual-spatial and fluid reasoning factors. However, despite the apparent confidence in a five-factor model, Dombrowski et al. (2015) and Canivez et al. (2016, 2017) have found the research by Wechsler (2014b) wanting. Some of the concerns raised include criticism of the WISC-V development team for not adequately describing their conceptualisation of the different latent factors. These authors also noted that the preferred five-factor model, which allows Arithmetic to cross-load on the Working Memory, Visual-Spatial, and Fluid Reasoning Indices, discards the tightness of a simple structure through this allowance. Furthermore, with a standardised path coefficient of .99 between fluid reasoning and the second-order *g*-factor, research suggests that the inclusion of both is excessive, and that the fluid reasoning factor is redundant.

Additionally, concerns were expressed about the unsubstantiated use of Weighted Least Squares estimation, as opposed to Maximum Likelihood estimation, which Dombrowski et al. (2015) and Canivez et al. (2016, 2017) considered more appropriate. Another concern raised was that other models, such as a bifactor model, were not tested alongside the higher-order model, despite support for this approach in the investigation of intelligence tests. Additionally, the WISC-V development team was criticised for relying solely on CFA at the expense of EFA, which they felt was an oversight, as EFA and CFA are considered complementary in the development of an intelligence test (Canivez & Watkins, 2016). This is considered especially true for an intelligence test such as the WISC-V, where specific subtests were omitted (Word Reasoning and Picture Completion) and others were added (Picture Span, Visual Puzzles, and Figure Weights) in the revision (Dombrowski et al., 2015; Canivez et al., 2016, 2017). In addressing these concerns, Dombrowski et al. (2015) conducted an Exploratory Bifactor Analysis with the WISC-V normative sample, allowing the *g*-factor and the latent factors to influence the subtests directly. The results of this analysis indicated that a three-factor model, incorporating perceptual reasoning, processing speed, and working memory group factors, would be more appropriate. As such, it did not highlight the need for verbal comprehension and fluid reasoning to be included as separate factors. Additionally, it was found that the four primary subtests comprising the existing Visual-Spatial and Fluid Reasoning Indices should be combined to form a single perceptual reasoning factor, rather than the two separate factors currently in use. Interestingly, many of the Verbal Comprehension Index subtests loaded negatively onto the processing speed factor. However, this negative loading was described as small and was seen as supportive of the character of the processing speed factor. Whereas the subtests included in the Processing Speed Index are all timed and thus rely on processing speed in their completion, the Verbal Comprehension Index subtests are not constrained in this manner. Thus, the small negative loading on the processing speed factor supported its evaluation as a processing speed factor (Dombrowski et al., 2015).

The most surprising finding of the Dombrowski et al. (2015) study is that verbal comprehension was not identified as a separate factor, especially given that it has been included as a distinct factor since the publication of the first Wechsler scales. These authors provide a possible explanation by mentioning that in a higher-order CFA, where the *g*-factor is considered a second-order factor, there is an indirect relationship between it and the different subtests. Consequently, subtests must first have a strong relationship with the first-order factors, which in turn must demonstrate a strong relationship with the second-order factor. If these conditions are not met, there will only be a slight relationship between any of the given subtests and the *g*-factor. As such, should the verbal subtests have strong loadings on other factors, or if the latent verbal comprehension factor was not appropriately defined, the relationship between these and the *g*-factor will be small, which can result in the outcome we see in the Exploratory Bifactor Analysis where subtests have a direct relationship with the *g*-factor. Dombrowski et al. (2015) admit, however, that this is contrary to other research, which shows a strong relationship between the *g*-factor and verbal abilities. For instance, in the standardisation of the WISC-IV and WISC-V, research indicated that some subtests loaded higher on the *g*-factor than others. Amongst these were Similarities and Vocabulary, two verbal subtests (Wechsler, 2003, 2014b). These findings were supported by Chen et al. (2015), Keith et al. (2006) and Weiss et al. (2013), who also found strong relations between the WISC-IV Vocabulary and Similarities subtests and the *g*-factor. Further, research by Reynolds and Keith (2017) discussed earlier also highlighted a distinct verbal comprehension factor. Regardless, Dombrowski et al. (2015) concluded that their findings support the WISC-V as primarily a measure of general ability, and that practitioners should take this into account in test interpretation.

Added to the above, Canivez et al. (2016) conducted an EFA with the entire WISC-V normative sample and found support for a four-factor model, but not for the present five-factor model. Similarly, Canivez et al. (2017) also found support for a four-factor model in their study where independent CFAs were conducted with Maximum Likelihood Estimation, also using the WISC-V normative sample. This was done in addition to bifactor analyses where the *g*-factor and the latent factors all directly influence the subtests. The reported CFAs (Canivez et al., 2017), as well as the EFA (Canivez et al., 2016) were conducted with the 16 primary and secondary subtests, whilst the CFAs also included the 10 primary subtests. Results from the latter however indicated similar results. Interesting, in one of the bifactor models (Canivez et al., 2017), permissible results were obtained, but here three of the Fluid Reasoning subtests did not load significantly on this factor, including Matrix Reasoning, Picture Concepts and Figure Weights. In addition, the four-factor bifactor and CFA models were similar; however, it was concluded that the four-factor bifactor model was the best representation of the WISC-V structure when all 16 subtests are considered. Fluid reasoning as a separate factor was not supported as this factor could not be disentangled from the *g*-factor (Canivez et al., 2017). Further, like the Dombrowski et al. (2015) study, Canivez et al. (2016, 2017) also found the WISC-V to be a good indicator of general ability, with processing speed the only other discernible factor found by Canivez et al. (2017). Here the EFA (Canivez et al., 2016, p. 982) further indicated that “the WISC–V *g*–factor accounted for 6 to 12 times more variance than any single group factor and twice the variance of all four group factors combined." The latter study also found fluid reasoning to be a redundant factor, with only one subtest loading on it (Figure Weights), whilst noting that the Picture Concepts subtest did not load sufficiently on any of the latent factors. Arithmetic did load, albeit weakly, on the working memory factor, where it was placed in the WISC-IV (Canivez et al., 2016).

In 2018 Canivez and colleagues published further findings from another EFA, this time conducted with specific age groups, 6-8-years, 9-11-years, 12-14-years, 15-16-years from the WISC-V normative sample. Again, a four-factor model was deemed sufficient in explaining the structure of the WISC-V across the different age-groups. Required evidence to support fluid reasoning as a fifth factor was again not found and whilst the Block Design and Visual Puzzles subtests proofed to be stronger indicators of perceptual reasoning when compared to Matrix Reasoning and Figure Weights, the findings of the Canivez et al. (2018) study did not support a separate factor for the latter subtests. Also corresponding to the findings of the Canivez et al. (2016) study, the 2018 study again found a strong relationship between most of the subtests and the *g*-factor. Except for the processing speed subtests, all other subtests were found to be stronger measurements of the *g*-factor than of their respective latent factors (Canivez et al., 2018). In addition, the study found that the Picture Concepts subtest, again did not load sufficiently on any of the latent factors in two of the four age groups (6 – 8 years and 9 – 11 years), which leads to questions regarding its inclusion as a possible substitute for Matrix Reasoning and Figure Weights on the Fluid Reasoning Index. Interesting, in the other age groups, this subtest loaded on the verbal comprehension factor; however, this is not considered consistent with the available theoretical understandings of this subtest and the verbal comprehension factor. Finally, Arithmetic again loaded on the working memory factor, but only for the youngest two age groups. In the group incorporating 12 – 14-year-olds, it loaded on the verbal comprehension factor, whereas it did not show sufficient relations to any specific factors in the oldest age group. Here, Arithmetic loaded on the verbal comprehension, working memory and perceptual reasoning factors (Canivez et al., 2018).

The importance of these studies (Dombrowski et al., 2015; Canivez et al., 2016, 2017, 2018), lies in the consistent finding that the WISC-V is primarily a measure of the general intelligence factor. Further, practitioners should take cognisance of the concerns regarding the fluid reasoning factor as it is not consistently found to be a separate factor, and the high correlation between this and the *g*-factor, makes the fluid reasoning in essence redundant. Caution should also be taken in interpreting the Arithmetic subtest as a measure of fluid reasoning. As such, researchers and practitioners alike should be very cautious in only using specific test manuals in the interpretation of test results (Canivez et al., 2018). It is also important to be cognisant that all of the reported studies, were done with data from the WISC-V normative sample. Research conducted with other samples to examine the construct validity of the WISC-V will be discussed below in considering the cross-cultural applicability of the WISC-V.

2.7.4.4. WISC-V: Socio-Cultural Considerations.

During the development of the WISC-V in the US, several steps were taken to enhance the cross-cultural applicability of this intelligence test. For example, test developers consulted researchers and practitioners across 10 counties to review the content of test items and ensure their suitability for different socio-cultural contexts (Van de Vijver et al., 2019). In addition, the potential impact of socio-cultural factors on test performance was investigated in the normative sample. Included in the socio-cultural factors investigated, were parent education level as a proxy for socio-economic status, and ethnicity. According to Kaufman et al. (2016e) and Weiss, Locke, et al. (2019), the results of these investigations indicated that parent education level strongly correlated with participants' performance on the WISC-V, more so their performance on the FSIQ, the General Ability Index, and the Verbal Comprehension Index. Conversely, children's performance on the Processing Speed and Naming Speed Indices was less influenced by parent education level, likely because these indices are less reliant on a child's verbal ability and crystallised intelligence. Interestingly, there was an approximate 20-point difference in the mean FSIQ scores of children whose parents had attained an education level of eight grade or less, compared to children whose parents had attained a bachelor's degree or higher.

In exploring the potential impact of ethnicity on WISC-V performance, the scores of White (*n* = 1,228), African American (*n* = 312), Hispanic (*n* = 458), Asian (*n* = 89), and *Other* (*n* = 113) participants of the normative sample were compared (Kaufman et al., 2016e). In understanding the results of these analysis, Weiss, Locke, et al. (2019) caution that ethnicity is a very complex construct as persons from different ethnic groups not only differ in terms of English language proficiency given their cultural and historic backgrounds, but also have more, or less access to resources, wealth, and education. Thus, when assessing the impact of ethnicity on intelligence tests, consideration should be given to numerous inter-related factors, including language differences, parent education level, quality of education, access to healthcare, and socio-economic status. Nonetheless, as the largest mean differences were observed between White and African American, and White and Hispanic participants, Kaufman et al. (2016e) and Weiss, Locke, et al. (2019) explored these further. In comparing the White and African American participants, ethnicity, together with parent education level and sex, accounted for 23% of the variance observed in the groups' FSIQ scores. However, after adjustments were made for sex and parent education level, ethnicity accounted for only 3.5% of variance in the FSIQ. Moreover, when exploring the impact of ethnicity on indices that rely less on crystallised intelligence, including Working Memory, Processing Speed, Cognitive Proficiency, and Naming Speed, almost no variance was found between African American and White participants. A similar trend was observed in comparing the scores of Hispanic and White children, where parent education level, ethnicity, and sex had the greatest impact on the children's performance on the FSIQ, General Ability Index, and Verbal Comprehension Index. After adjustments were made for sex and parent education level, ethnicity still accounted for more variance in the FSIQ than in the Nonverbal Index. Thus, ethnicity and parent education level had a greater impact on indices that rely on verbal ability, crystallised intelligence and acquired knowledge than those who are considered more nonverbal in nature (Kaufman et al., 2016e; Weiss, Locke, et al., 2019). Interestingly, in investigating differences in mean scores across 10 samples from Western countries, including Australia and New Zealand, Spain, Scandinavia, the UK, the US, French and English Canada, the Netherlands and Flanders, France and Germany, Van de Vijver et al. (2019) found small and inconsequential differences in mean scores. In addition, these differences could not be attributed to the social, economic and educational contexts of the different countries. However, cognisance should be taken that in this study, the WISC-V had been adapted for use in the different countries.

In 2023, Carvalho published further research which investigated the impact of language and socio-economic status on children's performance on the WISC-V, specifically the General Ability, Verbal Comprehension, Visual-Spatial, and Fluid Reasoning Indices. The sample included 79 African American children from lower socio-economic backgrounds. Although all the participants spoke English at home, the researcher stated that their English language proficiency would have been impacted by socio-cultural and historical factors that produce language differences, even in children who are considered to speak only English. The results of the study (Carvalho, 2023) confirmed that in comparing the performance of the sample to the WISC-V normative sample, statistically significant differences were observed. Specifically, while small and medium differences were observed between the two groups on the Visual-Spatial and Fluid Reasoning Indices respectively, a large difference was observed in the groups' scores on the General Ability Index and an exceptionally large difference in their scores on the Verbal Comprehension Index. The researcher concluded that language and socio-economic status have a significant impact on children's performance on the WISC-V, more so on indices and subtests that are more verbal in nature and rely more on crystallised intelligence and acquired knowledge.

In addition to these studies, which investigated the impact of socio-cultural factors on WISC-V test performance, the original WISC-V has also been adapted for use in various counties, including the

United Kingdom, Spain, France, Canada, New Zealand and Australia, the Netherlands and Flanders, and Scandinavia (Van de Vijver et al., 2019). In each of these countries, rigorous procedures were followed to ensure the suitability of the WISC-V for use in these countries. Included were investigations into the reliability and validity of the adapted versions of the WISC-V, with several countries employing factor analyses to examine the construct validity and theoretical structure of the WISC-V when administered to different populations. Van de Vijver et al. (2019) combined data from all these countries and examined the WISC-V five-factor model as a hierarchical structure, placing the variables represented by the five primary indices at the lowest level, and the general intelligence factor at the highest level. The five-factor structure was confirmed in all of these countries, supporting the construct validity of the WISC-V across different cultural and linguistic populations. However, following the publication of the Canadian version of the WISC-V (WISC-V^{CDN}) Watkins et al. (2018) conducted factor analyses, specifically CFA, with data obtained from 880 participants in the normative sample. According to the results of this study, the data obtained from the WISC-V^{CDN} five indices and its 16 subtests, best matched a bifactor model with four group factors (verbal comprehension, visual-spatial, processing speed and working memory) and one global factor (*g*-factor), rather than the WISC-V five-factor model. According to these authors, the WISC-V^{CDN} proposes the use of the five-factor model in its manual, however, it lacks discriminant validity. Watkins et al. (2018) further cautioned users of the WISC-V^{CDN} to exercise caution in using the group factors when interpreting children's performance and recommending interventions to address concerns, as the global factor accounted for a substantial portion of variance. Similarly, although the French WISC-V (WISC-V^{FR}) also proposes the use of a higher-order five-factor model in the interpretation of children's test scores, Lecerf and Canivez (2018) conducted EFA and CFA with data obtained from the normative sample (N = 1,049) and participants' performance on the WISC-V^{FR} five indices and its 15 subtests. The results of the study again did not find sufficient support for a five-factor model. Rather, support was also found for a bifactor model with four group factors and a dominant global factor.

Additional research by Lecerf and Canivez (2022) examining the factor structure of the WISC-V^{FR} using hierarchical EFA across five age groups from the normative sample, also did not find support for separate visual-spatial and fluid reasoning factors and saw the FSIQ accounting for the most variance in the five age groups. Similar results were found for the Spanish version of the WISC-V (Fenollar-Cortés & Watkins, 2019), the UK version (Canivez et al., 2019), and the German version (Canivez et al., 2020; Pauls & Daseking, 2021). In the Chilian version of the WISC-V, EFA (*N* = 853) conducted by Rodríguez-Cancino and Concha-Salgado (2024) evidence was also found for a four-factor model. However, these authors further conducted CFA and found that the hierarchical four- and five-factor models showed comparable fit.

In contrast to these factor analytic studies, research conducted by Wilson et al. (2023) with data from the Australian and New-Zealand normative sample (*N* = 528) and the original US sample (*N* = 2,200) found support for a first-order five-factor model, which showed excellent model fit in both samples. The results further indicated that in comparing the Australian and New Zealand version of the WISC-V (WISC-V A&NZ) with the original US WISC-V, the five indices had the same meaning across the samples. A small, but significant difference between the two samples' mean scores were found on the visual-spatial factor, however, only for females. The authors thus stressed the importance of developing country-specific normative data as these differences can affect the interpretation of test scores and recommendations made. Interestingly, a replication of this research by Wilson et al. (2024), but adding a UK sample (*N* = 415), found similar results. Again, the first-order five-factor model showed excellent fit in all three samples. Also, in this study, a small, but significant difference was once more found in comparing the mean visual-spatial scores of female participants, further highlighting the importance of country-specific research and norms. Egeland et al. (2022) also conducted CFAs with the Scandinavian version of the WISC-V, using data obtained from the assessment of 237 children referred to a Norwegian child psychiatry outpatient unit for a range of challenges. Included in these factor analyses were hierarchical, bifactor, and correlated candidate factor models. Although the manual of the Scandinavian version of the WISC-V actually proposes a four-factor model, the correlated candidate five-factor model showed the best fit. In the other analyses, all four- and five-factor bifactor and hierarchical models showed good fit, although, the four-factor bifactor models showed slightly better fit than the five-factor bifactor and hierarchical models. However, with the correlated candidate five-factor model showing superior fit, the authors recommended that clinicians use a five-factor model as represented by the five primary indices in interpreting children's test scores, especially as the WISC-V is more likely to be used in the assessment of children referred for various challenges rather than typically developing children.

Scheiber (2016a) investigated the factor structure of the WISC-V in an ethnically diverse sample (*N* = 2,637) of African American (*n* = 409), Hispanic (*n* = 621), and White (*n* = 1,607) male and female participants. Although these participants were all involved in the original standardisation of the WISC-V, the results are interesting and indicate that the five-factor structure of the US WISC-V was invariant across these ethnic groups. Although, as discussed earlier, differences in mean scores were observed between these ethnic groups (Kaufman et al., 2016e; Weiss, Locke, et al., 2019). Conversely, Graves et al. (2021) examined the factor structure of the WISC-V in a sample (*N* = 1184) of African American participants, which did not form part of the normative sample. These participants had all been referred for special education services; however, reasons for referral are not specified. Nonetheless, these authors conducted CFA and EFA, but did not find sufficient support for the WISC-V five-factor structure in this sample of African American children. Furthermore, the Picture Span and Figure Weights subtests did not show invariance, indicating that these two subtests operated differently in this African American sample when compared to the original normative sample. The authors thus stressed the importance of further research with diverse samples in exploring the construct validity of the WISC-V. Similarly, Taylor (2019) investigated the structure of the WISC-V in a sample of African American children (*N* = 473) referred for a variety of challenges, including specific learning disabilities, intellectual developmental disorder, health difficulties, emotional and behavioural challenges, speech and language impairments, and autism spectrum disorder. Using CFA, the results of the study indicated that the participants' data fitted both four- and five-factor hierarchical models, although, fit in the four-factor model was slightly more superior with goodness-of-fit indices not supporting the more complex five-factor model. Also, while variance was observed between males and females in the five-factor model, this was not the case for the four-factor model and thus, the latter was proposed as the better model. Further, with the FSIQ accounting for the majority of variance, the researcher did not find sufficient support for the interpretation of test scores at the index and subtest level.

In conclusion, Van de Vijver et al. (2019) found sufficient evidence for the construct validity of the WISC-V across several countries and observed small differences in the mean scores of the different samples from these countries. Support for the WISC-V five-factor structure was also found in Australia and New Zealand, the US and the UK (Wilson et al. 2023, 2024) and Scandinavia (Egeland et al., 2022) and across different ethnic groups by Scheiber (2016a). However, publications by several authors (Canivez et al., 2019, 2020; Carvalho, 2023; Fenollar-Cortés & Watkins, 2019; Graves et al., 2021; Kaufman et al., 2016e; Lecerf & Canivez, 2018, 2022; Pauls & Daseking, 2021; Rodríguez-Cancino & Concha-Salgado, 2024; Taylor, 2019; Watkins et al., 2018; Weiss, Locke, et al., 2019) question the construct validity of the WISC-V across countries, ethnic and language groups and found sufficient evidence for the influence of socio-cultural factors on test performance. As such, further research into the psychometric properties of the WISC-V when used with diverse samples, is indicated.

2.7.5. Psychometric Properties of the KABC-II

2.7.5.1. Reliability of the KABC-II.

Like the WISC-V, internal consistency reliability was also investigated for the KABC-II subtests through the Spearman-Brown split-half technique as outlined earlier. The KABC-II was originally standardised on a normative sample of 3,025 English-speaking children between the ages of 3 years and

18 years, 11 months (Kaufman & Kaufman, 2004; Kaufman et al., 2005b). The sample was stratified based on data from the 2001 US census according to sex, ethnicity, geographical region, and parent education level. An almost equal number of male and female children were included in the normative sample. In terms of ethnicity, the sample included 1861 White participants (61.5%), 545 Hispanic participants (18%), 465 African American participants (15.4%), and 154 participants (5.1%) who belonged to the *Other* group. Average and age-group internal consistency reliability coefficients for the KABC-II subtests are reported in Kaufman and Kaufman (2004) and Kaufman et al. (2005b). Internal consistency reliability coefficients were calculated for the core and supplementary subtests with the average subtest reliability coefficients across the age groups showing good internal consistency reliability. Specifically, in the 3 – 6 years age group, the average coefficient for the core subtests was .85 and for the 7 – 18 years age group, it was .87. The supplementary subtests showed slightly lower internal consistency reliability coefficients. Adding to this, the average subtest reliability coefficients for the 3 – 6-year age group, ranged from .69 for Hand Movements to .92 for Rebus. Other subtests in this age group, except for Hand Movements, Face Recognition and Gestalt Closure, showed reliability coefficients in the .80s. In the older age group of 7 – 18-year-olds, the mean subtest reliability coefficients ranged from .74 Gestalt Closure to .93 Rebus. Four subtests did not show reliability coefficients above .80, including Number Recall, Hand Movements, Gestalt Closure, and Story Completion.

2.7.5.2. KABC-II Intercorrelation Studies.

During the standardisation of the KABC-II, CFAs were done, as well as analysis for discriminant and convergent validity through intercorrelation studies (Kaufman & Kaufman, 2004; Kaufman et al., 2005b). With reference to the latter, Kaufman and Kaufman (2004) note that the intercorrelation of global, scale, and subtest scores were investigated for each age group, as well as combinations of age groups. These combinations were based on ages that share core subtests and as such, ages 3 and 4 years formed the first age group, ages 5 and 6 years the second, ages 7 years through 12 years the third and finally, ages 13 years through 18 years the fourth. As far as the first and second age groups are concerned, statistical analysis showed moderate intercorrelations between the scales, while the correlations between the global scores were very high (.96 for the first age group and .98 for the second age group). These high correlations were attributed to the fact that at these ages, the subtests of the MPI are all included in the FCI. However, cognisance should still be taken that despite these high correlations, the choice of one or the other global index can have a significant impact on the manner in which an individual child's profile will be interpreted (Kaufman & Kaufman, 2004).

As far as the third and fourth age groups are concerned, intercorrelation studies between pairs of scales showed correlations between .31 and .61 (Kaufman & Kaufman, 2004). The highest intercorrelation (.61) was found between the Planning and the Knowledge Scales. This is very interesting as the skills that underlie these, when considered from a CHC theoretical perspective, are often seen as two distinct forms of intelligence. Kaufman and Kaufman (2004) concluded that the intercorrelation score between these two scales shows that whilst they measure distinct constructs, they both also contribute to an underlying shared construct. Finally, with reference to the intercorrelations between the scales and global indices, the last two age groups displayed a consistent pattern of correlations. Here the correlations between the Sequential Scale and the MPI and FCI are the lowest (.60 to .70). Falling in the mid .70s, the Simultaneous and Learning Scales showed slightly higher correlations with the CFI and MPI. The highest correlations (about .80) were found between the Planning Scale and the two global scales, as well as between the Knowledge Scale and the FCI (Kaufman & Kaufman, 2004).

2.7.5.3. KABC-II Factor Analysis.

As noted, factor analyses were also conducted with the KABC-II, which Kaufman and Kaufman (2004) believe is one of the most important analyses to be conducted in exploring the validity of a particular assessment measure. Grounded in CHC theory, factor analyses conducted with the KABC-II standardisation data, investigated the least number of latent factors that could account for the noted intercorrelations. Further, it aimed to establish the extent to which each observed variable measured its associated latent factor, and finally, how the latent factors are related to one another. Whilst keeping these goals in mind, Kaufman and Kaufman (2004) confirm that they conducted CFAs in the analysis of the KABC-II data to first look at the strength of the relationships between latent factors and between latent factors and measured variables. In addition, goodness-of-fit statistics were included to confirm the fit between the prespecified factor model and the observed data. As a first step, CFA was conducted to establish whether the subtests from the Planning and Simultaneous Scales should be grouped in one Nonverbal Scale, or whether there is sufficient support for this division. Although a statistically significant distinction was found between these latent factors that supported the inclusion of both scales, the Rover subtest loaded stronger on the Simultaneous Scale than expected. It was thus decided to include this subtest there. After further investigation, the Block Counting subtest was added to further strengthen the Simultaneous Scale, whilst the subtests of the Planning Scale contain an element of response speed which in turn strengthens this scale (Kaufman & Kaufman, 2004).

Following the standardisation of the KABC-II, several CFAs were again conducted with narrower age ranges and like the WISC-V, with models that increased in complexity. Within the 3 years age group, results supported a one-factor model with subtests loading on either the CFI or the MPI (Kaufman & Kaufman, 2004; Kaufman et al., 2005b). For the age groups 4 years to 6 years, analysis indicated support for the Sequential, Simultaneous, Learning and Knowledge Scales, whilst the inclusion of the Planning Scale from the age of 7 years were also confirmed. However, at the age of 4 years, the Concept Formation subtest loaded on both the Knowledge and Simultaneous Scales, leading to an insignificant division between these scales. However, based on the measure's theoretical underpinnings and an understanding that these latent constructs measure sufficiently different content, the distinction was maintained in the publication of the KABC-II (Kaufman et al., 2005b). As far as the older age groups are concerned (7 years through 18 years), support was found for the inclusion of the Triangles subtest in the Simultaneous Scale for ages 7 – 12 years, whilst Block Counting was included in this scale for ages 13 – 18 years (Kaufman & Kaufman, 2004). This assisted in a clear divide between the constructs measured by the Simultaneous and Planning Scales (Kaufman et al., 2005b). In conclusion the CFAs conducted with the standardisation data of the KABC-II, showed that the subtests within a given scale had higher factor loadings on that specific latent factor. The only exception was Face Recognition, which was retained for clinical purposes for ages 3 – 4 years. Further, the CFAs also showed that all the latent factors loaded on the global indices to a greater or lesser extent; again, showing support for the inclusion of a *g*-factor in the assessment of a child's intellectual functioning. Finally, the core subtests indicated for the different age groups showed extremely good fit. Here the CFI values were between .997 and .999, whilst the RMSEA values fell between .025 and 0.55 (Kaufman & Kaufman, 2004; Kaufman et al., 2005b).

In 2007, Reynolds and colleagues published their findings of a CFA conducted with data obtained from the normative sample, which included all 16 subtests of the KABC-II. These researchers used the current FCI five-factor model proposed by Kaufman and Kaufman (2004) as a baseline and within the context of CHC theory, made certain adjustments to see if any of these improved the fit of the model. According to the results obtained, Reynolds et al. (2007) indicated that the KABC-II measures the same constructs across all its age groups. In addition, it further found support for the FCI five-factor model presented by Kaufman and Kaufman (2004) for ages 7 – 18 years. As such, these authors also supported the use of the current scales included in calculating the FCI, as well as the calculation of a general intelligence factor as a second-order factor (Reynolds et al., 2007). However, mention was made of the high correlation between the Planning Scale and the general intelligence factor, similar to what was found in some of the WISC-V factor analytic studies for the Fluid Reasoning Index (Chen et al., 2015; Reynolds & Keith, 2017; Wechsler, 2014b). This again supported the important role of fluid reasoning and planning in a person's overall performance. Other inconsistencies between the Reynolds et al.

(2007) model and the one proposed by Kaufman and Kaufman (2004) were found on three subtests: Hand Movements, Gestalt Closure and Pattern Reasoning. These subtests all appeared to load on more than one factor. As the last of these subtests were included in the current study, it is important to take cognisance of the fact that it loaded on both fluid reasoning and visual processing.

In contrast to the study by Reynolds et al. (2007), McGill and Spurgin (2017) published the results of an EFA, conducted with the normative sample within Luria's theoretical framework, which did not support the factor model conceptualised by Kaufman and Kaufman (2004). In fact, the findings of this study indicated that variance could be best explained by a second-order general intelligence factor and recommended that this is where assessors should place their focus. In 2018 McGill and Dombrowski published the results of another EFA conducted with all 16 subtests of the KABC-II. Their study was, however, done against the background of CHC theory, although it also included data from the original normative sample. Unlike the study by Reynolds et al. (2007), the EFA did not confirm the five-factor model proposed by Kaufman and Kaufman (2004). Rather, the 10 core subtests for ages 7-12 years and 13-18 years supported a four-factor model. These factors included crystallised intelligence, sequential processing, learning or long-term memory and retrieval, and a complex perceptual reasoning factor, where the subtests of the Simultaneous Scale are joined by Pattern Reasoning. Story Completion initially loaded on crystallised intelligence and perceptual reasoning; however, once a Schmid and Leiman orthogonalisation transformation procedure was conducted, it was found that this subtest did not significantly load on any of the latent factors (McGill & Dombrowski, 2018). In addition to an investigation of the core subtests, an analysis was also conducted with the 16 core and supplementary subtests combined. Here, a five-factor model was identified; however, this did not precisely fit the model proposed by Kaufman and Kaufman (2004). Once again, separate factors for the simultaneous/visual processing subtests and the planning/fluid reasoning subtests were not found. These combined into a perceptual reasoning factor, with the fluid reasoning factor not supported in the

EFA across the 7 – 12 years and 13 – 18 years groups. Interestingly, the Atlantis Delayed and Rebus Delayed subtests combined to form their own delayed recall factor. However, this factor was not supported in the 13 – 18-year-old group. Other factors included the three remaining scales: Sequential Processing, Learning, and Knowledge (McGill & Dombrowski, 2018).

2.7.5.4. KABC-II: Socio-Cultural Considerations.

Similar to the WISC-V development team, researchers involved in the development and standardisation of the KABC-II took several steps to increase the cross-cultural suitability of this intelligence test. For example, through the purposeful exclusion of knowledge-based items in the MPI, the KABC-II is considered more cross-culturally applicable than assessments which include items that assess acquired knowledge or crystallised intelligence. In addition, teaching items are included and during the standardisation process, experts were consulted to investigate item bias (Fletcher-Janzen & Lichtenberger, 2005). De Sousa et al. (2010) also noted that the Kaufman batteries are considered more culturally applicable as they have been designed to minimise verbal test instructions and verbal responses. Further, Kaufman and Kaufman (2004) not only ensured the representation of children from ethnic minority groups in the standardisation sample, but also published the results of investigations into the mean score differences between ethnic groups in the KABC-II manual to ensure transparency. As stated earlier, the normative sample included 3,025 English-speaking children, stratified based on data from the 2001 US census according to sex, ethnicity, geographical region, and parent education level as a proxy for socio-economic status. In terms of ethnicity, the sample included 1861 White participants (61.5%), 545 Hispanic participants (18%), 465 African American participants (15.4%), and then 154 participants (5.1%) who belonged to the *Other* group (Kaufman & Kaufman, 2004; Kaufman et al., 2005b). In their investigations, Kaufman and Kaufman (2004) found modest differences between the mean scores of these ethnic groups when assessed on the KABC-II. In fact, after they had controlled for the potential impact of parent education levels, they found the impact of ethnic difference on the KABC-

II to be minimal. After adjustments were made to accommodate sex and parent education, White children in the sample obtained a mean MPI score of 101.7, while Hispanic children obtained a mean score of 97.1 and African American children a mean score of 96.0. All these scores centre around the mean of 100, with small differences found between them.

Prior to the publication of the KABC-II, Fletcher-Janzen (n.d.) conducted a cross-cultural study comparing the results of 46 Taos Pueblo Indian participants between Grades R and 12 with the normative sample. Various socio-cultural differences, including dissimilarities in socio-economic status and English language proficiency, were observed between the two samples. However, despite these, the results of the Taos Pueblo sample were commensurate with the normative sample. The mean global scores of the Taos Pueblo sample, including their FCI, MPI and Nonverbal scores, all fell in the average range. A similar trend was observed in their mean scale scores, with the highest mean scale score (100) obtained on the Simultaneous Scale and the lowest mean scale score (95) on the Knowledge Scale. A study conducted by Dale et al. (2011) similarly found minimal differences in the mean scores of White (n = 49) and African American (n = 49) preschool children when assessed on the KABC-II. Participants' ages ranged from 4 years to 5 years, 11 months (M = 59.27, SD = 5.63). The FCI was used as the global score. After controlling for age, sex, and parent education level, minimal differences were observed between the participants' mean scores on the FCI and the Simultaneous, Sequential, Knowledge and Learning Scales. Although, White participants did score significantly higher on the Expressive Vocabulary subtest, likely due to the verbal demand of this subtest and the focus on lexical knowledge. In fact, earlier research by Gomez (2008) found evidence for the impact of English language proficiency on KABC-II test performance. This researcher used the KABC-II to examine the relationship between English language proficiency and intelligence test scores in a sample (N = 137) of Mexican American children. These participants spoke English as an additional language, were between the ages of 7 – 12 years (M = 9.67, SD = 1.60) from lower socio-economic backgrounds and had all been referred for special education provision. The MPI was used as an indication of global intellectual functioning. The mean MPI score of males (*n* = 92) participants was 82.14 (*SD* = 1.21) and the female (*n* = 45) participants 77.38 (*SD* = 1.26). These scores are both more than one standard deviation below the mean. The results further revealed a significant relationship between English language proficiency, as measured by the California English Language Development Test, and participants' MPI scores. The research thus concluded that limited English language proficiency had a significant impact on participants' performance on the MPI in this sample of Mexican American children, despite earlier findings of small mean differences between children from different ethnic groups. However, cognisance must be taken that the participants in the Gomez (2008) study all experienced academic challenges and were from lower socio-economic backgrounds, which could have further impacted test performance.

Scheiber and Kaufman (2015) investigated the predictive validity of the KABC-II; specifically the validity of the FCI, the MPI, and the Nonverbal Index in predicting the achievement of Hispanic, African American and Caucasian children. The sample of 2001 participants between Grades 1 and 12 were drawn from the original sample used to compare the results of the Kaufman Test of Educational Achievement – Second Edition (KTEA-II) and the KABC-II, however, the sample was stratified according to level of education and ethnicity. Most of the sample included White participants (*n* = 1313), followed by Hispanic participants (*n* = 376) and African American participants (*n* = 312). Structural equation modelling was used to determine which of the three KABC-II global scores, correlated best with the participants' performance on three of the KTEA-II achievement domains: Reading, Mathematics, and Written Language. Interestingly, the results of the study indicated that the FCI, which is considered to be more culturally loaded than the other two global scores, was the most valid predictor variable across all three ethnic groups. It further indicated that the MPI and the Nonverbal Index overpredicted the achievement of African American and Hispanic participants, particularly in the Grade 5 – 8 group. Further analysis by Scheiber (2017) of data from the same sample, moreover, indicated that the

Simultaneous, Sequential, Learning and Planning Scales also overpredicted the performance of African American and Hispanic participants on the Reading, Mathematics, and Written Language achievement domains. Thus, the Knowledge Scale, was the most accurate predictor of performance across the three samples. The author concluded that external factors, such as education and learning strategies, socio-economic status and psychosocial wellbeing, could potentially lead to under-achievement in children from different ethnic groups and thus lower than expected scores on measures of achievement.

Using data from the same sample, Scheiber (2016b) also conducted a CFA to determine whether the joint factor structure of the KABC-II and the KTEA-II showed invariance across the three ethnic groups. Given that a joint factor structure was explored, there were seven first-order factors based on CHC theory, including visual processing, fluid reasoning, long-term memory and retrieval, short-term memory, crystallised intelligence and acquired knowledge, reading and written language, and mathematics. Across the three ethnic groups, the mean KABC-II FCI and scale scores of the participants all fell in the average range, showing small differences. In addition, the joint factor-structure of the KABC-II and KTEA-II was invariant across the three ethnic groups, indicating good construct validity. KABC-II factor analysis was also conducted with a sample (*N* = 627) of participants from Germany, all assessed on the German adaptation of the KABC-II (Renner et al., 2023). These children, between the ages of 7 to 12 years, had been referred for numerous challenges, including speech and language, adjustment, health-related, emotional, or behavioural difficulties. CFA was done to examine the higher-order factor structure of the KABC-II core subtests. Given the high correlation between the fluid reasoning/planning and visual/simultaneous processing factors, and cross-loadings of the Pattern Reasoning subtest on both of these factors, a four-factor structure was considered a better fit for the data.

The KABC-II has also been adapted for use in other non-Western countries, including India (Malda et al., 2008). Fifty-seven children between the ages of 6 and 10 years (*M* = 8.08) from lower socio-economic backgrounds, who spoke Kannada as a first language, took part in the adaptation process. In adapting an intelligence test, standardised in a Western context to a non-Western context, extensive adaptations were required, including modifications to the instructions of the test, the content of the items, and the order of the items. Importantly, the modifications were required to the item content of both verbal and more nonverbal items. In later publication, Malda et al. (2010) investigated the psychometric properties of this adapted version of the KABC-II amongst a sample of 598 participants from Bangalore, India who also spoke Kannada as a first language and resided in lower socio-economic areas. Reliability coefficients found the adapted KABC-II to be reliable for use in this sample of children. In addition, the results of structural equation modelling further validated the test's CHC model across the entire age range, as well as sexes and different age ranges. Thus, while there is evidence of the cross-cultural applicability of the KABC-II (Dale et al., 2011; Fletcher-Janzen, n.d; Kaufman & Kaufman, 2004; Malda et al., 2010; Scheiber, 2016b), some research studies (Gomez, 2008; Renner et al., 2023) have questioned its factor structure in different cultural groups, as well as its suitability for different language groups. Further research into the psychometric properties of the KABC-II when used in diverse samples, is thus indicated.

2.7.6. Item Analysis and Intelligence Testing

Although reference is often made to IRT and item analysis in the development of questionnaires, it has also been used in the review, adaptation and revision of intelligence tests. For instance, in Brazil, Wechsler et al. (2013) used IRT in adapting the Woodcock-Johnson III Tests of Cognitive Abilities (WJ-III). Given that their sample included over 1000 children, involved in the standardisation of the WJ-III for use in Brazil, items were analysed based on discrimination, difficulty level, and misfit parameters. The outcome of their research indicated that items from all the WJ-III subtests, except for the Visual-Auditory Learning test, needed to be revised. Another study conducted in Malaysia, with the Malaysian UKM1 Intelligence Test, applied IRT and a Rasch model in the revision of the assessment measure

(Ariffin et al., 2010). Following an administration of the test to 772 participants, the researchers concluded that "items analysis is the best method for determining the quality of developed test items" (Ariffin et al., 2010, p. 208). Immekus and Maller (2009) investigated DIF in the Kaufman Adolescent and Adult Intelligence Test (KAIT). Using a 2PLM, situated in an IRT likelihood ratio approach, data obtained from the normative sample (*N* = 2,200) were investigated for items bias in terms of sex. The sample included 51.6% females (*n* = 1,135) and 48.4% males (*n* = 1,065). The results of the analyses indicated difficulty (uniform DIF) and discrimination (nonuniform DIF) differences in several of the subtests assessing crystallised intelligence. However, the authors (Immekus & Maller, 2009, p. 1007) concluded that "… the total number of KAIT items exhibiting DIF was relatively minor and should not affect test score interpretation across gender groups."

Using data from the WISC-III normative sample (*N* = 2,200), Maller (2001) investigated DIF in six of the WISC-III subtests. The item analysis again focussed specifically on sex-related item bias in a sample equally divided between male and female participants. The six subtests included in the analyses were Similarities, Vocabulary, Comprehension, Information, Arithmetic, and Picture Completion. A 2PLM, including both discrimination and difficulty was used. The results of the study indicated that from the 151 items included in the analyses, 52 showed DIF when the scores of male and female participants were compared. Similar to the KAIT study, both uniform and nonuniform DIF was observed. However, in the WISC-III study, Maller (2001) concluded that male and female participants correctly answered different subtest items and thus, some of the content in the items favoured females and some favoured males. Subsequently, while female and male participants did not demonstrate significant differences in mean scores, the scores obtained by the two groups could not "be assumed to have the same meaning" (p. 814). According to Reynolds and Hadorn (2016) DIF was also investigated during the standardisation of the WISC-V; specifically on a sample of 54 children from African American and Hispanic backgrounds. However, other than saying that this was done to investigate potential problematic items in administering the test to children from differently socio-cultural backgrounds, no further information is provided in the WISC-V Technical and Interpretive Manual (Wechsler, 2014b). Although, Miller and McGill (2016) mention that the WISC-V items did not demonstrate bias in terms of ethnicity.

In the South African context, De Beer (2004) used IRT, specifically a three parameter IRT model, to investigate the items of the Learning Potential Computerised Adaptive Test (LPCAT). Although the LPCAT is not technically considered an intelligence test, it is a valuable measure in assessing the learning potential of children as a component of cognitive ability (De Beer, 2013). Regardless, with a large sample of 2554 high school participants, the three parameter IRT model could be applied. Based on the analyses conducted, 82 of the original 270 items were discarded as they showed item bias. De Beer (2004) concluded by saying that the value of IRT lies in the fact that it provides valuable information that can assist test developers to easily recognise and discard potentially biased items.

Furthermore, during the adaptation and standardisation of the Wechsler Adult Intelligence Scale – Fourth Edition for South Africa (WAIS-IV-SA) Wechsler (2014d), a Rasch analysis was performed to investigate the item order of subtests of the WAIS-IV in a South African sample. The normative sample included 433 individuals between the ages of 16 and 59 years (*M* = 28.41, *SD* = 11.51). The majority of the sample (30.5%) had completed high school (Grade 12) and included slightly more females (*n* = 221) than males (*n* = 212). Stratified according to the 2011 South African national census, the greatest number of participants were African (*n* = 337), followed by White (*n* = 44), Coloured (*n* = 31) and Indian/Asian (*n* = 21) participants. Although further information on socio-economic status and living conditions are not provided, it is reported that the participants resided in urban (*n* = 365) and rural areas (*n* = 39). According to the outcome of the Rasch analysis, which was conducted with data from the SA research edition of the WAIS-IV, the order of the items of the Similarities, Vocabulary, Comprehension, Information and Picture Completion subtests had to be altered. This was because the order of the items of these subtests differed in difficulty when data from the South African sample was compared to the UK edition of the WAIS-IV.

2.7.7. African and South African Research on the WISC and the K-ABC

Several African, including South African, research studies have utilised the K-ABC or WISC in the assessment of children's intelligence. Included in these, are African studies by Bangirana et al. (2009), Baumgartner et al. (2012), Boivin et al. (1993, 2010, 2017), Chernoff et al. (2018), Conant et al. (1999), Gichovi et al. (2018), Holding et al. (2004), Ogunlade et al. (2011), Piper et al. (2023, 2024), Ruffieux et al. (2009), and Rushton and Jensen (2003). From these, four studies focussed on the psychometric properties of these tests and included the WISC-V or the KABC-II. Through factor analysis, Bangirana et al. (2009) investigated the construct validity of the KABC-II among a group of children with a history of cerebral malaria. Aged between 6 – 17 years (M = 9.90, SD = 2.46), their sample consisted of 65 children assessed on the KABC-II 44 months after their diagnoses. Most of the participants (n = 40) were male. Applying Luria's neuropsychological theory, the researchers investigated the participants' performance on four scales, Sequential, Simultaneous, Planning and Learning. Subtests from the Knowledge Scale were thus not included. In addition to the removal of seven items from the Gestalt Closure subtest, subtest instructions were also translated into Luganda. In investigating the factor structure of the KABC-II in this sample, a "principal component factor analysis with Varimax solution" (Bangirana et al., 2009, p. 189) was conducted. The results of this indicated that the KABC-II displayed good construct validity in this sample of children, with subtests measuring equivalent cognitive abilities loading on the same factors. Interestingly, five factors emerged from the analysis, including sequential processing, simultaneous processing, learning, planning, and an immediate and delayed recall factor, with Number Recall and Atlantis Delayed loading on this fifth factor.

Incorporating participants from South Africa, Zimbabwe, Malawi, and Uganda, Chernoff et al. (2018) included the KABC-II in a battery of assessments to investigate the impact of HIV and AIDS on children's neuropsychological functioning. At the baseline visit, the sample (*N* = 611) consisted of almost equal numbers of males (47%) and females (53%), with ages ranging from 5 – 11 years (*M* = 7.2, *SD* = 1.4). Although the participants from South Africa (*n* = 309), resided in urban settings, participants from the other countries (*n* = 302) resided in rural settings. The baseline sample included children who were infected with HIV (*n* = 246), children who had been exposed to HIV, but were not infected (*n* = 183), and children who had not been exposed and were not infected with HIV (*n* = 182). From the original sample, 98.7% (*n* = 603) were seen at the 48-week visit and 96.2% (*n* = 588) at the 69-week visit. The KABC-II was administered at weeks 0 and 48, with data from the baseline study (week 0) included in the factor analysis. The Knowledge Scale was again omitted and thus, only four of the five KABC-II scales were administered. Test instructions were relayed in the participants' local languages. Statistical analyses included a principal EFA, which identified four factors including learning, simultaneous processing, planning, and sequential processing. As such, the original factor-structure of the KABC-II was confirmed, however, the four factors identified described merely 47% of the variance.

The most recent study using the KABC-II to explore the intellectual functioning of children in Africa, were published by Piper et al. (2024). This study adapted and piloted two of the KABC-II Planning Scale subtests, Story Completion and Pattern Reasoning, for cross-cultural application in Zimbabwe. The sample (*N* = 80) was randomly selected from the children who participated in a pilot study at 7-years-old (*M* = 7.6, *SD* = 0.2) and included females (49%) and males (51%). This pilot study aimed to investigate the physical and cognitive development of children from rural and peri-urban areas in Zimbabwe. While 80 children were initially included in the sample, the results of 10 participants were excluded from analysis as these were used for standardisation among fieldworkers. As such, 70 participants were thus included in the final sample. However, after subtests from the KABC-II Learning, Sequential, Simultaneous, and Planning Scales were administered to the first 50 participants, researchers found that they consistently performed poorer on the Planning Scale subtests. A decision was thus made to adapt the two subtests of this scale and investigate the performance of 20 children on these adapted subtests. Details of the adaptations made to the two subtests can be found in Piper et al. (2024). Regardless, when comparing the results, the 20 participants assessed on the adapted subtests performed significantly better than the 50 children assessed on the original items. However, in comparing their performance on the Planning Scale to the other three scales (Learning, Sequential and Simultaneous), their scores on the modified Planning Scale were still lower. Although, a post-hoc Tukey's test revealed significant differences between the Planning and Learning scales only. The researchers concluded that the adaptation of subtests from etic intelligence tests, can improve their cross-cultural applicability when administered to children from low- and middle-income contexts.

While the three stated African studies included the KABC-II in the assessment of children's cognitive abilities and processes, a study by Gichovi et al. (2018) investigated the WISC-V global scores (FSIQ) of 83 children from Kenya. The sample again included an almost equal number of boys (n = 43) and girls (n = 40) between the ages of 10 and 13 years (M = 11.10, SD = 1.10) who attended private schools in the Embu West Sub County. The 10 WISC-V primary subtests were administered to all the participants. Results indicated a mean FSIQ score (M = 92.98, SD = 10.51) falling in the average range. Although socio-cultural factors such as parent education level and socio-economic status, were not reported, it was assumed that the children were from similar socio-economic backgrounds as they all attended private schools. Nonetheless, the performance of the children in the sample was seen as comparable to the normative sample.

In addition to these African studies, several South African research studies which include different editions of the WISC and K-ABC, have been published. Many of these studies, included research with different editions of the WISC, including Bickell (2015), Ferrett (2011), Shuttleworth-Edwards, Van der Merwe, et al. (2013), and Skuy et al. (2001). As stated, the research by Cassoojee (2020) was a preliminary study, completed under the supervision of the researcher, with 50 participants of the present research study and focussed specifically on comparing the scores of the participants on specific scales and indices of the WISC-V and KABC-II that are said to assess similar constructs. The only other research study, which compared the K-ABC and the Wechsler Intelligence Scale for Children – Revised (WISC-R) in a sample of South African children who experience barriers to learning and development, was published by Skuy et al. (2000). The remaining South African studies included published research studies conducted with the K-ABC by De Sousa et al. (2010) and Jansen and Greenop (2008), and the KABC-II by Mitchell (2015) and Mitchell et al. (2018).

As part of a battery of assessments that are often employed in assessing the neuropsychological functioning of individuals, Skuy et al. (2001) administered the WISC-R to a total of 252 high school children from Soweto, nearby Johannesburg, Gauteng province. These participants spoke a variety of languages, although, English was predominantly an additional language. The first 100 participants included children from Grades 1-12 between the ages of 12 years, 0 months and 24 years, 6 months (*M* = 17 years, 2 months). The second group of 152 participants were in Grades 7-9 and included 76 females and 76 males. The average age of the females was 14 years, 3 months, and the males, 14 years, 4 months. Children were from mostly under-resourced and lower socio-economic backgrounds. Specific subtests from the WISC-R administered to the first 100 participants, included “Similarities, Arithmetic, Digit Span, Picture Arrangement, Mazes, and Coding” (Skuy et al., 2001, p. 1416). The entire WISC-R was administered to the second group of 152 participants. The WISC-R mean scores obtained by the first group of participants were lower than those obtained by participants in the US normative sample and were identified as either *below average* or *borderline impaired*. Scores ranged from 5.52 (*SD* = 2.6) for the Similarities subtest to 7.84 (*SD* = 3.3) for the Mazes subtest. Mean scores of the second group of participants were again lower than that of the US normative sample and ranged from 2.85 (*SD* = 1.64) for the Vocabulary subtest to 7.60 (*SD* = 2.90) for the Mazes subtest. Overall, the results of this study (Skuy et al., 2001) highlighted the importance of developing norms for specific socio-cultural groups.

Similar to the Skuy et al. (2001) study, Ferrett (2011) included the WISC-IV Coding subtest in a battery of assessment measures aimed at investigating the potential impact of ethnicity, age, sex, and socio-cultural factors such as level and quality of education, and socio-economic status on children's performance on intelligence and neuropsychological tests. The participants (*N* = 215) all resided in the Western Cape province, were between the ages of 12 and 15 years (*M* = 13.91, *SD* = 1.23), and included 117 females (54.4%) and 98 males (45.6%). While 148 (68.8%) participants identified as Coloured, 67 (31.2%) identified as White. The sample was further stratified according to language and quality of education. Although instructions were provided in the participants' first languages, further adaptations to the Coding subtest were not made. The results of an Analysis of Covariance (ANCOVA) indicated that the participants' test performance was primarily influenced by age, sex, and quality of education. From these, age explained the highest proportion of variance (16.2%) observed.

Shuttleworth-Edwards, Van der Merwe, et al. (2013) conducted a within-group demographically stratified normative study with 69 Grade 7 participants between the ages of 12 and 13 years (*M* = 13.04, *SD* = 0.34) from the Eastern Cape province. The sample was stratified according to ethnicity (White, Coloured and Black), quality of education, and language (English, Afrikaans and Xhosa) and were almost equally distributed across male (*n* = 36) and female (*n* = 33) children. Children who spoke English as their first language, and Xhosa-speaking children were assessed in English. For Coloured and White children who spoke Afrikaans as a first language, test instructions were translated into Afrikaans, which could have impacted their performance. Researchers used the WISC-IV to assess the participants' intellectual functioning. Results of the study indicated that White, EFL participants, attending well-resourced schools offering a good quality of education, consistently obtained higher mean scores when compared to the other participants. In addition, their performance was equivalent to normative samples from the UK and US. However, children who attended under-resourced schools where they were likely exposed to a lower quality of education, consistently obtained lower scores when compared to the other participants who all attended well-resourced schools. In addition, their scores were 20 to 30 IQ points lower than the normative samples. The researchers (Shuttleworth-Edwards, Van der Merwe, et al., 2013) concluded that quality of education has a significant impact on a child's performance on intelligence tests such as the WISC-IV.

Following the research published by Shuttleworth-Edwards, Van der Merwe, et al. (2013), Bickell (2015) conducted follow-up research to obtain WISC-IV normative data for a sample of 8- and 9-year-old (*M* = 9 years, 10 months) Black, Xhosa-speaking participants (*N* = 32) in Grade 3 at an under-resourced school in the Eastern Cape province. Given that the children attended an under-resourced school, quality of education was likely lower than what it would be in better resourced schools. Included in the sample were 13 (41%) males and 19 (59%) females. The results of this study again highlighted the impact of quality of education on the performance of children on intelligence tests with participants obtaining lower scores of between 20 to 30 IQ points, when compared to the WISC-IV UK normative sample. The results of an independent sample *t*-test further indicated that aside from the Digit Span subtest, performance of the participants in the Bickell (2015) study was not statistically significantly different from participants in the Shuttleworth-Edwards, Van der Merwe, et al. (2013) study. Interestingly, the Grade 3 learners obtained significantly higher scores than the Grade 7 learners on the Digit Span subtest.

In 2020, Cassoojee used results from 50 participants of the present study to investigate the validity of the KABC-II and WISC-V. This was done by comparing the participants' results on scales and indices designed to measure the same underlying constructs. Specifically, the participants' results on the Verbal Comprehension Index were compared to the Knowledge Scale; their results on the Visual-Spatial Index were compared to the Simultaneous Scale; their results on the Working Memory Index were compared to the Sequential Scale; and their results on the Fluid Reasoning Index were compared to the Planning Scale. The participants were all from one private remedial school in Johannesburg, between the ages of 10 to 14 years (*M* = 11.9, *SD* = 1.18). Most of the participants (*n* = 17) were in Grade 6 at the time. Other participants were in Grade 4 (*n* = 6), Grade 5 (*n* = 11), Grade 7 (*n* = 10), and Grade 8 (*n* = 6). The results of a Pearson correlation coefficient indicated strong correlations between all the stated scale and index combinations (Verbal Comprehension Index and Knowledge Scale; Visual-Spatial Index and Simultaneous Scale; Working Memory Index and Sequential Scale; and Fluid Reasoning Index and Planning Scale). In addition, a strong correlation was also noted between the Working Memory Index and the Knowledge Scale. However, the results of a *t*-test indicated the participants' scores on the Working Memory Index were significantly lower when compared to their results on the Sequential Scale. In addition, their results on the Knowledge Scale were also significantly higher than their results on the Verbal Comprehension Index (Cassoojee, 2020).

A previous South African study, which compared the test performance of children on the K-ABC and WISC-R, was published by Skuy et al. (2000). In this study, researchers compared the performance of Black South African and White South African children on the WISC-R and K-ABC. A total of 21 Black and 35 White children in the third grade, attending two private schools for children experiencing barriers to learning and development in Johannesburg, South Africa, were included in the sample. The mean ages of the two groups ranged from 9.5 years for the Black children to 8.4 years for the White children. The sample consisted of 40 males and 16 females. The results of *t*-tests showed that while there were no statistically significant differences between the two groups on the K-ABC, there were statistically significant differences observed in the two groups' performance on the WISC-R Verbal Scale. Similarly, while significant differences between the two groups were not observed on the K-ABC MPI, significant differences were noted in the groups' performance on the WISC-R FSIQ. Additional significant differences between the two groups were also noted in their scores on the WISC-R Vocabulary and Information subtests. Researchers (Skuy et al., 2000) concluded that the results of this study indicated that the K-ABC was a more suitable measure for use with South African children when compared to the

WISC-R. Assessed on the WISC-R, the Black sample consistently obtained significantly lower scores than the White sample; however, on the K-ABC differences observed were non-significant.

Two additional South African studies, which included the K-ABC, were published by De Sousa et al. (2010) and Jansen and Greenop (2008). In the first of these, selected subtests from the K-ABC Sequential and Simultaneous Scales were administered to 199 children from Soweto, near Johannesburg in the Gauteng province. The sample included slightly more females (*n* = 102) than males (*n* = 97). In this longitudinal study (Jansen & Greenop, 2008), children were assessed at 5 years and again at 10 years. At both these points, the Number Recall, Hand Movements and Word Order subtests were administered as part of the Sequential Scale. At 5 years, Spatial Memory, Gestalt Closure, Triangles, and Matrix Analogies were assessed as part of the Simultaneous Scale, while Photo Series was added as an additional subtest at 10 years. Among the analyses conducted, the results of “a principal component analysis with varimax rotation” (Jansen & Greenop, 2008, p. 358) investigating the factor structure of the K-ABC, identified two factors at 5 and again at 10 years. The supported the factor structure of the K-ABC in this group of children from an urban area in South Africa. In the second of the two stated South African studies which included the K-ABC, De Sousa et al. (2010) aimed to investigate the potential impact of monolingualism and bilingualism on the processing skills and learning of a group of 60 children from Johannesburg. At the time of data collection, participants were in Grade 3, selected from two primary schools in the same area. The mean age of the monolingual group (*n* = 30), which included 10 males and 20 females, was 9 years, 8 months (*SD* = .90). The bilingual group (*n* = 30) included 16 males and 14 females, with a mean age 9 years, 9 months (*SD* = .85). While the monolingual group included only EFL children, the bilingual group spoke Afrikaans as their first language but learned to read and spell in both Afrikaans and English. Subtests administered as part of the Sequential Scale, included Hand Movements, Number Recall, and Word Order, and the Simultaneous Scale, Gestalt Closure, Triangles, Matrix Analogies, Spatial Memory, and Photo Series. The results of the study did not indicate significant differences in the performance of monolingual and bilingual children on the Sequential Scale, the Simultaneous Scale, or the MPI. In addition, the participants' mean Sequential and Simultaneous Scale scores fell in the average range, according to the criteria for interpreting the K-ABC. The researchers (De Sousa et al., 2010) concluded that the norms of the K-ABC were suitable for use when assessing the cognitive processing skills of this group of English and Afrikaans First Language children from an urban area.

In the first published research study of its kind in South Africa, Mitchell (2015) investigated the psychometric properties of the KABC-II in a sample of South African children. The sample consisted of 382 participants between the ages of 7 to 11 years (*M* = 8.96; *SD* = 0.82). Most of the participants were in Grade 3 (*M* = 3.37, *SD* = 1.08) at the time of data collection, with 65% being female (*n* = 250), and 35% male (*n* = 132). Given that the participants were from a rural part of the KwaZulu-Natal province, where English is spoken as an additional language, test instructions were translated from English to IsiZulu. Although the participants' quality of education is not specified, Shuttleworth-Edwards and Truter (2023) indicate that it was likely low provided the location of the schools which the participants attended. Following review by a panel of experts, three subtests were excluded, Rebus, Rebus Delayed, and Gestalt Closure. The content of these subtests was considered to show bias given the diverse cultural, linguistic, and educational experiences of South African children. With the exclusion of Rebus, Atlantis Delayed was included as a subtest in the Learning Scale. This is considered an appropriate substitution by Kaufman and Kaufman (2004). Finally, given the application of Luria's neuropsychological framework, subtests of the Knowledge Scale were also excluded.

According to Mitchell (2015) and Mitchell et al. (2018), the results of the study showed excellent internal consistency reliability calculated through split-half comparisons and the Spearman-Brown method. However, the participants' mean MPI score (*M* = 76.12, *SD* = 11.0) was lower than anticipated. Further analysis of their performance on the KABC-II scales, indicated that participants struggled most with subtests of the Planning Scale, more so with Story Completion (*M* = 4.38, *SD* = 2.47) compared to

Pattern Reasoning (*M* = 5.81, *SD* = 2.84). This was followed by the subtests of the Simultaneous Scale, where participants obtained a lower mean score on Triangles (*M* = 5.41, *SD* = 2.5) compared to Rover (*M* = 7.16, *SD* = 3.02). According to the outcome of the CFA, the originally validated factor-structure of the KABC-II was maintained. A second CFA, which excluded the global MPI score, fitted equally well with the data. It is interesting that this CFA indicated a strong correlation between the simultaneous processing and planning factors (.91). This was further investigated through an EFA, with results indicating that the subtests of these two scales, loaded on one factor. This led to additional investigations into the subtests of the Simultaneous and Planning Scales. The results of these analyses indicated that children may have been disadvantaged by the time-point bonusses which apply to the Story Completion and Triangles subtests. Block Counting was subsequently added as a supplementary subtest, with the results indicating that this subtest obtained the highest mean score (*M* = 8.00, *SD* = 2.8) of all the Simultaneous Scale subtests. Also, after Block Counting substituted Triangles in the CFA, results indicated acceptable fit statistics.

2.8. Conclusion

This literature review explored research related to the core concepts under investigation in this study, starting with barriers to learning and development and its use in the South African context. This incorporated a review of the history of education in South Africa, including a review of the manner in which children with barriers to learning and development were educated. The literature review continued with an exploration of the concept of intelligence in terms of the three theoretical models that underlie the development and interpretation of the WISC-V and the KABC-II, followed by an overview of these two intelligence tests, their indices/scales and subtests, and the cognitive abilities and processes assessed by these. Thereafter, the literature review explored the concepts of reliability, validity, and DIF and explored research related to the validity, reliability and cross-cultural applicability of the WISC-V and KABC-II. It also provided a review of research that has been conducted to investigate

DIF in intelligence tests. The literature review concluded with an overview of African and South African research incorporating different editions of the WISC and K-ABC, highlighting that South African research into the psychometric properties of the WISC-V and KABC-II remains scare. The following chapter will provide an overview of the methods employed to investigate the research questions and address the aims and objectives of this preliminary investigation into the psychometric properties of the WISC-V and KABC-II in a sample of South African children experiencing barriers to learning and development.

3. Methods

3.1. Introduction

The aims of the study necessitated a post-positivist research paradigm, underlined by the assumption that human characteristics, such as intelligence, can be measured by reliable and valid tools. This chapter will expand on these concepts by providing an overview of the research approach, paradigm, design, and methods utilised to achieve the study's aims and objectives. In addition, information will be provided on the sample that participated in the study as well as the research instruments and the procedure followed in conducting this study. Finally, information will be provided on the statistical techniques employed to analyse the data, as well as the ethical considerations considered in conducting this research.

3.2. Aims and Objectives of the Research Study

As stated in Chapter 1, the first aim of the research study was to conduct a preliminary investigation into the psychometric properties of the Wechsler Intelligence Scale for Children – Fifth Edition (WISC-V) and the Kaufman Assessment Battery for Children – Second Edition (KABC-II) in a sample of South African children experiencing barriers to learning and development that negatively impact the acquisition and development of essential academic skills. Aligned to this, two objectives were set:

1. Objective 1: To investigate the internal consistency reliability of the WISC-V and KABC-II subtests using the Spearman-Brown split-half reliability method within the framework of Classical Test Theory.
2. Objective 2: To investigate the construct validity of the WISC-V and KABC-II through the use of correlation coefficients and confirmatory factor analysis (CFA), within the framework of Classical Test Theory.

In addition, the study also aimed to investigate the suitability of these two measures in a South African sample through item analysis and a comparison of the performance of English First Language (EFL) and English Additional Language (EAL) participants. In line with this, two additional objectives were set:

3. Objective 3: To determine whether statistically significant differences in performance exist between EFL participants and EAL participants, using a General Linear Model (GLM) with participants' scores as the dependent variables and English language background as the independent variable.
4. Objective 4: To investigate differential item functioning (DIF) using Item Response Theory (IRT), focusing specifically on the subtest items of the WISC-V and KABC-II.

3.3. Research Questions

Corresponding research questions were posed for each of the four objectives:

1. Objective 1: To investigate the internal consistency reliability of the WISC-V and KABC-II subtests using the Spearman-Brown split-half reliability method within the framework of Classical Test Theory.
    - Is the WISC-V reliable for use with a sample of South African children who experience barriers to learning and development that negatively impact the acquisition and development of essential academic skills?
    - Is the KABC-II reliable for use with a sample of South African children who experience barriers to learning and development that negatively impact the acquisition and development of essential academic skills?
2. Objective 2: To investigate the construct validity of the WISC-V and KABC-II through the use of correlation coefficients and CFA, within the framework of Classical Test Theory.
    - Does the factor structure of the WISC-V in this sample of South African children who experience barriers to learning and development that negatively impact the acquisition and development of

essential academic skills align with its theoretical model based on correlation coefficients and CFA?

- Does the factor structure of the KABC-II in this sample of South African children who experience barriers to learning and development that negatively impact the acquisition and development of essential academic skills align with its theoretical model based on correlation coefficients and CFA?

3. Objective 3: To determine whether statistically significant differences in performance exist between EFL participants and EAL participants, using a GLM with participants' scores as the dependent variables and English language background as the independent variable.
    - Do statistically significant differences exist between the scores of EFL and EAL participants in this sample of South African children who experience barriers to learning and development that negatively impact the acquisition and development of essential academic skills?
4. Objective 4: To investigate DIF using IRT focusing specifically on the subtest items of the WISC-V and KABC-II.
    - Do the subtest items of the WISC-V show DIF when administered to a sample of South African children who experience barriers to learning and development that negatively impact the acquisition and development of essential academic skills?
    - Do the subtest items of the KABC-II show DIF when administered to a sample of South African children who experience barriers to learning and development that negatively impact the acquisition and development of essential academic skills?

3.4. Research Paradigm, Design, and Methods

The research approach outlines a general framework guiding how the study will be conducted and documented. Within any approach, however, there are interconnected components that warrant further examination: the philosophical paradigm, research design, and research methods (Creswell &

Creswell, 2020). Given the aims of the current study with their aligned objectives and research questions, a quantitative approach situated within a post-positivist paradigm was followed in conducting the research study. Moreover, given the different objectives, data was collected and analysed through both correlational and causal-comparative research designs.

The philosophical paradigm of a research study, also referred to as a philosophical worldview by Creswell and Creswell (2020), reflects the underlying philosophical assumptions about the world that the researcher brings to the study. Within the context of research, a paradigm dictates the way a researcher views and then studies a particular phenomenon. These beliefs are applied consistently throughout the research study and will thus influence the methods and data analyses techniques employed (Devlin, 2006). Eyisi (2016) and Taylor and Medina (2011) identify two factors within a person's philosophical paradigm that are important in the context of research. The first is ontology, which refers to the phenomenon or the problem under investigation, a person's beliefs about it, and a person's views about knowledge and reality (Scotland, 2012). The second is epistemology, which refers to the way the phenomenon or problem can be understood or how knowledge comes about, is created, developed, attained, uncovered, studied, and/or communicated (Eyisi, 2016; Krauss, 2005; Scotland, 2012).

A quantitative approach to conducting research is often seen as falling within a positivist paradigm (Cresswell & Cresswell, 2020; Eyisi, 2016). Positivism is grounded in the belief that a singular, measurable reality and absolute truth exists and can be identified and comprehended through objective study and the control of extraneous variables that can impact the outcome of the findings (Park et al., 2020). It maintains that appropriately conducted research reveals objective reality, necessitating a strict separation between researchers and participants to minimise bias and uphold validity (Ponterotto, 2005; Ryan, 2006; Taylor & Medina, 2011). However, this is not always possible in social research studies. For instance, the interaction between the researcher and participants could not be avoided in this study, as the researcher was an integral part of the data collection process and assessments had to be conducted one-on-one. Additionally, as reported previously, numerous factors can influence the outcome of an intelligence test, some of which are beyond the researcher's control. As such, this study fell more within a post-positivist paradigm.

Taylor and Medina (2011) argue that this paradigm, often regarded as a milder form of positivism, permits some interaction between the researcher and the participants. Panhwar et al. (2017) add that post-positivism was born out of the frustration expressed by educational researchers about the limitations of positivism. Whereas positivism rests on the assumption that one can only study that which can be directly observed and tested, and one can uncover the absolute truth of experience (Devlin, 2006), Cresswell and Cresswell (2020) state that post-positivism is underlined by the belief that absolute truth does not exist when we study humans and their behaviour. Thus, we cannot be sure that our findings represent an absolute truth about the human participants involved in the study. Panhwar et al. (2017) and Ponterotto (2005) agree that an absolute truth does not exist, and that post-positivism considers truth to be probabilistic and provisional as it can be impacted by several factors, including the researcher, the participants, and the factors affecting the researcher and the participants on any given day. This is particularly relevant in the context of intelligence testing, where various factors, including the assessor, the characteristics of the assessment tools, environmental conditions, fatigue, and illness, can influence a participant's performance. As such, the outcome of the assessment cannot be considered an absolutely accurate reflection of the participant’s intellectual functioning, but rather a snapshot taken on a specific day under specific circumstances.

It was previously noted that a research paradigm is based on ontological and epistemological assumptions, with Krauss (2005) and Cresswell and Cresswell (2020) adding that the research approach also guides the research methods employed in data collection and analysis. It seems understandable that within a quantitative research approach, quantitative research methods would be employed;

however, it is important to gain a more in-depth understanding of what this entails and how it was applied in the current study. In general terms, quantitative research seeks to understand and investigate the relationship/s between variables (Apuke, 2017; Ary et al., 2019; Morgan & Sklar, 2012a). To do so, it is however important that the variables of the study be clarified and explained. The preceding chapter introduced constructs, latent variables, and measured variables. To reiterate, constructs refer to variables that cannot be directly observed but are crucial in providing valuable insights, aiding in data interpretation, and supporting theory development (Apuke, 2017; Ary et al., 2019). Variables can also be considered constructs; however, Cresswell and Guetterman (2019) refer to them as attributes or characteristics and state that whilst constructs are abstract, variables are specific. In general, variables have been explained in terms of dependent and independent variables (Salkind, 2018).

Dependent variables are those tied to the study's outcomes, changing in response to the independent variable. The independent variable, by contrast, directly or indirectly influences the dependent variable and precedes it in the analysis (Apuke, 2017; Ary et al., 2019; Salkind, 2018). In this study, for example, English language proficiency serves as the independent variable, while participants' scores on the intelligence tests are the dependent variables. Notably, in the context of an ex-post facto design, the independent variable may also be termed an attribute variable, as it represents a specific, non-manipulable characteristic within the two comparison groups (Ary et al., 2019). By examining the relationship between the independent or attribute variable and the dependent variables, the study aimed to determine which intelligence test is more appropriate for a multilingual South African context. Salkind (2018) further explains that latent variables, or factors, are typically considered independent variables, whereas measured variables are regarded as dependent variables. This study also investigated correlations among these latent and measured variables to evaluate the construct validity of the WISC-V and KABC-II in a sample of South African children experiencing barriers to learning and development.

Quantitative research is also defined by its use of specific measurement tools to collect data expressed in numerical format (Apuke, 2017; Creswell & Guetterman, 2019; Eyisi, 2016). Conducted within a post-positivist paradigm, such research aims for objective data collection (Ary et al., 2019), though it acknowledges that numerical data does not represent an absolute truth. In this study, numerical data was gathered using two standardised intelligence tests and subsequently analysed through statistical procedures (Apuke, 2017; Creswell & Guetterman, 2019; Seabi, 2012b). However, the study also recognises that these results reflect a snapshot of participants' intellectual functioning at a particular time and under specific conditions, rather than a definitive assessment of their intellectual capacities.

The last characteristic of quantitative research is the fact that the research design is chosen upfront (Ary et al., 2019). Therefore, the selection of a research design, understood as a type or strategy of inquiry, is influenced by the study's aims, research objectives, and questions, but must also align with the research approach and paradigm (Creswell & Creswell, 2020). In true experimental research, the independent variable would be manipulated and extraneous variables controlled to establish clear cause-and-effect relationships (Gravetter & Forzano, 2009; Seabi, 2012b). However, as noted by Johnson (2001), "there are so many important but nonmanipulable independent variables needing further study in the field of education" (p.3). Thus, the attention of the researcher is turned to other post-positivist research designs where the independent variable cannot be manipulated, such as the correlational research design (Apuke, 2017; Ary et al., 2019; Cresswell & Cresswell, 2020) and the ex post facto design, also referred to as a causal-comparative research design (Ary et al., 2019; Seabi, 2012b). Given the four objectives set, both these non-experimental designs were included in the present study.

As implied by the name, correlational research seeks to measure the relationship, association, or correlation, between two or more variables and then describe it statistically (Apuke, 2017; Creswell & Cresswell, 2020; Gravetter & Forzano, 2018). However, according to Ary et al. (2019), it goes beyond this as it not only seeks to investigate a potential relationship but also the strength and the direction of that relationship and in doing so, describes the relationship and the nature thereof. A correlational research design was considered fitting for the first two objectives set, as investigations were conducted with data from a single group of participants and not between groups, as is the case with causal-comparative research (Ary et al., 2019). In addition, potential relationships between variables were investigated, for instance, the correlation between latent variables, measured variables, and latent and measured variables. In addition, in investigating the internal consistency of the subtests, a correlational design was employed to assess consistency by correlating the split-half scores.

Ex post facto or causal-comparative research seeks to investigate potential differences between two or more pre-determined groups (Gravetter & Forzano, 2009; Schenker & Rumrill, 2004). Instead of investigating cause-and-effect relationships by regulating "the effect of some conditions by manipulating the independent variable" (Seabi, 2012b, p.87), ex post facto or causal-comparative research does not involve the manipulation of any variables. In exploring the third and fourth objectives, differences in the performance of EFL and EAL participants were explored. Participants were divided into two groups based on an independent variable that already existed before the start of the study and the performance of these two groups were compared. This was done to investigate statistically significant differences and to determine whether the items of the WISC-V and the KABC-II functioned differently in these two language groups and thus, a causal-comparative design was deemed appropriate (Apuke, 2017; Ary et al., 2019).

3.5. Sample and Sampling

At the outset of the research project, a total sample of 120 children were envisioned. Sixty of these would have been from private remedial schools and another 60 from private mainstream schools. The latter would be used as a comparative sample to which the results of the participants in remedial schools could be contrasted. This was considered important as neither the KABC-II or the WISC-V have been standardised or normed for South African children. Further, it fitted with the initial aim of the study, which was to investigate the cognitive profiles of children with barriers to learning and development in South Africa. However, some difficulties were encountered in obtaining the anticipated sample and, although 12 schools in total had been contacted by the end of 2019, only three private remedial schools agreed to partake in the study. Several reasons were offered, including the assessment of children by in-house psychologists, new schools who wanted to first establish themselves before allowing a research project to commence, and schools who had already allowed several research projects and could not accommodate further studies. In addition, it became evident that some mainstream schools were reluctant as they did not want to have their children assessed on intelligence tests just for the sake of research. This might be especially true in the context of information that was provided in the school principal information sheet (Appendix C) which indicated that both intelligence tests have the potential to identify specific cognitive strengths and weaknesses in children, which can influence their scholastic progress and point towards a potential learning difficulty. Regardless, by the end of 2019, 59 children from private remedial schools had been assessed on both measures and it was hoped that the children from mainstream schools would be assessed in 2020. Consideration was also given to recruit participants from government mainstream schools in socio-economic areas that resembled the three schools who had already participated. Although one mainstream school agreed to participate in the study in the beginning of 2020, data collection was brought to a halt by the COVID-19 pandemic and subsequent national lockdown. Schools were not only closed for a significant part of 2020, but once re-opened, they were very cautious to allow people in to test children due to the risk of infection. Permission could also not be obtained from the Gauteng Department of Education (GDE) to access schools during the latter half of 2020 when a moratorium on all research activities in Gauteng schools were placed. Similarly, in the beginning of 2021, a significant number of schools only re-opened mid-February and there was once again concern about additional people being allowed into schools considering the then second COVID-19 wave. The moratoriums on all research activities in GDE schools also remained. As a result, the research project was reviewed, and it was decided to conduct the preliminary investigation into the psychometric properties of the WISC-V and KABC-II with children from remedial schools only. Following this an attempt was made to increase the sample of participants from remedial schools and with the assistance of one of the schools in particular, consent was obtained to assess an additional 31 participants. In the end, a total of 90 participants were assessed on the WISC-V and 89 on the KABC-II. One participant was thus not assessed on the KABC-II, once again due to the COVID-19 pandemic and its third wave which brought a halt to further testing.

The participating schools, as well as the 90 participants, were selected through non-probability, purposive sampling. Ary et al. (2010), Morgan and Sklar (2012a) and Salkind (2018) note that when non-probability sampling is employed, the probability of selecting a specific participant from the broader population cannot be determined. Purposive sampling is a form of non-probability sampling where the participants are selected in a purposeful or targeted manner, based on specific inclusion and exclusion criteria (Best & Khan, 2006; Creswell, 2012; Fraenkel et al., 2019). Given the aims and objectives of the study, children from remedial schools were approached as these children all present with some barriers to learning and development that impact the acquisition and development of essential academic skills. However, by nature of their placement in remedial schools, these children generally do not present with severe intellectual or general developmental delays, severe physical or neurological difficulties, severe internalising or externalising disorders, or other disorders that have a significant impact on children's functioning and warrant placement in more specialised or Full-Service schools. Given the multilingual nature of the South African context, children who are learning in an additional language would be encountered; however, through the inclusion of only private schools in the sample, and the language of learning and teaching in these selected schools being English, the researcher envisaged that most children would have had adequate formal exposure to English. However, the potential impact of second language testing was still investigated further through some of the objectives set.

As far as the selection of the schools is concerned, only private schools in the Johannesburg area were included in the study. This involved a component of convenience sampling as the researcher chose schools in the area where she resided and the schools were thus easily accessible (Cresswell & Guetterman, 2019; Fraenkel et al., 2019; Salkind, 2018). During the selection of individual participants, all children for whom parental consent was obtained were approached, and those who provided assent were included in the sample. By choosing participants from private schools in the Johannesburg area, the researcher hoped to minimise the impact of socio-economic status and quality of education as potential extraneous variables; however, it must be noted that the three schools included, fell in different locations across the broader Johannesburg area and thus the Living Standards Measure (LSM) (Liversage, 2013) as an indication of socio-economic status, was included in the parent questionnaire (see Appendix F). To allow increased exposure to English for children who were assessed and taught in English as an additional language, children in the intermediate (Grades 4-6) and early senior phase (Grades 7-8) of their schooling were approached. The participants' exposure to English as an additional language, was however also explored through the parent questionnaire.

Although it is recognised that this sample is not representative of the broader South African context, it is important to note that it has deliberately been chosen to address as many extraneous variables as possible at the outset of the study. Further, to minimise the possibility of sampling error, the sample size needed to be as large as possible (Creswell, 2012; Salkind, 2012). Mention has already been made of the difficulties experienced in recruiting a bigger sample; however, cognisance was given to the criteria set out by Best and Khan (2006), Cresswell and Guetterman (2019), Fraenkel et al. (2019) and Salkind (2018). Best and Khan (2006) and Salkind (2018) speak to 30 participants as an appropriate number, meaning that at least 30 participants need to partake in a quantitative study. However, Fraenkel et al. (2019) recommend a sample of at least 50 participants in correlational research and 30 participants per group in causal-comparative research, while Cresswell and Guetterman (2019) recommend a sample of approximately 70 participants in correlational research. With 90 participants in total and 32 in the EAL group and 56 in the EFL group (2 participants did not disclose their first language), the guidelines set out by the stated authors, were adhered to.

The characteristics of the sample, with information extracted from the parent questionnaire (Appendix F), are summarised in Table 4. As reported, the final sample was drawn from three private remedial schools in Johannesburg, Gauteng; however, the majority of the sample (*n* = 72) was drawn from a school in the south of Johannesburg. The participants' grade level ranged from Grade 4 to Grade 8, while ages ranged from 9 years, 4 months to 14 years, 9 months (*M* = 11 years, 10 months*; SD* = 1 year 3 months). Although the measurement instruments, the WISC-V and KABC-II, do not make a distinction between sexes in terms of norms, it was still important for reporting purposes, to make this distinction. Most of the participants (*n* = 56) in the sample were male (63.33%), which is not representative of the general population where 50,21% of children between the ages of 10 – 14 years were reportedly male in 2022 (Statistics South Africa, 2023).

As first language can play a role in the performance of children on intelligence tests, it was important to consider the constitution of the sample in terms of home language. The results indicate that most of the sample (*n* = 56) spoke English as a first language (Table 4). To investigate the role of home language, the sample was divided into EFL (63.64%) and EAL groups (36.36%) for data analysis and comparisons. Another demographic factor that may have an impact on test performance was parents' level of education. Levels ranged from Primary School to Doctoral Degree. From the 83 participants' mothers who had disclosed their level of education, a total of 77 participants' mothers (92.77%) had completed high school. Of the 71 participants' fathers who had disclosed their level of education, a total of 64 participants' fathers (90.14%) indicated that they had at least completed high school. Again, this is not representative of the broader South African population, were only 37.6% of the population completed high school in 2022 (Statistics South Africa, 2023).

The final demographic factor that had to be considered, was the socio-economic status of the sample. This was important to consider, not only because socio-economic factors may have an impact on test performance, but also due to the fact that the three schools that were included in the sample, were located in different areas in Johannesburg. In considering the socio-economic status of the sample, the LSM (Liversage, 2013) was used to explore the living standards of the participants as a component of socio-economic status. Although the levels ranged from 7 High to 10 High, the majority of the sample fell in the 9 High to 10 High ranges, indicating that 87% of the sample (*n* = 74) who completed the LSM (*n* = 85), fell within the highest LSM groups (9 and 10) and that none of the participants fell below the 7 High Group (Table 4). Individuals in the LSM 7–10 categories generally experience a higher standard of living, reside predominantly in urban areas, and have access to essential infrastructure, adequate sanitation, and a wide range of household appliances. They also tend to have higher levels of education and employment, access to private transportation, and greater exposure to technology, including television, smartphones, and the internet (Ntloedibe & Ngqinani, 2020). The LSM will be further discussed in the section on the measurement instruments below.

In terms of the barriers to learning and development experienced by the participants, a range of challenges were noted. The most prevalent barriers to learning and development identified in this sample, included specific learning disorder (SLD) or learning difficulties (*n* = 35) and attention-deficit/hyperactivity disorder (ADHD) (*n* = 38). Of these, 11 participants experienced comorbid ADHD and SLD or learning difficulties. It is interesting that although international literature often identifies the difficulties that impact children’s acquisition of essential academic skills as learning disabilities, impairments, or disorders (see Chapter 2), the majority of the children in this sample, had diagnoses of ADHD. Although, from personal experience and working as an Educational Psychologist in South African for over 20 years, the researcher is aware that South African children in remedial schools are often not formally diagnosed with learning difficulties or disorders as this is not required for placement. ADHD is however one of the most recognisable and diagnosed disorders amongst South African children (Kern et al., 2015). Also, there is a very high comorbidity between SLD or learning challenges and ADHD (Becker et al., 2021), with Nel and Grosser (2016) including difficulties in attention, impulsivity, and hyperactivity as part of their criteria for a learning disability in South Africa. Moreover, the Diagnostic and Statistical Manual of Mental Disorders – Fifth Edition, Text Revision (DSM-5-TR) (American Psychiatric Association [APA], 2022) notes that children diagnosed with ADHD, also regularly experience academic challenges, while children diagnosed with SLD often experiences difficulties in attention. Thus, the higher number of participants with learning disorders or difficulties and/or ADHD in this sample is understandable.

Table 4

Sample Characteristics

Sample Characteristics		Total	Percentage
Schools (*N* = 90)	Southern Johannesburg School	72	80%
	Northern Johannesburg School	14	15.56%
	Central Johannesburg School	4	4.44%
Grades (*N* = 90)	4	28	31.11%
	5	20	22.22%
	6	27	30%
	7	10	11.11%
	8	5	5.56%
Sex (*N* = 90)	Male	57	63.33%
	Female	33	36.67%
First Language (*n* = 88)	English	56	63.64%
	IsiZulu	13	14.77%
	Sotho	9	10.23%
	IsiXhosa	5	5.68%
	Setswana	3	3.41%
	Afrikaans	1	1.14%
	Shona	1	1.14%
	Unknown	2	
Language Groups (*n* = 88)	English First Language	56	63.64%
	English Additional Language	32	36.36%
	Unknown	2	
Mother's Education (*n* = 83)	Primary School (Grade 7 or less)	1	1.20%
	Grade 8 or 9	3	3.61%
	Grade 10 or 11	2	2.41%
	Grade 12 or the Equivalent	14	16.87%
	Certificate	3	3.61%
	Higher Certificate	8	9.64%
	Diploma	19	22.89%
	Postgraduate Diploma	6	7.23%
	Bachelor's Degree	14	16.87%
	Professional First-Degree Postgraduate)	3	3.61%
	Honour's Degree	4	4.82%
	Master's Degree	6	7.23%
	Unknown	7	

Sample Characteristics

Father's Education (*n* = 71)	Primary School (Grade 7 or less)	1	1.41%
	Grade 8 or 9	2	2.82%
	Grade 10 or 11	4	5.63%
	Grade 12 or the Equivalent	16	22.54%
	Certificate	1	1.41%
	Higher Certificate	7	9.86%
	Diploma	7	9.86%
	Postgraduate Diploma	6	8.45%
	Bachelor's Degree	12	16.90%
	Professional First-Degree Postgraduate)	2	2.82%
	Honour's Degree	3	4.23%
	Master's Degree	9	12.68%
	Doctoral Degree	1	1.41%
	Unknown	19	
Living Standards Measure (*n* = 85)	LSM – 7 High	1	1.18%
	LSM – 8 Low	1	1.18%
	LSM – 8 High	9	10.59%
	LSM – 9 Low	7	8.24%
	LSM – 9 High	14	16.47%
	LSM – 10 Low	27	31.76%
	LSM – 10 High	26	30.59%
	Unknown	5	
Medical Conditions (*n* = 22)	Yes	22	24.72%
	No	67	75.28%
	Unknown	1	
Physical Disorder/Difficulty (*n* = 6)	Yes	6	7.23%
	No	77	92.77%
	Unknown	7	
Language Disorder/Difficulty (*n* = 19)	Yes	19	23.17%
	No	63	76.83%
	Unknown	8	
Sensory Disorder/Difficulty (*n* = 16)	Yes	16	19.28%
	No	67	80.72%
	Unknown	7	
Emotional Disorder/Difficulty (*n* = 25)	Yes	25	29.07%
	No	61	70.93%
	Unknown	4	
ADHD (*n* = 38)	Yes	38	43.68%
	No	49	56.32%
	Unknown	3	
Specific Learning Disorder/Difficulty (*n* = 35)	Yes	35	40.70%
	No	51	59.30%
	Unknown	4	
Additional Barriers (*n* = 23)	Yes	23	27.71%
	No	60	72.29%
	Unknown	7	

3.6. Measures and Measurement Instruments

3.6.1. Parent Questionnaire

Guided by previous research (Amod, 2013; Amod & Seabi, 2013; Cockcroft et al., 2008; Nisbett et al., 2012; Fagan & Holland, 2007; Kaufman & Kaufman, 2004; Russell et al., 2008; Seabi & Amod, 2009; Shuttleworth-Edwards, Van der Merwe, et al., 2013), parents/legal guardians were asked to complete a parent questionnaire aimed at exploring various variables. Developed for the purpose of this study, the parent questionnaire (Appendix F) gathered information on participants' sex, age, grade, parent education level, and first and second language. In addition, information pertaining to the participants' medical and diagnostic history, emotional and intellectual wellbeing, and sensory development were gathered. The LSM (Liversage, 2013) was also included in the parent questionnaire to investigate living standards as a component of socio-economic status. The measure included several items and parents needed to indicate whether these were present in the home or not, e.g., Home Security Service, Vacuum Cleaner, and Electric Stove. In addition, it also includes statements such as, "My child is a city dweller" which parents just needed to mark as either true or false. The LSM is a South African measure and was scored according to the guidelines provided by the South African Audience Research Foundation (2010). Depending on their scores, participants could be divided into one of 10 LSM groups, with 1 being the lowest and 10 being the highest. Groups 7 – 10 are divided into two sub-groups, high and low (for example, 10 High and 10 Low). For the study, high groups were captured as for instance 10.5 and low groups just as 10.

3.6.2. Wechsler Intelligence Scale for Children – Fifth Edition

As the WISC-V was discussed in depth in the preceding chapter and an overview was provided of research related to its psychometric properties, including an overview of the original WISC-V's psychometric properties and research related to this, this section will only provide an overview of the subtests included in this study. As Wechsler (2014c) recommends that all 10 primary subtests be administered in obtaining a comprehensive profile of a child's intellectual functioning, all 10 primary subtests were included in the assessment of the participants, including Similarities, Vocabulary, Block Design, Visual Puzzles, Matrix Reasoning, Figure Weights, Digit Span, Picture Span, Coding and Symbol Search. Through the inclusion of the 10 primary subtests, the five primary indices could be calculated. These include the Verbal Comprehension, Visual-Spatial, Fluid Reasoning, Working Memory and Processing Speed Indices. It also allowed for the calculation of three of the five ancillary indices, namely the Nonverbal, General Ability, and Cognitive Proficiency Indices. Given their importance in calculating the remaining two of the five ancillary indices, the Quantitative Reasoning and Auditory Working Memory Indices, the Arithmetic and Letter-Number Sequencing subtests were also administered to the majority of participants. However, for some participants who had been assessed previously, these scores were not available and thus not included in the analyses. Some participants also completed the Comprehension and Information subtests, however, as most of the participants were assessed on two intelligence tests, time was a consideration and thus these were not included for all participants. Further, as it was felt that the inclusion of the Symbol Translation subtests made the assessment very long, these were also not included in the study. However, as the Naming Speed subtests, Naming Speed Literacy and Naming Speed Quantity, are quick to administer and complete, they were added were possible. The WISC-V administered to this sample, was the original WISC-V developed in the United States (US). The participants' scores were interpreted according to the guidelines provided by Flanagan and Alfonso (2017) (see Table 5 below). The FSIQ, the indices, and the two Naming Speed subtests, have a theoretical mean of 100 and a standard deviation of 15, while all the other subtests have a theoretical mean of 10 and a standard deviation of three.

Table 5

Interpretation of WISC-V Indices and Subtests

Composite Categories	Subtest Categories	Qualitative Interpretations
≥ 130	16 – 19	Extremely High
120 - 129	14 – 15	Very High
110 - 119	13	High Average
90 - 109	8 – 12	Average
80 - 89	7	Low Average
70 - 79	5 – 6	Very Low
≤ 69	1 – 4	Extremely Low

Note. Adapted from "How to Interpret the WISC-V," by D.P. Flanagan and V.C. Alfonso, *Essentials of WISC-V Assessment* (p. 172), by D.P. Flanagan and V.C. Alfonso (Eds.), 2017, John Wiley and Sons, Inc. Copyright 2017 by John Wiley & Sons, Inc.

3.6.3. Kaufman Assessment Battery for Children – Second Edition (KABC-II)

Similar to the WISC-V, Chapter 2 provided a more detailed description of the KABC-II and its psychometric properties, including data obtained from the normative sample. Here an overview of the subtests administered in this study will be provided. As discussed in Chapter 2, the KABC-II renders two global scores. Whereas the Mental Processing Index (MPI) is based on Luria's neuropsychological model and does not contain a measure of acquired knowledge or crystallised intelligence, the Fluid-Crystallised Index (FCI) is based on the Cattell-Horn-Carroll (CHC) theory and therefore contains the Knowledge Scale, a measure of acquired knowledge and verbal comprehension (Kaufman & Kaufman, 2004). Both of these were calculated for this sample and thus, all the core subtests for the two older age groups (ages 7 – 12 years and ages 13 – 18 years) were administered to the majority of participants. These included Atlantis, Rebus, Number Recall, Word Order, Story Completion, Pattern Reasoning, Rover, Triangles, Block Counting, Riddles, and Verbal Knowledge. As stated earlier, the KABC-II was administered to 89 of the total 90 participants because another COVID-19 wave brought the data collection process to a halt. Also, as the Triangles subtest is considered a core subtest in the 7 – 12-year-old group, this was administered to 87 participants. Similarly, as Block Counting is considered a core subtest in the 13 – 18-year-old group, this subtest was administered to only 84 participants. Where time and the participants' observed enthusiasm and energy allowed, these subtests were administered regardless of the participants ages. Another factor that was taken into consideration throughout, was the participants' emotional wellbeing. As such, one participant was not assessed on the entire KABC-II. The assessment was discontinued due to the participant's high levels of anxiety, observed by the researcher. This participant completed the Atlantis, Story Completion, Number Recall, Rover, Verbal Knowledge and Rebus subtests, which allowed for the calculation of the Learning Scale, but none of the other indices. In an individual feedback interview, the participant's heightened level of anxiety was discussed with their parents and recommendations for support made. Like the WISC-V, the US norms of the original KABC-II were used in this study. As can be seen from Table 6 below, the KABC-II does not follow the same 10-point intervals as the WISC-V in qualitatively interpreting children's scores. Rather, qualitative interpretations are based on the number of standard deviations from the mean of 100. However, the two KABC-II global scores, the MPI and the FCI, as well as the five scale scores, also have a theoretical mean of 100 and a standard deviation of 15 and the subtest scores, a theoretical mean of 10 and a standard deviation of three. For the purpose of this study, subtest categories were calculated as they were for the WISC-V, although, cognisance must be taken that these were not explicitly described in the KABC-II manual (Kaufman & Kaufman, 2004; Kaufman et al., 2005c).

Table 6

Interpretation of KABC-II Indices and Subtests

Composite Categories	Subtest Categories	Qualitative Interpretations
131 – 160	16 – 19	Upper Extreme
116 – 130	13 – 15	Above Average
85 – 115	8 – 12	Average
70 – 84	5 – 7	Below Average
40 - 69	1 – 4	Lower Extreme

Note. Adapted from "How to interpret the KABC-II: Step by Step," by A.S. Kaufman, E.O. Lichtenberger, E. Fletzer-Jansen, and N.L. Kaufman, *Essentials of KABC-II Assessment* (p. 85), by A.S. Kaufman, E.O. Lichtenberger, E. Fletcher-Janzen, and N.L. Kaufman, 2005, John Wiley & Sons, Inc. Copyright 2005 by John Wiley & Sons, Inc.

3.7. Procedure

Provisional permission was obtained from the three schools. As the schools are all private schools, the researcher did not need to obtain permission from the Gauteng Department of Education. Following this, ethical clearance could be applied for and obtained (see Appendixes B and C). The ethics committee required that the researcher obtain provisional permission from the schools prior to applying for ethical clearance. Thereafter, the parents/legal guardians of the participants were approached through the parent information and consent forms (see Appendix D). Only once consent had been obtained from parents/legal guardians, were participants approached. Their assent was obtained through an information and assent letter (see Appendix E). By virtue of their placement in private schools, some of the participants had undergone an intelligence test prior to admission, which were confirmed by their school records. If participants had been assessed on the WISC-V or KABC-II 18 months prior to the date of data collection, these intelligence tests were not re-administered. Although some children had been assessed on the WISC-V, none of the children had been assessed on the KABC-II and thus all participants, aside from one, were assessed on this measure.

Participants were assessed individually by either registered educational or intern educational psychologists. Intern psychologists worked under the supervision of a registered educational psychologist. Participants were all assessed in private offices/rooms at their respective schools. Each of the assessment measures took approximately 70-90 minutes to administer; however, the duration varied depending on the participants' ages, pace, number of breaks required, and cognitive abilities. As the participants' assessment results could potentially be influenced by the different practitioners involved, the number of practitioners in the study was limited in any of the given schools. In addition, all practitioners were briefed to ensure consistency in the administration of the intelligence tests, as far as possible. As recommended by Wechsler (2014b) the order of the two intelligence tests was changed across the sample, with 46 participants (51.69%) assessed first on the KABC-II and then on the WISC-V, and 43 participants (48.31%) first on the WISC-V and then on the KABC-II (keeping in mind that one participant was not assessed on the KABC-II). The average length of time between the two assessment measures was 66 days, taking cognisance that some of the participants had been assessed before. Once the assessments had been conducted, the researcher checked the scoring on all the assessments to ensure accuracy and consistency, and then transferred raw data obtained into digital format for the purpose of data analysis. In digital format, all identifying particulars were removed. Parents were also offered feedback, either by the researcher or by the different psychologists / intern psychologists at the participating schools.

3.8. Statistical Analysis

Data analysis was carried out using SAS version 9.4 for Windows. Prior to investigating the four objectives of this study, descriptive statistics which included means, standard deviations, minimums, and maximums were calculated for the global, index/scale, and subtest scores. Thereafter, the internal consistency reliability of the WISC-V and KABC-II subtests was determined by the Spearman-Brown split-half reliability method. Split-half comparisons is a convenient way to measure reliability as it requires only one test administration. The disadvantage of this approach is that it can only be used for subtests that consist of multiple items. For this reason, split-half comparisons could not be calculated for the

WISC-V Coding, Symbol Search, Naming Speed Quantity and Naming Speed Literacy subtests. Reliabilities for these four subtests can only be calculated by means of a test-retest method, which was not possible, for logistical reasons, in this study.

The second objective of the study required an investigation of intercorrelations between subtests, indices/scales and global scores and were determined by Pearson's correlation coefficients. Thereafter, the extent to which the WISC-V and KABC-II fitted their respective models was formally tested using a second-order CFA. The level of significance of the path weights between each observed variable and each latent factor was set to $\alpha = 0.05$. Univariate and multivariate normality in the data was assessed through the Henze-Zirkler and Shapiro-Wilk tests. The null hypothesis significance of the chi square test ($\chi 2$) was considered for each model, which compared the degree to which predicted and observed covariances differed. Small and non-significant chi-square values indicated a good fit. As sample size and data distribution could have a meaningful impact on this statistic, model adequacy was determined through additional global fit indices. Indices that are more susceptible to model specification included the "Comparative Fit Index (CFI) as a goodness-of-fit index based on model comparisons, the Normed Fit Index (NFI), the Bollen Fit Index (IFI), the Goodness of Fit Index (GFI), and the Adjusted Goodness of Fit Index" (Hu & Bentler, 1999, pp. 2-3). Fit indices with values equal to or higher than .90 demonstrated a good fit. The Root Mean Square Error of Approximation (RMSEA) was also considered as it provides a descriptive measure of the overall model fit, with a RMSEA of .06 or lower seen as a good fit (Brown, 2015; Hu & Bentler, 1999).

In relation to the third objective, selected overall and index/scale scores were compared to each other using a repeated measures GLM with Tukey-Kramer adjustment for multiple comparisons. These scores were compared according to English language status using a GLM with score as the dependent variable and English language status as the independent variable. Finally, in terms of the fourth objective, construct validity of the KABC-II and WISC-V test items were investigated through a graded

IRT model or a Graded Response Model (GRM) for all items (since most subtests consisted of both dichotomous and ordered items), with all slopes constrained to be equal. In the absence of existing studies on the item validity of both the WISC-V and the KABC-II within the South African context, item validity was calculated for the total sample. Subsequently, DIF was calculated separately for the two groups to determine the extent to which language impacts item performance. DIF was determined by a linear mixed model with items and English language status as fixed effects, and the participant as random effect. First, a null model (intercept-only) was applied to the data, which specified that all item difficulty parameters are the same. Second, all items were added to the model and thus the difficulty parameters for items were allowed to vary. Third, the interaction effects between English language status and items were added to the model to detect DIF. Since the models were nested, the likelihood ratio test was used to compare the likelihood of pairs of models. Items showed DIF in overall tests if the model fit statistics were improved; however, specific items with DIF were identified by inspecting the significance of the item-English status interactions. Cross-tabulations were then inspected to determine which language group was favoured.

3.9. Ethical Considerations

Research involving human participants, especially those younger than the age of 18 years, should adhere to strict ethical guidelines (Ary et al., 2019; Clark-Carter, 2004). These include "informed consent, confidentiality, anonymity, protection from harm and access to results" (Morgan & Sklar, 2012b, p. 121). Ethical clearance was obtained from the University of the Witwatersrand, Human Research Ethics Committee – Non-Medical (Protocol Number 16/02/37) (see Appendix B). Informed and written consent was obtained through several information and consent forms, which stipulated the study's aims, procedures, researcher's expectations, and roles and responsibilities of all persons involved. In addition to voluntary participation, these information letters also emphasises confidentiality and anonymity. Participants were afforded the right to withdraw from the study and not respond to any questions they did not feel comfortable with, without any consequences to them or their schools, until the analyses of the data and the inclusion of the results in this thesis. The details included in the participant information letter were also discussed in-person with all participants, who were given an opportunity to raise any concerns.

In terms of confidentiality, participants' actual test results, the raw data, were only seen by the researcher and the educational psychologists or intern educational psychologists who assisted in data collection and provided feedback to parents as needed and requested. Raw data is stored in a locked cupboard at the University of the Witwatersrand and electronic data on a password protected computer. Both intelligence tests have the potential to identify cognitive strengths and weaknesses in children, which can point towards a potential learning difficulty and influence their scholastic progress. Parents/legal guardians were informed of this prior to providing consent and feedback was offered to all parents/legal guardians. Moreover, should any concerns regarding the participants' intellectual functioning arise, parents/legal guardians would be informed accordingly. Feedback sessions were offered at a convenient time for the parents/legal guardians. When the researcher provided feedback to parents/legal guardians, their written consent was obtained to share this feedback with the participants' schools so that children could be supported appropriately. In instances where school-based educational psychologists/intern educational psychologists provided feedback, parents were informed prior to the assessment.

There were no anticipated risks, benefits or exposure to harm for the participants. Although, as stated, parents/legal guardians could access their children's results for the purpose of supporting them academically. With the collection of data during the COVID-19 pandemic, all school procedures were strictly adhered to for the protection of the researcher, the intern educational psychologist who assisted with data collection at the time, and the participants. Additional measures, such as the sanitisation of equipment, test materials and desks were also put in place. Further, as stated earlier, in preventing distress, one assessment was discontinued given the participant's heightened levels of test anxiety. In addition, the situation was immediately managed by a registered educational psychologist, the parents were informed, and the participant was referred for further psychotherapy with the school's educational psychologist as such test anxiety could also impact her performance on tests and exams.

Data collection was scheduled at a time that would not negatively impact on the academic or other related activities of the schools or the participants. Raw data was transferred into digital format, with any information that could identify the schools, the participants, and/or their parents/legal guardians removed. In digital format, participants' tests scores and personal information were linked by means of codes. As such, anonymity of the schools, parents/legal guardians, and the participants in the final thesis and other potential publications, were ensured.

The results of the study have been compiled into this thesis, which will be stored in the University of the Witwatersrand main library and made available online on the WITS Institutional Repository on WIReDSpace. Publication of a peer reviewed journal article is considered. This was disclosed to all relevant parties in the stated information letters.

3.10. Conclusion

A quantitative research thesis follows a specific procedure, starting with the aims, objectives, research questions, and a review of literature appropriate to the topic. This is then followed by the information presented in this chapter, the research methods, and data analyses techniques. As noted, these are situated within a specific approach, which incorporates the research paradigm, the methodology, and the research design. Based on the aims, objectives and research questions, this study followed a quantitative research approach, situated within a post-positivist research paradigm. Quantitative methods were employed in data collection through the administration of two intelligence tests. Finally, appropriate research designs were selected to guide the study. Given the four objectives of the study, a correlational research design and a causal-comparative design, were deemed appropriate. In turn these designs guided the data analysis techniques, which focussed on investigating the research questions through reliability coefficients, correlations, factor analyses, GLM comparisons, and item analysis. The chapter to follow will present the findings of these analyses.

4. Results

4.1. Introduction

The results of the study are broadly divided into descriptive and inferential statistics. Building on the statistical analysis presented in Chapter 3, which provided a more detailed description of the sample, the descriptive statistics here offer an overview of the participants' scores obtained on the Wechsler Intelligence Scale for Children – Fifth Edition (WISC-V) and the Kaufman Assessment Battery for Children – Second Edition (KABC-II). These include the means, standard deviations, minimums, and maximums calculated for the participants' global, index/scale, and subtest scores. Thereafter, the results of inferential analyses are presented in relation to the study's four objectives, as outlined in Chapters 1 and 3. These included reliability coefficients determined through the Spearman-Brown split-half reliability method, Pearson correlation coefficients, confirmatory factor analyses (CFA), General Linear Models to investigate differences in language status and graded item response theory models to further investigate the subtests' items and determine whether these functioned differently in the two language groups, English First Language (EFL) and English Additional Language (EAL).

4.2. WISC-V Descriptive Statistics

The WISC-V global, index, and subtest scores, including their mean, standard deviation, minimum, and maximum values, are presented below (see Table 7). The participants of this study (N = 90) achieved a mean Full Scale Intelligence Quotient (FSIQ) score of 78.4 (SD = 13.7), which falls in the very low range and is almost 22 standard points lower than the mean of 100. Since a person's global cognitive ability encompasses a range of specific cognitive skills, attention is directed to the five primary indices. Although all of the mean index scores fall within the low average range, the highest mean score was recorded on the Visual-Spatial Index (M = 86.2, SD = 17.5), which is more nonverbal and visual in nature, and the lowest mean score was observed on the Verbal Comprehension Index (M = 79.9, SD = 15.1), which is more verbal in nature. Similarly, as far as the ancillary indices are concerned, the lowest mean score was obtained on the Auditory Working Memory Index (*M* = 82.0, *SD* = 13.3), which is again more verbal in nature, while the highest mean score was obtained on the Nonverbal Index (*M* = 82.7, *SD* = 16.2), a measure of nonverbal and visual-spatial processing and reasoning. Moreover, the participants' performance on the WISC-V subtests also aligns with these findings. The highest mean subtest score (*M* = 7.9) was obtained on three visual and more nonverbal subtests: Visual Puzzles (*SD* = 3.2), Matrix Reasoning (*SD* = 3.3), and Picture Span (*SD* = 3.8). In contrast, the lowest rounded mean scores included three subtests that depend on English language proficiency, crystallised intelligence, and acquired knowledge. These subtests were Similarities (*M* = 6.3, *SD* = 2.9), Vocabulary (*M* = 6.3, *SD* = 3.0), and Comprehension (*M* = 5.5, *SD* = 2.9). In addition, participants also performed poorly on the Arithmetic subtest (*M* = 5.6, *SD* = 2.6). Although this forms part of the Fluid Reasoning Index, it relies heavily on auditory working memory and processing of verbal information. The distinction between components of the Digit Span subtest also highlighted differences in cognitive abilities that impact performance on these components. Participants struggled the most with the Digit Span Backward (*M* = 7.1, *SD* = 2.7) and Digit Span Sequencing (*M* = 6.9, *SD* = 2.6) components, where the rounded mean scores were in the (very low range). In contrast to the Digit Span Forward component, these components rely more on auditory working memory than on memory span. Performance on the Digit Span Backward and Digit Span Sequencing components was however consistent with the Letter-Number Sequencing subtest (*M* = 7.1, *SD* = 2.6), another measure of auditory working memory.

Considering that the theoretical mean index score of the WISC-V is 100, with a standard deviation of 15, most rounded mean index scores fell one standard deviation below the theoretical mean. Exceptions were the Visual-Spatial Index (*M* = 86.2, *SD* = 17.5), the Fluid Reasoning Index (*M* = 86.0, *SD* = 14.4), and the Naming Speed Index (*M* = 84.5, *SD* = 13.6). Participants' performance on the Naming Speed Index appeared to be supported more by their performance on the Naming Speed Quantity subtest, with a mean score of 86.4 (*SD* = 17.1) than by the Naming Speed Literacy subtest with a mean score 82.8 (*SD* = 17.0). Moreover, with a theoretical mean score of 10 for each subtest and a standard deviation of 3, the rounded mean scores of most subtests were also one standard deviation below the theoretical mean, except for the three highest scoring subtests, Visual Puzzles (*M* = 7.9, *SD* = 3.2), Matrix Reasoning (*M* = 7.9, *SD* = 3.3) and Picture Span (*M* = 7.9, *SD* = 3.8) and the Digit Span Forward component (*M* = 7.7, *SD* = 2.9).

Table 7

Descriptive Statistics for the WISC-V

Indices and Subtests	*N/n*	*Min*	*Max*	*M*	*SD*
Full Scale	90	47	111	78.4	13.7
Verbal Comprehension	90	45	130	79.9	15.1
Similarities	90	1	17	6.3	2.9
Vocabulary	90	1	15	6.3	3.0
Information	44	1	12	6.9	2.3
Comprehension	45	0	13	5.5	2.9
Visual Spatial	90	49	122	86.2	17.5
Block Design	90	1	15	7.2	3.6
Visual Puzzles	90	2	15	7.9	3.2
Fluid Reasoning	90	55	123	86.0	14.4
Matrix Reasoning	90	1	14	7.9	3.3
Figure Weights	90	3	14	7.3	2.3
Arithmetic	75	1	12	5.6	2.6
Working Memory	90	45	122	84.0	15.7
Digit Span	90	1	12	6.5	2.5
Digit Span Forward	90	1	18	7.7	2.9
Digit Span Backward	90	1	13	7.1	2.7
Digit Span Sequencing	90	1	14	6.9	2.6
Picture Span	90	1	18	7.9	3.8
Letter Number Sequencing	75	1	12	7.1	2.6
Processing Speed	90	45	108	82.7	15.4
Coding	90	1	12	6.8	2.9
Symbol Search	90	1	13	7.1	3.0
Quantitative Reasoning	75	57	117	80.4	12.0
Auditory Working Memory	75	45	108	82.0	13.3
Nonverbal	90	49	121	82.7	16.2
General Ability	90	56	123	80.9	14.5
Cognitive Proficiency	90	40	114	80.4	15.4
Naming Speed	72	48	122	84.5	13.6
Naming Speed Literacy	72	45	124	82.8	17.0
Naming Speed Quantity	72	24	118	86.4	17.1

4.3. KABC-II Descriptive Statistics

The KABC-II was administered to 89 participants; however, as stated, the assessment was discontinued for one participant. Consequently, the global Fluid-Crystallised Index (FCI) and Mental Processing Index (MPI) scores, along with scores for four of the five scales (Knowledge, Simultaneous, Sequential, and Planning), were calculated for 88 participants. On the other hand, statistics for the Learning Scale included all 89 participants. Table 8 summarises the global (FCI and MPI), scale, and subtest scores, reporting their means, standard deviations, minimums, and maximums. The data indicate that participants (*n* = 88) obtained a higher mean MPI score of 86.4 (*SD* = 12.6) compared to their FCI mean score of 84.3 (*SD* = 12.1). The highest mean scale score was 96.6 (*SD* = 13.3) on the Learning Scale. In contrast, the lowest mean scale score, 82.2 (*SD* = 11.5) was obtained on the Knowledge Scale, a measure strongly influenced by English language proficiency, crystallised intelligence, and acquired knowledge. As stated in Chapter 3, the qualitative interpretations used for the KABC-II (FCI, MPI, and scales), differ slightly from those of the WISC-V. Most rounded mean scale scores, as well as the MPI, were within the average range (85 – 115). Notably, only the FCI and the Knowledge Scale fell in the below average range (70 – 84). Regarding subtest scores, most rounded mean scores fell within the average range (8 – 12). However, the rounded mean scores for Riddles (*M* = 6.1, *SD* = 2.3), Verbal Knowledge (*M* = 7.0, *SD* = 2.4), and Pattern Reasoning (*M* = 7.4, *SD* = 2.8) were in the very low range, one standard deviation below the theoretical mean.

Table 8

Descriptive Statistics for the KABC-II

Indices and Subtests	*N/n*	*Min*	*Max*	*M*	*SD*
Fluid Crystallised Index	88	60	112	84.3	12.1
Mental Processing Index	88	60	113	86.4	12.6
Sequential Scale	88	63	112	87.5	13.6
Number Recall	89	3	14	8.0	2.8
Word Order	88	2	14	7.7	2.5
Planning Scale	88	51	128	86.7	14.5
Story Completion	89	1	16	8.0	3.2
Pattern Reasoning	88	1	13	7.4	2.8
Simultaneous Scale	88	43	127	88.0	18.2
Rover	89	1	17	8.1	3.8
Triangles	87	1	15	7.8	3.1
Block Counting	84	0	16	8.0	3.5
Learning Scale	89	61	126	96.6	13.3
Atlantis	89	4	18	10.2	3.1
Rebus	89	1	15	8.7	2.4
Knowledge Scale	88	48	117	82.2	11.5
Verbal Knowledge	89	1	14	7.0	2.4
Riddles	88	1	13	6.1	2.3

4.4. Objective 1: Internal Consistency Reliability

Reliability of each subtest was determined by the Spearman-Brown split-half reliability method, with the exception of the Coding, Symbol Search, Naming Speed Literacy, and Naming Speed Quantity subtests of the WISC-V which are not divided into individual subtest items. In addition, the three components of the Digit Span subtest were considered individually. Guidelines for the interpretation of reliability scores were provided in Chapter 2.

4.4.1. WISC-V Internal Consistency Reliability

Appendix G, G1 summarises the number of participants whose subtest items could be considered for inclusion in these analyses, the number of items per subtest, and the Spearman-Brown split-half reliability scores for each of the WISC-V subtests included for these analyses. As reported in Chapter 3, all participants completed the 10 primary WISC-V subtests. However, only 83.3% (*n* = 75) completed the Arithmetic and Letter-Number Sequencing subtests, 50% (*n* = 45) the Comprehension subtest, and 48.9% (n = 44) the Information subtest. Despite this, not all of the participants' subtests could be included for these analyses as item scores were not available for some participants who had been previously assessed. Although the researcher tried to obtain this missing item data, this was not possible for all participants. From the 14 subtests included in these analyses, 10 showed excellent internal consistency reliability (≥ .90), whilst the remaining four showed good internal consistency reliability (≥. 80). It is interesting to note that the three components of the Digit Span subtest, included Digit Span Forward (.84), Digit Span Backward (.87), and Digit Span Sequencing (.87) are included amongst the latter, with Comprehension (.89) being the fourth subtest.

4.4.2. KABC-II Internal Consistency Reliability

As was done with the WISC-V, Appendix G, Table G2 summarises the number of participants, whose subtests' items could be considered for inclusion in these analyses, the number of items per subtest, and the Spearman-Brown split-half reliability scores for each of the included KABC-II subtests. As stated in Chapter 3, the KABC-II was administered to 89 participants; however, the assessment of one participant was not completed. Thus, for most of the KABC-II subtests, the data of 88 participants were included in these analyses. The exceptions are Atlantis, Number Recall, Rover, Verbal Knowledge, and Rebus. The items of these subtests could be included for the total number of participants to whom the KABC-II was administered (n = 89). Two other exceptions are Triangles (n = 87) and Block Counting (n = 84). Of the 11 subtests administered as part of the KABC-II, eight achieved excellent internal consistency reliability scores (≥ .90), whilst the other three obtained good internal consistency reliability scores (≥.80). The lowest scores were obtained on the two Sequential Scale subtests, Number Recall (.81) and Word Order (.88), although these are still considered good. Regardless, it is interesting that, similar to the WISC-V where the Digit Span components obtained some of the lowest reliability scores, the lowest reliability scores for the KABC-II are again for subtests that rely on sequential processing, working memory, and memory span.

4.5. Objective 2: Construct Validity: Correlation Coefficients and Confirmatory Factor Analysis

In investigating the construct validity of the WISC-V and KABC-II in this sample, intercorrelations between global, indices/scales, and subtests were determined through Pearson correlation coefficients. Thereafter, factorial validity was investigated through second-order confirmatory factor analysis (CFA). The criteria used in the interpretation of the correlation coefficients are presented in Appendix G Table G3.

4.5.1. WISC-V Correlation Coefficients

Appendix G, Table G4 summarises the correlations between the WISC-V FSIQ, primary indices, General Ability Index and Cognitive Proficiency Index. The WISC-V FSIQ and General Ability Index were very strongly correlated ($r = .96$, $p < .0001$). In addition, these two indices were also very strongly correlated with three of the primary indices, namely Fluid Reasoning, Verbal Comprehension, and Visual Spatial. While the FSIQ showed a substantial correlation with the Working Memory Index ($r = .69$, $p < .0001$), the latter was only moderately correlated with the General Ability Index ($r = .57$, $p < .0001$). Both the FSIQ ($r = .60$, $p < .0001$) and the General Ability Index ($r = .42$, $p < .0001$) showed only moderate correlations with the Processing Speed Index. However, very strong correlations were observed between the Processing Speed ($r = .85$, $p < .0001$) and Working Memory Indices ($r = .86$, $p < .0001$) and the Cognitive Proficiency Index. The Cognitive Proficiency Index showed a substantial correlation with the FSIQ ($r = .75$, $p < .0001$).

In terms of the primary indices, Appendix G, Table G5 indicates that while the Verbal Comprehension Index showed moderate correlations with the Visual Spatial ($r = .59$, $p < .0001$), Fluid Reasoning ($r = .58$, $p < .0001$) and Working Memory Indices ($r = .54$, $p < .0001$), it correlated weakly with the Processing Speed Index ($r = .34$, $p < .0001$). The Visual Spatial Index showed a substantial correlation with the Fluid Reasoning Index ($r = .70$, $p < .0001$). These two indices were moderately correlated with the Working Memory Index; however, while the Visual Spatial Index also showed a moderate correlation with the Processing Speed Index ($r = .45$, $p < .0001$), the latter was weakly correlated with the Fluid Reasoning Index ($r = .39$, $p < .0001$). The Working Memory and Processing Speed Indices were moderately correlated ($r = .48$, $p < .0001$). The Naming Speed Index correlated weakly with all the primary and ancillary indices included in this sample, aside from the Cognitive Proficiency Index where a moderate correlation was noted ($r = .46$, $p < .0001$). Conversely, the Nonverbal Index showed substantial to very strong correlations with all the other primary and ancillary indices, except for the Auditory Working Memory Index where only a moderate correlation ($r = .56$, $p < .0001$) was noted. The strongest correlations were noted between the Nonverbal Index and the General Ability Index ($r = .87$, $p < .0001$), the Visual Spatial Index ($r = .87$, $p < .0001$), and the Fluid Reasoning Index ($r = .85$, $p < .0001$). Similarly, the Quantitative Reasoning Index also showed substantial to very strong correlations with most of the other primary and ancillary indices. Here, the exception was the Processing Speed Index were only a moderate correlation ($r = .47$, $p < .0001$) was observed. The strongest correlations were between the Quantitative Reasoning Index and the Fluid Reasoning Index ($r = .82$, $p < .0001$), and the General Ability Index ($r = .81$, $p < .0001$). Substantial correlations were also found between the Auditory Working Memory Index and the Working Memory Index ($r = .79$, $p < .0001$), as well as the Cognitive Proficiency Index ($r = .73$, $p < .0001$).

Appendix G Tables G6 and G7 present the correlations between the WISC-V indices and the primary, ancillary, and complementary subtests included in this study. Both complementary subtests administered to this sample, Naming Speed Literacy ($r = .91$, $p < .0001$) and Naming Speed Quantity ($r = .82$, $p < .0001$) correlated very strongly with the Naming Speed Index. From the four Verbal Comprehension Index subtests, only Similarities ($r = .92$, $p < .0001$) and Vocabulary ($r = .92$, $p < .0001$) correlated very strongly with this index. Although a substantial correlation was noted between the Verbal Comprehension Index and the Information subtest ($r = .75$, $p < .0001$), only a moderate correlation was noted with the Comprehension subtest ($r = .57$, $p < .0001$). While both Visual Spatial

Index subtests, Block Design ($r = .93$, $p < .0001$) and Visual Puzzles ($r = .91$, $p < .0001$) correlated very strongly with their index, some very strong and substantial correlations were also noted between these subtests and some of the other indices. Most notably is the very strong correlation between the Visual Puzzles subtest and the Nonverbal Index ($r = .82$, $p < .0001$), with which the Block Design subtest also showed a substantial correlation ($r = .77$, $p < .0001$). Other indices that showed substantial correlations with these subtests, were the Fluid Reasoning Index (Block Design: $r = .60$, $p < .0001$; Visual Puzzles: $r = .69$, $p < .0001$) and the General Ability Index (Block Design: $r = .77$, $p < .0001$; Visual Puzzles: $r = .77$, $p < .0001$). Finally, Visual Puzzles also showed a substantial correlation with the Quantitative Reasoning Index ($r = .68$, $p < .0001$).

As far as the Fluid Reasoning Index and Quantitative Reasoning Index subtests in this sample are concerned, Figure Weights showed very strong correlations with the Fluid Reasoning Index ($r = .87$, $p < .0001$) and the Quantitative Reasoning Index ($r = .87$, $p < .0001$), in addition to substantial correlations with the Visual Spatial Index ($r = .63$, $p < .0001$), Nonverbal Index ($r = .73$, $p < .0001$), and the General Ability Index ($r = .75$, $p < .0001$). Whilst the Arithmetic subtest showed a very strong correlation with the Quantitative Reasoning Index ($r = .91$, $p < .0001$), it also showed substantial correlations with a number of other indices, including the Nonverbal Index ($r = .69$, $p < .0001$), General Ability Index ($r = .68$, $p < .0001$), Auditory Working Memory Index ($r = .65$, $p < .0001$), Cognitive Proficiency Index ($r = .63$, $p < .0001$), Fluid Reasoning Index ($r= .62$, $p < .0001$), Working Memory Index ($r = .62$, $p < .0001$) and the Verbal Comprehension Index ($r = .62$, $p < .0001$). Although the final Fluid Reasoning Index subtest, Matrix Reasoning does not contribute directly to the calculation of the Quantitative Reasoning Index, a substantial correlation was found between these ($r = .64$, $p < .0001$). However, a very strong correlation was found between this subtest and the Fluid Reasoning Index ($r = .93$, $p < .0001$), as well as the General Ability Index ($r = .81$, $p < .0001$) and the Nonverbal Index ($r = .81$, $p < .0001$). In addition, a substantial correlation was also found between Matrix Reasoning and the Visual Spatial Index ($r = .65$, $p < .0001$).

Whilst the Digit Span subtest showed very strong correlations with the Auditory Working Memory Index in this sample (r = .89, p < .0001) and the Working Memory Index (r = .81, p < .0001), as well as a substantial correlation with the Cognitive Proficiency Index (r = .67, p < .0001), the Digit Span Forward component did not show strong or very strong correlations with any of the indices. The Digit Span Backward component did however show a substantial correlation with the Auditory Working Memory Index (r = .68, p < .0001) and the Working Memory Index (r = .65, p < .0001), whilst the Digit Span Sequencing component showed a very strong correlation with the Auditory Working Memory Index (r = .83, p < .0001) and substantial correlations with the Working Memory Index (r = .72, p < .0001) and the Cognitive Proficiency Index (r = .62, p < .0001). The other Auditory Working Memory Index subtest, Letter-Number Sequencing, showed a similar pattern with a very strong correlation with the Auditory Working Memory Index (r = .90, p < .0001) and substantial correlations with the Working Memory Index (r = .62, p < .0001) and the Cognitive Proficiency Index (r = .62, p < .0001). The visual working memory subtest, Picture Span, showed a very strong correlation with the Working Memory Index (r = .92, p < .0001) and the Cognitive Proficiency Index (r = .82, p < .0001), as well as substantial correlations with the Nonverbal Index (r = .72, p < .0001). The first of the two Processing Speed Index subtests, Coding, showed a very strong correlation with its index (r = .89, p < .0001), as well as a substantial correlation with the Cognitive Proficiency Index (r = .72, p < .0001). The Symbol Search subtest however showed very strong correlations with the Processing Speed Index (r = .90, p < .0001) and the Cognitive Proficiency Index (r = .81, p < .0001), and a substantial correlation with the Nonverbal Index (r = .63, p < .0001).

As far as the intercorrelations between the WISC-V subtests in this sample are concerned (see Appendix G, Table G7), the Naming Speed Literacy and Naming Speed Quantity subtests were substantially correlated (r = .61, p < .0001). Amongst the Verbal Comprehension Index subtests, Similarities and Vocabulary were substantially correlated (r = .71, p < .0001), as were Vocabulary and

Information (r = .71, p < .0001). Whilst Information and Similarities were moderately correlated (r = .58, p < .0001), the Comprehension subtest did not correlate strongly with any of the other three Verbal Comprehension Index subtests. Its highest correlation was a moderate correlation with Vocabulary (r = .51, p < .001). Of particular interest here is the substantial correlation between Arithmetic and Information (r = .64, p = < .001). The Arithmetic subtest also correlated substantially with the Digit Span (r = .62, p < .0001) and Vocabulary (r = .61, p < .0001) subtests, whilst only showing moderate correlations with the other two Fluid Reasoning Index subtests, Figure Weights (r = .59, p < .0001) and Matrix Reasoning (r = .54, p < .0001). It also showed moderate correlations with a number of other subtests, including Visual Puzzles (r = .59, p < .0001), Digit Span Sequencing (r = .57, p < .0001), Letter-Number Sequencing (r = .55, p < .0001), Similarities (r = .53, p < .0001), and Picture Span (r = .51, p < .0001). A substantial correlation was however noted between Figure Weights and Matrix Reasoning (r = .64, p < .0001), as well as between Matrix Reasoning and Visual Puzzles (r = .63, p < .0001). Similarly, Figure Weights also showed a substantial correlation with Visual Puzzles (r = .65, p < .0001), as did Visual Puzzles to Block Design (r = .71, p < .0001). Block Design was moderately correlated with Matrix Reasoning (r = .58, p < .0001) and Figure Weights (r = .51, p < .0001). Interestingly, Matrix Reasoning also showed moderate correlations with the Vocabulary (r = .54, p < .0001) and Picture Span subtests (r = .53, p < .0001). Visual Puzzles also showed a moderate correlation with Vocabulary (r = .59, p < .0001).

In this sample, the two Auditory Working Memory subtests, Letter-Number Sequencing and Digit Span showed a substantial correlation (r = .63, p < .0001). Amongst the three components of the Digit Span subtest, the strongest correlation was seen between Digit Span and Digit Span Sequencing (r = .85, p < .0001); however, substantial correlations were also noted between Digit Span Backward (r = .72, p < .0001) and Digit Span Forward (r = .65, p < .0001). Whilst Letter-Number Sequencing also showed a substantial correlation with Digit Span Sequencing (r = .63, p < .0001), its correlations with the other two Digit Span components are weak (Digit Span Forward, r = .37, p < .001), Digit Span Backward (r

= .45, *p* < .0001). Both Digit Span (*r* = .53, *p* < .0001) and Letter-Number Sequencing (*r* = .51, *p* < .0001) showed moderate correlations with Picture Span. In addition, the Picture Span subtest also showed moderate correlations to the Vocabulary (*r* = .46, *p* < .0001), Visual Puzzles (*r* = .47, *p* < .0001) and Symbol Search (*r* = .46, *p* < .0001) subtests. Further, although its correlations to the Coding (*r* = .39, *p* < .0001) and Figure Weights (*r* = .40, *p* < .0001) subtests were significant, the strength of these correlations, was weak. Finally, the two Processing Speed Index subtests, Symbol Search and Coding showed a substantial correlation (*r* = .62, *p* < .0001). Significant, albeit weak to moderate correlations, were also observed between the Symbol Search and the Naming Speed Literacy (*r* = .44, *p* < .0001), Naming Speed Quantity (*r* = .45, *p* < .0001), Block Design (*r* = .39, *p* < .0001), Similarities (*r* = .39, *p* < .0001), Matrix Reasoning (*r* = .43, *p* < .0001), Letter-Number Sequencing (*r* = .42, *p* < .0001), Information (*r* = .54, *p* < .0001) and Arithmetic (*r* = .47, *p* < .0001) subtests. Similarly, the correlation between the Coding and the Naming Speed Quantity (r = .46, *p* < .0001) subtests was significant, but moderate.

4.5.2. KABC-II Correlation Coefficients

According to the information presented in Appendix G Table G8 a number of very strong correlations were noted between the KABC-II global and scale scores. The first of these is between the two global scores, the FCI and the MPI (*r* = .98, *p* < .0001). The Simultaneous Scale also showed very strong correlations with both of these (MPI, *r* = .82, *p* < .0001; FCI, *r* = .81, *p* < .0001). Although moderate, the Sequential Scale showed significant correlations with the FCI (*r* = .43, *p* < .0001) and the MPI (*r* = .42, *p* > .0001). The Learning Scale showed a substantial correlation with the MPI (*r* = .61, *p* < .0001) and a moderate correlation with the FCI (*r* = .58, *p* < .0001). The Planning Scale showed substantial correlations with the MPI (*r* = .78, *p* < .0001), the FCI (*r* = .77, *p* < .0001), and the Simultaneous Scale (*r* = .68, *p* < .0001). The Knowledge Scale also showed substantial correlations with the FCI (*r* = .76, *p* < .0001) and the MPI (*r* = .65, *p* < .0001), and a moderate correlation with the Simultaneous Scale (*r* = .51, *p* < .0001).

As far as the KABC-II Scale and subtests' correlations in this sample are concerned, the information provided in Appendix G, Table G9 indicate that the two subtests of the Sequential Scale, Number Recall (r = .89, p < .0001) and Word Order (r = .88, p < .0001) correlate very strongly with this scale. Although only a moderate correlation (r = .60, p < .0001) was observed between these two subtests. Whilst the Rover (r = .92, p < .0001) and Triangles (r = .84, p < .0001) subtests correlated very strongly with the Simultaneous Scale, a substantial correlation was noted between this scale and the Block Counting subtest (r = .69, p < .0001). Further, a substantial correlation was also noted between the Simultaneous Scale and the Pattern Reasoning subtest (r = .65, p < .0001), as well as a moderate relationship between this scale and the Story Completion subtest (r = .57, p < .0001). Whilst the Rover subtest showed substantial correlations with the Triangles (r = .64, p < .0001) and Pattern Reasoning (r = .62, p < .0001) subtests, it showed a moderate correlation with the Block Counting subtest (r = .59, p < .0001). The Triangles subtest also showed a moderate correlation with the Block Counting subtest (r = .55, p < .0001), but also too other subtests, including Story Completion (r = .54, p < .0001) and Verbal Knowledge (r = .51, p < .0001).

Interesting is the fact that all three Simultaneous Scale subtests, showed moderate to substantial correlations with the Planning Scale: Rover (r = .61, p < .0001), Triangles (r = .59, p < .0001) and Block Counting (r = .54, p < .0001). However, the two Planning Scale subtests, Story Completion (r = .88, p < .0001) and Pattern Reasoning (r = .85, p < .0001) showed very strong correlations with this index, whilst only showing a moderate correlation with each other (r = .53, p < .0001). As was seen with the three Simultaneous Scale subtests, the Planning Scale subtests also showed moderate to substantial correlations with the Simultaneous Scale: Pattern Reasoning (r = .65, p < .0001) and Story Completion (r = .57, p < .0001). The Learning Scale showed a very strong correlation with its Atlantis subtest (r = .89, p < .0001) and a substantial correlation with its Rebus subtest (r = .73, p < .0001). Nevertheless, these two subtests only showed a weak correlation with each other (r = .40, p < .0001). Finally, the Knowledge

Scale subtests, Verbal Knowledge ($r = .93$, $p < .0001$) and Riddles ($r = .92$, $p < .0001$) showed very strong correlations with this index, as well as a substantial correlation with each other ($r = .76$, $p < .0001$). Interesting is that the Verbal Knowledge subtest showed a moderate correlation with the Simultaneous Scale ($r = .52$, $p < .0001$), as well as with the Triangles subtest stated above.

4.5.3. WISC-V Confirmatory Factor Analysis

In addition to the intercorrelation studies, second-order CFA was done to investigate the present WISC-V five-factor model in the current sample, and to further explain the stated intercorrelations. The results of this factor analysis are presented in Appendix G Table G10 and Figure G1. As reported in Chapter 2, fit indices with values equal to or higher than .90 demonstrate a good fit. Also, as a root mean square error of approximation (RMSEA) further provides a descriptive measure of the overall model fit, a RMSEA of .06 or lower will be considered good fit, and a RMSEA of .08 will be considered adequate fit.

As can be seen form the information presented in Table G10 the data obtained from the study sample showed good fit to the current WISC-V five-factor model on almost all fit indices: Comparative Fit Index (CFI = .98), Normed Fit Index (NFI = .91), Bollen Fit Index (IFI = .98), and Goodness of Fit Index (GFI = .92). The only exception here was the Adjusted Goodness of Fit Index (AGFI = .86). In addition, the RMSEA was equal to .06 (RMSEA = .060). Further, from the path coefficients included in Figure 1, it is evident that the fluid reasoning factor correlated most strongly with the second-order *g*-factor (.94). This is followed by the visual-spatial factor, which also correlated very strongly with the *g*-factor (.89). In addition, both the working memory (.79) and the verbal comprehension factors (.78) correlated substantially with the *g*-factor, whilst a moderate correlation was noted between the latter and the processing speed factor (.58). As far as the five first-order factors and their subtests are concerned, the strongest correlation was noted between the processing speed factor and the Symbol Search subtest (.95). Other very strong correlations included the Visual Puzzles subtest and the visual-spatial factor

(.92), the Vocabulary subtest and the verbal comprehension factor (.90), and the Matrix Reasoning subtest and the fluid reasoning factor (.83). All of the other subtests and their respective indices showed substantial correlations: Similarities and the verbal comprehension factor (.79), Picture Span and the working memory factor (.79), Figure Weights and the fluid reasoning factor (.77), Block Design and the visual-spatial factor (.77), Digit Span and the working memory factor (.67), and Coding and the processing speed factor (.66).

4.5.4. KABC-II Confirmatory Factor Analysis

In contrast to the WISC-V, the KABC-II FCI model could not be fitted as proposed as there were too many parameters and too little data. As a result, the SAS software fixed two of the parameters (Number Recall and Planning Scale) to 1.00 (see Appendix G, Figure G2). With this in mind, the highest correlation noted (see Appendix G, Figure G2), was between the simultaneous processing or visual processing factor and the second-order FCI (.93). A substantial correlation was noted between the FCI and the knowledge or crystallised reasoning factor (.66), and a moderate correlation between the FCI and the learning or long-term storage and retrieval factor (.51). The sequential processing or short-term memory factor correlated very weakly (.16) with the second-order FCI. As far as the first-order factors and their subtests are concerned, cognisance should again be taken that the correlation between the sequential processing factor and the Number Recall subtest was fixed to 1.0. With this in mind, very strong correlations were noted between the knowledge factor and its two subtests, Riddles (.84) and Verbal Knowledge (.91). In addition, a very strong correlation was noted between the Rover subtest (.82) and the simultaneous processing factor, with substantial correlations between the latter and its other subtests, Block Counting (.70) and Triangles (.79). Substantial correlations were also noted between the Story Completion (.75) and Pattern Reasoning (.72) subtests and planning or fluid reasoning factor, as well as between the Rebus subtest (.75) and the learning factor. Further, moderate correlations were noted between the Atlantis subtest (.54) and the learning factor and between the

Word Order subtest (.59) and the sequential processing factor. Finally, as can be seen from Table G11 the model fit statistics were generally not good, with only the CFI (.94) and the IFI (.95) showing good fit. This is likely the result of the over-fitting of the model.

4.6. Objective 3: Comparison of the WISC-V and KABC-II Indices and Scales

The third objective of this study was to investigate whether the WISC-V or the KABC-II were more suitable for use with the present sample through a comparison of the global and index/scale mean scores of the two language groups, EFL and EAL. As noted in Chapter 3, 88 participants reported their first languages. The majority, 63.6% of these reported English (*n* = 56) as their first language, whilst 36.4% (*n* = 32) reported English as an additional language. Table G Appendix G12 includes the global and index/scale scores of the EFL group in terms of mean, standard deviation, minimum and maximum. Whilst the WISC-V FSIQ, General Ability Index, Cognitive Proficiency Index and primary index scores could be included for all EFL (*n*=56) participants, the slightly lower number of participants who completed the whole KABC-II, resulted in the FCI, MPI, and scale scores of 54 EFL participants being included in the analyses. In comparison, the results of all 32 EAL participants were included in the analyses (Appendix G, Table G13).

As can be seen from Appendix G, Tables G12 and G13, as well as Figure G3 (Appendix G), the mean WISC-V scores of the EFL participants, ranged from 81.0 (*SD* = 13.6) for the FSIQ, to 89.6 (*SD* = 17.9) for the Visual-Spatial Index. In comparison, the WISC-V means scores of the EAL participants (Appendix G, Table G13), ranged from 73.7 (*SD* = 12.7) for the FSIQ to 81.1 (*SD* = 12.2) for the Fluid Reasoning Index. When comparing the scores on the indices for the two groups, it is evident that the EFL group generally performed better on all of the WISC-V Index scores than the EAL group. As far as the KABC-II performance of the two groups is concerned, the mean scores of the EFL group ranged from 84.1 (*SD* = 12.8) for the Knowledge Scale to 97.9 (*SD* = 12.0) for the Learning Scale. In contrast, the mean KABC-II scores of the EAL group ranged from 78.9 (*SD* = 8.4) for the Knowledge Scale to 94.2 (*SD* = 15.4) for the Learning Scale. Similar to the WISC-V, the EFL group generally obtained higher mean scores on all the KABC-II indices and scale, with the exception of Sequential Scale where the EAL group (*M* = 89.1, *SD* = 16.5) performed between than the EFL group (*M* = 86.9, *SD* = 14.0).

In determining whether the differences between the EFL and EAL groups were statistically significant, participants' scores were compared using a General Linear Model (GLM) with WISC-V and KABC-II scores as the dependent variables and English language status as the independent variable. The results of these analyses indicated that the effect of English language status on the mean FSIQ score was significant (F (1,86) = 6.16, p = .015) (see Appendix G, Table G14). Similarly, English language also had a significant effect on participants' performance on the General Ability Index (F(1,86) = 8.4, p = .005) (see Appendix G, Table G15), Verbal Comprehension Index (F(1,86) = 5.95, p = .017) (see Appendix G, Table G17), Visual Spatial Index (F(1,86) = 6.97, p = .009) (see Appendix G, Table G18), and Fluid Reasoning Index (F(1,86) = 5.75, p = .019) (see Appendix G, Table G19). However, whereas English language status had a significant effect on a number of the WISC-V mean global and index scores, it only impacted significantly on the Knowledge Scale of the KABC-II (F (1,84) = 4.21, p = .043) as reported in Appendix G, Table G28. Thus, English language status had a wider impact on the WISC-V indices.

4.7. Objective 4: Item Response Theory and Differential Item Functioning

4.7.1. Graded Response Model: WISC-V and KABC-II Discrimination Parameters

As reported in the Chapter 3, a Graded Response Model (GRM) was used to investigate the items administered in this study, which included two parameters in its investigation of the subtests' items. The first parameter was the discriminant parameter or *a*-value, which is referred to as the slope in a GRM model. The second parameter was the difficulty parameter or the *b*-value and is referred to as the threshold/s in the GRM model. Focussing specifically on the discriminant parameter, an investigation of Appendix G, Table G29 indicated that most of the WISC-V subtests' discriminatory values fell in the moderate to very high categories. However, given the SAS cut-off point of 1 (SAS

Institute, 2018), it is important to note that the *a*-values of Picture Span, Information, and Comprehension are all <1, thus, raising concern about the information that assessors could extract from these subtests in this sample. When compared to the other subtests, the three components of the WISC-V Digit Span subtest, Digit Span Forward (*a* = 2.97), Digit Span Backward (*a* = 2.47) and Digit Span Sequencing (*a* = 2.34), showed higher discriminatory sensitivity. As stated, the Picture Span (*a* = .92), Information (*a* = .93), and Comprehension (*a* = .94), indicated lower discriminatory sensitivity. As far as the KABC-II subtests are concerned, it is interesting to note that most of the subtests fell in the moderate range, with none in the very high range. The only *a*-values falling in the high range were Block Counting (*a* = 1.27) and Rover (*a* = 1.37). Also, there were again three subtests that had *a*-values < 1, Atlantis (*a* = .76), Word Order (*a* = .97), and Story Completion (*a* = .98). Further, despite the fact that it showed the highest mean subtest score amongst the KABC-II subtests, the Atlantis subtest displayed the lowest discriminatory sensitivity. Unlike the WISC-V subtests, none of the KABC-II subtests had *a*-values higher than 2.

4.7.2. Graded Response Model: WISC-V Difficulty Parameters

As stated in Chapter 2, the *b*-value or difficulty parameter is expressed as thresholds in the GRM. Some subtest items may contain more than one threshold because not all KABC-II and WISC-V item responses are dichotomous in nature. Regardless, more difficult items will show higher thresholds or *b*-values than easier items. The tables summarising these are presented in Appendix G. With reference to Appendix G, Table G30 and the first WISC-V subtest, Block Design, it is important to note that item 13 was not included in the analyses as none of the participants (*N* = 90) provided a correct response to this item. This would thus be considered a particularly difficult item for this sample; however, cognisance should be taken of the sample's age groupings. As for the other items, the data from this sample indicated that items 1–5 follow the expected difficulty pattern. However, between items 6 and 9, the order changes slightly, with item 7 presenting as the most difficult among these four items and item 8 being the easiest. From item 10, the expected difficulty pattern is again observed. None of the items' *p*-values are greater than 0.05; thus, raising concerns about model fit. In addition, the *b*-values calculated for the Block Design subtest in this sample range from item 1 (lower threshold, $b = -2.72$) to item 12 (higher threshold, $b = 2.72$). Being < -2.5, the first item appears noticeably easy; however, cognisance should be taken that the first two items are generally not administered in this sample's age group. Conversely, being > +2.5, items 11 (higher threshold) and 12 (higher threshold, $b = 2.72$) appeared to be particularly difficult for this sample.

Regarding the Similarities subtest's items, an interesting pattern of difficulty is observed in this sample (see Appendix G, Table G31), with item 3 being easier than item 2. Also, item 7 is less difficult than items 5 and 6, and items 9, 11, 12, and 13 are less difficult than item 8. Following this are items 10 and 15, which are less difficult than items 14 and 16. Whilst items 18 to 20 remain in the expected order of difficulty, item 22 is less difficult than item 21. Item 23 is the most challenging item on the Similarities subtests. Thus, in this sample, the order of the Similarities subtest's items according to their lower thresholds is markedly different from the order of the items as they are currently presented. Unlike the Block Design subtest, items 8 (lower threshold, $p = .28$; upper threshold, $p = .37$), 10 (lower threshold, $p = .31$), 11 (higher threshold, $p = .33$), and 12 (lower threshold, $p = .21$) showed better model fit (p = > .05). For the remaining items and thresholds, $p < .05$. As far as their *b*-values are concerned, the difficulty levels of the Similarities subtest's items range from item 1 ($b = -2.27$) to item 21 (upper threshold, $b = 2.95$). None of the *b*-values thus fell below -2.5, whilst the values for items 21 (upper threshold, $b = 2.95$), 23 ($b = 2.79$), 20 (upper threshold, $b = 2.88$), and 19 (upper threshold, 2.82) are all > +2.5. These items have thus been identified as very difficult in this sample.

In analysing the items of the Matrix Reasoning subtest, items 31 and 32 were excluded as none of the participants in this study ($N = 90$) responded correctly to these. In addition, item 2 was excluded as it was completely confounded with item 6. Whilst taking cognisance that these items were excluded, the first few items of the Matrix Reasoning subtest take on a different pattern to the current order in which these items are presented (see Appendix G, Table G32). For example, item 6 follows item 1 in being the two easiest items. This is followed by item 4 and then 3, item 8 and then 7, and then items 5 and 9, making item 5 one of the more difficult items amongst the first 9 items. The order of items 10 to 30 show less variance, although item 13 is less difficult than 12, item 16 is less difficulty than 15, item 21 is less difficult than 20, item 23 is less difficult than item 22, and finally, item 28 is less difficult than 27. Thus, in terms of their level of difficulty, there is a difference in the order of the Matrix Reasoning subtest's items when compared to the current order in which they are administered. Also, four items obtained *b*-values < -2.5. These included items 1 (b = -3.59), 6 (b = -3.06), 3 (b = -2.74), and 4 (b = -2.75); however, again cognisance needs to be taken that most of the participants in this sample would have started at item 5, and some at item 9. Thus, items 1 – 4 would be easier given the expected ability levels of this age group. The anomaly is item 6, which was also highlighted as particularly easy in this sample. Items 29 (b = 2.67) and 30 (b = 2.93) achieved *b*-values > +2.5, thus indicating that these presented as particularly difficult. Together with items 31 and 32, participants' struggles with these two items may be the results of this sample's age range. Finally, items 15 (p = .20), 17 (p = .48), and 18 (p = .39) showed better model (p = > .05) fit than the majority of the other items.

As far as the three components of the Digit Span subtest is concerned (see Appendix G, Tables G33-G35), items 8 and 9 of Digit Span Forward was excluded from the analyses as none of the participants in this sample provided a correct response to item 9 (N = 90) and item 8 was highly confounded with item 7. Similarly, items 7, 8, and 9 were excluded from both Digit Span Backward and Digit Span Sequencing as none of the participants (N = 90) in this sample managed to correctly respond to these items. Model fit was observed in two items. The first of these is item 3 (p = .17) in Digit Span Backward (upper threshold), and the second is item 4 (p = .18) in Digit Span Sequencing (lower threshold). The order of the items in these components also remained unchanged to a large extent. The first exception is Digit Span Forward, where item 1 was found to be more difficult in this sample than item 2. The second exception was Digit Span Backward, where item 2 was also found to be easier than item 1 in this sample. Digit Span Sequencing retained its current order of difficulty. Also interesting is the fact that none of the Digit Span subtests obtained *b*-values < -2.5 or > +2.5.

In the analyses of the Vocabulary subtest's items (see Appendix G, Table G36), items 1, 2, and 4 were removed because all participants (*N* = 90) obtained scores of 1. The exclusion of these earlier items is most likely due to the fact that most of the participants in this study, started with either items 5 or 9, depending on their ages. This most likely also contributed to some items obtaining *b*-values < -2.5. Included are items 3 (*b* = -3.21), 6 (lower threshold, *b* = -3.21), and 7 (lower threshold, *b* = -2.90). Items 24 to 29 were excluded as none of the participants (*N* = 90) provided appropriate responses here. Items 22 (*b* = 2.97) and 23 (*b* = 3.40) obtained b-values > +2.5, indicating that these, in addition to items 24-29 were experienced as particularly difficult by the participants in this sample. Again, variability was observed in the order of the items according to level of difficulty, especially amongst the first few items. For instance, item 7 is less difficult than 5 and 6, whilst item 10 is also less difficult than 8 and 9. Similarly, item 14 is less difficult than 12 and 13. From item 15 to 23 the current order of difficulty is followed, except for item 19 which is less difficult than 18. Items 12 (lower threshold, *p* = .22), 13 (lower threshold, *p* = .45; upper threshold, *p* = .32), 14 (lower threshold, *p* = .19) and 15 (lower threshold, *p* = .09) showed better model fit with p > .05. Aside from that, all *p*-values < .05.

Similar to the Vocabulary subtest, in the analysis of the Figure Weights subtest's items (see Appendix G, Table G37) the first 3 items were excluded from the analyses, most likely due to the fact that most of the participants in this sample started with item 4 given their ages. Item 4 (*b* = -2.57) was also identified as particularly easy in this sample. Conversely, item 34 was excluded as none of the participants answered this item correctly (*N* = 90). Again, as with all the other subtests, one needs to take cognisance of the sample's mean age and the fact that the WISC-V has been standardised for children between the ages of 6 years to 16 years, 11 months. Younger children are thus not necessarily expected to respond correctly to the more difficult items at the end of a subtest. Items 29 (b = 2.50), 26 (b = 2.68), 32 (b = 2.78) and 33 (b = 2.93) obtained b-values > +2.5. What is interesting, is the exclusion of items 28 and 31, because none of the participants responded correctly to these, but some did respond correctly to seemingly more difficult items (in terms of the present order of administration), specifically items 29 (n = 3), 30 (n = 3), 32 (n = 1) and 33 (n = 1). Although the data indicates that only a few participants answered these last items correctly, it still impacts the difficulty ranking of the items. Finally, as far as the Figure Weights subtest is concerned, items 16 (p = .26), 17 (p = .19), 18 (p = .30) and 19 (p = .31) showed better model fit with p-values > .05. As far as the difficulty ranking of the Figure Weights subtest in this sample is concerned, variance is again observed from the current order, with item 9 being easier than items 8, then 6, then 5 and then 7 being the most difficult item in this group. Similarly, item 11 is easier than item 10, item 13 easier than item 12, and item 15 easier than item 14. As far as items 16 – 20 are concerned, item 17 is the easiest, followed then by item 19, item 18, item 16 and then item 2. Item 16 is thus more difficult than items 17 – 19. Following this, item 21 is more difficult than 22, and item 23 more difficult than item 24. Amongst items 27 to 33, item 27 is the easiest, then followed by item 30, item 25, item 29, item 26, item 32, and then item 33.

For the purpose of analysing the Visual Puzzles subtest (see Appendix G, Table G38) in this sample, items 1 – 2 were removed, which can be understood when taken into consideration that the starting point for this subtest in this age group would have been item 5 and for a few participants, item 8. If item 5 was not answered correctly, the easier items would have been administered in reverse order until two consecutive correct responses were given; hence, the inclusion of items 3 and 4. These two items were however identified as particularly easy with b-values < -2.5; item 3 (b = -3.51) and item 4 (b = -3.51). Items 25 to 29 were also excluded as none of the participants who completed this subtest (N = 90) answered these items correctly. Aside from these, none of the items obtained b-values > +2.5. As far as the difficulty ranking of the Visual Puzzles subtest's items are concerned, items 3 – 6 follow the expected order, followed by items 8 and 9. Item 7 is however more difficult than items 8 and 9. Item 13 is easier than items 11 and 12, and item 17 easier than item 16. From there the expected order is indicated; except for item 22 that is easier than items 20 and 21. Better model fit was observed in items 12 (p = .25), 14 (p = .18), 15 (p = .09) and 17 (p = .06) with p-values > .05.

With reference to the items of the Picture Span subtest (see Appendix G, Table G39), items 1 – 3 were removed as the scores for all of these were 1, which is the highest score that can be obtained on these items. Thereafter, it is interesting to note that items 5 and 6 were calculated to be easier in this sample than item 4, especially since all of these items require examinees to retain and identify two visual images. Furthermore, item 8 was calculated as being easier than item 7, which appears incongruent as item 8 requires examinees to retain and identify three images, whilst item 7 only presents two images. Similarly, item 14 was identified as more difficult than items 12 and 13, with the latter two items requiring examinees to retain and identify three images, whilst item 14 requires them to retain and identify four images. Whilst item 18 was identified as easier than 17, both of these present examinees with four images. Likewise, item 21 was also identified as easier than 20; however, both of these present five images to examinees. From item 22 onwards, the pattern remains as it currently stands. The lower thresholds of items 4 (b = -2.59), 5 (b = -3.64), and 6 (b = -2.59) recognise these as particularly easy items, with all the b-values falling below -2.5. Conversely, items 25 (b = 3.16) and 26 (b = 3.57) were identified as particularly difficulty, which can be understood from the fact that they are the last two items and whilst item 25 presents examinees with seven images, item 26 requires them to retain and identify eight images. Items 11 (higher threshold, p = .06), 12 (higher threshold, p = .36), 13 (higher threshold, p = .24), 15 (higher threshold, p = .06) and 16 (lower threshold, p = .19) showed better model fit with p > .05.

A number of items were excluded in analysing the Information subtest items (see Appendix G, Table G40). Items 1 – 3 and item 5 were excluded because all of these values were 1, whilst items 25 – 31 were excluded because all these values were 0. In addition, items 4 and 7 were completely confounded with item 12 and were thus excluded. Being a measure of crystallised intelligence, which is likely to be impacted by cultural-linguistic factors, it is interesting to note that the difficulty ranking of the items in this sample, is quite varied from the order currently used. For instance, item 12 was recognised as the easiest item, keeping in mind that items 1 – 3 and 5 were excluded and items 4 and 7 were confounded with 12. This is followed by items 6, 9, 10 and 11 and then 8. Item 13 is followed by 15 and then 14. Item 18 appears next but is then followed by 17, 19, 16, and 22. The last few items considered in this analysis, 20, 21, 23, and 24 retain that order, with none of the participants being able to respond correctly to items 25 to 31. Item 12 (b = -3.12) was the only item with a threshold < -2.5, whilst item 24 (b = 2.92) the only one with a threshold < +2.5. Item 18 (p = .35) showed better model fit with p-values > .05.

In contrast to the Information subtest, Letter-Number Sequencing retained most of its current order of difficulty with only item 2 presenting as more difficult than items 3 and 4 (see Appendix G, Table G41). This again appears incongruent because whereas item 2 requires the retention of two stimuli, items 3 and 4 requires the retention of three stimuli and one would expect children to struggle more with these. Whilst only item 1 (b = -3.08) had a lower threshold number < -2.5, none of the items had thresholds > +2.5; although, cognisance must be taken that none of the participants were able to complete items 8 – 10. Only item 5 (p = .11) showed better model fit with a p-value > .05.

The last WISC-V verbal subtest, that also relies on crystallised intelligence, Comprehension, again showed variance in the difficulty ranking of the items (see Appendix G, Table G42). Item 2 presented as the easiest in this sample, with threshold values < -2.5 (lower threshold, b = -3.25; upper threshold, b = -2.69). This was followed by item 1, which also presented with a threshold level < -2.5

(lower threshold, b = -2.79). Although its threshold values were not quite as low, item 4 presented as easier than 3. Following this, great variance is noted, starting with item 8 which is followed by items 7, 5, 9, and then 6. Similarly, item 10 is next, followed by items 12, 15, 13, 17, 11, 14, 16, and then 18. Item 19 was excluded from the analysis as none of the participants responded correctly to this. Although, it is interesting to note that item 14 (higher threshold, $b = 2.55$) was the only value > +2.5. Better model fit was noted in items 3 (higher threshold, $p = .17$) and 10 (lower threshold, $p = 2.55$).

The final WISC-V subtest administered to this sample was Arithmetic. Item 27 was excluded as it was completely confounded with item 26 (see Appendix G, Table 43). Following items 1 – 3, which were also excluded from the analyses because all the participants ($N = 90$) answered these correctly, items 4 ($b = -3.15$), 6 ($b = -2.62$), and then 5 ($b = -2.61$) presented as the easiest in this sample with values < -2.5. Thereafter follows items 7 – 9 in their current order; however, item 12 succeeds these and then items 10, 14, 13, and 11. Items 15 – 20 again retain their current order, but is followed by items 22, 21, 23, 25, 24, and lastly 26. The last three items, 25 ($b = 2.92$), 24 ($b = 2.94$), and 26 ($b = 3.03$) all had *b*-values that exceeded +2.5. None of the participants managed to respond correctly to items 28 – 34; thus, making these the most difficult items in this sample. However, again cognisance needs to be taken of the current sample's age group. Items 11 ($p = .35$), 13 ($p = .09$), and 15 ($p = .13$) showed better model fit with *p*-values < .01.

4.7.3. Graded Response Model: KABC-II Difficulty Parameters

As stated above, the first of the KABC-II subtests, Atlantis, presented with the highest mean subtest score; however, obtained the lowest discriminatory value. As far as its difficult parameters are concerned, item 3 was excluded and thus, 53 items were included in the analyses. Great variance was observed in the difficulty levels of the items compared to their current standing (see Appendix G, Table G44). However, here cognisance needs to be taken that while all the participants started with item 1, different discontinuation points were adhered to based on each participant's cumulative score following items 11, 20, 29, or 42. Additionally, it is important to note that items 1–42 are all considered teaching items, and thus, participants received continuous feedback on any errors made. Regardless, as far as the difficulty levels of the items are concerned, item 8 is considered the most straightforward (*b* = -4.14), whilst item 47 was considered the most difficult (lower threshold, *b* = -.12). In addition to item 8, several items presented with b-values > -2.5, including items 1 (*b* = -3.38), 2 (*b* = -3.76), 5 (*b* = -3.5), 6 (lower threshold, *b* = -3.63), 7 (*b* = -3.18), 8 (lower threshold, *b* = -4.14), 9 (lower threshold, *b* = -3.35), 10 (*b* = -2.79), 11 (lower threshold (*b* = -3.45), 14 (lower threshold, *b* = -2.51), 15 (lower threshold, *b* = -4.08), 16 (lower threshold, *b* = -3.37), and 22 (lower threshold, *b* = -3.27; upper threshold, *b* = -2.79). None of the items received a *b*-value > +2.5, thus indicating that the items of the Atlantis subtest presented as easier than the items of other subtests administered to this sample, which could explain the higher mean score, but lower discriminatory value. Finally, items that showed better model fit (*p* > .05) were 26 (upper threshold, *p* = .36), 29 (upper threshold, *p* = .15), 31 (upper threshold, *p* = .41), 33 (upper threshold, *p* = .08), 34 (upper threshold, *p* = .34), 35 (upper threshold, *p* = .14), 40 (upper threshold, *p* = .29), 41 (upper threshold, *p* = .34), 44 (upper threshold, *p* = .09), 46 (upper threshold, *p* = .41), 47 (lower threshold, *p* = .29), 49 (upper threshold, *p* = .39), 50 (lower threshold, *p* = .15), 51 (upper threshold, *p* = .07), 52 (upper threshold, *p* = .15), and 53 (upper threshold, *p* = .49).

Across the 18 items of the Story Completion subtest, different scores were awarded to the participants based on the answers provided and from item 8 onwards, also the time it took to complete the item. Thus, item scores ranged from 0 – 4, which is reflected in the different difficulty thresholds. Most of the participants in this sample would have started with item 4. Based on the information provided in Appendix G, Table G45 four items presented with difficulty levels < -2.5. These, presented in the order of difficulty as measured in this sample, included items 3 (*b* = -3.83), 1 (*b* = -3.83), 2 (*b* = -3.05), and then 6 (*b* = -2.60). None of the items achieved *b*-values > +2.5. Although the order of difficulty, as measured in this sample, differ somewhat from the current order in which items are presented, there is a closer resemblance than what has been seen in most of the other subtests. Following item 6, are items 4, 5, 7, 9, and then 8. This is interesting as the first seven subtests all required participants to only place two cards in their correct order and, most of the participants only started with item 4 and would have been awarded the maximum points if they managed to provide correct responses to items 4 – 6. Nevertheless, following these were items 10, 12, 11, and 13. With all the subtest's items included in the analyses, the last few subtests are ordered 15, 14, 16, 17 and then 18. Item thresholds that showed better model fit, included 8 (upper threshold, $p = .48$), 9 (upper threshold, $p = .07$), 10 (second threshold, $p = .06$), 11 (second threshold, $p = .14$), 12 (third threshold, $p = .48$), 13 (second threshold, $p = .06$), 14 (lower threshold, $p = .39$; second threshold, $p = .06$), and 15 (lower threshold, $p = .44$; second threshold, $p = .37$). All other item threshold *p*-values < .05.

In the next subtest, Number Recall, items 1 – 3 were excluded as these obtained scores of 1, whilst items 17, and then 19 – 22 were also excluded as none of the participants ($n = 89$) provided correct responses to these. Aside from item 13, which presented as more difficult than item 14, the order of difficulty as measured in this sample, remains the same as the current order (see Appendix G, Table G46). Both of these require participants to retain and then name six numbers. The first three items included in the final analyses, all obtained difficulty values < -2.5. These are items 4 ($b = -2.74$), 5 ($b = -2.73$), and 6 ($b = -2.69$). Whilst taking cognisance that several difficult items were excluded from this analysis, item 18 ($b = 2.66$) was the only one to obtain a difficulty value > +2.5. Items 11 ($p = .08$) and 12 ($p = .06$) showed better model fit.

Similar to the Story Completion subtest, none of the items of the Rover subtest were excluded from the analyses with scores from 0 – 2 based on the number of correct moves made within a specific time limit (see Appendix G, Table G47). Variance was again observed in the order of difficulty as measured in this sample, with item 2 being easier than 1, and item 5 being more difficult than 4, 6 and then 9. This is followed by items 7, 8, 10 and then 15. Items 11 and 12 are thus followed by 13 and 14.

The variance continues with items 16, 18, 19, 17, 20, 22 and then 21. None of the items obtained *b*-values < -2.5. or > +2.5. However, a number of items showed better model fit, including 5 (upper threshold, *p* = .19), 8 (upper threshold, *p* = .16), 11 (lower threshold, *p* = .06), 13 (lower threshold, *p* = .28), 14 (lower threshold, *p* = .09), and 15 (lower threshold, *p* = .38). None of the other item threshold *p*-values > .05.

With a total of 90 items, and most participants starting with either items 21 or 28, it is understandable that a number of items were excluded from the analysis of the Verbal Knowledge subtest's items, including Items 1-4, 6-9, 12-14, and 17, which all the participants (*n* = 89) responded to correctly (see Appendix G, Table G48). Conversely, items 69, 78-79, and 82-90 were excluded as none of the participants responded correctly to these. Also, items 5, 10 and 11 were confounded and thus omitted; as were items 16, 18 and 20 and then 15 and 19. Whilst taking these omissions into consideration, great variance was observed in the difficulty order of the items as measured in this sample; however, none of the items had difficulty values < -2.5. In contrast, several items presented with difficulty values > +.2.5, including 77 (*b* = 2.51), 80 (*b* = 2.52), 73 (*b* = 2.54), 56 (*b* = 2.54), 74 (*b* = *2*.72), 72 (*b* = 2.97), and 81 (*b* = 3.17). The four items that showed better fit were 38 (*p* = .42), 39 (*p* = .15), 41 (*p* = .39) and 44 (*p* = .07).

Similar to the Atlantis subtest, the other Learning Scale subtest, Rebus, also allowed some feedback to participants on their answers provided, and different discontinuation points based on the cumulative scores obtained at items 7, 10, 12, 15, and 18. All participants started at item 1 and as can be seen from Appendix G, Table G49 all 21 items were included in the analyses. Each of these items allotted participants scores between 0 and 5; hence, the different difficultly thresholds. It is interesting to note that together with the variance in the difficulty order of the items, none of the items obtained difficulty values > +2.5. However, several items obtained difficulty values > -2.5, including items 4 (lower threshold, *b* = -2.88), 7 (lower threshold, *b* = -2.84), 3 (lower threshold, *b* = -2.83), 11 (lower threshold, *b*

= -2.79), 10 (lower threshold, b = -2.74) and 1 (b = -2.69). Items that showed better model fit were 8 (upper threshold, p = .24), 9 (upper threshold, p = .21), 12 (third threshold, p = .25), 13 (upper threshold, p = .07), 14 (third threshold, p = .08), 16 (second threshold, p = .47), 17 (second threshold, p = .41), 18 (second threshold, p = .07), 20 (upper threshold, p = .06), and 21 (second threshold, p = .35).

Although a core subtest for ages 3 – 12-years, the Triangles subtest was administered to 87 of the 90 participants are norms are available for all age groups. The majority of the participants in this sample would have started with item 15; however, if they provided incorrect responses to items 15 – 17, they would have returned to item 11. Regardless, most participants would have received full credit for items 1 – 1. It is thus not surprising that items 2 – 5 were excluded as all participants (n = 87) provided correct responses to these. Items 1 and 6 were also excluded as these were completed confounded with items 8 and 7 respectively (see Appendix G, Table G50). Amongst the six items that obtained b-values > -2.5, were items 7 (b = -3.54), 9 (b = -3.51), 8 (b = -3.47), 12 (b = -2.72), 11 (b = -2.69), and 10 (b = -2.58). None of the subtests obtained difficulty values > +2.5. Some did however show better model fit, including items 17 (upper threshold, p = .2119), 18 (upper threshold, p = .42) and 19 (upper threshold, p = .12), 21 (lower threshold, p = .20), and 23 (lower threshold, p = .15).

Although only a core subtest for ages 13 – 18-years, the Block Counting subtest was again administered to the majority of the participants in this sample as norms are available for all age groups (see Appendix G, Table G51). Most of the participants in this sample (n = 84) would have started with item 7; however, the only item excluded from the earlier items, is item 1 which all the participants answered correctly. Thereafter, all items are included with scores ranging from 0 – 2. Across the entire subtest, variance is observed in the difficulty order of the items; however, only item 2 obtained a difficulty value > -2.5 (b = -2.66), with none of the items obtaining difficulty values > +2.5. Also, only item 16 showed better model fit (p = .28).

Word Order is the second of the Sequential Scale subtests and, unlike Number Recall, there was some variance observed in the difficulty order of the items (see Appendix G, Table G52). Cognisance does need to be taken that some items were omitted from the analyses, including item 3 which all of the participants (n = 88) answered correctly, and item 26 which they all provided incorrect responses. Items 1 and 2 were also completely confounded with item 5 and thus excluded, as were item 27 which was completely confounded with item 25. From the remaining items, items 5 (b = -3.16), 8 (b = -2.78), and 4 (b = -2.75) the three easiest items, with difficulty values < -2.5. The variance observed here continues for the remainder of the subtest's items, with items 23 (b = 2.86) and 25 (b = 3.38) being the most difficult amongst those included in the analyses, presenting with difficulty values > +2.5. Two items, 14 (p = .18) and 15 (p = .22) showed better model fit.

Only item 1 was excluded amongst the Pattern Reasoning subtest as all the participants (n = 88) who completed this subtest, responded correctly (see Appendix G, Table G53). The remaining 35 items were thus included in the analyses and with the participants in this study, starting with item 6, it is understandable that items 4 (b = -3.19), 2 (b = -2.93), 5 (b =-2.93), and then 3 (b = -2.91) were amongst the easiest presenting with difficulty values < -2.5. It is however interesting that item 7 (b = -3.25) presented as the easiest amongst all the items analysed. This variance is again observed throughout the subtest, with items 33 (b = 2.53), 35 (upper threshold, b = 2.59; lower threshold, b = 2.95), 34 (b = 2.78) and 36 (b = 2.80) being the most difficult, presenting with difficulty thresholds > +2.5. There are a number of item thresholds that showed better model fit, including items 12 (lower threshold, p = .29), 13 (lower threshold, p = .34), 14 (lower threshold, p = .06; upper threshold, p = .30), 16 (lower threshold, p = .18), 17 (lower threshold, p = .11) and 19 (lower threshold, p = .48).

The final KABC-II subtest administered to this subtest, Riddles, has a total of 51 items. From these items 1-5 and 9 were excluded as all the participants responded correctly to these, whilst items 41, and 47 to 51 were omitted as none responded correctly (see Appendix G, Table G54). In addition, items 6 and 7 were completely confounded with item 8 and were thus excluded. Similarly, item 44 was completely confounded with item 45 and thus also excluded. From the remaining items, the four easiest ones that all obtained difficulty values < -2.5 were items 8 (b = -3.16), 10 (b = -3.12), 11 (b = -2.81), and 13 (b = -2.64). Conversely, the two most difficult items with values > +2.5 were 45 (b = 2.80) and 46 (b = 2.74). Only three items showed slightly better model fit, including 24 (p = .17), 26 (p = .15) and 32 (p = .35).

4.7.4. Differential Item Functioning: WISC-V and KABC-II Subtests' Items

In investigation Differential Item Functioning (DIF) in the WISC-V and KABC-II subtests' items, the sample was divided into EFL and EAL participants to identify items that functioned differently or showed bias in these language groups. The results of these analysis are presented in Appendix G, Tables G55 – G77). Amongst the items of the Verbal Comprehension Index subtests administered to this sample, DIF was detected in eight items of the two primary subtests; however, seven of these are Vocabulary subtest items. The only Similarities subtest item that showed DIF was item 6 (p = .011) (see Appendix G, Table G55) whilst the following seven items from the Vocabulary subtest presented with DIF, items 5 (p = .002), 13 (p = .002), 14 (p = .004), 15 (p = .001), 16 (p = .009), 17 (p = .018), and 18 (p = .009) (see Appendix G, Table G56). All of these were in favour of the EFL participants. Of interest though is that none of the Information or Comprehension subtest items showed DIF (see Appendix G, Tables G57 and G58). With reference to the two Visual-Spatial Index subtests (see Appendix G, Tables G59 and G60), DIF was detected in four of the Block Design subtests, again all in favour of the EFL participants. These included items 5 (p = .027), 7 (p = .022), 9 (p = .005), and 11 (p = .035). Conversely, no DIF was detected in any of the Visual Puzzle's subtests. Similarly, neither of the two primary Fluid Reasoning Index subtests showed any DIF. However, DIF was detected in three items from the Arithmetic subtest, including items 23 (p < .001), 26 (p < .001), and 27 (p < .001) (see Appendix G, Table G63). These were also in favour of the EFL participants. In analysing DIF amongst the Working Memory Index subtests that were administered to this sample, none of the Digit Span Forward, Digit Span Backward, or Letter-Number Sequencing items showed any DIF. Nevertheless, two items from Digit Span Sequencing showed DIF (see Appendix D, Table G64). These were items 3 (p = .033) and 6 (p = .040), also in favour of the EFL participants. In addition, four of the Picture Span subtest items also showed DIF in favour of the EFL participants. Included here were items 7 (p = .003), 12 (p = .002), 14 (p = .034), and 19 (p = .049). Finally, as was the case with the other analysis, DIF could not be investigated amongst the two Processing Speed Index subtests as neither of these are divided into individual items.

Like the WISC-V subtest items, EFL and EAL participants were divided into their respective groups to determine the extent to which the KABC-II subtests' items may negatively dispose participants from either group. The items of the first of the two Learning Scale subtests, Atlantis, did not show any DIF (see Appendix G, Table G67). However, in the Rebus subtest, three items that showed DIF in favour of the EFL participants were identified. These were items 14 (p = .0019), 20 (p = .038), and 21 (p = .032). Amongst the Planning Scale subtest items, 6 items presented with DIF in favour of the EFL participants. Two of these were from the Story Completion subtest and included items 12 (p = .017) and 15 (p = .002), whilst the other four were from the Pattern Reasoning subtest. These were items 13 (p = .039), 17 (p = .006), 18 (p = .020), and 23 (p = .026). Conversely, none of the items of the Sequential Scale subtests, Number Recall or Word Order, showed any DIF. The same trend was found in the Knowledge Scale subtests, with none of the items of the Verbal Knowledge or Riddles subtests showing any DIF. Finally, amongst the Simultaneous Scale subtests administered to the participants of this sample, six items from the Rover subtest presented with DIF, all in favour of the EFL participants. Included here were items 5 (p = .002), 14 (p = .011), 16 (p = .009), 17 (p = .041), 21 (p = .021), and then 22 (p = .035). In the other two subtests, Triangles and Block Counting, slightly fewer items presented with DIF; however, these were again in favour of the EFL participants. The three items identified in the Triangles subtest were items 20

(p = .023), 23 (p = .001), and 24 (p = .038). In the Block Counting subtest, two items were identified, including items 16 (p = .031) and 30 (p = .006).

4.8. Conclusion

This chapter presented the results of the study, first in terms of descriptive statistics and then inferential statistics. The descriptive statistics presented provided an overview of the participants global, index/scale and subtests' scores in terms of mean, standard deviation, minimum, and maximum. Inferential statistics were presented according to the four objectives set at the outset of the study. The first of these presented the internal consistency reliability statistics of selected WISC-V and KABC-II subtests. The second objective, focussed on construct validity, included intercorrelations and confirmatory factor analyses. The third objective compared the WISC-V and KABC-II scores of EFL and EAL participants to determine whether language differences may have impacted the results of the participants in this study. The fourth objective included an investigation of specific subtests' items in the context of IRT and DIF. The next chapter will discuss these results in more depth through a comparison to existing research and an analysis of the results in the context of the Wechsler, CHC and Lurian conceptualisations of intelligence.

5. Discussion

5.1. Introduction

South African psychologists, who utilise intelligence tests such as the Wechsler Intelligence Scale for Children – Fifth Edition (WISC-V) and the Kaufman Assessment Battery for Children – Second Edition (KABC-II), have a responsibility to understand the psychometric properties and cultural-applicability of these measures, especially since they have not been standardised and normed for use in the South African context. With limited available research findings on the psychometric properties of these two intelligence tests in the South African context, this study aimed to conduct a preliminary investigation of the reliability and validity of the WISC-V and the KABC-II in a sample of South African children experiencing barriers to learning and development. In addition, the study also aimed to investigate the suitability of these two measures in a South African sample through item analysis and a comparison of the performance of English First Language (EFL) and English Additional Language (EAL) participants. In line with these aims, four objectives were set. Structured around these four objectives, this chapter provides a more in-depth discussion of the study's findings, starting with an overview of the main findings of the research study. This will be followed by a discussion of the results of the study in the context of the theoretical perspectives reported in Chapter 2 and comparisons to available literature on the psychometric properties of the WISC-V and KABC-II. The chapter will conclude the study by discussing the implications and strengths of the study, its limitations and recommendations for future research.

5.2. Summary of Findings

The first objective of this research study aimed to investigate the internal consistency reliability of the WISC-V and KABC-II subtests using the Spearman-Brown split-half reliability method. Reliability coefficients calculated for the WISC-V and the KABC-II subtests, indicated that they were reliable in assessing the cognitive abilities and processes they were designed to assess in this sample of South

African children with barriers to learning and development. Similar to results obtained during the standardisation of the WISC-V and KABC-II, the reliability coefficients indicated either excellent or good reliability for the subtests included in these analyses, signifying the dependability of these subtests in assessing the cognitive abilities and processes of this sample of South African children.

The second objective of the study aimed to investigate the construct validity of the WISC-V and KABC-II through intercorrelations and second-order confirmatory factor analysis (CFA). Construct validity in the form of convergent and discriminant validity of the WISC-V indices and subtests in this sample, were supported by intercorrelations which highlighted that the WISC-V primary subtests correlated very strongly with their respective indices, more so than they did with any of the other primary indices. Furthermore, for the majority of the WISC-V subtests, stronger correlations were observed between subtests within an index than with subtests from other indices. The only exception was the Digit Span subtest, which demonstrated a stronger correlation with other auditory working memory subtests, Letter-Number Sequencing and Arithmetic, than it did with its counterpart on the Working Memory Index, the Picture Span subtest. Similarly, in this sample of South African children with barriers to learning and development, evidence of convergent and discriminant validity was also found for the KABC-II scales and subtests, where subtests demonstrated stronger correlations with their respective scales than with any of the other scales. As far as comparisons between subtests within specific scales were concerned, evidence of convergent and discriminant validity was found for the majority of the subtests, including subtests in the Knowledge Scale, the Sequential Scale, and the Simultaneous Scale. However, in the Planning Scale, a moderate correlation between the Story Completion and Pattern Reasoning subtests were observed, with Pattern Reasoning showing a stronger correlation with Rover, and Story Completion a stronger correlation with Triangles. Moreover, subtests within the Learning Scale, Rebus and Atlantis showed a weak correlation.

Further investigations into the construct validity of the WISC-V were conducted through second-order CFA. The results of the WISC-V CFA provided additional evidence for the construct validity of this test by confirming the five-factor model proposed by the Wechsler development team in relation to the five primary indices and their respective subtests. The WISC-V CFA further indicated that in this sample, fluid reasoning had the highest *g*-loading, confirming previous research which indicated the importance of fluid reasoning in a child's general cognitive functioning. Conversely, processing speed had the lowest *g*-loading, indicating that in this sample, processing speed contributed the least to the participants' overall intellectual functioning when compared to the other first-order factors. With reference to the WISC-V subtests, the results of the current study revealed that the Symbol Search subtest would be considered a stronger indicator of processing speed compared to Coding, Vocabulary was a stronger indicator of verbal comprehension compared to Similarities, Visual Puzzles was a stronger indicator of visual-spatial processing compared to Block Design, Matrix Reasoning was a stronger indicator of fluid reasoning compared to Figure Weights, and Picture Span was a stronger indicator of working memory compared to Digit Span. In investigating the construct validity of the KABC-II in this sample through CFA, the KABC-II Fluid-Crystallised model proposed by the development team could not be fitted as stated as there were too many parameters and too little data. With the Number Recall and Planning Scale parameters fixed to 1.00, the visual processing factor had the highest *g*-loading, followed by the crystallised reasoning factor, the long-term storage and retrieval factor and then the short-term memory factor. Moreover, the results further indicated that the Verbal Knowledge subtest was a stronger indicator of crystallised reasoning, compared to Riddles, Rover was a stronger indicator of visual processing compared to Triangles and Block Counting, Story Completion was a stronger indicator of planning compared to Pattern Reasoning, and Rebus was a stronger indicator of long-term storage and retrieval compared to Atlantis.

Objectives three and four aimed to investigate the suitability of the WISC-V and KABC-II in this sample of South African children. This was achieved by comparing the results of EFL and EAL participants to investigate whether English language status had a significant impact on participants' performance on the WISC-V and KABC-II global indices and scales. In addition, item analysis of WISC-V and KABC-II subtests' items using a Graded Response Model (GRM) was conducted to investigate the difficulty of the items, as well as their discriminatory ability in this sample. Finally, evidence of differential item functioning (DIF) based on English language status was investigated. The results of the comparisons of the KABC-II and WISC-V global and index/scale scores indicated that EFL participants consistently obtained higher scores than EAL participants, except for the KABC-II Sequential Scale. In addition, participants' English language status had a significant impact on two of the WISC-V global indices, the Full-Scale Intelligence Quotient (FSIQ) and the General Ability Index, as well as three of its primary indices: the Verbal Comprehension, Visual-Spatial, and Fluid Reasoning Indices. However, only one KABC-II scale seemed to be significantly influenced by English language status, the Knowledge Scale. Participants also consistently obtained higher scores on the KABC-II compared to the WISC-V. The findings of this study thus substantiated earlier South African research and indicated that, in this sample, the KABC-II was more suitable for assessing the intellectual functioning of the participants compared to the WISC-V. However, the results of the item analyses did not support the current order of the majority of the WISC-V and KABC-II subtests' items based on their level of difficulty. Although fewer adaptations were indicated to the order of the KABC-II and WISC-V subtests' items that measure auditory memory span and working memory, more adaptations were indicated to the order of subtests' items that measure crystallised reasoning, English language proficiency, and acquired knowledge. With this in mind, Differential Item Functioning (DIF) in the items of the WISC-V and KABC-II subtests was investigated by comparing the results of the EFL and EAL participants in this study. In the WISC-V, DIF was observed in all of the indices, while the KABC-II only demonstrated DIF in three of its scales.

Nonetheless, practitioners and researchers must be aware of the results, especially if the WISC-V or KABC-II is to be adapted for use in South Africa.

5.3. Objective 1: Reliability of the WISC-V and the KABC-II

Standardised intelligence tests, such as the WISC-V and KABC-II, are generally expected to show excellent reliability where measurement error has virtually no effect on the individual's performance (Murphy & Davidshofer, 2005; Sattler, 2014). In line with this, investigations into the internal consistency reliability of the WISC-V subtests in this South African sample revealed that the majority of the WISC-V subtests showed excellent reliability. Only four subtests, which included the three Digit Span components and the Comprehension subtest, showed good reliability. While minimal research is available on the reliability of the WISC-V subtests in other samples, these findings are commensurate with those obtained during the standardisation of the WISC-V (Flanagan & Alfonso, 2017; Reynolds & Hadorn, 2016; Wechsler, 2014b). The majority of the subtest reliability coefficients calculated for the overall normative sample, as well as smaller samples of children diagnosed with specific learning disorder (SLD) or attention-deficit/hyperactivity disorder (ADHD), also ranged from good to excellent. Similarly, from the 11 KABC-II subtests administered to this sample, eight showed excellent reliability, while the remaining three showed good reliability. Included in the latter were the two Sequential Scale subtests, Word Order and Number Recall, as well as one Planning Scale subtest, Story Completion. Again, limited research is available reporting the reliability of the KABC-II in different samples. However, comparing these reliability coefficients to those obtained for the KABC-II normative sample (ages 7–18 years) (Kaufman & Kaufman, 2004; Kaufman et al., 2005b) and the Mitchell et al. (2018) sample, similarity in the coefficients is observed. Across these samples, the majority of the reliability coefficients for the KABC-II subtests were once again either good or excellent. Thus, the reliability coefficients for the KABC-II and WISC-V subtests obtained in this sample were similar to those reported for other samples, indicating that the WISC-V and KABC-II subtests were reliable in assessing the cognitive abilities and processes of this sample of South African children.

5.4. Objective 2: Construct Validity of the WISC-V and the KABC-II

Objective 2 of this research study aimed to examine the construct validity of the WISC-V and KABC-II through intercorrelations and second-order CFA. As such, investigations into the construct validity of the WISC-V and KABC-II aimed to examine whether the indices, scales, and subtests of the WISC-V and the KABC-II accurately assessed the cognitive abilities and processes they were designed to evaluate in this sample of South African children.

5.4.1. WISC-V Intercorrelations

During the development of the WISC-V, it was assumed that intercorrelations would be noted between the FSIQ and the primary Indices as the latter measure the *g*-factor to a greater or lesser extent (Reynolds & Hadorn, 2016; Wechsler, 2014b). In line with this assumption, as well as the results obtained by Wechsler (2014b), the results of the current research study highlighted very strong correlations between the WISC-V FSIQ and the Verbal Comprehension, Fluid Reasoning and Visual-Spatial Indices, as well as substantial correlations between the FSIQ and the Working Memory and Processing Speed Indices. Both the Cattell-Horn-Carroll (CHC) theory and Wechsler's conceptualisation of intelligence, support the assumption that a person's global or overall intellectual ability is constituted by several underlying cognitive abilities (Flanagan & Dixon, 2014; McGrew, 2009; Sternberg et al., 2008; Van Eeden & De Beer, 2018; Wasserman, 2018). In the context of these theories, the results of this study emphasised that the participants' global intellectual functioning, as measured by the WISC-V FSIQ, was predominantly influenced by crystallised reasoning, fluid reasoning, and visual-spatial processing and reasoning. While processing speed and working memory also contributed to overall intellectual functioning, these contributions were secondary to those of the other stated cognitive abilities (Flanagan et al., 2018; Flanagan & Dixon, 2014; Kaufman et al., 2016d; Schneider et al., 2017).

A second assumption made by the WISC-V development team was that convergent validity would be supported, with indices and subtests measuring specific cognitive abilities and processes showing stronger correlations with each other than with indices and subtests measuring different cognitive abilities and processes (Reynolds & Hadorn, 2016; Wechsler, 2014b). While intercorrelations among the WISC-V primary indices were predominantly moderate or weak, support for the convergent validity of the WISC-V was found in a substantial correlation between the Visual-Spatial Index and the Fluid Reasoning Index, which is again commensurate with the findings of Wechsler (2014b). In previous editions of the WISC, including the Wechsler Intelligence Scale for Children – Fourth Edition (WISC-IV), these two indices were combined into the Perceptual Reasoning Index (Wechsler, 2003). It was only with the publication of the WISC-V that they were separated into two distinct indices. From CHC and Lurian perspectives, however, they measure similar cognitive abilities and processes, including visual-spatial processing and reasoning, attention and concentration, as well as fluid reasoning and problem-solving (Flanagan et al., 2018; Kaufman et al., 2016d; Schneider et al., 2017). Based on the fact that they seem to measure similar cognitive abilities and processes, an argument can be made from a theoretical perspective that they can be combined into one index. This argument is further supported by a substantial correlation between the Fluid Reasoning Index and the two primary subtests of the Visual-Spatial Index, as well as by a substantial correlation between the Visual-Spatial Index and the two primary subtests of the Fluid Reasoning Index.

Convergent validity was further supported by very strong correlations between the primary subtests and their respective indices: the Verbal Comprehension Index and the Similarities and Vocabulary subtests, the Visual-Spatial Index and the Block Design and Visual Puzzles subtests, the Fluid Reasoning Index and the Matrix Reasoning and Figure Weights subtests, the Working Memory Index and the Digit Span and Picture Span subtests, and the Processing Speed Index and the Coding and Symbol Search subtests. Rendering support for the discriminant validity of the WISC-V, the primary subtests demonstrated stronger correlations with their respective indices than with any of the other primary indices. Moreover, the primary subtests within an index demonstrated stronger correlations with each other than with subtests from other primary indices. As such, substantial correlations were noted between the Similarities and Vocabulary subtests, the Visual Puzzles and Block Design subtests, the Matrix Reasoning and Figure Weights subtests, and the Coding and Symbol Search subtests. However, while the Digit Span subtest rendered a moderate correlation with the Picture Span subtest, it showed slightly stronger correlations with the Arithmetic and Letter-Number Sequencing subtests. These three subtests, Arithmetic, Digit Span, and Letter-Number Sequencing, all rely on auditory working memory, whereas the Picture Span subtest relies more on visual working memory (Flanagan et al., 2018; Kaufman et al., 2016d; Schneider et al., 2017). As such, it provides support for assessing different components of an individual's working memory through the two subtests included in the Working Memory Index. Similar findings were observed by the WISC-V development team (Wechsler, 2014b).

Of further interest is the fact that the Picture Span subtest showed the exact moderate correlation with the Matrix Reasoning subtest, as it did to the Digit Span subtest. From a Lurian perspective, Matrix Reasoning is generally considered a measure of simultaneous, rather than sequential processing. However, Kaufman et al. (2016d) acknowledge that some of its items rely more on sequential processing than simultaneous processing. Thus, the moderate correlation between the Picture Span and Matrix Reasoning subtests may suggest that participants in this sample employed sequential processing skills when completing these subtests. It may also indicate the application of visual working memory in completing the Matrix Reasoning subtest. These subtests also rely on other essential cognitive abilities and processes, such as visual processing, attention, and concentration, as substantiated by the observed correlations (Flanagan et al., 2018; Kaufman et al., 2016d; Schneider et al., 2017).

It is worth noting that, although the Visual Puzzles subtest showed a strong correlation with the Block Design subtest, it also demonstrated substantial correlations with the Figure Weights and Matrix Reasoning subtests. This can be understood against the backdrop that both the Matrix Reasoning and Visual Puzzles subtests rely on an understanding of part-whole relationships, visual-perceptual discrimination, and attention and concentration (Schneider et al., 2017). In addition, Visual Puzzles and Figure Weights both rely on attention, concentration, visual processing, and working memory. Conversely, while Block Design provides an individual with an indication of the accuracy of their responses through concrete visual feedback, Visual Puzzles relies more on mental processing and solutions in a similar manner to Figure Weights and Matrix Reasoning (Wechsler, 2014b).

The array of cognitive abilities and processes assessed by the Arithmetic subtest (Wechsler, 2014b), was supported by the results of the current study where this subtest demonstrated a very strong correlation with the Quantitative Reasoning Index, as well as substantial correlations with the Nonverbal, General Ability, Auditory Working Memory, Working Memory, Cognitive Proficiency, Fluid Reasoning, and Verbal Comprehension Indices. Additionally, it also showed substantial correlations with the Information, Digit Span, and Vocabulary subtests. In contrast, it only showed a moderate correlation with its counterpart on the Quantitative Reasoning Index, specifically the Figure Weights subtest. This can likely be explained by the fact that the Figure Weights subtest relies more on quantitative reasoning and general sequential reasoning. In contrast, Arithmetic is more dependent on mathematical knowledge and achievement. Additionally, as stated earlier, the Arithmetic subtest relies heavily on auditory working memory, which can explain its correlation with the Auditory Working Memory, Working Memory, and Cognitive Proficiency Indices. Moreover, given that the items all contain word problems, an individual will require an understanding of verbal information and quantitative concepts, to complete the subtest, which can explain its correlation with the Verbal Comprehension and General

Ability Indices (Flanagan et al., 2018; Flanagan & Dixon, 2014; Kaufman et al., 2016d; Schneider et al., 2017).

5.4.2. KABC-II Intercorrelations

Like the WISC-V, this study also investigated convergent and discriminant validity through an examination of intercorrelations between the KABC-II global indices, scales and subtests administered to this sample. Within the context of CHC theory and Luria's neuropsychological framework, which stipulate the interrelated nature of cognitive abilities and processes respectively (Amod, 2013; Flanagan & Dixon, 2014; McGrew, 2009; Naglieri, 1999; Sternberg et al., 2008; Wasserman, 2018), it was expected that the KABC-II global indices and scales would display some correlations. Similar to the findings by Kaufman and Kaufman (2004), a robust correlation was noted between the Fluid Crystallised Index (FCI) and the Mental Processing Index (MPI). This is understandable given that these global indices share four scales and thus several cognitive abilities and processes (Kaufman et al., 2005b). In addition, very strong correlations were also observed between the Simultaneous Scale and these two global indices, in line with similar findings by Kaufman and Kaufman (2004). Theoretically, these strong correlations suggest that the participants' performance on both global indices was significantly influenced by their visual and simultaneous processing skills. In contrast, none of the other scales showed strong correlations with the two global indices. However, substantial correlations were observed between the FCI and the Planning and Knowledge Scales, as well as between the MPI and the Planning and Learning Scales. This finding is supported by the fact that these scales all contribute subtests, thereby encompassing cognitive abilities and processes within these global indices (Kaufman et al., 2005b). However, it does not explain why the Sequential Scale showed moderate correlations with the global indices, as it also contributes to both. In the development of the KABC-II, it showed substantial correlations across the older age groups (Kaufman & Kaufman, 2004). This can be explained theoretically and is substantiated by the findings of this study that the participants' short-term and working memory skills contributed the least to their overall intelligence when assessed on the KABC-II (Drozdick et al., 2018; Flanagan et al., 2018; Kaufman et al., 2005c). This could be because only one of the Sequential Processing subtests, Word Order, measures working memory and only in the completion of its more complex items. Compared to the findings of the WISC-V, the KABC-II thus appeared less effective in measuring working memory in this sample.

Although the Knowledge Scale is omitted in the calculation of the MPI, a substantial correlation was observed between this scale and the MPI. Kaufman and Kaufman (2004) reported similar findings where the Knowledge Scale and the two global indices correlated. Explicitly based on CHC theory, the Knowledge Scale was incorporated as an indication of acquired knowledge and as such, crystallised reasoning, but also as a measure of fluid reasoning which it shares with the Planning and Simultaneous Scales (Drozdick et al., 2018; Flanagan et al., 2018; Kaufman et al., 2005c). Additionally, it shares cognitive abilities and processes with other scales, including auditory processing with the Sequential Scale and long-term storage, retrieval, and learning abilities with the Learning Scale. CHC theory further acknowledges that a person's intellectual functioning is the result of several interrelated cognitive abilities that contribute to overall intellectual functioning. It emphasises that these cognitive abilities are functionally integrated because, in real-world contexts, the execution of cognitive tasks relies on several cognitive abilities (Cormier et al., 2017; Evans et al., 2002; McGill, 2017). Similarly, from a Lurian perspective, Amod (2013) and Naglieri (1999) note that the functional units identified by Luria operate in an interdependent and integrated manner, constantly interacting with a person's level of acquired knowledge, thereby incorporating elements of crystallised intelligence. Therefore, while the Knowledge Scale primarily relies on crystallised reasoning and acquired knowledge, the findings of this study confirmed that other cognitive abilities and processes will also be integrated to perform the tasks posed to the individual.

Convergent and discriminant validity of the KABC-II subtests in this sample were further substantiated by the fact that the majority of the subtests within a scale showed stronger correlations with their respective scales than with the other scales. As such, very strong and substantial correlations were observed between the Sequential Scale and the Number Recall and Word Order subtests, the Simultaneous Scale and the Rover, Triangles, and Block Counting subtests, the Planning Scale and the Story Completion and Pattern Reasoning subtests, the Learning Scale and the Atlantis and Rebus subtests, and the Knowledge Scale and the Verbal Knowledge and Riddles subtests (Drozdick et al., 2018; Flanagan et al., 2018; Kaufman et al., 2005c). Similar to what was observed between the WISC-V Fluid Reasoning and Visual-Spatial Indices and their subtests, the KABC-II Simultaneous and Planning Scales, along with their subtests, also yielded substantial correlations. Again, this is evidence that these scales and subtests share cognitive abilities and processes, including planning, fluid reasoning, attention and concentration, simultaneous processing, and visual-perceptual reasoning, organisation, and discrimination (Drozdick et al., 2018; Flanagan et al., 2018; Kaufman et al., 2005c).

Further evidence of convergent and discriminant validity in the KABC-II subtests was observed in correlations between the subtests within a scale, which were generally higher than their correlations with subtests from other scales. Although within the Planning Scale, Story Completion and Pattern Reasoning showed a moderate correlation, Pattern Reasoning rendered a substantial correlation with the Rover subtest. Conversely, Story Completion showed a moderate correlation with the Triangles subtest. The moderate correlation between the two Planning Scale subtests, was likely evidence of the fact that while they both assess inductive reasoning, the Pattern Reasoning subtest relies more on nonverbal, visual-spatial reasoning, while Story Completion relies more on general sequential reasoning and general information and knowledge (Drozdick et al., 2018; Flanagan et al., 2018; Kaufman et al., 2005c). Finally, the two Learning Scale subtests, Rebus and Atlantis, showed a weak correlation. In the standardisation of the KABC-II, these subtests showed a moderate correlation (Kaufman & Kaufman, 2004). Although they would theoretically rely on long-term storage and retrieval, associative memory, and learning (Drozdick et al., 2018; Flanagan et al., 2018; Kaufman et al., 2005c), the results of this study indicated that the participants utilised different cognitive abilities and skills in their completion of these subtests, which would explain the weak correlation.

5.4.3. WISC-V Factor Analysis

Following the intercorrelations, factorial validity was investigated through second-order factor analyses (Roodt & De Kock, 2018c). As discussed in Chapter 2, factor analysis aims to further investigate and explain the theoretical relationships between latent variables or factors and measured variables (Byrne, 2005; MacCallum, 2012; Schreiber et al., 2006). In investigating the factorial validity of the WISC-V and KABC-II in this sample of South African children, second-order CFA was employed. CFA is not only deductive and based on empirical research, but also theory-driven, focusing on confirming or investigating a prespecified model that can be theoretically labelled and understood (Byrne, 2005; Wechsler, 2014b; Westen & Rosenthal, 2003). It enables researchers to investigate the extent to which proposed factor models, such as those developed by the KABC-II and WISC-V teams, hold across different socio-cultural contexts and apply similarly in diverse populations (Kaufman et al., 2005b; Roodt & De Kock, 2018c). In second-order factor analyses, the indices or scales of the WISC-V and KABC-II would represent the first-order factors. In contrast, the global indices would represent the second-order factors.

Following research conducted into the WISC-V standardisation sample, Wechsler (2014b) conducted CFA with the ten primary subtests and found sufficient evidence for a five-factor model where the five primary indices load onto the FSIQ as a global factor. As stated in Chapter 4, the results of this study confirmed this five-factor model, providing evidence for the construct validity of the WISC-V in this sample of South African children with barriers to learning and development. Similar findings were observed in additional factor analyses conducted with data from the normative sample (Chen et al., 2015; Reynolds & Keith, 2017). Moreover, Egeland et al. (2022), Sheiber (2016a), Van de Vijver et al. (2019), and Wilson et al. (2023, 2024) investigated the factorial validity of standardisations of the WISC-V across several countries and ethnic groups. Additionally, they found sufficient evidence to support the five-factor model. In contrast to this, other factor analyses conducted with data from the normative sample, as well as international and ethnically diverse samples found support for either a three-factor (Dombrowski et al., 2015) or four-factor model (Canivez et al., 2016, 2017, 2018; Fenollar-Cortés & Watkins, 2019; Graves et al., 2021; Lecerf & Canivez, 2018, 2022; Taylor, 2019; Watkins et al., 2018).

The results discussed in Chapter 4 further revealed that the fluid reasoning factor had the highest *g*-loading, followed by the visual-spatial factor, the working memory factor, the verbal comprehension factor, and then the processing speed factor. Similarly, research conducted by the Wechsler development team (Wechsler, 2014b), Chen et al. (2015), and Reynolds and Keith (2017) also found that fluid reasoning was the strongest indicator of general intelligence. This provides evidence for the importance of fluid reasoning in an individual’s overall intellectual functioning. Regarding the subtests, the current study's results revealed that Symbol Search had the highest correlation among all subtests and first-order factors, loading on the processing speed factor. Conversely, Coding had the lowest correlation, also loading on the processing speed factor. Thus, in the current sample, Symbol Search would be considered a stronger indicator of processing speed compared to Coding. Theoretically, this provides evidence that Symbol Search, as a measure of perceptual speed is a stronger indicator of processing speed than Coding, as a measure of the rate of test taking (Drozdick et al., 2018; Flanagan et al., 2018; Kaufman et al., 2005c). Similar findings were observed in the studies by Wechsler (2014b) and Reynolds and Keith (2017). However, in both of these studies, the differences in the factor loadings of these two subtests on the processing speed factor were slightly smaller than those observed in the current study. Research by Chen et al. (2015) and Egeland et al. (2022) indicated minimal differences between the factor loadings of these two subtests on the processing speed factor.

5.4.4. KABC-II Factor Analysis

Research conducted by the KABC-II development team during the standardisation of this test (Kaufman & Kaufman, 2004; Kaufman et al., 2005b) found support for the current factor structure of the test through CFA. Later research by Reynolds et al. (2007) with data from the normative sample, as well as cross-cultural research conducted in India (Malda et al., 2010), Africa (Bangirana et al., 2009; Chernoff et al., 2018), and South Africa (Mitchell et al., 2018), also found sufficient support for the KABC-II factor models in these diverse samples. However, factor analyses conducted by McGill and Spurgin (2017) and McGill and Dombrowski (2018) rendered contradictory results. As reported in Chapter 4, the KABC-II FCI model proposed by Kaufman and Kaufman (2004) could not be fitted as proposed in this study, due to the excessive number of parameters and limited data. As a result, the Number Recall and Planning Scale parameters were fixed to 1.00. With this in mind, the visual-spatial or simultaneous processing factor had the highest *g*-loading, followed by the crystallised reasoning (knowledge), long-term storage and retrieval (learning) and short-term memory (sequential processing) factors. Research conducted from both the CHC and Lurian perspectives yielded similar results, with the fluid reasoning or planning factor consistently showing the highest *g*-loading, followed by the simultaneous processing or visual-spatial factor (Mitchell et al., 2018; Renner et al., 2023; Reynolds et al., 2007). Similar to the WISC-V, these results again emphasise the importance of fluid reasoning and planning in an individual's intellectual functioning, as well as visual-spatial and simultaneous processing.

As stated, a weak correlation was observed between the two subtests of the Learning Scale, Rebus and Atlantis. While being aware that Number Recall was fixed at 1.00, the results of the current study indicated that Rebus was a stronger indicator of long-term storage and retrieval, as well as learning, compared to Atlantis. This suggests that the participants of this study relied more on forming verbal and visual associations in storing information in their long-term memory than on associations between verbal and auditory information.

5.5. Objective 3: Comparisons in Performance of English First Language and English Additional Language Participants

In addition to the investigations conducted into the reliability and validity of the KABC-II and the WISC-V in this sample, the last two objectives related to the second aim of the research study, which was to investigate the suitability of the WISC-V and KABC-II in this sample of South African children. The first of these two objectives aimed to compare the results of a group of EFL participants with those of a group of EAL participants, to investigate whether English language status impacted the participants' performance.

The first indication of the impact of English language proficiency on performance on both intelligence tests was that EFL participants consistently obtained higher mean scores compared to EAL participants on the FSIQ, General Ability, Cognitive Proficiency, and primary indices of the WISC-V. Similarly, when comparing the mean KABC-II global indices and scale scores of the participants, it was once again evident that the EFL participants obtained higher mean scores than the EAL participants. However, the Sequential Scale was an exception, with the results of this study indicating that EAL participants obtained a higher mean score than EFL participants. In investigating whether the differences between the two groups were statistically significant, the results of a General Linear Model (GLM) revealed that language had a more significant impact on participants' performance on the WISC-V compared to the KABC-II. As far as the WISC-V is concerned, statistically significant differences between the performance of the language groups were observed on both the global indices included in these analyses, the FSIQ and the General Ability Index, as well as three of the five primary indices, Verbal Comprehension Index, Visual-Spatial Index, and Fluid Reasoning Index. However, as far as the KABC-II is concerned, a statistically significant difference was only observed on the Knowledge Scale.

Although published research on the impact of English language status on children's performance on the WISC-V and KABC-II is limited, several studies have investigated other socio-cultural factors. For instance, research by Kaufman et al. (2016e) and Weiss, Locke, et al. (2019) indicated that parent education level was strongly correlated with participants' performance on the WISC-V, particularly their performance on the FSIQ, the General Ability Index, and the Verbal Comprehension Index. Conversely, children's performance on the Processing Speed Index was less influenced by parent education level, likely because this index is less reliant on a child's verbal ability and crystallised intelligence. Additional research investigated the role of ethnicity on children's WISC-V scores (Kaufman et al., 2016e). In this regard, Weiss, Locke, et al. (2019) cautioned that ethnicity is a very complex construct as persons from different ethnic groups not only differ in terms of English language proficiency given their cultural and historic backgrounds, but also have more, or less access to resources, wealth, and quality education. Nonetheless, the outcome of these investigations indicated that the most considerable mean score differences were observed between White and African American participants and White and Hispanic participants. However, after adjustments were made for sex and parental education level, ethnicity accounted for only 3.5% of the variance in the FSIQ when comparing the scores of African American and White children. Moreover, when exploring the impact of ethnicity on indices that rely less on crystallised intelligence, including Working Memory, Processing Speed, and Cognitive Proficiency, almost no variance was observed between African American and White participants. A similar outcome was observed in comparing the scores of Hispanic and White children. Although after adjustments were made for sex and parent education level, ethnicity still accounted for more variance in the FSIQ than in the Nonverbal Index. Thus, ethnicity and parent education level had a more significant impact on indices that rely on verbal ability, crystallised intelligence and acquired knowledge than those who are considered more nonverbal in nature, more so the Processing Speed, Working Memory and Cognitive Proficiency Indices (Kaufman et al., 2016e; Weiss, Locke, et al., 2019). Carvalho (2023) investigated the impact of language and socio-economic status on children's performance on the WISC-V, specifically the General Ability, Verbal Comprehension, Visual-Spatial, and Fluid Reasoning Indices. The results of the study, which included African American children from lower socioeconomic backgrounds, indicated small to medium differences between the normative sample and the African American sample on the Visual-Spatial and Fluid Reasoning Indices, respectively. However, a large difference was observed in the groups' scores on the General Ability Index and an exceptionally large difference in their scores on the Verbal Comprehension Index. Carvalho (2023) concluded that language and socio-economic status have a significant impact on children's performance on the WISC-V, more so on indices and subtests that are more verbal and rely more on crystallised intelligence and acquired knowledge.

The only other African study conducted with the WISC-V (Gichovi et al., 2018) revealed that the mean primary index and subtest scores of a group of children from Kenya were commensurate with those of the normative sample. Although socio-cultural factors such as parental education level, English language proficiency, and socio-economic status were not reported, it was assumed that the children were from similar socio-economic backgrounds, as they all attended private schools where the medium of instruction was English. However, South African research conducted with previous editions of the Wechsler Intelligence Scale for Children (WISC) (Bickell, 2015; Ferrett, 2011; Shuttleworth-Edwards, Van der Merwe, et al., 2013; Skuy et al., 2001) indicated that demographic and socio-cultural factors, such as socio-economic status, English language proficiency, quality of education, age, and sex had a significant impact on the performance of South African participants. The results of this South African study substantiated the findings of these previous South African studies and Carvalho (2023), indicating that English language proficiency had a significant impact on the performance of participants who reported English as their additional language. These differences were not only observed in indices relying on English language proficiency, acquired knowledge and crystallised reasoning, but also on those requiring more nonverbal, visual-spatial and fluid reasoning (Flanagan et al., 2018; Flanagan & Dixon, 2014; Kaufman et al., 2016d; Schneider et al., 2017). The most significant effect sizes were observed in EFL and EAL participants' performance on the General Ability Index, followed by the Visual-Spatial Index and then the FSIQ, Verbal Comprehension Index and Fluid Reasoning Index. Research by Galindo et al. (2013) and Roberson et al. (2002) confirmed that socio-cultural factors can impact the development of visual processing and planning skills and processes. Moreover, Kuthria and Serpell (1998), Matafwali and Serpell (2014) and Shuttleworth-Edwards (2016) emphasise that socio-cultural factors can impact children's performance on both verbal and nonverbal tasks of intelligence tests, a finding substantiated by this research. However, in line with the findings of Kaufman et al. (2016e) and Weiss, Locke, et al. (2019), the results of this study confirmed that English language status had less impact on the Cognitive Proficiency, Processing Speed and Working Memory Indices, which rely more on a person's cognitive proficiency than on reasoning and English language skills (Flanagan et al., 2018; Flanagan & Dixon, 2014; Kaufman et al., 2016d; Schneider et al., 2017).

In the development of the KABC-II, several steps were taken to increase its cross-cultural suitability, including the exclusion of knowledge-based items in the calculation of the MPI, the inclusion of teaching items, the reduction in verbal test instructions and verbal responses, and the inclusion of ethnic minority groups in the standardisation sample (De Sousa et al., 2010; Fletcher-Janzen & Lichtenberger, 2005; Kaufman & Kaufman, 2004). In their investigations, Kaufman and Kaufman (2004) found modest differences between the mean scores of the four ethnic groups included in the normative sample when assessed on the KABC-II. After controlling for the potential impact of parental education, they found the effect of ethnic difference on the KABC-II to be minimal. In addition, Fletcher-Janzen (n.d.) also found that the results obtained from a Taos Pueblo sample were commensurate with those of the normative sample, despite several socio-cultural differences that could have impacted the results of the Taos Pueblo sample. Similarly, Dale et al. (2011) also found minimal differences in the mean scores of White and African American preschool children when assessed on the KABC-II. However, research by Gomez (2008), which included the MPI as a global score, indicated that language proficiency significantly impacted the performance of a group of Mexican American children on the KABC-II, although mention was made that the lower socio-economic status of this group, and their barriers to learning and development, could have impacted test results further. In addition to these international studies, South African research conducted using the Kaufman Assessment Battery for Children (K-ABC) by De Sousa et al. (2010) concluded that the K-ABC norms were suitable for use when assessing the cognitive processing skills of a group of bilingual children from an urban area. In contrast to the findings from the WISC-V, the results of this study (see Chapter 4) indicated that although the EFL group obtained higher mean scores than the EAL group on most of the KABC-II scales, the only significant difference was observed on the Knowledge Scale. This substantiates previous South African research that English language proficiency has a more significant impact on measures of language, crystallised reasoning, and acquired knowledge when compared to more nonverbal measures (Bickell, 2015; Gomez, 2008; Laher, 2024; Laher & Cockcroft, 2017; Shuttleworth-Edwards, 2016; Shuttleworth-Edwards, Van der Merwe, et al., 2013).

There are limited published research findings comparing the results of the WISC-V and the KABC-II. Those that are available (Breaux et al., 2017; Wechsler, 2014b) tend to focus on test comparisons in investigating the validity of these measures. The only other South African comparative study included former editions of the WISC and the K-ABC. In this study of Skuy et al. (2000), which compared the performance of White and Black South African children on the Wechsler Intelligence Scale for Children – Revised (WISC-R) and the K-ABC, support was found for the use of the K-ABC as a more culturally fair test compared to the WISC. The outcome of the current study supported these findings. While several statistically significant differences were noted in the EFL and EAL participants’ mean scores on the WISC-V, only one statistically significant difference was found in their mean scores on the KABC-II. Also, similar to the Skuy et al. (2000) study, results of the current study indicated that while a significant difference was found in the two groups’ mean scores on the WISC-V FSIQ, significant differences were not observed in the groups’ mean scores on the KABC-II MPI or FCI. It is important to be cognisant that while the present study included the Knowledge Scale, this was not included in the Skuy et al. (2000) study as it was not available in the original K-ABC (Greenop et al., 2013). Although these findings indicate that the KABC-II was more suitable in the assessment of this sample of South African children with barriers to learning and development's intellectual functioning compared to the WISC-V, Ariffin et al. (2010) and Toland (2014) emphasise the importance of item analysis in examining the cross-cultural fairness of intelligence tests for use in a specific context, because the statistical analyses conducted happens not only at the level of the overall scale, but also at an item level.

5.6. Objective 4: Item Analysis and Differential Item Functioning

5.6.1. Item Analysis of WISC-V and KABC-II Subtests' Items

The last of the study's objectives aimed to further investigate the construct validity of the KABC-II and WISC-V subtests by specifically focussing on the items of these subtests. This was done through item analysis situated in Item Response Theory (IRT). It also investigated whether the items of the WISC-V and KABC-II showed item bias, known as Differential Item Functioning (DIF), when the performance of EFL and EAL participants were compared. In their article on construct validity, specifically convergent and discriminant validity, Campbell and Fiske (1959) reported the importance of item analysis in investigating the construct validity of test items. Through item analysis, one is able to investigate whether the items of the WISC-V and KABC-II subtests are able to distinguish between the ability levels of participants in a particular sample in relation to the constructs these items have been designed to measure. In addition, the researcher is able to obtain insight into the difficulty of subtests' items, especially important when items are administered to children outside of the normative sample.

Chapters 3 and 4 stated that in this study, a graded Item Response Theory (IRT) model, also known as a GRM, was used to investigate the discriminatory ability and difficulty level of the KABC-II and WISC-V subtests' items. The purpose of an intelligence test is to assess children's intellectual functioning by measuring various cognitive abilities and processes. For the test to be effective, the subtests' items must be designed to differentiate between different ability levels of the cognitive abilities or processes assessed. As reported in Chapter 4 of the study, the three components of the WISC-V Digit Span subtest —Digit Span Forward, Digit Span Backward, and Digit Span Sequencing —showed greater sensitivity in discriminating between participants based on their ability levels. In the context of CHC theory and Luria's neuropsychological theory, the latent constructs measured would be memory span, auditory working memory, and sequential processing (Flanagan et al., 2018; Flanagan & Dixon, 2014; Kaufman et al., 2016d; Schneider et al., 2017). Similarly, among the KABC-II subtests, the Number Recall subtest, another measure of auditory memory span and sequential processing, was again amongst the subtests with higher discriminatory ability (Drozdick et al., 2018; Flanagan et al., 2018; Kaufman et al., 2005c). Conversely, despite showing the highest mean subtest score among the KABC-II subtests, the Atlantis subtest had the lowest discriminatory value. Combined with earlier concerns about its contribution to long-term storage and retrieval, further research is needed to assess the suitability of this subtest.

About the difficulty levels of the WISC-V and KABC-II subtests' items, the results of the statistical analyses conducted (see Chapter 4) indicated that the current order of several subtests' items was deemed inappropriate in this sample. However, it must be emphasised that the normative sample included a wider range of ages and a population of typically developing children. Regardless, starting with the items of the WISC Verbal Comprehension Index subtests, the results of the GRM indicated that substantial differences from the standardised version of the test in the order of the items in the Similarities, Vocabulary, Information, and Comprehension subtests, given their difficulty (see section). During the adaptation and standardisation of the Wechsler Adult Intelligence Scale – Fourth Edition for South Africa (WAIS-IV-SA), Wechsler (2014d) performed a Rasch analysis to investigate the item order of subtests of the WAIS-IV in a South African sample. According to the outcome of their analyses, most of the subtests retained their order. However, similar to this study, the order of the items of the Similarities, Vocabulary, Comprehension, and Information subtests differed. This is likely the result of the fact that all of these subtests rely on English language proficiency, crystallised reasoning, and acquired general knowledge (Flanagan et al., 2018; Flanagan & Dixon, 2014; Kaufman et al., 2016d; Schneider et al., 2017).

Although some difference was observed in the order of the items in the Visual-Spatial Index subtests for this South African sample, specifically in Block Design and Visual Puzzles, this variance was less than that observed for the Verbal Comprehension subtests. Results for the Fluid Reasoning Index items indicated several adjustments to the order of difficulty of these items in this sample. The Arithmetic subtest showed the most significant variance in this index, with three items identified as very difficult and another seven excluded because none of the participants were able to respond correctly to these. Compared to the other two subtests, Arithmetic relies more on acquired knowledge, specifically mathematical knowledge, which could have made it more difficult for the participants of this study to complete this subtest (Flanagan et al., 2018; Flanagan & Dixon, 2014; Kaufman et al., 2016d; Schneider et al., 2017).

Subtests' items that showed the least variance in this South African sample belong to the Working Memory and Auditory Working Memory Indices, including the three components of the Digit Span subtest and Letter-Number Sequencing. Thus, not only are these subtests generally more sensitive to the ability levels of the participants about the latent constructs measured, but they also showed the least variance in difficulty when compared to the original US normative sample and the current order of item administration. A counterintuitive finding is that, on both the Digit Span Forward component and the Picture Span subtest, participants correctly responded to some items with a higher number of stimuli but incorrectly to some items with fewer stimuli. As measures of memory span and working memory, these subtests would be sensitive to attention (Flanagan et al., 2018; Flanagan & Dixon, 2014; Kaufman et al., 2016d; Schneider et al., 2017). Thus, participants' results could have been influenced by fluctuations in attention, mainly because several participants reported difficulties with attention as their primary barrier to learning.

As far as the KABC-II subtests' items are concerned, the results from this South African sample indicated some adjustment in the order of items for both the Atlantis and Rebus subtests. However, none of the items of these subtests were presented as particularly difficult in this sample. Conversely, a number of items in these subtests were identified as being very easy. In both of these Learning Scale subtests, the consistent corrective feedback provided and the large number of teaching items could have played a role (Kaufman & Kaufman, 2004). According to Fletcher-Janzen and Lichtenberger (2005), these teaching items contribute to the socio-cultural fairness of the KABC-II.

The Story Completion subtest in the KABC-II required fewer modifications, as indicated by the results of this South African sample, since none of the items were considered very difficult. In contrast, the Pattern Reasoning subtest, also part of the Planning Scale, required more modifications to its item order, with some items proving very difficult in this sample. Despite this, all items, except the first, were included in the analysis. This is interesting, given the findings by Piper et al. (2024), which indicate that both Planning Scale subtests were found to be very difficult and required adaptation. Similar to its counterparts on the WISC-V, the Digit Span and Letter-Number Sequencing subtests on the KABC-II required the least adaptations, with the subtest items maintaining most of their current difficulty order. As was the case for the three Digit Span components and Letter-Number Sequencing, the items of the Number Recall subtest also measure auditory memory span (Flanagan et al., 2018; Flanagan & Dixon, 2014; Kaufman et al., 2016d; Schneider et al., 2017). In contrast, the other Sequential Scale subtest, Word Order, required more adaptations than Number Recall (see Chapter 4). Thus, as also noted in their mean scores, the participants in this sample found the items of the Word Order subtest generally more difficult than those of the Number Recall subtest. This can be explained by the fact that Word Order relies more on working memory than Number Recall (Flanagan et al., 2018; Flanagan & Dixon, 2014; Kaufman et al., 2016d; Schneider et al., 2017).

According to the results of this study, several differences were observed in the order of the Simultaneous Scale subtests with the difficulty order of these subtests' items measuring visual and simultaneous processing differing from those in the current administration (Flanagan et al., 2018; Flanagan & Dixon, 2014; Kaufman et al., 2016d; Schneider et al., 2017). Finally, regarding the items of the Knowledge Scale subtests, modifications to the order of the items were suggested, and several items in these subtests were considered very difficult or excluded from the analyses, as none of the participants were able to answer them correctly. These results are likely indicative of the fact that the Knowledge Scale subtests measure crystallised reasoning, English language proficiency, and acquired knowledge, which can differ significantly from one sample to the next (Flanagan et al., 2018; Flanagan & Dixon, 2014; Kaufman et al., 2016d; Schneider et al., 2017).

5.6.2. Differential Item Functioning of WISC-V and KABC-II Subtests' Items

The results of these analyses indicated that the DIF observed in the WISC-V and KABC-II was consistently in favour of the EFL group, implying item bias in the EAL group. As far as the WISC-V is concerned, a total of eight items showed DIF in the Verbal Comprehension Index, although most of these were observed in the Vocabulary subtest. In addition, DIF was also observed in six of the Working Memory Index subtest items, with the majority of these instances occurring in the Picture Span subtest. In the Visual-Spatial Index, DIF was observed in four of the Block Design subtest's items, while in the Fluid Reasoning Index, DIF was observed in three of the Arithmetic subtest's items. As far as the KABC-II is concerned, DIF was observed in 11 of the Simultaneous Scale subtests' items, spread across the items of the Rover (six items), Triangles (three items) and Block Counting (two items) subtests. The Planning Scale also demonstrated DIF across its two subtests' items, which included two items from the Story Completion subtest and four items from the Pattern Reasoning subtest. Three of the Learning Scale subtests' items showed DIF, all of which formed part of the Rebus subtest. Finally, unlike the WISC-V, where DIF was observed in all of the indices, the KABC-II Sequential and Knowledge Scales subtests' items did not show any DIF. Thus, once again, the KABC-II was indicated to be the more culturally fair test as English language proficiency showed less impact on participants' performance.

As stated in Chapter 2, limited research has been conducted into DIF in the items of the WISC-V and the KABC-II. Although Fletcher-Janzen and Lichtenberger (2005) and Reynolds and Hadorn (2016) mention that item analysis was conducted with items from both the WISC-V and KABC-II during test development, the results of these analyses are not available. Immekus and Maller (2009) investigated DIF in the Kaufman Adolescent and Adult Intelligence Test (KAIT). However, investigations focused on item bias limited to sex. Nonetheless, the results of these analyses indicated discrimination and difficulty with DIF in several items assessing crystallised intelligence. Maller (2001) investigated DIF in six of the WISC-III subtests, examining sex differences and finding that among the 151 items included in the analyses, 52 showed DIF.

5.7. Strengths and Implications of the Research Study

With minimal published research findings available on the psychometric properties of the WISC-V and KABC-II in samples of South African children, this research study contributed to an understanding of the validity, reliability and cultural-applicability of these measures in the South African context, especially in a sample of a South African children with barriers to learning and development which impact the acquisition and development of their essential scholastic skills. Given that children with barriers to learning and development are more regularly referred for assessment compared to typically developing children, this was deemed important. Also, with only one previous South African study investigating the cross-cultural suitability of former editions of these intelligence tests through comparisons between different ethnic groups, the investigations conducted into the impact of English language proficiency on the EAL and EFL language groups, which provided evidence for the cross-cultural suitability of the KABC-II when compared to the WISC-V, was deemed a strength and contribution to an understanding of these two measures. Moreover, with very limited published research findings using item analysis to explore the discriminatory ability, difficulty, and differential item functioning of the KABC-II and WISC-V subtests' items, this study further contributed to an understanding of the cross-cultural suitability of these two measures, not only at a global and index level, but at an item level. Although the study focused primarily on the psychometric properties of these intelligence tests, the analysis of the results also guided South African practitioners on factors to consider when interpreting children's results on the WISC-V and KABC-II. For example, given the high correlation between measures of fluid reasoning and planning, and visual-spatial and simultaneous processing, practitioners should be cautious in considering the scores children obtain on the WISC-V Fluid Reasoning and Visual-Spatial Indices, and the KABC-II Simultaneous and Planning Scales, as salient indictors of either fluid reasoning and planning or visual-spatial and simultaneous processing. In addition, practitioners should be cautious in interpreting children's performance on the WISC-V Coding subtest as a measure of processing speed and consider the other cognitive abilities assessed by this subtest and their impact on performance, including the rate at which an individual can complete a test influenced by their psychomotor skills and fine motor dexterity. Another factor to consider is the weak correlation between the subtests of the KABC-II Learning Scale and the more significant contribution of the Rebus subtest to long-term storage and retrieval when compared to Atlantis. While participants obtained a higher mean score on the Atlantis subtest compared to any of the other KABC-II subtests, there is concern about the subtest's ability to accurately assess what it was designed to measure. Additionally, given the role that verbal and visual associations, as well as contextual clues, as measured by the Rebus subtest, played in the sample's retention and retrieval of information from long-term memory, these could be utilised in interventions to support storage and retrieval. The study further highlighted the importance of practitioners having a theoretical understanding of the intelligence tests they administer, as this enables a more in-depth understanding of a person's intellectual functioning and the cognitive abilities and processes that contribute to it. Should standardisation of either the WISC-V or KABC-II in the South African context be considered, it would be important to take cognisance of the findings of the item analysis, precisely the order in which the items of the subtests are presented based on level of difficulty and the evidence of item bias in certain items. In addition to its theoretical and practical implications, the study further underscored the need for the allocation of financial and human resources to research aimed at developing and standardising valid, reliable, and culturally appropriate South African intelligence tests. Such tests would benefit all South African children by identifying cognitive strengths and weaknesses, thereby informing appropriate interventions to support scholastic progress. Further, in line with research referenced in the introductory chapter, the results of this study - which demonstrated the impact of English language proficiency on children's performance across both verbal and nonverbal subtests and items - emphasise the importance of incorporating culturally relevant materials and items in the development of emic intelligence tests. Finally, the study also draws attention to the importance of teaching psychometrics at university level to ensure that future psychologists develop a clear understanding of the practical implications of validity, reliability, and bias when administering and interpreting psychological assessment measures.

5.8. Limitations of the Research Study

Although the study has strengths, it is also important to acknowledge its limitations. First, the small sample size limited the statistical analyses that could be conducted, impacting the researcher's ability to investigate the construct validity of these intelligence tests through further factor analyses, including Exploratory Factor Analysis. A larger sample size would have enabled the researcher to conduct Exploratory Factor Analysis and examine alternative factor structures. A more complex factor structure may have emerged. Second, a smaller sample size increases the risk of incorrect results; therefore, caution must be exercised when interpreting the results. Third, it must be acknowledged that items analysis typically requires larger sample sizes, and that the observed differences – particularly in item difficulty – may have been influenced by the relatively small sample size used in this study. Fourth, given that the study was conducted with a sample of children from private, remedial schools in the greater Johannesburg area, from well-resourced backgrounds, the generalisability of the results is limited. Related to this is the fact that the majority of participants were from one of the three participating schools, which further limits generalisability and may have impacted the test results. It must also be acknowledged that the barriers to learning and development experienced by the participants likely influenced their overall performance, and that the mean scores presented in the preceding chapters should be interpreted with this context in mind. Fifth, as explained in Chapter 3, multiple assessors were employed in administering the intelligence tests. This creates several potential limitations, including inconsistencies in data collection due to varying testing styles. Different assessors may interpret instructions differently or emphasise different aspects of the instructions. Additionally, varying levels of expertise may have impacted the quality and consistency of the data collected. Assessors may also, unintentionally, introduce their own biases or expectations during the administration of these intelligence tests. Related to this is how assessors interact with participants and how factors, such as tone of voice, body language, and facial expressions, can impact the comfort and engagement of participants. To mitigate these concerns, workshops were conducted with all assessors to ensure consistency in administering the intelligence tests and to share clear protocols for the research. Additionally, the researcher verified all scoring to ensure consistency. However, the researcher could not ensure that all subtests were administered to all participants; therefore, data from a limited number of participants were obtained for some of the WISC-V subtests. Sixth, given the original aim of the study as stated in Chapter 3, a test-retest method could not be employed in examining the reliability of the WISC-V Symbol Search, Coding, and Naming Speed subtests. Finally, as reported in Chapter 3, a comparative South African sample could not be obtained and included in the analyses of the participants' results. Thus, it is challenging to determine the extent to which the participants' barriers to learning and development affected their performance, particularly given their lower scores on several WISC-V and KABC-II indices, scales, and subtests. About this, the study viewed the participants as a homogeneous group with barriers to learning and development. It is, however, important to be cognisant that different challenges can diversely impact the results children obtain on intelligence tests.

5.9. Recommendations for Future Research

As stated in Chapter 1, it is essential to continue researching the psychometric properties of etic intelligence tests when administered to different subgroups of the South African population. Specifically, research studies that include larger and more diverse samples are recommended, allowing for further investigation into the impact of several socio-cultural factors on test performance, as well as samples of typically developing children from various linguistic and socio-economic backgrounds. In addition, further confirmatory and exploratory factor analyses are recommended with larger samples to investigate various and, where possible, more complex factor structures. Also, given the findings of the item analyses and the identification of DIF in the KABC-II and WISC-V subtests, it would be important to further investigate the discriminatory ability, difficulty and bias of the subtests' items, remove bias items, and make adjustment to item order to improve the cross-cultural suitability of these measures consistently. This is particularly important, given the limited research studies that have investigated the potential impact of English language proficiency on the items of the WISC-V and KABC-II. Additional studies with specific subgroups of children in the broader domain of barriers to learning and development would highlight the specific impact of particular challenges, for example, ADHD, on children's test performance. Finally, education policymakers and the Health Professions Council of South Africa (HPCSA) should collaborate with institutions of higher education, practitioners, and researchers to continually develop intelligence tests that are reliable, valid, and fair. This will also require funding initiatives to ensure the continuation of South African research.

5.10. Conclusion

This study set out to conduct a preliminary investigation into the psychometric properties of the WISC-V and the KABC-II in a sample of South African children with barriers to learning and development. In doing so, it hoped to contribute to the available research on the psychometric properties of etic intelligence tests when used with specific subgroups of the broader South African context and contribute to the calls by Mitchell et al. (2018) and Van der Merwe et al. (2022; 2024) for South African research studies that can further an understanding of the validity, reliability, and fairness of recently developed etic intelligence tests when used with South African children to determine appropriate school placement and suitable intervention programmes. Despite the study's limitations, it made a valuable contribution to the field, where research is scarce. Practically, it also contributed to an understanding of the socio-cultural suitability of these tests when used with South-African children. Nonetheless, given the use of intelligence tests in recognising barriers to learning and development in children, determining appropriate school placement, and recommending suitable interventions that can enhance educational outcomes, continued research into available etic intelligence tests is required.

References

Abdool Gafoor, L., Burke, A., & Fourie, J. (2021). The efficacy of the Senior South African Individual Scale Revised in distinguishing between attention deficit hyperactivity disorder, normal and sluggish cognitive tempo children. *African Journal of Psychological Assessment*, *3*, Article 45. https://doi.org/10.4102/ajopa.v3i0.45

Adelman, H. S, & Taylor, L. L. (2018). Addressing barriers to learning: In the classroom and schoolwide. University of California, Los Angeles (UCLA). Retrieved from https://escholarship.org/uc/item/55w7b8x8

Ainscow, M., & César, M. (2006). Inclusive education ten years after Salamanca: Setting the agenda. *European Journal of Psychology of Education, 21*(3), 231-238. https://www.jstor.org/stable/23421604

Akinboboye, J. T., Ayanwale, M. A., Adewuni, D. A., & Vincent, Y. I. (2024). Impact of unanswered questions on examinees' latent traits: An item response theory perspective. *African Journal of Psychological Assessment*, *6*, Article 161. https://doi.org/10.4102/ajopa.v6i0.161

Aldridge, V. K., Dovey, T. M., & Wade, A. (2017). Assessing test-retest reliability of psychological measures: Persistent methodological problems. *European Psychologist*, *22*(4), 207-218.

Alfonso, V. C., Flanagan, D. P., & Radwan, S. (2005). The impact of the Cattell-Horn-Carroll theory on test development and interpretation of cognitive and academic abilities. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (2nd ed., pp. 185-202). Guilford Press.

Alnaim, F. A. (2016). History of Learning Disabilities: Reflection on the development of the Concept and Assessment. *Global Journal of Human-Social Science: An Arts & Humanities–Psychology*, *16*(3), 33-36.

Al-Yagon, M., Cavendish, W., Cornoldi, C., Fawcett, A. J., Grünke, M., Hung, L.Y., Jiménez, J. E., Karande, S., Van Kraayenoord, C. E., Lucangeli, D., Margalit, M., Montague, M., Sholapurwala, R., Sideridis, G., Tressoldi, P. E., & Vio, C. (2013). The proposed changes for DSM-5 for SLD and ADHD: International perspectives—Australia, Germany, Greece, India, Israel, Italy, Spain, Taiwan, United Kingdom, and United States. *Journal of learning disabilities*, *46*(1), 58-72. https://doi.org/10.1177/0022219412464353

American Educational Research Association (AERA), American Psychological Association, and National Council on Measurement in Education. (2014). *Standards for Educational and Psychological Testing*. American Psychological Association.

American Psychiatric Association. (2022). *Diagnostic and statistical manual of mental disorders* (5th ed., text rev.). https://doi.org/10.1176/appi.books.9780890425787

Amod, Z. (2013). The Das-Naglieri Cognitive Assessment System. In S. Laher & K. Cockcroft (Eds.), *Psychological assessment in South Africa: Research and applications* (pp. 104-119). WITS University Press. https://doi.org/10.18772/22013015782

Amod, Z., Heafield, D., & Seabi, J. (2018). Assessing a remedial intervention programme in developing the planning skills of grade 4 and 5 learners. *International Journal of Disability, Development and Education*, *65*(4), 428-441. https://doi.org/10.1080/1034912X.2017.1406067

Amod, Z., & Seabi, J. (2013). Dynamic Assessment in South Africa. In S. Laher & K. Cockcroft (Eds.), *Psychological assessment in South Africa: Research and applications* (pp. 120-136). WITS University Press. https://doi.org/10.18772/22013015782

An, X., & Yung, Y-F. (2014). Item Response Theory: What It Is and How You Can Use the IRT Procedure to Apply It. *SAS Institute Inc*. Paper SAS364-2014.

Anunciacao, L. (2018). An overview of the history and methodological aspects of psychometrics-history and methodological aspects of psychometrics. *Journal for ReAttach Therapy and Developmental Diversities*, *1*(1), 44-58. https://doi.org/10.26407/2018jrtdd.1.6

Apuke, O. D. (2017). Quantitative research methods: A synopsis approach. *Kuwait Chapter of Arabian Journal of Business and Management Review, 33*(5471), 1-8. https://doi.org/10.12816/0040336

Ariffin, S. R., Asari, S. M., Mohamed, S., Shahar S. N., Ishak, N. M., Din, R., Ahmad, A. G., Yamat, H., Majid, R., Yasin, S. F. M. (2010). Validity of UKM1 intelligence test using Rasch analysis. *Procedia Social and Behavioral Sciences*, *7*(C), 205–209. https://doi.org/10.1016/j.sbspro.2010.10.029

Ary, D., Jacobs, L. C., & Sorensen Irvine, C. K. (2010). *Introduction to research in education* (8th ed.). Cengage Learning.

Ary, D., Jacobs, L. C., Sorenson Irvine, C. K., & Walker, D. A. (2019). *Introduction to research in education* (10th ed.). Cengage Learning.

Astle, D. E., & Fletcher-Watson, S. (2020). Beyond the core-deficit hypothesis in developmental disorders. *Current Directions in Psychological Science*, *29*(5), 431-437. https://doi.org/10.1177/0963721420925518

August, J. O. (2017). *A normative study of the Raven's Coloured Progressive Matrices for South African children with cognitive barriers to learning* [Master research report, Nelson Mandela Metropolitan University]. South East Academic Libraries System (SEALS). http://hdl.handle.net/10948/19419

Bangirana, P., Seggane-Musisi, P., Allebeck, P., Giordani, B., Chandy, C. J., Opoka, O. R., Byarugaba, J., Ehnvall, A., & Boivin, M. J. (2009). A preliminary examination of the construct validity of the KABC-II in Ugandan children with a history of cerebral malaria. *African Health Sciences*, *9*(3), 186 - 192.

Baumgartner, J., Smuts, C. M., Malan, L., Kvalsvig, J., Van Stuijvenberg, M. E., Hurrell, R. F., & Zimmermann, M. B. (2012). Effects of iron and n-3 fatty acid supplementation, alone and in combination, on cognition in school children: a randomized, double-blind, placebo-controlled intervention in South Africa. *The American Journal of Clinical Nutrition*, *96*(6), 1327-1338. https://doi.org/10.3945/ajcn.112.041004

Becker, A., Daseking, M., & Kerner auch Koerner, J. (2021). Cognitive profiles in the WISC-V of children with ADHD and specific learning disorders. *Sustainability*, *13*(17), Article 9948. https://doi.org/10.3390/su13179948

Best, J. W., & Kahn, J. V. (2006). *Research in education* (10th ed.). Pearson Education Inc.

Bickell, A. (2015). *WISC-IV test performance of Grade 3 Xhosa-speaking children: An extension of a prior South African normative database* [Master research report, University of Fort Hare]. South East Academic Libraries System (SEALS). http://hdl.handle.net/10353/5730

Boake, C. (2002). From the Binet–Simon to the Wechsler–Bellevue: Tracing the history of intelligence testing*. Journal of Clinical and Experimental Neuropsychology*, *24*(3), 383-405. https://doi.org/10.1076/jcen.24.3.383.981

Bodin, D., Pardini, D. A., Burns, T. G., & Stevens, A. B. (2009). Higher order factor structure of the WISC-IV in a clinical neuropsychological sample. *Child Neuropsychology*, *15*(5), 417-424. https://doi.org/10.1080/09297040802603661

Boivin, M. J., Bangirana, P., & Smith, R. C. (2010). The relationship between visual-spatial and auditory-verbal working memory span in Senegalese and Ugandan children. *PLoS One*, *5*(1), Article 8914. https://doi.org/10.1371/annotation/b9e2fb60-c9b6-4d00-9f09-69394cf2da10

Boivin, M. J., Giordani, B., Ndanga, K., Maky, M. M., Manzeki, K. M., Ngunu, N., & Muamba, K. (1993). Effects of treatment for intestinal parasites and malaria on the cognitive abilities of schoolchildren in Zaire, Africa. *Health Psychology*, *12*(3), 220-226.

Boivin, M. J., Okitundu, D., Makila-Mabe, B., Sombo, M. T., Mumba, D., Sikorskii, A., Mayambu, B., & Tshala-Katumbay, D. (2017). Cognitive and motor performance in Congolese children with konzo during 4 years of follow-up: a longitudinal analysis. *The Lancet Global Health*, *5*(9), e936-e947.

Bolhuis, J. J., Crain, S., & Roberts, I. (2023). Language and learning: The cognitive revolution at 60-odd. *Biological Reviews*, *98*(3), 931-941. https://doi.org/10.1111/brv.12936

Boone, W. J. (2016). Rasch analysis for instrument development: Why, when, and how? *CBE-Life Sciences Education, 15*(rm4), 1-7. https://doi.org/10.1187/cbe.16-04-0148

Booyse, J. J. (2011). The provision of education during the first half of the 20th century. In J. J. Booyse, C. S. le Roux, J. Seroto, & C. C. Wolhuter (Eds.), *A history of schooling in South Africa: Method and context* (pp. 169-214). Van Schaik Publishers.

Breaux, K. C., Avitia, M., Koriakin, T., Bray, M. A., DeBiase, E., Courville, T., Pan, X., Withold, T., & Grossman, S. (2017). Patterns of strengths and weaknesses on the WISC-V, DAS-II, and KABC-II and their relationship to students' errors in oral language, reading, writing, spelling, and math. *Journal of Psychoeducational Assessment*, *35*(1-2), 168-185. https://doi.org/10.1177/07342829166696

British Psychological Society [BPS]. (2013). *Classification of behaviour and experience in relation to functional psychiatric diagnosis: Time for a paradigm shift. DCP position statement*. British Psychological Society. https://www.bps.org.uk/sites/www.bps.org.uk/files/Member%20Networks/Divisions/DCP/Classification%20of%20behaviour%20and%20experience%20in%20relation%20to%20functional%20psychiatric%20diagnoses.pdf

Bronfenbrenner, U. (1979). *The ecology of human development: Experiments by nature and design*. Harvard University Press.

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

Brown, T. A. (2015). *Confirmatory factor analysis for applied research* (2nd ed.). Guilford Press.

Büttner, G., & Hasselhorn, M. (2011). Learning disabilities: Debates on definitions, causes, subtypes, and responses. *International Journal of Disability, Development and Education*, *58*(1), 75-87. https://doi.org/10.1080/1034912X.2011.548476

Byrne, B. M. (2005). Factor analytic models: Viewing the structure of an assessment instrument from three perspectives. *Journal of Personality Assessment*, *85*(1), 17-32. https://doi.org/10.1207/s15327752jpa8501_02

Byrne, M. (2013). *The Relationship between parental educational level and children's cognitive functioning* [Master research report, University of the Witwatersrand]. Wits Institutional Repository on DSpace (WIReDSpace). http://hdl.handle.net10539/13919

Caemmerer, J. M., Keith, T. Z., & Reynolds, M. R. (2020). Beyond individual intelligence tests: application of Cattell-Horn-Carroll theory. *Intelligence*, *79*, Article 101433. https://doi.org/10.1016/j.intell.2020.101433

Canivez, G. L., Dombrowski, S. C., & Watkins, M. W. (2018). Factor structure of the WISC-V in four standardization age groups: Exploratory and hierarchical factor analyses with the 16 primary and secondary subtests. *Psychology in the Schools*, *55*(7), 741-769. https://doi.org/10.1002/pits.22138

Canivez, G. L., McGill, R. J., Dombrowski, S. C., Watkins, M. W., Pritchard, A. E., & Jacobson, L. A. (2020). Construct validity of the WISC-V in clinical cases: Exploratory and confirmatory factor analyses of the 10 primary subtests. *Assessment*, *27*(2), 274-296. https://doi.org/10.1177/1073191118811609

Canivez, G.L., & Watkins, M.W. (2016). Review of the Wechsler Intelligence Scale for Children – Fifth Edition: Critique, commentary, and independent analysis. In A.S. Kaufman, S.E. Raiford, & D.L. Coalson (Eds.), *Intelligent testing with the WISC-V* (pp. 683-702). John Wiley & Sons, Inc.

Canivez, G. L., Watkins, M. W., & Dombrowski, S. C. (2016). Factor structure of the Wechsler Intelligence Scale for Children–Fifth Edition: Exploratory factor analyses with the 16 primary and secondary subtests. *Psychological Assessment*, *28*(8), 975–986. https://doi.org/10.1037/pas0000238

Canivez, G. L., Watkins, M. W., & Dombrowski, S. C. (2017). Structural validity of the Wechsler Intelligence Scale for Children–Fifth Edition: Confirmatory factor analyses with the 16 primary and secondary subtests. *Psychological Assessment*, *29*(4), 458–472. https://doi.org/10.1037/pas0000358

Canivez, G. L., Watkins, M. W., & McGill, R. J. (2019). Construct validity of the Wechsler Intelligence Scale for Children–Fifth UK Edition: Exploratory and confirmatory factor analyses of the 16 primary and secondary subtests. *British Journal of Educational Psychology*, *89*(2), 195-224. https://doi.org/10.1111/bjep.12230

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, *56*(2), 81–105. https://doi.org/10.1037/h0046016

Carvalho, F. (2023). *The effects of socioeconomic status and language difference on specific cognitive test performance: A comparative study of African American children* [Doctoral dissertation, Saint John’s University]. Saint John’s Scholar. https://scholar.stjohns.edu/theses_dissertations/711

Cassoojee, Z. (2020). *Comparative analysis of test performance of South African learners on indices/scales of the Wechsler Intelligence Scale for Children, Fifth Edition (WISC-V) and the Kaufman Assessment Battery for Children, Second Edition (KABC-II)* [Master research report, University of the Witwatersrand]. Wits Institutional Repository on DSpace (WIReDSpace). https://hdl.handle.net/10539/31388

Chen, F. F. (2007). Sensitivity of goodness of fit indices to lack of measurement invariance. *Structural Equation Modelling: A Multidisciplinary Journal*, *14*(3), 464-504. https://doi.org/10.1080/10705510701301834

Chen, H., Zhang, O., Raiford, S. E., Zhu, J., & Weiss, L. G. (2015). Factor invariance between genders on the Wechsler Intelligence Scale for Children–Fifth Edition. *Personality and Individual Differences*, *86*, 1-5. http://dx.doi.org/10.1016/j.paid.2015.05.020

Chen, H., & Zhu, J. (2012). Measurement invariance of WISC-IV across normative and clinical samples. *Personality and Individual Differences*, *52*(2), 161-166. https://doi.org/10.1016/j.paid.2011.10.006

Chernoff, M. C., Laughton, B., Ratswana, M., Familiar, I., Fairlie, L., Vhembo, T., Kamthunzi, P., Kabugho, E., Joyce, C., Zimmer, B., Ariansen, J.L., Jean-Philippe, P., & Boivin, M. J. (2018). Validity of neuropsychological testing in young African children affected by HIV. *Journal of Pediatric Infectious Diseases*, *13*(3), 185-201. https://doi.org/10.1055/s-0038-1637020

Children and Families Act. (2014). Children and Families Act 2014 (c. 6). https://www.legislation.gov.uk/ukpga/2014/6/part/3/enacted

Claassen, N. W. (1997). Cultural differences, politics and test bias in South Africa. *European Review of Applied Psychology*, *47*(4), 297-307.

Clark, K. (2007). *The social worker as facilitator in inclusive education* [Master thesis, University of Stellenbosch]. SUNScholar. http://hdl.handle.net/10019.1/1789

Clark-Carter, D. (2004). *Quantitative psychological research: A student's handbook*. Psychology Press.

Cockcroft, K. (2013). The Senior South African Individual Scales – Revised: A review. In S. Laher & K. Cockcroft (Eds.), *Psychological assessment in South Africa: Research and applications* (pp. 48-59). WITS University Press. https://doi.org/10.18772/22013015782

Cockcroft, K., Alloway, T., Copello, E., & Milligan, R. (2015). A cross-cultural comparison between South African and British students on the Wechsler Adult Intelligence Scales Third Edition (WAIS-III). *Frontiers in Psychology*, *6*, Article 297. https://doi.org/10.3389/fpsyg.2015.00297

Cockcroft, K., Amod, Z., & Soellaart, B. (2008). Level of maternal education and performance of Black, South African infants on the 1996 Griffiths Mental Development Scales. *African Journal of Psychiatry*, *11*, 44-50. https://doi.org/ 10.4314/ajpsy.v11i1.30254

Cody, R., & Smith, J. K. (2014). *Test scoring and analysis using SAS*. SAS Institute Inc.

Colliver, J. A., Conlee, M. J., & Verhulst, S. J. (2012). From test validity to construct validity and back? *Medical Education*, *46*(4), 366-371. https://doi.org/10.1111/j.1365-2923.2011.04194.x

Conant, L. L., Fastenau, P. S., Giordani, B., Boivin, M. J., Opel, B., & Dia Nseyila, D. (1999). Modality Specificity of Memory Span Tasks among Zaïrian Children: A Developmental Perspective. *Journal of Clinical and Experimental Neuropsychology*, *21*(3), 375–384. https://doi.org/10.1076/jcen.21.3.375.912

Cooper, S. (2014). A synopsis of South African psychology from apartheid to democracy. *American Psychologist*, *69*(8), 837-847. https://doi.org/10.1037/a0037569

Cormier, D. C., McGrew, K. S., Bulut, O., & Funamoto, A. (2017). Revisiting the relations between the WJ-IV measures of Cattell-Horn-Carroll (CHC) cognitive abilities and reading achievement during the school-age years. *Journal of Psychoeducational Assessment*, *35*(8), 731-754. https://doi.org/10.1177/07342829166592

Creswell, J. W. (2012). *Educational research: Planning, conducting and evaluating quantitative and qualitative research* (4th ed.). Pearson Education, Inc.

Cresswell, J. W., & Cresswell, J. D. (2020). *Research design: Qualitative, quantitative, and mixed-methods approaches* (5th ed.). SAGE Publications Inc.

Cresswell, J. W., & Guetterman, T. C. (2019). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research* (6th ed.). Pearson Education, Inc.

Croker, S. (2012). *The development of cognition*. Cengage Learning EMEA.

Dale, B. A., McIntosh, D. E., Rothlisberg, B. A., Ward, K. E., & Bradley, M. H. (2011). Profile analysis of the Kaufman Assessment Battery for Children, with African American and Caucasian preschool children. *Psychology in the Schools*, *48*(5), 476-487. https://doi.org/10.1002/pits.20571

Das, J. P. (1973). Structure of cognitive abilities: Evidence for simultaneous and successive processing. *Journal of Educational Psychology*, *65*(1), 103-108. https://doi.org/10.1037/h0034744

Das, J. P. (2002). A better look at intelligence. *Current Directions in Psychological Science*, *11*(1), 28-33. https://doi.org/10.1111/1467-8721.00162

Das, J. P., Kirby, J., & Jarman, R. F. (1975). Simultaneous and successive synthesis: An alternative model for cognitive abilities. *Psychological Bulletin*, *82*(1), 87–103. https://doi.org/10.1037/h0076163

Das, J. P., Naglieri, J. A., & Kirby, J. R. (1994). *Assessment of Cognitive Processes: The PASS Theory of Intelligence*. Allyn and Bacon.

De Beer, M. (2004). Use of differential item functioning (DIF) analysis for bias analysis in test construction. *South African Journal of Industrial Psychology*, *30*(4), 52-58. https://hdl.handle.net/10520/EJC89024

De Beer, M. (2013). The Learning Potential Computerised Adaptive Test in South Africa. In S. Laher & K. Cockcroft (Eds.), *Psychological assessment in South Africa: Research and applications* (pp. 137-157). WITS University Press. https://doi.org/10.18772/22013015782

De Kock, F., & Foxcroft, C. (2018). Cross-cultural test adaptation, translation and tests in multiple languages. In C. Foxcroft and G. Roodt (Eds.), *Introduction to psychological assessment in the South African context* (5th ed., pp. 100-122). Oxford University Press Southern Africa (Pty) Ltd.

Department of Basic Education. (2010). *Guidelines for Full-service/Inclusive Schools*. Department of Basic Education, Republic of South Africa.

Department of Basic Education. (2011a*). Curriculum and Assessment Policy Statement, Grades R-3, English home language*. Department of Basic Education, Republic of South Africa.

Department of Basic Education. (2011b). *Curriculum and Assessment Policy Statement, Grades 1-3, Mathematics*. Department of Basic Education, Republic of South Africa.

Department of Basic Education. (2014). *The policy on screening, identification, assessment and support*. Department of Basic Education, Republic of South Africa.

Department of Education. (1997). *Quality education for all: Overcoming barriers to learning and development. Report of the National Commission on Special Needs in Education and Training (NCSNET) and the National Committee on Education Support Services (NCESS)*. Department of Education, Republic of South Africa.

Department of Education. (2001). *Education White Paper 6: Special needs education: Building an inclusive education and training system*. Department of Education, Republic of South Africa.

Department of Education. (2005). *Conceptual and operational guidelines for the implementation of inclusive education: District-Based Support Teams*. Department of Education, Republic of South Africa.

Department for Education and Department of Health and Social Care. (2015). *SEND code of practice: 0 to 25 years. Guidance on the special educational needs and disability (SEND) system for children and young people aged 0 to 25*. United Kingdom Government Publications.

De Sousa, D. S., Greenop, K., & Fry, J. (2010). Simultaneous and sequential cognitive processing in monolingual and bilingual children in South Africa. *South African Journal of Psychology*, *40*(2), 165-173. https://hdl.handle.net/10520/EJC98582

Devlin, A.S. (2006). *Research method: Planning, conducting and presenting research.* Thomson Wadsworth.

De Wet, C., & Wolhuter, C. (2009). A transitiological study of some South African educational issues. *South African Journal of Education*, *29*(3), 359-376. https://doi.org/10.15700/saje.v29n3a272

D'Intino, J. S. (2017). Learning disabilities in Canada: Definitions and accommodations. *Canadian Psyhcology / Psychologie Canadienne*, *58*(3), 228-237. https://doi.org/10.1037/cap0000116

Dombrowski, S. C., Canivez, G. L., Watkins, M. W., & Beaujean, A. A. (2015). Exploratory bifactor analysis of the Wechsler Intelligence Scale for Children—Fifth Edition with the 16 primary and secondary subtests. *Intelligence*, *53*, 194-201. https://doi.org/10.1016/j.intell.2015.10.009

Donald, D., Lazarus, S., Moolla, N. (2014). *Educational psychology in social context: Ecosystemic applications on southern Africa* (5th ed.). Oxford University Press Southern Africa (Pty) Limited.

Donohue, D., & Bornman, J. (2014). The challenges of realising inclusive education in South Africa. *South African Journal of Education*, *34*(2), Article 806. https:/doi.org/10.15700/201412071114

Dreyer, L. M. (2008). *An evaluation of a learning support model in primary schools in the West Coast/Winelands area* [Doctoral thesis, University of Stellenbosch]. SUNScholar. http://hdl.handle.net/10019.1/1448

Drost, E. A. (2011). Validity and reliability in social science research. *Education Research and Perspectives, 38*(1), 105-123.

Drozdick, L. W., Kaufman Singer, J., Lichtenberger, E. O., Kaufman, J. C., Kaufman, A. S., & Kaufman, N. L. (2018). The Kaufman Assessment Battery for Children – Second Edition and the KABC-II normative update. In D. P. Flanagan & E. M. McDonough (Eds.), *Contemporary intellectual assessment* (4th ed., pp. 333-359). Guilford Publications.

Dudley-Marling, C. (2004). The social construction of learning disabilities. *Journal of Learning Disabilities*, *37*(6), 482-489. https://doi.org/10.1177/00222194040370060201

Du Plessis, P. (2013). Legislation and policies: Progress towards the right to inclusive education. *De Jure*, *46*(1), pp. 76-92. ISSN 2225-7160

Du Plessis, A-B., & Dednam, A. (2019). Learning impairment and attention-deficit/hyperactivity disorder (ADHD). In E. Landsberg (Ed.), Krüger, D., & E. Swart (Asst. Eds.), *Addressing barriers to learning: A South African perspective* (4th ed., pp. 451-469). Van Schaik Publishers.

Egeland, J., Andreassen, T. & Lund, O. (2022). Factor structure of the new Scandinavian WISC-V version: Support for a five-factor model. *Scandinavian Journal of Psychology*, *63*, 1–7. https://doi.org/10.1111/sjop.12780

Embretson, S. E. (2004). The second century of ability testing: Some predictions and speculations. *Interdisciplinary Research and Perspectives*, *2*(1), 1-32. https://doi.org/10.1207/s15366359mea0201_1

Engelbrecht, P. (2006). The implementation of inclusive education in South Africa after ten years of democracy. *European Journal of Psychology of Education*, *21*, 253-264. https://doi.org/10.1007/BF03173414

Eyisi, D. (2016). The usefulness of qualitative and quantitative approaches and methods in researching problem-solving ability in science education curriculum. *Journal of Education and Practice*, *7*(15), 91-100. ISSN 2222-288X

Evans, J. J., Floyd, R. G., McGrew, K. S., & Leforgee, M. H. (2002). The relations between measures of Cattell-Horn-Carroll (CHC) cognitive abilities and reading achievement during childhood and adolescence. *School Psychology Review*, *31*(2), 246-262. https://doi.org/10.1080/02796015.2002.12086154

Fagan, J. F., & Holland, C. R. (2007). Racial equality in intelligence: Predictions from a theory of intelligence as processing. *Intelligence*, *35*, 319-334. https://doi.org/10.1016/j.intell.2006.08.009

Fenollar-Cortés, J., & Watkins, M. W. (2019). Construct validity of the Spanish version of the Wechsler intelligence scale for children fifth edition (WISC-V[Spain]). *International Journal of School & Educational Psychology*, *7*(3), 150-164. https://doi.org/10.1080/21683603.2017.1414006

Ferguson, D. L. (2008) International trends in inclusive education: The continuing challenge to teach each one and everyone. *European Journal of Special Needs Education*, *23*(2), pp. 109-120. https://doi.org/10.1080/08856250801946236

Ferrett, H. L. (2011). *The adaptation and norming of selected psychometric tests for 12- to 15-year-old urbanized Western Cape adolescents* [Doctoral thesis, University of Stellenbosch]. SUNScholar. http://hdl.handle.net/10019.1/17799

Fiske, E., & Ladd, H. (2006). Racial equity in education: How far has South Africa come? *Perspectives in Education*, *24*(2), 95-108. https://hdl.handle.net/10520/EJC87372

Flack, P. (2005). Towards new understandings of learning disability. *Africa Education Review*, *2*(2), 318-328. https://doi.org/10.1080/18146620508566308

Flanagan, D. P., & Alfonso, V. C. (2017). How to interpret the WISC-V. In D. P. Flanagan & V. C. Alfonso (Eds.), *Essentials of WISC-V assessment* (pp. 167-178). John Wiley and Sons, Inc.

Flanagan, D. P., Alfonso, V. C., Costa, M., Palma, K., & Leahy, M. A. (2018). Use of ability tests in the identification of specific learning disabilities within the context of an operational definition. In In D. P. Flanagan & E. M. McDonough (Eds.), *Contemporary intellectual assessment: theories, tests, and issues* (4th ed., pp. 608-642). Guilford Publications.

Flanagan, D. P., & Dixon, S. G. (2014). The Cattell-Horn-Carroll Theory of Cognitive Abilities. *Encyclopaedia of Special Education: A reference for the education of children, adolescents, and adults with disabilities and other exceptional individuals.* Wiley Online Library: https://doi.org/10.1002/9781118660584.ese0431

Flanagan, D. P., & Kaufman, A. S. (2004). *Essentials of WISC-IV assessment*. John Wiley & Sons, Inc.

Fletcher-Janzen, E. (n.d.). *KABC-II validity study: A validity study of the Kaufman Assessment Battery for Children, Second Edition (KABC-II) and the Taos Pueblo Indian children of New Mexico.* Pearson Assessments US. https://www.pearsonassessments.com/campaign/kabc-ii-validity-study.html

Fletcher-Janzen, E., & Lichtenberger, E. O. (2005). Strengths and Weaknesses of the KABC-II. In A. S. Kaufman, E. O. Lichtenberger, E. Fletcher-Janzen, & N. L. Kaufman, N.L. (Eds.). *Essentials of KABC-II Assessment* (pp. 168-175). John Wiley & Sons, Inc.

Floyd, R. G., Evans, J. J., & McGrew, K. S. (2003). Relations between measures of Cattell-Horn-Carroll (CHC) cognitive abilities and mathematics achievement across the school-age years. *Psychology in the Schools*, *40*(2), 155-171. https://doi.org/10.1002/pits.10083

Forlin, C., & Chambers, D. (2023). Is a whole school approach to inclusion really meeting the needs of all learners? Home-schooling parents' perceptions. *Education Sciences*, *13*(6), Article 571. https://doi.org/10.3390/educsci13060571

Foxcroft, C. D. (1997). Psychological testing in South Africa: Perspectives regarding ethical and fair practices. *European Journal of Psychological Assessment*, *13*(3), 229-235. https://doi.org/10.1027/1015-5759.13.3.229

Foxcroft, C., Paterson, H., Le Roux, N., & Herbst, D. (2004). *Psychological assessment in South Africa: A needs analysis* [The Test Use Patterns and Needs of Psychological Assessment Practitioners: Final Report]. Human Sciences Research Council. http://hdl.handle.net/20.500.11910/7498

Foxcroft, C., Roodt, G., & Abrahams, F. (2018). Psychological assessment: A brief retrospective overview. In C. Foxcroft & G. Roodt (Eds.), *Introduction to psychological assessment in the South African context* (5th ed., pp. 12-36). Oxford University Press Southern Africa (Pty) Limited.

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

Franck, K. (2022). The educational context in expert assessments. A study of special education documents of children in ECEC institutions. *European Journal of Special Needs Education*, *37*(5), 819-833. https://doi.org/10.1080/08856257.2021.1954346

Frischkorn, G. T., & Schubert, A. L. (2018). Cognitive models in intelligence research: Advantages and recommendations for their application. *Journal of Intelligence*, *6*(3), Article 34. https://doi.org/10.3390/jintelligence6030034

Frischkorn, G. T., Wilhelm, O., & Oberauer, K. (2022). Process-oriented intelligence research: A review from the cognitive perspective. *Intelligence*, *94*, Article 101681. https://doi.org/10.1016/j.intell.2022.101681

Furr, R. M. (2021). *Psychometrics: An introduction*. SAGE publications.

Galindo, G., Machinskaya, R., Basilio, C., & Solovieva, S. (2013). Specifics of Visuo-spatial Processing in Urban and Rural Primary School Children. *International Journal of Advances in Psychology*, *2*(4), 186-196. https://doi.org/10.14355/ijap.2013.0204.02

Gates, B., & Mafuba, K. (2016). Use of the term 'learning disabilities' in the United Kingdom: Issues for international researchers and practitioners. *Learning Disabilities: A Contemporary Journal*, *14*(1), 9-23.

George, D., & Mallery, P. (2003). *SPSS for Windows step by step: A simple guide and reference. 11.0 update* (4th ed.). Allyn & Bacon Publishers.

Gichovi, F. M., Rono, R. C., & & Basnight-Brown, D. M. (2018). Adoption and efficacy of Wechsler Intelligence Scale for Children, Fifth Edition (WISC-V) among learners in selected private primary schools in Embu West Sub County, Kenya. *African Journal of Clinical Psychology*, *1*(2). https://doi.org/10.1891/1945-8959.16.3.304

Gilmore, L., & Campbell, M. (2009). Competence in intelligence testing: A training model for postgraduate psychology students. *The Australian Educational and Developmental Psychologist*, *26*(2), 165-173. https://doi.org/10.1375/aedp.26.2.165

Glas, C. A. W. (2008). Item response theory in educational assessment and evaluation. *Mesure et évaluation en éducation*, *31*(2), 19-34. https://doi.org/10.7202/1025005ar

Goldstein, S. (2013). The science of intelligence testing: Commentary on the evolving nature of interpretations of the Wechsler scales. *Journal of Psychoeducational Assessment*, *31*(2), 132-137. https://doi.org/1 0.1177/0734282913478033

Gomez, M. T. (2008). *The performance evaluation of low achieving Mexican American students on the Kaufman Assessment Battery for Children II (KABC-II): The role of English language proficiency* (Publication No. 3339017) [Doctoral dissertation, Capella University]. ProQuest Dissertations and Theses Global.

Graves, S. L., Smith, L. V., & Nichols, K. D. (2021). Is the WISC-V a fair test for black children: Factor structure in an urban public-school sample. *Contemporary School Psychology*, *25*, 157-169. https://doi.org/10.1007/s40688-020-00306-9

Gravetter, F. J., & Forzano, L. B. (2009). *Research methods for the behavioral sciences* (3rd ed.). Cengage Learning.

Gravetter, F. J., & Forzano, L. B. (2018). *Research methods for the behavioral sciences* (6th ed.). Cengage Learning.

Green, L., & Engelbrecht, P. (2007). An introduction to inclusive education. In P. Engelbrecht & L. Green (Eds.), *Responding to the challenges of inclusive education in southern Africa* (pp. 2 – 9). Van Schaik Publishers.

Greenfield, P. M. (2020). Historical evolution of intelligence. In R. Sternberg (Ed.), *The Cambridge handbook of intelligence* (2nd ed., pp. 916-939). Cambridge University Press.

Greenop, K., Rice, J., & De Sousa, D. (2013). The Kaufman Assessment Battery in South Africa. In S. Laher & K. Cockcroft (Eds.), *Psychological assessment in South Africa: Research and applications* (pp. 86-103). WITS University Press. https://doi.org/10.18772/22013015782

Grégoire, J. (2013). Measuring components of intelligence: Mission impossible? *Journal of Psychoeducational Assessment*, *31*(2), 138-147. https://doi.org/10.1177/0734282913478034

Grigorenko, E.L., Wenzel Geissler, P., Prince, R., Okatcha, F., Nokes, C., Kenny, D.A., Bundy, D.A., & Sternberg, R. J. (2001). The organisation of Luo conceptions of intelligence: A study of implicit theories in a Kenyan village. *International Journal of Behavioral Development*, *25*(4), 367-378. https://doi.org/10.1080/01650250042000348

Grünke, M., & Cavendish, W. M. (2016). Learning disabilities around the globe: Making sense of the heterogeneity of the different viewpoints. *Learning Disabilities: A Contemporary Journal*, *14*(1), 1-8.

Hall, K. (2024, August). *Child poverty*. Children's Institute; University of Cape Town. https://childrencount.uct.ac.za/indicator.php?domain=2&indicator=98

Hannant, P. (2024). Holistically bridging the gap between education and healthcare: A case study model of assessment. *British Journal of Special Education*, *51*(2), 211-220. https://doi.org/10.1111/1467-8578.12512

Holden, L. R., & Tanenbaum, G. J. (2023). Modern assessments of intelligence must be fair and equitable. *Journal of Intelligence*, *11*(6), Article 126. https://doi.org/10.3390/jintelligence11060126

Holding, P. A., Taylor, H. G., Kazungu, S. D., Mkala, T., Gona, J., Mwamuye, B., Mbonani, L., & Stevenson, J. (2004). Assessing cognitive outcomes in a rural African population: development of a neuropsychological battery in Kilifi District, Kenya. *Journal of the International Neuropsychological Society*, *10*(2), 246-260. https://doi.org/ 10.1017/S1355617704102166

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indices in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1-55. https://doi.org/10.1080/10705519909540118

Individuals with Disabilities Education Improvement Act [IDEIA]. (2004). Public Law No. 108-446, Stat. 446. https://www.congress.gov/108/plaws/publ446/PLAW-108publ446.pdf

Immekus, J. C., & Maller, S. J. (2009). Item parameter invariance of the Kaufman Adolescent and Adult Intelligence Test across male and female samples. *Educational and Psychological Measurement*, *69*(6), 994-1012. https://doi.org/10.1177/0013164409344489

International Test Commission. (2017). *The ITC Guidelines for Translating and Adapting Tests* (2nd ed.). International Test Commission. https://www.intestcom.org/files/guideline_test_adaptation_2ed.pdf

Jansen, P., & Greenop, K. (2008). Factor analyses of the Kaufman Assessment Battery for Children assessed longitudinally at 5 and 10 years. *South African Journal of Psychology*, *38*(2), 355-365. https://hdl.handle.net/10520/EJC98487

Johnson, B. (2001). Toward a new classification of nonexperimental quantitative research. *Educational Researcher*, *30*(2), 3-13. https://doi.org/10.3102/0013189X03000200

Kaufman, A. S., & Kaufman, N. L. (2004). *KABC-II: Kaufman Assessment Battery for Children, Second Edition Manual*. NCS Pearson, Inc.

Kaufman, A. S., Lichtenberger, E. O., Fletcher-Janzen, E., & Kaufman, N. L. (2005a). Clinical applications. In A. S. Kaufman, E. O. Lichtenberger, E. Fletcher-Janzen, & N. L. Kaufman (Eds.). *Essentials of KABC-II Assessment* (pp. 176-281). John Wiley & Sons, Inc.

Kaufman, A. S., Lichtenberger, E. O., Fletcher-Janzen, E., & Kaufman, N. L. (2005b). Overview. In A. S. Kaufman, E. O. Lichtenberger, E. Fletcher-Janzen, & N. L. Kaufman (Eds.). *Essentials of KABC-II Assessment* (pp. 1-32). John Wiley & Sons, Inc.

Kaufman, A. S., Lichtenberger, E. O., Fletcher-Janzen, E., & Kaufman, N. L. (2005c). How to interpret the KABC-II: Step by step. In A. S. Kaufman, E. O. Lichtenberger, E. Fletcher-Janzen, & N. L. Kaufman (Eds.). *Essentials of KABC-II Assessment* (pp. 79-137). John Wiley & Sons, Inc.

Kaufman, A. S., Raiford, S. E., & Coalson, D. L. (2016a). Intelligent testing. In A. S. Kaufman, S. E. Raiford, & D. L. Coalson (Eds.), *Intelligent testing with the WISC-V* (pp. 3-34). John Wiley & Sons, Inc.

Kaufman, A. S., Raiford, S. E., & Coalson, D. L. (2016b). Basic steps for WISC-V interpretation. In A. S. Kaufman, S. E. Raiford, & D. L. Coalson (Eds.), *Intelligent testing with the WISC-V* (pp. 225-247). John Wiley & Sons, Inc.

Kaufman, A. S., Raiford, S. E., & Coalson, D. L. (2016c). Intelligent WISC-V administration: Test kit version. In A. S. Kaufman, S. E. Raiford, & D. L. Coalson (Eds.), *Intelligent testing with the WISC-V* (pp. 35-89). John Wiley & Sons, Inc.

Kaufman, A. S., Raiford, S. E., Coalson, D. L. (2016d). Interpreting the WISC-V from the Perspective of Cattell-Horn-Carrol Theory. In A. S. Kaufman, S. E. Raiford, & D. L. Coalson (Eds.), *Intelligent testing with the WISC-V* (pp. 250-328). John Wiley & Sons, Inc.

Kaufman, A. S., Raiford, S. E., Coalson, D. L. (2016e). WISC-V sex, ethnic, and socio-economic status (SES) differences. In A. S. Kaufman, S. E. Raiford, & D. L. Coalson (Eds.), *Intelligent testing with the WISC-V* (pp. 157-174). John Wiley & Sons, Inc.

Kaya, H. O., & Seleti, Y. N. (2013). African indigenous knowledge systems and relevance of higher education in South Africa. *International Education Journal: Comparative Perspectives*, *12*(1), 30-44.

Keith, T. Z., Fine, J. G., Taub, G. E., Reynolds, M. R., & Kranzler, J. H. (2006). Higher order, multisample, confirmatory factor analysis of the Wechsler Intelligence Scale for Children—Fourth Edition: What does it measure. *School Psychology Review*, *35*(1), 108-127.

Keith, T. Z., & Reynolds, M. R. (2010). Cattell-Horn-Carroll abilities and cognitive tests: What we've learned from 20 years of research. *Psychology in the Schools*, *47*(7), 635-650. https://doi.org/10.1002/pits.20496

Kern, A. (2022). Using a combined bio-ecological and capability perspective to understand inclusive education in South Africa. *Prospects*, *52*(3), 469-485. https://doi.org/10.1007/s11125-021-09584-4

Kern, A., Amod, Z., Seabi, J., & Vorster, A. (2015). South African foundation phase teachers' perceptions of ADHD at private and public schools. *International Journal of Environmental Research and Public Health*, *12*(3), 3042-3059. https://doi.org/10.3390/ijerph120303042

Kinderman, P., Read, J., Moncrieff, J., & Bentall, R.P. (2013). Drop the language of disorder. *Evidence Based Mental Health*, *16*(1), 2–3. http://dx.doi.org/10.1136/eb-2012-100987

Kirby, J. R., & Das, J. P. (1978). Information processing and human abilities. *Journal of Educational Psychology*, *70*(1), 58–66. https://doi.org/10.1037/0022-0663.70.1.58

Knoetze, J., Bass, N., & Steele, G. (2005). The Raven's Coloured Progressive Matrices: Pilot norms for isiXhosa-speaking primary school learners in peri-urban Eastern Cape. *South African Journal of Psychology*, *35*(2), 175-194. https://hdl.handle.net/10520/EJC98326

Kovacs, K., & Pléh, C. (2023). William Stern: The relevance of his program of 'Differential Psychology' for contemporary intelligence measurement and research. *Journal of Intelligence*, *11*(3), Article 41. https://doi.org/10.3390/jintelligence11030041

Krabbe, P. F. M. (2017). *The measurement of health and health status: Concepts, methods and applications from a multidisciplinary perspective*. Elsevier Inc. https://doi.org/10.1016/C2013-0-19200-8

Krauss, S. E. (2005). Research paradigms and meaning making: A primer. *The Qualitative Report*, *10*(4), 758-770.

Kumar, P. A. (2017). *A changing role? Educational psychologists in South Africa: Two decades beyond democracy* [Master research report, University of the Witwatersrand]. Wits Institutional Repository on DSpace (WIReDSpace). https://hdl.handle.net/10539/24446

Kuthria, R. & Serpell, R. (1998). Standardization of the Panga Munthu Test – A non-verbal cognitive test developed in Zambia. *The Journal of Negro Education*, *67*(3), 228-241. https://doi.org/10.2307/2668192

Laher, S. (2024). Advancing an agenda for psychological assessment in South Africa. *South African Journal of Psychology*, *54*(4), 515-530. https://doi.org/10.1177/008124632412685

Laher, S., & Cockcroft, K. (2013). Contextualising psychological assessment in South Africa. In S. Laher & K. Cockcroft (Eds.), *Psychological assessment in South Africa: Research and applications* (pp. 1-14). WITS University Press. https://doi.org/10.18772/22013015782

Laher, S., & Cockcroft, K. (2014). Psychological assessment in post-apartheid South Africa: The way forward. *South African Journal of Psychology*, *44*(3), 303-314. https://doi.org/10.1177/008124631453363

Laher, S., & Cockcroft, K. (2017). Moving from culturally biased to culturally responsive assessment practices in low-resource, multicultural settings. *Professional Psychology: Research and Practice*, *48(2)*, 115-121. https://doi.org/10.1037/pro0000102

Lai, J. S., Teresi, J., & Gershon, R. (2005). Procedures for the analysis of differential item functioning (DIF) for small sample sizes. *Evaluation & the Health Professions*, *28*(3), 283-294. http://dx.doi.org/10.1177/0163278705278276

Learning Difficulties Australia. (n.d.). *Learning difficulties*. https://www.ldaustralia.org/disabilities-and-dyslexia.html

Learning Disabilities Association of America. (2019, December 14). *Core principles: Best practices in the use of cognitive assessment in learning disability identification*. https://ldaamerica.org/info/best-practices-cognitive-assessment-ld/

Learning Disabilities Association of Canada. (2015). *Official definition of learning disabilities*. https://www.ldac-acta.ca/official-definition-of-learning-disabilities/

Learning Disabilities Association of New Zealand. (n.d.). *Specific learning disabilities*. https://ldanz.org.nz/specific-learning-disabilities/

Lecerf, T., & Canivez, G. L. (2018). Complementary exploratory and confirmatory factor analyses of the French WISC–V: Analyses based on the standardization sample. *Psychological Assessment*, *30*(6), 793-808. https://doi.org/10.1037/pas0000526

Lecerf, T., & Canivez, G. L. (2022). Exploratory Factor Analyses of the French WISC-V (WISC-VFR) for Five Age Groups: Analyses Based on the Standardization Sample. *Assessment*, *29*(6), 1117-1133. https://doi.org/10.1177/10731911211005170

Le Roux, C. S. (2011a). European foundations shaping schooling in South Africa: Early Dutch and British colonial influence at the Cape. In J. J. Booyse, C. S. le Roux, J. Seroto, & C. C. Wolhuter (Eds.), *A history of schooling in South Africa: Method and context* (pp. 57-86). Van Schaik Publishers.

Le Roux, C. S. (2011b). The influences of western education on education in the Cape and Natal: 1807-1899. In J. J. Booyse, C. S. le Roux, J. Seroto, & C. C. Wolhuter (Eds.), *A history of schooling in South Africa: Method and context* (pp. 87-126). Van Schaik Publishers.

Lewis, A., & Steyn, J. C. (2003). A critique of mission education in South Africa according to Bosch's mission paradigm theory. *South African Journal of Education, 23*(2), 101-106.

Lichtenberger, E. O., Volker, M. A., Kaufman, A. S., & Kaufman, N. L. (2006). Assessing gifted children with the Kaufman assessment battery for children—second edition (KABC-II). *Gifted Education International*, *21*(2-3), 99-126. https://doi.org/10.1177/026142940602100030

Liversage, L. (2013). Working memory and multiple language proficiency. [Master research report, University of the Witwatersrand]. Wits Institutional Repository on DSpace (WIReDSpace).http://hdl.handle.net/10539/18367

Lloyd, J. W., Keller, C., & Hung, L. Y. (2007). International understanding of learning disabilities. *Learning Disabilities Research & Practice*, *22*(3), 159–160. https://doi.org/10.1111/j.1540-5826.2007.00240.x

Lu, Y., & Treiman, D. J. (2011). Migration, remittances and educational stratification among blacks in apartheid and post-apartheid South Africa. *Social Forces*, *89*(4), 1119-1143. https://doi.org/10.1093/sf/89.4.1119

MacCallum, R. C. (2012). Factor analysis. In R. E. Millsap & A. Maydeu-Olivares (Eds.), *The SAGE handbook of quantitative methods in psychology* (pp. 123-147). SAGE Publications Ltd. https://dx.doi.org/10.4135/9780857020994

Malahlela, M. K., & Johnson, E. (2024). South African teachers' application of inclusive education policies and their impact on learners with learning disabilities: Implications for teacher education. *Education Sciences*, *14*(7), Article 743. https://doi.org/10.3390/educsci14070743

Malda, M., van de Vijver, F. J., Srinivasan, K., Transler, C., & Sukumar, P. (2010). Traveling with cognitive tests: testing the validity of a KABC-II adaptation in India. *Assessment*, *17*(1), 107-115. https://doi.org/10.1177/10731911093414

Malda, M., Van De Vijver, F. J., Srinivasan, K., Transler, C., Sukumar, P., & Rao, K. (2008). Adapting a cognitive test for a different culture: An illustration of qualitative procedures. *Psychology Science Quarterly*, *50*(4), 451-468.

Maller, S. J. (2001). Differential item functioning in the WISC-III: Item parameters for boys and girls in the national standardization sample. *Educational and Psychological Measurement*, *61*(5), 793-817. https://doi.org/10.1177/00131640121971527

Mandal, S. (2021). The cognitive revolution. In T. K. Shackelford & V. A. Weekes-Shackelford (Eds.), *Encyclopaedia of evolutionary psychological science* (pp. 1167-1177). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-19650-3_1309

Matafwali, B., & Serpell, R. (2014). Design and validation of assessment tests for young children in Zambia. *New Directions for Child and Adolescent Development*, *146*, 77-96. https://doi.org/10.1002/cad.20074

Matteucci, M., & Stracqualursi, L. (2006). Student assessment via graded response model. *Statistica*, *66*(4), 435 – 447. https://doi.org/10.6092/issn.1973-2201/1216

McGill, R. J. (2015). Interpretation of KABC-II scores: An evaluation of the incremental validity of Cattell-Horn-Carroll (CHC) factor scores in predicting achievement. *Psychological Assessment*, *27*(4), 1417–1426. https://doi.org/10.1037/pas0000127

McGill, R. J. (2017). Re (examining) relations between CHC broad and narrow cognitive abilities and reading achievement. *Journal of Educational and Developmental Psychology*, *7*(1), 265-282. https://doi.org/10.5539/jedp.v7n1p265

McGill, R. J., & Dombrowski, S. C. (2018). Factor structure of the CHC model for the KABC-II: Exploratory factor analyses with the 16 core and supplementary subtests. *Contemporary School Psychology*, *22*(3), 279-293. https//doi.org/ 10.1007/s40688-017-0152-z

McGill, R. J., & Spurgin, A. R. (2017). Exploratory higher order analysis of the Luria interpretive model on the Kaufman Assessment Battery for Children–Second Edition (KABC-II) school-age battery. *Assessment*, *24*(4), 540-552. https://doi.org/10.1177/1073191115614081

McGrew, K. S. (2023). Carroll's three-stratum (3S) cognitive ability theory at 30 years: Impact, 3S-CHC theory clarification, structural replication, and cognitive–achievement psychometric network analysis extension. *Journal of Intelligence*, *11*(2), Article 32. https://doi.org/10.3390/jintelligence11020032

McGrew, K. S., Schneider, W. J., Decker, S. L., & Bulut, O. (2023). A psychometric network analysis of CHC intelligence measures: Implications for research, theory, and interpretation of broad CHC scores “Beyond g”. *Journal of Intelligence*, *11*(1), Article 19. https://doi.org/10.3390/jintelligence11010019

McKeever, M. (2017). Educational inequality in apartheid South Africa. *American Behavioral Scientist*, *61*(1), 114-131. https://doi.org/10.1177/0002764216682988

McKown, D. M. (2010). *A comparison study: Kaufman Assessment Battery for Children, Second Edition (KABC-II) and Wechsler Intelligence Scale for Children, Fourth Edition (WISC-IV) with referred students*. [School Psychology thesis. Marshall University]. Theses, Dissertations and Capstones. https://mds.marshall.edu/etd/423

Meiring, D., Van de Vijver, A. J. R., Rothmann, S., Barrick, M. R. (2005). Construct, item, and method bias of cognitive and personality tests in South-Africa. *South African Journal of Industrial Psychology*, *31*(1), 1-8. https://hdl.handle.net/10520/EJC89040

Miller, G. A. (2003). The cognitive revolution: A historical perspective. *Trends in Cognitive Sciences*, *7*(3), 141-11. https://doi.org/10.1016/S1364-6613(03)00029-9

Miller, D. C., & McGill, R. J. (2016). Review of the WISC-V. In A. S. Kaufman, S. E. Raiford, & D. L. Coalson (Eds.), *Intelligent testing with the WISC-V* (pp. 644-661). John Wiley & Sons, Inc.

Mitchell, J. M. (2015). *Psychometric evaluation of the Kaufman Assessment Battery for Children, Second Edition (KABC-II) in rural South Africa* [Master thesis, University of Stellenbosch]. SUNScholar. http://hdl.handle.net/10019.1/98043.

Mitchell, J. M., Tomlinson, M., Bland, R. M., Houle, B., Stein, A., & Rochat, T. J. (2018). Confirmatory factor analysis of the Kaufman assessment battery in a sample of primary school-aged children in rural South Africa. *South African Journal of Psychology*, *48*(4), 434-452. https://doi.org/10.1177/0081246317741822

Moerdyk, A. (2022). *The principles and practice of psychological assessment* (3rd ed.). Van Schaik Publishers.

Morgan, B., & Sklar, R. H. (2012a). Sampling and research paradigms. In J. G. Maree (Ed.), *Complete your thesis or dissertation successfully: Practical guidelines* (pp.69-80). Juta & Company Ltd.

Morgan, B., & Sklar, R. H. (2012b). Writing the quantitative research method chapter. In J. G. Maree (Ed.), *Complete your thesis or dissertation successfully: Practical guidelines* (pp.107-126). Juta & Company Ltd.

Moses, T. (2017). A review of developments and applications in item analysis. In R. E. Bennett & M. von Davier (Eds.), *Advancing human assessment: The methodological, psychological and policy contributions of ETS* (pp. 19-46). Springer Open.

Murphy, K. R., & Davidshofer, C. O. (2005). *Psychological testing: Principles and applications* (6th int. ed.). Pearson Education International.

Msila, V. (2007). From apartheid education to the Revised National Curriculum Statement: Pedagogy for identity formation and nation building in South Africa. *Nordic Journal of African Studies*, *16*(2), 146-160.

Naglieri, J. A. (1999). *Essentials of CAS Assessment*. John Wiley & Sons, Inc.

Naglieri, J. A. (2009). Intelligent intelligence testing: The influence of Alan S. Kaufman. In J. C. Kaufman (Ed.), *Intelligent testing. Integrating psychological theory and clinical practice* (pp. 73-96). Cambridge University Press.

Naglieri, J. A. (2015). Hundred years of intelligence testing: Moving from traditional IQ to second-generation intelligence tests. In S. Goldstein, D. Princiotta, & J. A. Naglieri (Eds.), *Handbook of intelligence: Evolutionary theory, historical perspective, and current concepts* (pp. 295-316). Springer Publishing Company.

Naglieri, J. A., & Das, J. P. (1990). Planning, Attention, Simultaneous, and Successive (PASS) Cognitive Processes as a model for intelligence. *Journal of Psychoeducational Assessment*, *8*, 303-337. https://doi.org/10.1177/073428299000800308

Naglieri, J. A., & Das, J. P. (2005). Planning, Attention, Simultaneous, Successive (PASS) Theory: A revision of the concept of intelligence. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (2nd ed., pp. 120-135). Guilford Press.

Naglieri, J. A., Das, J. P., Stevens, J.J ., & Ledbetter, M. F. (1991). Confirmatory factor analysis of Planning, Attention, Simultaneous, and Successive Cognitive Processing Tasks. *Journal of School Psychology*, *29*, 1-17. https://doi.org/10.1016/0022-4405(91)90011-F

Naglieri, J. A., Prewett, P. N., & Bardos, A. N. (1989). An exploratory study of Planning, Attention, Simultaneous, and Successive Cognitive Processes. *Journal of School Psychology*, *27*, 347-364. https://doi.org/10.1016/0022-4405(89)90012-5

Naicker, S. M. (2000). From apartheid education to inclusive education: The challenges of transformation. *International Education Summit for a Democratic Society*, 26-28.

Nel, M., Engelbrecht, P., Nel, N., & Tlale, D. (2014). South African teachers' views of collaboration within an inclusive education system. *International Journal of Inclusive Education*, *18*(9), 903-917. https://doi.org/10.1080/13603116.2013.858779

Nel, M., & Grosser, M. M. (2016). An appreciation of learning disabilities in the South African context. *Learning Disabilities: A Contemporary Journal*, *14*(1), 79-92.

Nel, M., Nel, N., & Hugo, A. (2016). Inclusive education: An introduction. In M. Nel, N. Nel, & A. Hugo (Eds.), *Learner support in a diverse classroom: A guide for Foundation, Intermediate and Senior Phase teachers* (2nd ed., pp. 3-34). Van Schaik Publishers.

Nell, V. (1994). Interpretation and misinterpretation of the South African Wechsler-Bellevue Adult Intelligence Scale: A history and a prospectus. *South African Journal of Psychology*, *24*(2), 100-109. https://doi.org/10.1177/008124639402400208

Neisser, U., Boodoo, G., Bouchard Jr, T. J., Boykin, A. W., Brody, N., Ceci, S. J., Halpern, D.E., Loehlin, J.C., Perloff, R., Sternberg, R.J., & Urbina, S. (1996). Intelligence: knowns and unknowns. *American Psychologist*, *51*(2), 77-101. https://doi.org/10.1037/0003-066X.51.2.77

New Zealand Federation of Specific Learning Disabilities. (n.d.). *Specific learning disability definitions*. https://www.speld.org.nz/specific-learning-disability-definitions

New Zealand Government. (2020). *Education and Training Act, No. 38 of 2020*. New Zealand Government. https://www.legislation.govt.nz/act/public/2020/0038/latest/LMS170676.html

New Zealand Ministry of Education. (2021, April 27). *Students with learning support needs.* https://www.education.govt.nz/school/student-support/special-education/

Nguyen, T. H., Han, H. R., Kim, M. T., & Chan, K. S. (2014). An introduction to item response theory for patient-reported outcome measurement. *The Patient-Patient-Centered Outcomes Research*, *7*(1), 23-35. https://doi.org/10.1007/s40271-013-0041-0

Niileksela, C. R., & Reynolds, M. R. (2019). Enduring the tests of age and time: Wechsler constructs across versions and revisions. *Intelligence*, *77*, Article 101403. https://doi.org/10.1016/j.intell.2019.101403

Nisbett, R. E., Aronson, J., Blair, C., Dickens, W., Flynn, J., Halpern, D. F., & Turkheimer, E. (2012). Intelligence: New findings and theoretical developments. *American Psychologist*, *67*(2), 130-159. https://doi.org/10.1037/a0026699

Ntloedibe, M., & Ngqinani, P. (2020, November 16). *Understanding the Living Standards Measure segmentation in South Africa* (Report No. SF2020-0061). US Department of Agriculture, Foreign Agricultural Service.

Oakland, T. (2004). Learning disabilities internationally and in Brazil: issues to consider in developing services for Brazilian students with learning disabilities. *Avaliação Psicológica, 3*(2), 115-120. ISSN 2175-3431

Ogunlade, A. O., Kruger, H. S., Jerling, J. C., Smuts, C. M., Covic, N., Hanekom, S. M., Mamabolo, R.L., & Kvalsvig, J. (2011). Point-of-use micronutrient fortification: lessons learned in implementing a preschool-based pilot trial in South Africa. *International Journal of Food Sciences and Nutrition*, *62*(1), 1-16. https://doi.org/10.3109/09637486.2010.495710

Panhwar, A. H., Ansari, S., & Shah, A. A. (2017). Post-positivism: An effective paradigm for social and educational research. *International Research Journal of Arts and Humanities*, *45*(45), 253-259. https://0-www-proquest-com.innopac.wits.ac.za/scholarly-journals/post-positivism-effective-paradigm-social/docview/2044301228/se-2?accountid=15083

Pauls, F., & Daseking, M. (2021). Revisiting the factor structure of the German WISC-V for clinical interpretability: An exploratory and confirmatory approach on the 10 primary subtests. *Frontiers in Psychology*, *12*, Article 710929. https://doi.org/10.3389/fpsyg.2021.710929

Pearson Assessments US. (2024). Wechsler Adult Intelligence Scale – Fifth Edition (WAIS-5). Pearson Assessments US. https://www.pearsonassessments.com/store/usassessments/en/Store/Professional-Assessments/Cognition-%26-Neuro/Wechsler-Adult-Intelligence-Scale-%7C-Fifth-Edition/p/P100071002.html

Piper, J. D., Mazhanga, C., Mapako, G., Mapurisa, I., Mashedze, T., Munyama, E., Mwapaura, M., Chidhanguro, D., Gerema, G., Tavengwa, N. V., Ntozini, R., Langhaug, L. F., Smuk, M., Rochat, T., Kaufman, A., Kaufman, N., Gladstone, M., Allen, E., & Prendergast, A. J. (2024). Piloting the adaptation of the Kaufman Assessment Battery for Children—2nd edition (KABC-II) to assess school-age neurodevelopment in rural Zimbabwe. *Wellcome Open Research*, *7*, Article 274. https://doi.org/10.12688/wellcomeopenres.17902.2

Piper, J.D., Mazhanga, C., Mapako, G., Mapurisa, I., Mashedze, T., Munyama, E., Mwapaura, M.,Chidhanguro, D., Mpofu, E., Mutasa, B., Gladstone, M. J., Wells, J. C., Langhaug, L. F., Tavengwa, N. V., Ntozini, R., & Prendergast, A. J. (2023). Characterising school-age health and function in rural Zimbabwe using the SAHARAN toolbox. *Plos One*, *18*(5), e0285570. https://doi.org/10.1371/journal.pone.0285570

Pluker, J. A., & Esping, A. (2014). *Intelligence 101*. Springer Publishing Company.

Poletti, M. (2016). WISC-IV intellectual profiles in Italian children with specific learning disorder and related impairments in reading, written expression, and mathematics. *Journal of Learning Disabilities*, *49*(3), 320-335. https://doi.org/10.1177/002221941455541

Ponterotto, J. G. (2005). Qualitative research in counseling psychology: A primer on research paradigms and philosophy of science. *Journal of Counseling Psychology*, *52*(2), 126-136. https: //doi.org/ 10.1037/0022-0167.52.2.126

Reid, C. A., Kolakowsky-Hayner, S. A., Lewis, A. N., & Armstrong, A. J. (2007). Modern psychometric methodology: Applications of item response theory. *Rehabilitation Counseling Bulletin*, *50*(3), 177-188. https://doi.org/ 10.1177/00343552070500030501

Reindal, M. R. (2008). A social relational model of disability: A theoretical framework for special needs education? *European Journal of Special Needs Education*, *23*(2), 135-146, https://doi.org/10.1080/08856250801947812

Reise, S. P., & Waller, N. G. (2009). Item response theory and clinical measurement. *Annual Review of Clinical Psychology*, *5*, 27-48. https://doi.org/10.1146/annurev.clinpsy.032408.153553

Renner, G., Schroeder, A., & Irblich, D. (2023). Factorial validity of the German KABC-II at ages 7 to 12 in a clinical sample: Four factors fit better than five. *Journal of Intelligence*, *11*(7), Article 148. https://doi.org/10.3390/jintelligence11070148

Republic of South Africa (RSA). (1988). *Education Affairs Act (House of Assembly) 70 of 1988*. South African Government.

Republic of South Africa (RSA). (1996a). *Constitution of the Republic of South Africa, 1996.* South African Government.

Republic of South Africa (RSA). (1996b). *South African Schools Act, No. 84 of 1996*. South African Government.

Revelle, W., & Condon, D. M. (2019). Reliability from α to ω: A tutorial. *Psychological Assessment*, *31*(12), 1395-1411.

Reynolds, M. R., & Hadorn, M. B. (2016). Our WISC-V Review. In A. S. Kaufman, S. E. Raiford, & D. L. Coalson (Eds.), *Intelligent testing with the WISC-V* (pp. 615-635). John Wiley & Sons, Inc.

Reynolds, M. R., & Keith, T. Z. (2017). Multi-group and hierarchical confirmatory factor analysis of the Wechsler Intelligence Scale for Children—Fifth Edition: What does it measure? *Intelligence*, *62*, 31-47. http://dx.doi.org/10.1016/j.intell.2017.02.005

Reynolds, M. R., Keith, T. Z., Fine, J. G., Fisher, M. E., & Low, J. A. (2007). Confirmatory factor structure of the Kaufman Assessment Battery for Children: Consistency with Cattell-Horn-Carroll theory. *School Psychology Quarterly*, *22*(4), 511-539. https://doi.org/10.1037/1045-3830.22.4.511

Roberson, D., Davidoff, J., & Shapiro, L. (2002). Squaring the circle: The cultural relativity of 'good' shape. *Journal of Cognition and Culture*, *2*(1), 29-51. https://doi.org/10.1163/156853702753693299

Rodríguez-Cancino, M., & Concha-Salgado, A. (2024). The Internal Structure of the WISC-V in Chile: Exploratory and Confirmatory Factor Analyses of the 15 Subtests. *Journal of Intelligence*, *12*(11), 105-123. https://doi.org/10.3390/jintelligence12110105

Roodt, G., & De Kock, F. (2018a). Reliability: Basic concepts and measures. In C. Foxcroft and G. Roodt (Eds.), *Introduction to psychological assessment in the South African context* (5th ed., pp. 59-68). Oxford University Press Southern Africa (Pty) Ltd.

Roodt, G., & De Kock, F. (2018b). Basic measurement and statistical concepts. In C. Foxcroft and G. Roodt (Eds.), *Introduction to psychological assessment in the South African context* (5th ed., pp. 37-58). Oxford University Press Southern Africa (Pty) Ltd.

Roodt, G., & De Kock, F. (2018c). Validity: Basic concepts and measures. In C. Foxcroft and G. Roodt (Eds.), *Introduction to psychological assessment in the South African context* (5th ed., pp. 70-81). Oxford University Press Southern Africa (Pty) Ltd.

Ruffieux, N., Njamnshi, A. K., Mayer, E., Sztajzel, R., Eta, S. C., Doh, R. F., Kengne, A. M., Ngamaleu, R. N., Chanal, J., Verdon, V., & Hauert, C. A. (2009). Neuropsychology in Cameroon: first normative data for cognitive tests among school-aged children. *Child Neuropsychology*, *16*(1), 1-19. https://doi.org/10.1080/09297040902802932

Rushton, J. P., & Jensen, A. R. (2003). African–White IQ differences from Zimbabwe on the Wechsler Intelligence Scale for Children-Revised are mainly on the g factor. *Personality and Individual Differences*, *34*(1), 177-183. https://doi.org/10.1016/S0191-8869(02)00192-7

Russell, C., Amod, Z., & Rosenthal, L. (2008). The effects of parent-child Mediated Learning Experience (MLE) interaction on young children's cognitive development. *Perspectives in Education*, 26(4), 28-41. https://hdl.handle.net/10520/EJC87499

Ryan, A. B. (2006). Post-positivist approaches to research. In M. Antonesa (Ed.) *Researching and writing your thesis: A guide for postgraduate students* (pp. 12-26). National University of Ireland.

Saklofske, D.H., Breaux, K., Beal, A.L., Raiford, S.E., & Weiss, L.G. (2019). WISC-V and the evolving role of intelligence testing in the assessment of learning disabilities. In L. G. Weiss, D. H. Saklofske, J. A.

Holdnack, & A. Prifitera (Eds.), *WISC-V Clinical use and interpretation* (2nd ed., pp. 271-329). Elsevier Inc.

Salkind, J. (2012). *Exploring research* (8th ed.). Pearson Education, Inc

Salkind, J. (2018). *Exploring research* (9th ed., global ed.). Pearson Education Limited

SAS Institute. (2018). The IRT procedure. In SAS Institute (Ed.), *SAS/STAT® 15.1 User's Guide* (pp. 5230 – 5319. SAS Institute Inc.

Sattler, J. M. (2014). *Foundations of behavioural, social and clinical assessment of children* (6th ed.). Jerome M Sattler, Publisher, Inc.

Scheiber, C. (2016a). Is the Cattell–Horn–Carroll-based factor structure of the Wechsler Intelligence Scale for Children—Fifth Edition (WISC-V) construct invariant for a representative sample of African American, Hispanic, and Caucasian male and female students ages 6 to 16 years? *Journal of Pediatric Neuropsychology*, *2*, 79-88. https://doi.org/10.1007/s40817-016-0019-7

Scheiber, C. (2016b). Do the Kaufman tests of cognitive ability and academic achievement display construct bias across a representative sample of Black, Hispanic, and Caucasian school-age children in grades 1 through 12? *Psychological Assessment*, *28*(8), 942-952. https://doi.org/10.1037/pas0000236

Scheiber, C. (2017). Does the KABC-II Display Ethnic Bias in the Prediction of Reading, Math, and Writing in Elementary School Through High School? *Assessment*, *24*(6), 729-745. https://doi.org/10.1177/1073191115624545

Scheiber, C., & Kaufman, A. S. (2015). Which of the three KABC-II global scores is the least biased? *Journal of Pediatric Neuropsychology*, *1*, 21-35. https://doi.org/10.1007/s40817-015-0004-6

Schneider, W. J., Flanagan, D. P., & Alfonso, V. C. (2017). Overview of the WISC-V. In D. P. Flanagan & V. C. Alfonso (Eds.), *Essentials of WISC-V assessment* (pp. 1 -52). John Wiley and Sons, Inc.

Schneider, W. J., & McGrew, K. S. (2018). The Cattell-Horn-Carrol Theory of Cognitive Abilities. In D. P. Flanagan & E. M. McDonough (Eds.), *Contemporary intellectual assessment: theories, tests, and issues* (4th ed., pp. 56-72). Guilford Publications.

Schreiber, J. B., Nora, A., Stage, F. K., Barlow, E. A., & King, J. (2006). Reporting structural equation modeling and confirmatory factor analysis results: A review. *The Journal of Educational Research*, *99*(6), 323-338. https://doi.org/10.3200/JOER.99.6.323-338

Schwartz, D. M. (2013). A four-and five-factor structural model for Wechsler tests: Does it really matter clinically? *Journal of Psychoeducational Assessment*, *31*(2), 175-185. https://doi.org/10.1177/0734282913478038

Scotland, J. (2012). Exploring the philosophical underpinnings of research: Relating ontology and epistemology to the methodology and methods of the scientific, interpretive, and critical research paradigms. *English Language Teaching, 5*(9), 9-16.

Seabi, J. (2012a). Feuerstein's mediated learning experience as a vehicle for enhancing cognitive functioning of remedial school learners in South Africa. *Australian Journal of Educational & Developmental Psychology*, *12*, 35-45. https://eric.ed.gov/?id=EJ1002245

Seabi, J. (2012b). Research designs and data collection techniques. In J. G. Maree (Ed.), *Complete your thesis or dissertation successfully: Practical guidelines* (pp.81-95). Juta & Company Ltd.

Seabi, J., & Amod, Z. (2009). Effects of mediated intervention on cognitive functions: A pilot study with Grade 5 learners at a remedial school. *Australian Journal of Learning Difficulties*, *14*(2), 185-198. https://doi.org/10.1080/19404150903264328

Sellbom, M., & Tellegen, A. (2019). Factor analysis in psychological assessment research: Common pitfalls and recommendations. *Psychological Assessment*, *31*(12), 1428-1441. http://dx.doi.org/10.1037/pas0000623

Seroto, J. (2011a). Indigenous education during the pre-colonial period in southern Africa. *Indilinga African Journal of Indigenous Knowledge Systems*, *10*(1), 77-88.

Seroto, J. (2011b). Indigenous education in the pre-colonial era. In J. J. Booyse, C. S. le Roux, J. Seroto, & C. C. Wolhuter (Eds.), *A history of schooling in South Africa: Method and context* (pp. 37-56). Van Schaik Publishers.

Serpell, R. (2011). Social responsibility as a dimension of intelligence, and as an educational goal: Insights from programmatic research in an African society. *Child Development Perspectives*, *5*(2), 126-133. https://doi.org/10.1111/j.1750-8606.2011.00167.x

Schenker, J. D., & Rumrill, P. D. (2004). Perspectives on scientific inquiry: Causal-comparative research designs. *Journal of Vocational Rehabilitation*, *21*, 117-121.

Schirmer, S., & Visser, R. (2023). *The silent crisis: South Africa's failing education system*. The Centre for Development and Enterprise (CDE). https://www.cde.org.za/wp-content/uploads/2023/03/The-Silent-Crisis-South-Africas-failing-education-system.pdf

Shuttleworth-Edwards, A. B. (2016). Generally representative is representative of none: Commentary on the pitfalls of IQ test standardization in multicultural settings. *The Clinical Neuropsychologist*, *30*(7), 975-998. https://doi.org/10.1080/13854046.2016.1204011

Shuttleworth-Edwards, A. B., Gaylard, E. K., & Radloff, S. E. (2013). WAIS-III test performance in the South African context: Extension of a prior cross-cultural normative database. In S. Laher & K. Cockcroft (Eds.), *Psychological assessment in South Africa: Research and applications* (pp. 17-32). WITS University Press. https://doi.org/10.18772/22013015782

Shuttleworth-Edwards, A. B., Kemp, R. D., Rust, A. L., Muirhead, J. G., Hartman, N. P., & Radloff, S. E. (2004). Cross-cultural effects on IQ test performance: A review and preliminary normative indications on WAIS-III test performance. *Journal of Clinical and Experimental Neuropsychology*, *26*(7), 903-920. https://doi.org/10.1080/13803390490510824

Shuttleworth-Edwards, A. B., & Truter, S. (2023). *Cross-cultural cognitive test norms: An advanced collation from Africa*. Inter-ED Publishers.

Shuttleworth-Edwards, A. B., Van der Merwe, A. S., Van Tonder, P., & Radloff, S. E. (2013). WISC-IV test performance in the South African context: A collation of cross-cultural norms. In S. Laher & K. Cockcroft (Eds.), *Psychological assessment in South Africa: Research and applications* (pp. 33-47). WITS University Press. https://doi.org/10.18772/22013015782

Sideridis, G. D. (2007). International approaches to learning disabilities: More alike or more different? *Learning Disabilities Research & Practice*, *22*(3), 210-215. https://doi.org/10.1111/j.1540-5826.2007.00249.

Skuy, M., Schutte, E., Fridjhon, P., & O'Carroll, S. (2001). Suitability of published neuropsychological test norms for urban African secondary school students in South Africa. *Personality and Individual Differences*, *30*(8), 1413-1425. https://doi.org/10.1016/S0191-8869(00)00123-9

Skuy, M., & Skuy, M. (2005). Contribution of intelligence and cognitive-affective variables to university grades among African, Indian, and White Engineering students in South Africa. *Journal of Cognitive Education and Psychology*, *5*(1), 25-46. https://doi.org/10.1891/194589505787382568

Skuy, M., Taylor, M., O'Carroll, S., Fridjhon, P., & Rosenthal, L. (2000). Performance of black and white South African children on the Wechsler Intelligence Scale for Children—Revised and the Kaufman Assessment Battery. *Psychological Reports*, *86*(3), 727-737. https://doi.org/10.2466/pr0.2000.86.3.7

South African Audience Research Foundation. (2010). *Living Standards Measure*. South African Audience Research Foundation. http://www.saarf.co.za/lsm/lsm-presentations.asp

South African Litigation Centre. (2016, July). *SALC Booklet: Laws and Policies Affecting Transgender Persons in Southern Africa*. South African Litigation Centre. Downloaded from https://www.southernafricalitigationcentre.org/wp-content/uploads/2017/08/Transgender-Rights-Booklet.pdf

Spaull, N. (2015). Schooling in South Africa: How low-quality education becomes a poverty trap. In A. de Lannoy, S. Swartz, L. Lake, & C. Smith (Eds.), *The South African Child Gauge* (pp. 34-41). Children's Institute, University of Cape Town. https://ci.uct.ac.za/sites/default/files/content_migration/health_uct_ac_za/533/files/Child_Gauge_2015-Schooling.pdf

Statistics South Africa. (2023, October 10). *Census 2022 statistical release*. https://census.statssa.gov.za/assets/documents/2022/P03014_Census_2022_Statistical_Release.pdf

Stegemann, K. C. (2016). Learning disabilities in Canada. *Learning Disabilities: A Contemporary Journal, 14*(1), 53-62.

Sternberg, R. J., & Grigorenko, E. L. (2006). Cultural Intelligence and Successful Intelligence. *Group & Organization Management*, *31*(1), 27-39. https://doi.org/10.1177/1059601105275255

Sternberg, R. J., Kaufman, J. C., & Grigorenko, E. L. (2008). *Applied intelligence*. Cambridge University Press.

Sunderaraman, P., Zahodne, L. B., & Manly, J. J. (2016). A commentary on 'generally representative is representative of none: pitfalls of IQ test standardization in multicultural settings' by A.B. Shuttleworth-Edwards. *The Clinical Neuropsychologist*, *30*(7), 999–1005. https://doi.org/10.1080/13854046.2016.1211321

Swart, E., & Pettipher, R. (2019). A framework for understanding inclusion. In E. Landsberg (Ed.), Krüger, D., & E. Swart (Asst. Eds.), *Addressing barriers to learning: A South African perspective* (4th ed., pp. 3-27). Van Schaik Publishers.

Swart, E., & Phasha, T. (2019). Family and community partnerships. In E. Landsberg (Ed.), Krüger, D., & E. Swart (Asst. Eds.), *Addressing barriers to learning: A South African perspective* (4th ed., pp. 269-291). Van Schaik Publishers.

Taber, K. S. (2018). The use of Cronbach's alpha when developing and reporting research instruments in science education. *Research in Science Education*, *48*, 1273-1296. https://doi.org/10.1007/s11165-016-9602-2

Taylor, N. (2009). *Construct, item and response bias across cultures in personality measurement* [Doctoral dissertation, University of Johannesburg]. University of Johannesburg Institutional Repository (UJ IR). http://hdl.handle.net/10210/3250

Taylor, R. Y. (2019). Structural validity evidence of the Wechsler Intelligence Scale for Children- Fifth Edition with African American students who have been referred for evaluation [Doctoral dissertation, Georgia State University]. ScholarWorks @ Georgia State University. https://scholarworks.gsu.edu/eps_diss/211

Taylor, T. (2013). APIL and TRAM learning potential assessment instruments. In S. Laher & K. Cockcroft (Eds.), *Psychological assessment in South Africa: Research and applications* (pp. 158-168). WITS University Press. https://doi.org/10.18772/22013015782

Taylor, P. C., & Medina, M. (2011). Educational research paradigms: From positivism to pluralism. *Assumption College of Nabunturan Research Journal*, *1*(1), 1-16.

Te Nijenhuis, J., Murphy, R., & van Eeden, R. (2011). The Flynn effect in South Africa. *Intelligence*, *39*(6), 456-467. https://doi.org/10.1016/j.intell.2011.08.003

Thomas, M. L. (2019). Advances in applications of item response theory to clinical assessment. *Psychological Assessment*, *31*(12), 1442–1455. https://doi.org/10.1037/pas0000597

Tobin, R. M. (2013). An Introduction to the Wechsler Intelligence Tests: Revisiting Theory and Practice. *Journal of Psychoeducational Assessment*, *31*(2), 91-93. https://doi.org/ 10.1177/0734282913478029

Toland, M. D. (2014). Practical guide to conducting an Item Response Theory analysis. *Journal of Early Adolescence*, *34*(1), 120–151. https://doi.org/10.1177/0272431613511332

Truijens, F. L., Cornelis, S., Desmet, M., De Smet, M. M., & Meganck, R. (2019). Validity beyond measurement: Why psychometric validity is insufficient for valid psychotherapy research. *Frontiers in Psychology*, *10*, Article 532. https://doi.org/10.3389/fpsyg.2019.00532

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

Underwood, J. E. (2007). *Empirically supported interpretation of the WISC-IV: A commonality analysis approach* (Publication No. 3288764) [Doctoral dissertation, Walden University]. ProQuest Dissertations and Theses Global.

United Nations. (1948). *Universal Declaration of Human Rights*. United Nations. https://www.un.org/sites/un2.un.org/files/udhr.pdf

United Nations. (1989). *Convention on the Rights of the Child*. United Nations. https://www.ohchr.org/Documents/ProfessionalInterest/crc.pdf

United Nations. (2006). *Convention on the Rights of People with Disabilities*. United Nations. https://www.un.org/disabilities/documents/convention/convoptprot-e.pdf

United Nations Children's Fund (UNICEF). (1999). *The African Charter on the Rights and Welfare of the Child*. UNICEF. https://www.achpr.org/public/Document/file/English/achpr_instr_charterchild_eng.pdf

United Nations Educational, Scientific, and Cultural Organization (UNESCO). (1990). *World Declaration on Education for All and Framework for Action to Meet Basic Learning Needs*. UNESCO. https://www.right-to-education.org/sites/right-to-education.org/files/resource-attachments/UNESCO_World_Declaration_For_All_1990_En.pdf

United Nations Educational, Scientific and Cultural Organisation (UNESCO). (1994). *The Salamanca Statement and Framework for Action on Special Needs Education*. UNESCO. https://www.european-agency.org/sites/default/files/salamanca-statement-and-framework.pdf

United Nations Educational, Scientific and Cultural Organisation (UNESCO). (2000). *Framework for Action: Education for All: Meeting our Collective Commitments.* UNESCO. https://www.right-to-education.org/sites/right-to-education.org/files/resource-attachments/Dakar_Framework_for_Action_2000_en.pdf

United Nations Educational, Scientific and Cultural Organisation (UNESCO). (2001). *Inclusive Education: The Way of the Future*. UNESCO. http://www.ibe.unesco.org/fileadmin/user_upload/Policy_Dialogue/48th_ICE/CONFINTED_48-3_English.pdf

United Nations Educational, Scientific and Cultural Organisation (UNESCO). (2005). *Guidelines for Inclusion: Ensuring Access to Education for All*. UNESCO. https://files.eric.ed.gov/fulltext/ED496105.pdf

Van de Vijver, F. J. R. (2002). Cross-cultural assessment: Value for money? *Applied Psychology: An International Review*, *51*(4), 545-566. https://doi.org/10.1111/1464-0597.00107

Van de Vijver, F. J. R., Weiss, L. D., Saklofske, D. H., Batty, A., & Prifitera, A. (2019). A cross-cultural analysis of the WISC-V. In L. G. Weiss, D. H. Saklofske, J. A. Holdnack, & A. Prifitera (Eds.), *WISC-V Clinical use and interpretation* (2nd ed., pp. 223-242). Elsevier Inc.

Van der Merwe, I., de Klerk, W., & Erasmus, P. (2022). Intelligence instruments applied to South African school learners: A critical review. *Frontiers in Psychology*, *13*, Article 853239. https://doi.org/10.3389/fpsyg.2022.853239

Van der Merwe, I., Erasmus, P., & De Klerk, W. (2024). Experiences of experts in intelligence measurement of South African school learners. *African Journal of Psychological Assessment*, *6*, Article 148. https://doi.org/10.4102/ajopa.v6i0.148

Van Eeden, R. (1991). *Manual for the Senior South African Individual Scale – Revised (SSAIS-R). Part 1: Background and standardisation*. Human Sciences Research Council.

Van Eeden, R., & De Beer, M. (2018). Assessment of cognitive functioning. In C. Foxcroft & G. Roodt (Eds.), *Introduction to psychological assessment in the South African context* (5th ed., pp. 169-199). Oxford University Press Southern Africa (Pty) Ltd.

Van Eeden, E. S., & Vermeulen, L. M. (2005). Christian National Education (CNE) and People's Education (PE): Historical perspectives and some broad common grounds. *New Contree*, *50*, 77-205. http://hdl.handle.net/10394/5320

Vanheule, S. (2012). Diagnosis in the field of psychotherapy: A plea for an alternative to the DSM-5. x. *Psychology and Psychotherapy: Theory, Research and Practice*, *85*(2), 128-142. https://doi.org/10.1111/j.2044-8341.2012.02069.x

Van Kraayenoord, C. E., & Chapman, J. W. (2016). Learning Disabilities in Australia and New Zealand. *Learning Disabilities: A Contemporary Journal*, *14*(1), 93-101.

Van Zyl, C. J., & Taylor, N. (2012). Evaluating the MBTI® Form M in a South African context. *SA Journal of Industrial Psychology*, *38*(1), 11-26. http://dx.doi.org/10.4102/sajip.v38i1.977

Vauclair, J., & Perret, P. (2003). The cognitive revolution in Europe: Taking the developmental perspective seriously. *Trends in Cognitive Sciences*, *7*(7), 284-285. https://hal.science/hal-01779361v1

Venter, R., & Venter, H. (2019). Addressing socio-economic barriers to learning. In E. Landsberg (Ed.), Krüger, D., & E. Swart (Asst. Eds.), *Addressing barriers to learning: A South African perspective* (4th ed., pp. 51-65). Van Schaik Publishers.

Vergottini, M., & Weyers, M. (2020). The foundations and nature of south African school social work: An overview. *Social Work*, *56*(2), 125-138. http://dx.doi.org/10.15270/52-2-816

Vetter, T. R., & Cubbin, C. (2019). Psychometrics: trust, but verify. *Anaesthesia & Analgesia*, *128*(1), 176-181. https://doi.org/10.1213/ANE.0000000000003859

Walton, E. (2011). Getting inclusion right in South Africa. *Intervention in School and Clinic*, *46*(4), 240-245. https://doi.org/10.1177/10534512103890

Walton, E., & Engelbrecht, P. (2024). Inclusive education in South Africa: Path dependencies and emergences. *International Journal of Inclusive Education*, *28*(10), 2138-2156. https://doi.org/10.1080/13603116.2022.2061608

Warne, R. T., Burton, J. Z., Gibbons, A., & Melendez, D. A. (2019). Stephen Jay Gould's analysis of the Army Beta test in The Mismeasure of Man: Distortions and misconceptions regarding a pioneering mental test. *Journal of Intelligence*, *7*(1), Article 6. https://doi.org/10.3390/jintelligence7010006

Wasserman, J. D. (2018). A history of intelligence assessment: The unfinished tapestry. In D. P. Flanagan & E. M. McDonough (Eds.), *Contemporary intellectual assessment: theories, tests, and issues* (4th ed., pp. 3-55). Guilford Publications.

Watkins, M. W., Dombrowski, S. C., & Canivez, G. L. (2018). Reliability and factorial validity of the Canadian Wechsler Intelligence Scale for Children–Fifth Edition*. International Journal of School and Educational Psychology*, *6*(4), 252-265. https://doi.org/10.1080/21683603.2017.1342580

Wechsler, D. (1939). The nature of intelligence. In D. Wechsler (Ed.), *The measurement of adult intelligence* (pp. 3–12). Williams & Wilkins Co. https://doi.org/10.1037/10020-001

Wechsler, D. (1943). Non-intellective factors in general intelligence. *The Journal of Abnormal and Social Psychology*, *38*(1), 101–103. https://doi.org/10.1037/h0060613

Wechsler, D. (1950). Cognitive, conative, and non-intellective intelligence. *American Psychologist*, *5*(3), 78–83. https://doi.org/10.1037/h0063112

Wechsler, D. (1975). Intelligence defined and undefined: A relativistic appraisal. *American Psychologist*, *30*(2), 135–139. https://doi.org/10.1037/h0076868

Wechsler, D. (1981). The psychometric tradition: Developing the Wechsler Adult Intelligence Scale. *Contemporary Educational Psychology*, *6*(2), 82–85. https://doi.org/10.1016/0361-476X(81)90035-7

Wechsler, D. (2003). *WISC-IV: Wechsler Intelligence Scale for Children – Fourth Edition*. Psychological Corporation.

Wechsler, D. (2014a). *WAIS-IV-SA: Wechsler Adult Intelligence Scale - Fourth South African Edition, Technical and Interpretive Manual*. JvR Psychometrics (Pty) Ltd.

Wechsler, D. (2014b). *WISC-V: Wechsler Intelligence Scale for Children – Fifth Edition, Technical and Interpretive Manual*. Pearson Clinical Assessment.

Wechsler, D. (2014c). *WISC-V: Wechsler Intelligence Scale for Children – Fifth Edition, Administration and Scoring Manual*. Pearson Clinical Assessment.

Wechsler, D. (2014d). *WAIS-IV-SA: Wechsler Adult Intelligence Scale - Fourth South African Edition, Administration and Scoring Manual*. JvR Psychometrics (Pty) Ltd.

Wechsler, S. M., Nunes, C. S., Schelini, P. W., Pasian, S. R., Homsie, S. V., Moretti, L., & Anache, A. A. (2013). Brazilian adaptation of the Woodcock-Johnson III Cognitive Tests. *School Psychology International*, *31*(4), 409-421. https://doi.org/ 0.1177/0143034310377165

Weiss, L. G., Keith, T. Z., Zhu, J., & Chen, H. (2013). WISC-IV and clinical validation of the four-and five-factor interpretative approaches. *Journal of Psychoeducational Assessment*, *31*(2), 114-131.https://doi.org/10.1177/0734282913478032

Weiss, L. G., Locke, V., Pan, T., Harris, J. G., Saklofske, D. H., & Prifitera, A. (2019). Wechsler Intelligence Scale for Children – Fifth Edition: Use in societal context. In L. G. Weiss, D. H. Saklofske, J. A. Holdnack & A. Prifitera (Eds.), *WISC-V Clinical use and interpretation* (2nd ed., pp. 129-186). Elsevier Inc.

Weiss, L. G., Saklofske, D. H., Holdnack, J. A., & Prifitera, A. (2016). *WISC-V assessment and interpretation: Scientists-Practitioner Perspectives*. Elsevier Inc.

Weiss, L. G., Saklofske, D. H., Holdnack, J. A., & Prifitera, A. (2019). Advances in the assessment of intelligence. In L. G. Weiss, D. H. Saklofske, J. A. Holdnack & A. Prifitera (Eds.), *WISC-V Clinical use and interpretation* (2nd ed., pp. 1-18). Elsevier Inc.

Westen, D., & Rosenthal, R. (2003). Quantifying construct validity: Two simple measures. *Journal of Personality and Social Psychology*, *84*(3), 608–618. https://doi.org/10.1037/0022-3514.84.3.608

Wilson, C. J., Bowden, S. C., Batty, A. M., Byrne, L. K., & Weiss, L. G. (2024). Factorial invariance of the Wechsler Intelligence Scale for Children, across the UK, US and Australia & New Zealand. *British Journal of Clinical Psychology*, *63*(4), 603-626. https://doi.org/10.1111/bjc.12491

Wilson, C. J., Bowden, S. C., Vannier, L. C., Byrne, L. K., & Weiss, L. G. (2023). Measurement invariance of the Wechsler Intelligence Scale for children, in Australian and New Zealand and US standardization samples. *Psychological Assessment*, *35*(6), 510-521. https://doi.org/10.1037/pas0001221

World Bank. (2023, April). Poverty and Equity Brief, South Africa. https://databankfiles.worldbank.org/public/ddpext_download/poverty/987B9C90-CB9F-4D93-AE8C-750588BF00QA/current/Global_POVEQ_ZAF.pdf

World Health Organisation. (2022). *International Classification of Diseases for Mortality and Morbidity Statistics* (11th revision) [ICD-11]. World Health Organisation.

Yang, S. Y., & Sternberg, R. J. (1997). Taiwanese Chinese people's conceptions of intelligence. *Intelligence*, *25*(1), 21-36. https://doi.org/10.1016/S0160-2896(97)90005-2

Zanon, C., Hutz, C. S., Yoo, H., Hambleton, R. K. (2016). An application of item response theory to psychological test development. *Psicologia: Reflexão e Crítica*, *29*(18), 1-10. https://doi.org/10.1186/s41155-016-0040-x

Appendix A

Cognitive Abilities and Processes Assessed by WISC-V and KABC-II

Table A1

Cognitive Abilities Assessed by the Verbal Comprehension Index and Its Subtests

CHC Broad Ability	CHC Narrow Ability	Additional Cognitive Abilities Assessed
Crystallised Intelligence and Fluid Intelligence (CHC)	Similarities: Lexical Knowledge & Induction	• Language Development • Verbal Concept Formation • Abstract Verbal Reasoning • Verbal and Auditory Comprehension • Verbal Expression or Spoken Language • Associative and Categorical Thinking • Cognitive Flexibility • Long-Term Storage and Retrieval • Distinction between Essential and Non-Essential Verbal Information • Receptive and Expressive Language
	Vocabulary: Lexical Knowledge	• Language Development • Knowledge Base • Learning Ability • Long-Term Storage and Retrieval • Verbal Comprehension • Verbal Concept Formation • Verbal Fluency • Expressive and Receptive Language
	Information: General (Verbal) Information and Knowledge	• Long-Term Storage and Retrieval • Verbal Comprehension • Expressive and Receptive Language
	Comprehension: General Information and Knowledge	• Language Development • Long-Term Storage and Retrieval • Social Judgement • Common Sense and Logical Reasoning • Knowledge of Societal and Cultural Norms • Moral and Ethical Judgement • Verbal Comprehension • Expressive and Receptive Language • Verbal Reasoning and Conceptualisation

Note. Information in this table was compiled from the following sources: "Use of Ability Tests in the Identification of Specific Learning Disabilities within the Context of an Operational Definition," by D.P. Flanagan, V.C. Alfonso, M. Costa, K. Palma, and M.A. Leahy, *Contemporary Intellectual Assessment: Theories, Tests, and Issues* (4th ed., p. 608), by D.P. Flanagan and E.M. McDonough (Eds.), 2018, Guilford Publications. Copyright 2018 by Guilford Publications. "Interpreting the WISC-V from the Perspective of Cattell-Horn-Carrol Theory," by A.S. Kaufman, S.E. Raiford, and D.L. Coalson, *Intelligent Testing with the WISC-V* (p. 256), by A.S Kaufman, S.E. Raiford, and D.L. Coalson (Eds.), 2016, John Wiley & Sons, Inc. Copyright 2016 by John Wiley & Sons, Inc. "Overview of the WISC-V," by W.J. Schneider, D.P. Flanagan, and V.C. Alfonso, *Essentials of WISC-V Assessment* (pp. 24-25), by D.P. Flanagan and V.C. Alfonso (Eds.), 2017, John Wiley and Sons, Inc. Copyright 2017 by John Wiley & Sons, Inc.

Table A2

Cognitive Abilities and Processes Assessed by the Visual-Spatial Index and Its Subtests

CHC Broad Ability/Lurian Term	CHC Narrow Ability	Additional Cognitive Abilities Assessed
Visual Processing (CHC) Simultaneous Processing (Luria)	Block Design: Visualisation	• Learning Ability • Non-Verbal Concept Formation • Visual-Perceptual Reasoning, Organisation, and Discrimination • Simultaneous Processing • Visual-Motor Integration • Visual Analysis and Synthesis • Visual-Motor Processing Speed • Planning • Attention and Concentration
	Visual Puzzles: Visualisation	• Visual-Perceptual Discrimination • Visual Acuity • Attention and Concentration • Visual Working Memory • Integration and Synthesis of Part-Whole Relationships

Note. Information in this table was compiled from the following sources: "Use of Ability Tests in the Identification of Specific Learning Disabilities within the Context of an Operational Definition," by D.P. Flanagan, V.C. Alfonso, M. Costa, K. Palma, and M.A. Leahy, *Contemporary Intellectual Assessment: Theories, Tests, and Issues* (4th ed., p. 608), by D.P. Flanagan and E.M. McDonough (Eds.), 2018, Guilford Publications. Copyright 2018 by Guilford Publications. "Interpreting the WISC-V from the Perspective of Cattell-Horn-Carrol Theory," by A.S. Kaufman, S.E. Raiford, and D.L. Coalson*, Intelligent Testing with the WISC-V* (p. 257), by A.S Kaufman, S.E. Raiford, and D.L. Coalson (Eds.), 2016, John Wiley & Sons, Inc. Copyright 2016 by John Wiley & Sons, Inc. "Overview of the WISC-V," by W.J. Schneider, D.P. Flanagan, and V.C. Alfonso, *Essentials of WISC-V Assessment* (pp. 25-26), by D.P. Flanagan and V.C. Alfonso (Eds.), 2017, John Wiley and Sons, Inc. Copyright 2017 by John Wiley & Sons, Inc.

Table A3

Cognitive Abilities and Processes Assessed by the Fluid Reasoning Index and Its Subtests

CHC Broad Ability/Lurian Term	CHC Narrow Ability	Additional Cognitive Abilities Assessed
	Matrix Reasoning: Induction	• Visual-Perceptual Reasoning, Organisation, and Discrimination • Visual and Visual-Spatial Processing and Reasoning • Simultaneous `Processing • Classification Ability • Understanding of Part-Whole Relationships • Attention and Concentration
Fluid Reasoning & Quantitative Knowledge (CHC) Planning (Luria)	Figure Weights: Quantitative Reasoning and General Sequential Reasoning	• Induction • Working Memory • Visual Processing • Attention and Concentration
	Arithmetic: Mathematical Knowledge, Mathematical Achievement, Working Memory, and Working Memory Capacity	• Ability to apply computational skills to solve numerical problems • Long-Term Storage and Retrieval • Numerical Reasoning • Attention and Concentration

Note. Information in this table was compiled from the following sources: "Use of Ability Tests in the Identification of Specific Learning Disabilities within the Context of an Operational Definition," by D.P. Flanagan, V.C. Alfonso, M. Costa, K. Palma, and M.A. Leahy, *Contemporary Intellectual Assessment: Theories, Tests, and Issues* (4th ed., p. 608), by D.P. Flanagan and E.M. McDonough (Eds.), 2018, Guilford Publications. Copyright 2018 by Guilford Publications. "Interpreting the WISC-V from the Perspective of Cattell-Horn-Carrol Theory," by A.S. Kaufman, S.E. Raiford, and D.L. Coalson*, Intelligent Testing with the WISC-V* (p. 257), by A.S Kaufman, S.E. Raiford, and D.L. Coalson (Eds.), 2016, John Wiley & Sons, Inc. Copyright 2016 by John Wiley & Sons, Inc. "Overview of the WISC-V," by W.J. Schneider, D.P. Flanagan, and V.C. Alfonso, *Essentials of WISC-V Assessment* (pp. 26-27), by D.P. Flanagan and V.C. Alfonso (Eds.), 2017, John Wiley and Sons, Inc. Copyright 2017 by John Wiley & Sons, Inc.

Table A4

Cognitive Abilities and Processes Assessed by the Working Memory Index and Its Subtests

CHC Broad Ability/Lurian Term	CHC Narrow Ability	Additional Cognitive Abilities Assessed
Short-Term Memory, which includes Working Memory (CHC) Sequential Processing (Luria)	Digit Span: Memory Span, Working Memory, and Working Memory Capacity	• Auditory Processing and Discrimination • Numerical Ability • Auditory Sequential Processing • Auditory Rehearsal • Mental Manipulation • Temporary Storage in Short-Term Memory • Rote Learning • Attention and Concentration
	Picture Span: Working Memory and Working Memory Capacity	• Visual working memory and sequential processing • Rote Learning • Attention and Concentration
	Letter Number Sequencing: Working Memory and Working Memory Capacity	• Auditory Processing and Discrimination • Numerical Ability • Auditory Sequential Processing • Mental Manipulation • Attention and Concentration

Note. Information in this table was compiled from the following sources: "Use of Ability Tests in the Identification of Specific Learning Disabilities within the Context of an Operational Definition," by D.P. Flanagan, V.C. Alfonso, M. Costa, K. Palma, and M.A. Leahy, *Contemporary Intellectual Assessment: Theories, Tests, and Issues* (4th ed., p. 608), by D.P. Flanagan and E.M. McDonough (Eds.), 2018, Guilford Publications. Copyright 2018 by Guilford Publications. "Interpreting the WISC-V from the Perspective of Cattell-Horn-Carrol Theory," by A.S. Kaufman, S.E. Raiford, and D.L. Coalson*, Intelligent Testing with the WISC-V* (p. 257), by A.S Kaufman, S.E. Raiford, and D.L. Coalson (Eds.), 2016, John Wiley & Sons, Inc. Copyright 2016 by John Wiley & Sons, Inc. "Overview of the WISC-V," by W.J. Schneider, D.P. Flanagan, and V.C. Alfonso, *Essentials of WISC-V Assessment* (p. 27), by D.P. Flanagan and V.C. Alfonso (Eds.), 2017, John Wiley and Sons, Inc. Copyright 2017 by John Wiley & Sons, Inc.

Table A5

Cognitive Abilities Assessed by the Processing Speed Index and Its Subtests

CHC Broad Ability	CHC Narrow Ability	Additional Cognitive Abilities Assessed
Processing Speed (CHC)	Coding: Rate-of-Test Taking	• Visual-Motor Integration • Attention and Concentration • Flexible Thinking • Fine Motor Coordination and Dexterity • Associative Memory • Procedural and Incidental Learning • Psychomotor Speed • Short-Term Visual Memory • Visual Scanning • Visual Perceptual Discrimination
	Symbol Search: Perceptual Speed	• Visual-Motor Integration • Attention and Concentration • Inhibitory Control • Fine Motor Coordination • Psychomotor Speed and Coordination • Visual Scanning • Decision Speed • Visual-Perceptual Discrimination

Note. Information in this table was compiled from the following sources: "Use of Ability Tests in the Identification of Specific Learning Disabilities within the Context of an Operational Definition," by D.P. Flanagan, V.C. Alfonso, M. Costa, K. Palma, and M.A. Leahy, *Contemporary Intellectual Assessment: Theories, Tests, and Issues* (4th ed., p. 608), by D.P. Flanagan and E.M. McDonough (Eds.), 2018, Guilford Publications. Copyright 2018 by Guilford Publications. "Interpreting the WISC-V from the Perspective of Cattell-Horn-Carrol Theory," by A.S. Kaufman, S.E. Raiford, and D.L. Coalson*, Intelligent Testing with the WISC-V* (p. 258), by A.S Kaufman, S.E. Raiford, and D.L. Coalson (Eds.), 2016, John Wiley & Sons, Inc. Copyright 2016 by John Wiley & Sons, Inc. "Overview of the WISC-V," by W.J. Schneider, D.P. Flanagan, and V.C. Alfonso, *Essentials of WISC-V Assessment* (p. 28), by D.P. Flanagan and V.C. Alfonso (Eds.), 2017, John Wiley and Sons, Inc. Copyright 2017 by John Wiley & Sons, Inc.

Table A6

Cognitive Abilities and Processes Assessed by the Sequential Scale and Its Subtests

CHC Broad Ability/Lurian Term	CHC Narrow Ability	Additional Cognitive Abilities Assessed
Short-Term Memory (CHC) Sequential Processing (Luria)	Number Recall: Memory Span Word Order: Memory Span (easier items) and Working Memory (with colour interference)	• Auditory Processing and Discrimination • Numerical Ability • Auditory Sequential Processing • Mental Manipulation • Temporary Storage in Short-Term Memory • Rote Learning • Attention and Concentration • Working Memory

Note. Information in this table was compiled from the following sources: "The Kaufman Assessment Battery for Children – Second Edition and the KABC-II Normative Update," by L.W. Drozdick, J. Kaufman Singer, E.O. Lichtenberger, J.C. Kaufman, A.S. Kaufman, and N.L. Kaufman, *Contemporary Intellectual Assessment: Theories, Tests, and Issues* (4th ed., p. 333), by D.P. Flanagan and E.M. McDonough (Eds.), 2018, Guilford Publications. Copyright 2018 by Guildford Publications. "Use of Ability Tests in the Identification of Specific Learning Disabilities within the Context of an Operational Definition," by D.P. Flanagan, V.C. Alfonso, M. Costa, K. Palma, and M.A. Leahy, *Contemporary Intellectual Assessment: Theories, Tests, and Issues* (4th ed., p. 608), by D.P. Flanagan and E.M. McDonough (Eds.), 2018, Guilford Publications. Copyright 2018 by Guildford Publications. "How to interpret the KABC-II: Step by Step," by A.S. Kaufman, E.O. Lichtenberger, E. Fletzer-Jansen, and N.L. Kaufman, *Essentials of KABC-II Assessment* (p. 129), by A.S. Kaufman, E.O. Lichtenberger, E. Fletcher-Janzen, and N.L. Kaufman, 2005, John Wiley & Sons, Inc. Copyright 2005 by John Wiley & Sons, Inc.

Table A7

Cognitive Abilities and Processes Assessed by the Simultaneous Scale and Its Subtests

CHC Broad Ability/Lurian Term	CHC Narrow Ability	Additional Cognitive Abilities Assessed
Visual Processing (CHC) Simultaneous Processing (Luria)	Triangles: Spatial Relations and Visualisation Block Counting: Visualisation and Math Achievement Rover: General Sequential Reasoning, Spatial Scanning, and Math Achievement	• Non-Verbal Concept Formation • Visual-Perceptual Reasoning, Organisation, and Discrimination • Simultaneous Processing • Visual-Motor Integration • Visual Analysis and Synthesis • Planning • Attention and Concentration • Visual-Perceptual Discrimination • Visual Acuity • Integration and Synthesis of Part-Whole Relationships

Note. Information in this table was compiled from the following sources: “The Kaufman Assessment Battery for Children – Second Edition and the KABC-II Normative Update,” by L.W. Drozdick, J. Kaufman Singer, E.O. Lichtenberger, J.C. Kaufman, A.S. Kaufman, and N.L. Kaufman, *Contemporary Intellectual Assessment: Theories, Tests, and Issues* (4th ed., p. 333), by D.P. Flanagan and E.M. McDonough (Eds.), 2018, Guilford Publications. Copyright 2018 by Guildford Publications. “Use of Ability Tests in the Identification of Specific Learning Disabilities within the Context of an Operational Definition,” by D.P. Flanagan, V.C. Alfonso, M. Costa, K. Palma, and M.A. Leahy, *Contemporary Intellectual Assessment: Theories, Tests, and Issues* (4th ed., p. 608), by D.P. Flanagan and E.M. McDonough (Eds.), 2018, Guilford Publications. Copyright 2018 by Guildford Publications. “How to interpret the KABC-II: Step by Step,” by A.S. Kaufman, E.O. Lichtenberger, E. Fletzer-Jansen, and N.L. Kaufman, *Essentials of KABC-II Assessment* (p. 129), by A.S. Kaufman, E.O. Lichtenberger, E. Fletcher-Janzen, and N.L. Kaufman, 2005, John Wiley & Sons, Inc. Copyright 2005 by John Wiley & Sons, Inc.

Table A8

Cognitive Abilities and Processes Assessed by the Planning Scale and Its Subtests

CHC Broad Ability/Lurian Term	CHC Narrow Ability	Additional Cognitive Abilities Assessed
Fluid Reasoning (CHC) Planning (Luria)	Pattern Reasoning: Induction and Visualisation Story Completion: Induction, General Sequential Reasoning and General Information and Knowledge	• Visual-Perceptual Reasoning, Organisation, and Discrimination • Simultaneous Processing • Attention and Concentration

Note. Information in this table was compiled from the following sources: “The Kaufman Assessment Battery for Children – Second Edition and the KABC-II Normative Update,” by L.W. Drozdick, J. Kaufman Singer, E.O. Lichtenberger, J.C. Kaufman, A.S. Kaufman, and N.L. Kaufman, *Contemporary Intellectual Assessment: Theories, Tests, and Issues* (4th ed., p. 333), by D.P. Flanagan and E.M. McDonough (Eds.), 2018, Guilford Publications. Copyright 2018 by Guildford Publications. “Use of Ability Tests in the Identification of Specific Learning Disabilities within the Context of an Operational Definition,” by D.P. Flanagan, V.C. Alfonso, M. Costa, K. Palma, and M.A. Leahy, *Contemporary Intellectual Assessment: Theories, Tests, and Issues* (4th ed., p. 608), by D.P. Flanagan and E.M. McDonough (Eds.), 2018, Guilford Publications. Copyright 2018 by Guildford Publications. “How to interpret the KABC-II: Step by Step,” by A.S. Kaufman, E.O. Lichtenberger, E. Fletzer-Jansen, and N.L. Kaufman, *Essentials of KABC-II Assessment* (p. 129), by A.S. Kaufman, E.O. Lichtenberger, E. Fletcher-Janzen, and N.L. Kaufman, 2005, John Wiley & Sons, Inc. Copyright 2005 by John Wiley & Sons, Inc.

Table A9

Cognitive Abilities and Processes Assessed by the Learning Scale and Its Subtests

CHC Broad Ability /Lurian Term	CHC Narrow Ability	Additional Cognitive Abilities Assessed
Long-Term Storage and Retrieval (CHC) Learning (Luria)	Atlantis: Associative Memory Rebus: Associative Memory	• Visual Memory and Working Memory Capacity • Visual Processing and Discrimination, • Attention and Concentration, • Scanning Ability

Note. Information in this table was compiled from the following sources: "The Kaufman Assessment Battery for Children – Second Edition and the KABC-II Normative Update," by L.W. Drozdick, J. Kaufman Singer, E.O. Lichtenberger, J.C. Kaufman, A.S. Kaufman, and N.L. Kaufman, *Contemporary Intellectual Assessment: Theories, Tests, and Issues* (4th ed., p. 333), by D.P. Flanagan and E.M. McDonough (Eds.), 2018, Guilford Publications. Copyright 2018 by Guildford Publications. "Use of Ability Tests in the Identification of Specific Learning Disabilities within the Context of an Operational Definition," by D.P. Flanagan, V.C. Alfonso, M. Costa, K. Palma, and M.A. Leahy, *Contemporary Intellectual Assessment: Theories, Tests, and Issues* (4th ed., p. 608), by D.P. Flanagan and E.M. McDonough (Eds.), 2018, Guilford Publications. Copyright 2018 by Guildford Publications. "How to interpret the KABC-II: Step by Step," by A.S. Kaufman, E.O. Lichtenberger, E. Fletzer-Jansen, and N.L. Kaufman, *Essentials of KABC-II Assessment* (p. 129), by A.S. Kaufman, E.O. Lichtenberger, E. Fletcher-Janzen, and N.L. Kaufman, 2005, John Wiley & Sons, Inc. Copyright 2005 by John Wiley & Sons, Inc.

Table A10

Cognitive Abilities Assessed by the Knowledge Scale and Its Subtests

<table>
<tr><th>CHC Broad Ability</th><th>CHC Narrow Ability</th><th>Additional Cognitive Abilities Assessed</th></tr>
<tr><td rowspan="2">Crystallised Intelligence and Fluid Intelligence (CHC)</td><td>Riddles: Language Development and Lexical Knowledge</td><td rowspan="2">• Language Development
• Verbal Concept Formation
• Abstract Verbal Reasoning
• Verbal and Auditory Comprehension
• Verbal Expression
• Cognitive Flexibility
• Long-Term Storage and Retrieval
• Distinction between Essential and Non-Essential Verbal Information
• Receptive and Expressive Language
• Knowledge Base
• Learning Ability
• Verbal Fluency</td></tr>
<tr><td>Verbal Knowledge: General (Verbal) Information and Knowledge and Lexical Knowledge (items that measure vocabulary)</td></tr>
</table>

Note. Information in this table was compiled from the following sources: “The Kaufman Assessment Battery for Children – Second Edition and the KABC-II Normative Update,” by L.W. Drozdick, J. Kaufman Singer, E.O. Lichtenberger, J.C. Kaufman, A.S. Kaufman, and N.L. Kaufman, *Contemporary Intellectual Assessment: Theories, Tests, and Issues* (4th ed., p. 333), by D.P. Flanagan and E.M. McDonough (Eds.), 2018, Guilford Publications. Copyright 2018 by Guildford Publications. “Use of Ability Tests in the Identification of Specific Learning Disabilities within the Context of an Operational Definition,” by D.P. Flanagan, V.C. Alfonso, M. Costa, K. Palma, and M.A. Leahy, *Contemporary Intellectual Assessment: Theories, Tests, and Issues* (4th ed., p. 608), by D.P. Flanagan and E.M. McDonough (Eds.), 2018, Guilford Publications. Copyright 2018 by Guildford Publications. “How to interpret the KABC-II: Step by Step,” by A.S. Kaufman, E.O. Lichtenberger, E. Fletzer-Jansen, and N.L. Kaufman, *Essentials of KABC-II Assessment* (p. 129), by A.S. Kaufman, E.O. Lichtenberger, E. Fletcher-Janzen, and N.L. Kaufman, 2005, John Wiley & Sons, Inc. Copyright 2005 by John Wiley & Sons, Inc.

Appendix B
Ethics Clearance Certificate and Permission for Extension

Research Office

HUMAN RESEARCH ETHICS COMMITTEE (NON-MEDICAL)
R14/49 Vorster

CLEARANCE CERTIFICATE	**PROTOCOL NUMBER: H16/02/37**
PROJECT TITLE	The utility of the WISC-V and KABC-II in identifying South African children with specific learning difficulties realted to reading and written express
INVESTIGATOR(S)	Mrs A Vorster
SCHOOL/DEPARTMENT	Human and Community Development/
DATE CONSIDERED	19 February 2016
DECISION OF THE COMMITTEE	Approved
EXPIRY DATE	17 November 2019

DATE 18 November 2016 **CHAIRPERSON** ______________
(Professor J Knight)

cc: Supervisor : Dr Z Amod

DECLARATION OF INVESTIGATOR(S)

To be completed in duplicate and **ONE COPY** returned to the Secretary at Room 10004, 10th Floor, Senate House, University. Unreported changes to the application may invalidate the clearance given by the HREC (Non-Medical)

I/We fully understand the conditions under which I am/we are authorized to carry out the abovementioned research and I/we guarantee to ensure compliance with these conditions. Should any departure to be contemplated from the research procedure as approved I/we undertake to resubmit the protocol to the Committee. **I agree to completion of a yearly progress report.**

______________________ ______/______/______
Signature Date

PLEASE QUOTE THE PROTOCOL NUMBER ON ALL ENQUIRIES

UNIVERSITY OF THE WITWATERSRAND, JOHANNESBURG

Research Office

HUMAN RESEARCH ETHICS COMMITTEE (NON-MEDICAL)
R14/49 Vorster

CLEARANCE CERTIFICATE	**PROTOCOL NUMBER: H16/02/37**
PROJECT TITLE	The utility of the WISC-V and KABC-II in identifying South African children with specific learning difficulties realted to reading and written express
INVESTIGATOR(S)	Mrs A Vorster
SCHOOL/DEPARTMENT	Human and Community Development/
DATE CONSIDERED	19 February 2016
DECISION OF THE COMMITTEE	Approved Two Year Extension Only
EXPIRY DATE	17 November 2021

DATE 18 November 2016 **CHAIRPERSON** ______________________
(Professor J Knight)

cc: Supervisor : Dr Z Amod

DECLARATION OF INVESTIGATOR(S)

To be completed in duplicate and **ONE COPY** returned to the Secretary at Room 10004, 10th Floor, Senate House, University. Unreported changes to the application may invalidate the clearance given by the HREC (Non-Medical)

I/We fully understand the conditions under which I am/we are authorized to carry out the abovementioned research and I/we guarantee to ensure compliance with these conditions. Should any departure to be contemplated from the research procedure as approved I/we undertake to resubmit the protocol to the Committee. **I agree to completion of a yearly progress report.**

______________________ Date ____/____/____
Signature

PLEASE QUOTE THE PROTOCOL NUMBER ON ALL ENQUIRIES

Appendix C
Principal Information Letter

Dear Sir/Madam

Good Day. My name is Adri Vorster, and I am conducting research for the purpose of completing my PhD. The focus of my research study is on the usefulness of two cognitive assessment measures (intelligence tests) in the identification of South African children with specific learning difficulties. These include the Wechsler Intelligence Scale for Children – Fifth Edition (WISC-V) and the Kaufman Assessment Battery for Children – Second Edition (KABC-II).

In order to conduct my research I will need two groups of participants from various remedial and mainstream private schools in Johannesburg. Taking this into consideration, I would like to invite your school to partake in my research.

Participation in this research will require the administration of the above mentioned assessments with each of the participants. In addition, teachers will be asked to complete the Working Memory Rating Scale for each of the children who will be assessed. If any child has been assessed on the WISC-V and the KABC-II in the past 18 months, the specific assessment will not be repeated. In these instances, permission will be obtained from you, as well as the child's parents/legal guardians, to access these results for the purpose of the study. Each of the assessment measures takes between 35 – 100 minutes to administer and will thus be conducted with each of the participants on separate days. These assessments will be conducted at a time that will not negatively impact on the participants' or the schools' academic or related activities. The assessments will be administered by either registered educational psychologists, or registered educational intern psychologists who have been specifically trained in the administration of these measures and will work under the supervision of a registered psychologist. As far as the rating scale is concerned, teachers will be asked to complete this at a time convenient for them. It should take between 5-10 minutes to complete.

In addition to the administration of the intelligence tests and the teacher's rating scale, all parents/legal guardians will also be asked to complete a parent questionnaire, which includes questions that will assist me in appropriately describing the sample (group of participants). Further, consent will be obtained from you and the parents/legal guardians to access the participants' school records. However, it is important to note that all information of the participants and the school will be treated as confidential. Although anonymity cannot be ensured in the collection of data as the participants will be assessed individually, no particulars that can identify the school or any of the participants will be included in the final research thesis or any publications/presentations that might arise from it. Only my supervisor and I will have access to the raw data/assessment results, and it will not be disclosed to anyone without the written and informed consent of the parents/legal guardians.

It should be noted that both of the intelligence tests have the potential to identify specific cognitive strengths and weaknesses in children, which can influence their scholastic progress and point towards a potential learning difficulty. Parents/legal guardians will be advised of this and feedback will be offered to all parents/legal guardians at a time convenient for them. In addition, parents/legal guardians will be told that should any concerns regarding their children's cognitive functioning arise, they will be informed accordingly. Further, with the written consent of the parents/legal guardians, feedback will be provided in collaboration with the School Learner Support Specialist/s so that, where necessary, children can be supported academically within their educational environments.

Participation in this study is completely voluntary and there are no foreseeable risks or benefits for partaking in it; other than the parents/legal guardians having access to their children's assessment results for the purpose of supporting them academically. Also, the research does not expose any of the participants or schools to any harm. All participants and schools have the right to withdraw from the study at any time; without any consequence. Participants also do not have to answer any questions that they do not feel comfortable with.

Once data collection is completed, all raw data will be transformed into digital format with all identifying particulars removed. This digital data will be stored on a password protected file on a password protected computer. The raw data will be stored in a locked cupboard at the University of the Witwatersrand. In digital format, the participants' assessment results, personal and biographical information will be linked by means of codes. In this manner anonymity of the schools and the participants in the final thesis and other potential publications/presentations, can be ensured.

The results of the study will be compiled into a thesis, which will be stored in the University of the Witwatersrand main library, and made available online on the WITS Institutional Repository environment on WIReDSpace, where an electronic copy of all research reports, dissertations and theses are kept. Results of the study might also be published in the form of peer reviewed journal articles and/or book chapters. Likewise it might be presented at conferences. The school will also be provided with an executive summary of the research results.

Should you consent for your school to partake in this study, please indicate this in writing. Should you have any questions or concerns with regards to this research study, please contact me directly.

Thank you

____

Adri Vorster
Researcher: Educational Psychologist
011 717 4554
adri.vorster@wits.ac.za

____

Dr. Zaytoon Amod
Research Supervisor: Educational & Clinical Psychologist
011 717 8326
zaytoon.amod@wits.ac.za

Principal Consent Form

I, ________________________________, hereby give consent for Adri Vorster to conduct her research on the usefulness of specific cognitive assessment measures in the identification of children with specific learning difficulties in the South African context, at _________________________ (name of school).

I understand that:

- Participation in this study is completely voluntary and there are no foreseeable risks or benefits for partaking in it; other than the parents/legal guardians having access to their children's assessment results for the purpose of supporting them academically.
- The research does not expose any of the participants or schools to any harm.
- All participants and schools have the right to withdraw from the study at any time; without any consequence.
- Participants do not have to answer any questions that they do not feel comfortable with.
- The school's and participants' information will be treated as confidential.
- Anonymity and confidentiality will be ensured in the final research thesis or any publications/presentations that might arise from this research.
- The results of the study will be compiled into a thesis, which will be stored in the University of the Witwatersrand main library, and made available online on the WITS Institutional Repository environment on WIReDSpace, where an electronic copy of all research reports, dissertations and theses are kept.
- Results of the study might also be published in the form of peer reviewed journal articles and/or book chapters. Likewise it might be presented at conferences.
- The school will be provided with an executive summary of the research results.
- Feedback will be offered to all parents/legal guardians at a time that is convenient for them and they will be informed if any concerns regarding their children's cognitive functioning arise. With the written consent of the parents, this will be done in collaboration with the school's learner support specialist/s so that specific children can be supported within their educational environments.

______________________ ______________________

School Principal's Signature Researcher's Signature

____

Date

Appendix D
Parent/Legal Guardian Information Letter

Dear Sir/Madam

My name is Adri Vorster, and I am conducting research for the purpose of completing my PhD. The focus of my research study is on the usefulness of two cognitive assessment measures (intelligence tests) in the identification of South African children with specific learning difficulties. The two assessment measures include the Wechsler Intelligence Scale for Children – Fifth Edition (WISC-V) and the Kaufman Assessment Battery for Children – Second Edition (KABC-II). In order to conduct my research I will need two groups of participants from various remedial and mainstream private schools in Johannesburg. Taking this into consideration, I would like to invite your child/ward to partake in my research.

Should you give consent for your child/ward to participate in this study, he/she will be assessed on both these tests. However, if your child/ward has been assessed on either of these assessments in the past 18 months, the specific assessment will not be repeated. In these instances, permission will be obtained from you, as well as the school, to access these results for the purpose of the study. Each of the assessment measures takes between 35 – 100 minutes to administer and will thus be conducted with each of the children on separate days. These assessments will be done at a time that will not negatively impact on the children's or the schools' academic or related activities. The assessments will be done by either registered educational psychologists, or registered educational intern psychologists who have been specifically trained in doing these assessments. The intern educational psychologists will work under the supervision of a registered psychologist.

In addition to the administration of the intelligence tests, you will also be asked to complete a parent questionnaire, which includes questions that will assist me in appropriately describing the group of children who participate in the study. This questionnaire includes a variety of questions related to the children. It also probes whether your child/ward has ever been diagnosed with any difficulties that can impact his/her performance on these assessment measures.

In addition to the parent questionnaire, your child's/ward's teacher will be asked to complete the Working Memory Rating Scale. Further, with your consent, I will access your child's school records. However, it is important to note that all information of the children and the school will be treated as confidential. Although I cannot keep your child's/ward's information anonymous in the collection of the data as the assessments will be done individually with each child/ward, no information that can identify your child/ward or the school will be included in the final research thesis or any publications/presentations that might arise from it. Only my supervisor and I will have access to the raw data/assessment results and it will not be disclosed to anyone without your written and informed consent as the parent/legal guardian.

Both intelligence tests have the potential to identify your child's/ward's cognitive strengths and weaknesses, which can influence his/her progress at school and point towards a potential learning difficulty. Taking this into consideration, feedback will be offered on your child's results and you will be advised if any concerns regarding your child's cognitive functioning arise. This feedback will be done at a time that is convenient for you. With your written permission, I will ask the school's Learner Support Specialist/s to be available at this feedback session so that your child/ward can be assisted from the school's side where needed.

Participation in this study is completely voluntary and it is your choice whether your child/ward will participate or not. I do not foresee that you, your child/ward or the school will be subjected to any risks whilst involved in this study. The only benefit is the fact that you will have access to your child's/ward's assessment results and that he/she can be supported scholastically based on these results. All parents, children and schools have the right to withdraw from the study at any time; without any consequence. Children/wards and parents/legal guardians also do not have to answer any questions that they do not feel comfortable with.

Once data collection is completed, I will transfer all raw data into digital format to be stored on a password protected computer. In this format, any information that can identify you, your child/ward or the school will be removed. The raw assessment results in paper format will be stored in a locked cupboard at the University of the Witwatersrand. In digital format, the children's assessment results, personal and biographical information will be linked by means of codes. In this manner anonymity of the schools and the children in the final thesis and other potential publications/presentations, can be ensured.

The results of the study will be compiled into a thesis, which will be stored in the University of the Witwatersrand main library, and made available online on the WITS Institutional Repository environment on WIReDSpace, where an electronic copy of all research reports, dissertations and theses are kept. Results of the study might also be published in the form of peer reviewed journal articles and/or book chapters. Likewise it might be presented at conferences. The school will also be provided with an executive summary of the research results.

Should you consent for your child/ward to be approached to partake in this research, please read and sign the attached consent form. Also, please complete the attached parent questionnaire as honestly and comprehensively as possible. Should you have any questions or concerns with regards to this research study, please contact me directly.

Please detach and keep these sheets

Thank you

____

Adri Vorster
Researcher: Educational Psychologist
011 717 4554
adri.vorster@wits.ac.za

____

Dr. Zaytoon Amod
Research Supervisor: Educational & Clinical Psychologist
011 717 8326
zaytoon.amod@wits.ac.za

Parent/Legal Guardian Consent Form

I, ____________________________________, parent/legal guardian of ____________________, currently in Grade ____, hereby give consent for Adri Vorster to include my child/ward in her research on the usefulness of specific cognitive assessment measures in the identification of children with specific learning difficulties in the South African context.

I understand that:

- My child/ward will be assessed on all or some of the following assessment measures; Wechsler Intelligence Scale for Children – Fifth Edition (WISC-V) and the Kaufman Assessment Battery for Children – Second Edition (KABC-II).

- The researcher will access my child's/ward's school records. My child's/ward's teacher will be asked to complete the Working Memory Rating Scale. I will be asked to complete a parent questionnaire.

- Participation in this study is completely voluntary and there are no foreseeable risks or benefits for partaking in it; other than me having access to my child's/ward's assessment results for the purpose of supporting him/her academically.

- The research does not expose my child/ward or his/her school to any harm.

- My child/ward and his/her school have the right to withdraw from the study at any time; without any consequence. Also, my child/ward does not have to answer any questions that he/she does not feel comfortable with.

- The school's and my child's/ward's information will be treated as confidential. Anonymity and confidentiality will be ensured in the final research thesis or any publications/presentations that might arise from this research.

- The results of the study will be compiled into a thesis, which will be stored in the University of the Witwatersrand main library, and made available online on the WITS Institutional Repository environment on WIReDSpace, where an electronic copy of all research reports, dissertations and theses are kept.

- Results of the study might also be published in the form of peer reviewed journal articles and/or book chapters. Likewise it might be presented at conferences.

- The school will be provided with an executive summary of the research results.

- I will offered feedback regarding my child's/ward's assessment results and informed if any concerns regarding my child's/ward's functioning arises. With my permission the school's Learner Support Specialist will be included in this feedback session so that my child can be supported within the classroom.

_________________________ _________________________

Parent/Legal Guardian's Signature Researcher's Signature

____

Date

Appendix E
Participant Information Letter and Assent Form

Dear Learner

Hello. My name is Adri Vorster, and I am a student just like you. As part of my studies, I have to conduct some research, just like you have to do some projects. I would like to ask if you would be willing to help me with my project. If you are willing, one of my helpers or I will do some tasks with you. During our time together, we will do different tasks with you. Sometimes we will ask you to answer questions and sometimes we will ask you to build things or to look at pictures and answer questions about them. We might ask you to remember things for a bit and then to tell us what you remember. Some of the tasks might be easy and some might be more difficult. We only ask that you do your best and answer as many questions as you can. If you really don't want to do anything, you can tell us.

These tasks will be done with you alone and we will probably need at least two days on which to meet with you. On each of these days, you will be busy for about an hour and a half. We will meet you at a time that is convenient for you, your school and your parents. This means that my helpers and I will not ask you to do anything at a time that interferes with your schoolwork or other activities that you might be busy with. You can decide whether you want to be a part of my project or not. Also, if you no longer want to be part of the project, you can also tell us.

Remember that my project is not part of your schoolwork and your teachers, or anyone else in the school, will not see any of your answers. Your parents might ask me to tell them how you did with the different tasks and then I will tell them so that they and your teachers can help you with any of your schoolwork that is a bit difficult. Other than that, no-one but my supervisor, who is my teacher, and I will know how you did with the different tasks.

If you are willing to help me with my project, please fill in your name and sign this letter and one my helpers or myself will contact you soon to do the tasks. These might take more than one day and if you are tired, you can tell us and we will take a break.

Thank you very much

______________________	______________________
Learner's Name	Learner's Signature

____
Date

Appendix F
Parent Questionnaire

Dear Parent/Legal Guardian

Thank you very much for your willingness to complete this questionnaire and for allowing your child to take part in this study.

Please complete the questions below as comprehensively as possible by adding a cross (×) or a tick (√) in the spaces provided. Where possible, please elaborate on your answers as I value your insight into your child's current learning and functioning.

1. **Child's Name and Surname** ____

2. **Child's Date of Birth** ____

3. **Child's Grade**

Grade 4	
Grade 5	
Grade 6	
Grade 7	

4. **Child's Gender**

Male	
Female	

5. **What is your child's first language?** ____________________________________

6. **What is your child's second language?** ____________________________________

7. **Other languages that your child is regularly exposed to?** ______________________________

8. **If English is your child's second or additional language, how long has he/she been exposed to it? (please elaborate if necessary)**

9. Mother's Highest Level of Education

Primary School (Grade 7 or less?)	
Grade 8 or 9	
Grade 10 or 11	
Grade 12 or the Equivalent	
Certificate	
Higher Certificate	
Diploma	
Postgraduate Diploma	
Bachelor's Degree	
Professional First Degree Postgraduate	
Honour's Degree	
Master's Degree	
Doctoral Degree	

10.Father's Highest Level of Education

Primary School (Grade 7)	
Grade 8 or 9	
Grade 10 or 11	
Grade 12 or the Equivalent	
Certificate	
Higher Certificate	
Diploma	
Postgraduate Diploma	
Bachelor's Degree	
Professional First Degree Postgraduate	
Honour's Degree	
Master's Degree	
Doctoral Degree	

11.Has your child ever been diagnosed with a serious medical condition that impacts on his/her learning? (If yes, please elaborate)

Yes	
No	

12. Has your child ever been diagnosed with a Specific Learning Disorder or Difficulty that impacts on his/her learning? (If yes, please elaborate)

Yes	
No	

13. Has your child ever been diagnosed with an Emotional Disorder or Difficulty that impacts on his/her learning? (If yes, please elaborate)

Yes	
No	

14. Has your child ever been diagnosed with ADHD? (If yes, please elaborate)

Yes	
No	

15. Has your child ever been diagnosed with a Physical Disorder or Difficulty that impacts on his/her learning? (If yes, please elaborate)

Yes	
No	

16. Has your child ever been diagnosed with a Language Disorder or Difficulty that impacts on his/her learning? (If yes, please elaborate)

Yes	
No	

17. Has your child ever been diagnosed with a Sensory Disorder or Difficulty that impacts on his/her learning? (If yes, please elaborate)

Yes	
No	

18. Are there any other factors, that in your opinion, impact on your child's learning at present? (If yes, please elaborate)

Yes	
No	
