Neuropsychological Evaluation Among School-Aged Children in the Context of HIV in an Urban Kenyan Setting

Maina, R.W.

Publication date:
2024

Document Version
Publisher's PDF, also known as Version of record

NEUROPSYCHOLOGICAL EVALUATION AMONG SCHOOL-AGED CHILDREN IN THE CONTEXT OF HIV IN AN URBAN KENYAN SETTING

RACHEL W. MAINA

Neuropsychological Evaluation Among School-Aged Children in the Context of HIV in an Urban Kenyan Setting

Proefschrift ter verkrijging van de graad van doctor aan Tilburg University op gezag van de rector magnificus, prof. dr. W.B.H.J. van de Donk, in het openbaar te verdedigen ten overstaan van een door het college voor promoties aangewezen commissie in de Aula van de Universiteit op vrijdag 9 februari 2024 om 13.30 uur door Rachel Wanjiru Maina, geboren te Nairobi, Kenia

Promotores:	em. prof. dr. A.J.R. van de Vijver † (Tilburg University) prof. dr. J.M. Wicherts (Tilburg University)
Copromotores:	prof. dr. A. Abubakar (Aga Khan University) dr. M. Kumar (University of Nairobi) dr. J. He (Tilburg University)
leden promotiecommissie:	prof. dr. M.J. Gladstone dr. P. Bangirana prof. dr. M.M. Sitskoorn dr. N. Nakasujja prof. dr. I. Kvestad

This study was part of a larger study that was funded through a seed grant for early career researchers organized by Partnerships for Mental Health Development in Sub-Saharan Africa (PaMD) (NIMH award number U19MH98718) and the Kenyatta National Hospital's Research & Programs Department (KNH/R&P/23F/55/13). The funding institutions had no role in the study conceptualization, data collection, analysis, data interpretation and manuscript preparation.

ACKNOWLEDGEMENTS

I am tremendously grateful for the scientific and personal support by my promoters, Prof. Jelte M. Wicherts and Prof. Fons J. Van De Vijver, as well as co-promoters, Prof. Abubakar Amina, Dr. Jia He and Dr. Kumar Manasi. Their encouragement and guidance over the PhD program duration ensured I understood new concepts, refined my skills and knowledge, interacted with renowned and gifted experts in my field and plotted a career trajectory that would impact the lives of children living with neurocognitive deficits. Sadly, Prof. Fons J. Van De Vijver passed on in June 2019. I did not get the chance to meet him in person, but his encouragement and clear good intentions for me to pursue this work in cross-cultural measurements was felt 3800 miles away. I am the researcher I am today because of the invaluable support of my promoters and co-promoters.

I would also like to acknowledge the mentorship and generous support of Prof. Miguel Perez-Garcia in giving me unrestricted access to BENCI. The ideas I had in designing a contextually appropriate neurocognitive tool could not have been fruitful without access to a digitized, free access battery with all DSM V cognitive domains and whose items and stimuli I could re-design with the help of his team in Granada. Thank you.

I also acknowledge the valuable support I got from the KEMRI Wellcome Trust staff in Kilifi namely Khamisi Katana, Eva Mwangombe and Paul Mwangi. They helped with preliminary information and training on the Kilifi Toolkit. In addition, my thanks go to the staff in the HIV Clinics and Schools with much appreciation to Elder Mrs. Roseann Maina and Sister Owen. Thank you for trusting me with your patients and students and the support in collecting data for the studies.

I am grateful to my parents for keeping me in prayers even when it took a while to understand what exactly I was doing for my PhD program. Thank you for trusting me to make independent decisions and unswervingly supporting me in prioritising the PhD work.

I thank the Lord Almighty for his grace in conducting this PhD work up to completion. This far I have experienced his mercy and might. Thank you

TABLE OF CONTENTS

CHAPTER ONE: INTRODUCTION

Children growing up in lower- and middle-income countries (LMICs) are at significant risk of experiencing neurocognitive impairment due to exposure to multiple risk factors (1, 2) such as poor social-economic status, low maternal education, limited schooling, and early exposure to malnutrition, poor healthcare, and infectious diseases including HIV (Human Immunodeficiency Virus) (3-6). Among lower school students, early exposure to trauma, prenatal complications, stunting, and HIV infection may impair their attention, executive functioning, memory, reasoning, and perceptual and language capabilities, among other neurocognitive functions (6-10). Yet little is known about the extent of cognitive impairment in this population in LMICs where risk factors including malnutrition and infectious diseases are highly prevalent (11-14). The true burden of neurocognitive impairments among children in LMICs is not documented partly due to a shortage of adequately standardized, easy-to-implement (non-invasive) neurocognitive assessments within these contexts (15-19). The aim of this thesis is threefold. First, through a review of the currently available neurocognitive tools for use with lower school students (children aged 6 – 12 years), we highlight the need and urgency to adapt/test tools in resource-limited contexts. Second, we aim to empirically test the psychometric appropriateness of a preferred neurocognitive battery in a cohort of Kenyan children. Third, we use the battery to study the effects of stunting and HIV on cognitive outcomes in the same cohort.

Current neurocognitive tools and the need for adaptation

While a host of neurocognitive tools for 6–12-year-olds are currently available in Kenya, unfortunately, these are not culturally validated, and their psychometric properties are not well documented, thus creating a lack of systematic knowledge in the literature (18, 19). The few existing tools are neither standardized nor tested for psychometric appropriateness in African settings, given limited resources and expertise (e.g., non-existing training programs) among

researchers in these settings (20). The few earlier studies (18) on these tools in Africa often involved small sample sizes, were restricted to clinical cohorts, often lacked normative data, and often showed mixed findings regarding reliability, validity, and measurement invariance findings (19, 21).

Neurocognitive assessment remains vital in evaluating, monitoring, and managing cognitive illnesses, especially among children whose cognitive functions are impaired by HIV and/or stunting (22, 23). Indeed, HIV and often co-occurring stunting are associated with lower performance in perceptual information processing, memory, reasoning, language, verbal information processing, numerical abilities, and overall cognitive functioning (5, 6, 9, 10). Developing tools for low-resource settings with norms that tap into cognitive functions is important for lower- and middle-income countries, considering that these settings have a large population of children living with HIV and stunting at risk of neurocognitive deficits.

Individual differences in neurocognitive performance are multidimensional; the Cattell–Horn–Carroll (CHC) model of differences describes a wide range of cognitive functions that show individual differences, including working memory (Gsm), processing speed (Gs), long-term memory encoding and retrieval (Glr), visuospatial ability (Gv), acquired knowledge or crystallized ability (Gc), and fluid reasoning (Gf) as main factors (24). The CHC model continues to evolve as additional factors of previously unmeasurable and unknown abilities are integrated. The CHC model stresses the need for multidimensional batteries (25) in neurocognitive assessment. Because cultural factors [i.e., knowledge and behaviours that characterize a particular group of people (26)] can affect neurocognitive assessment, adaptation and standardization of non-invasive and validated tools could aid in appraising the true burden of cognitive impairment among children who suffer from HIV and/or stunting.

Evidence suggests that importing Western measures into non-Western settings without adequate attention to adaptation, standardization, and validation can yield invalid and unreliable results (27, 28). Cultural adaptation and validation of tools measuring cognitive functions are recommended in light of cultural influences on mental and neural-biological processes and effects on measurements themselves, like familiarity with items and instructions and alignment of test difficulty to test takers' ability levels (4, 29, 30). Research suggests that how humans think, learn, and behave is affected by a host of cultural factors (3, 4, 30-32). For example, cultural adaptations could lead various parts of the brain, such as the occipital and temporal lobes, to process information differently (33). It is challenging to have a culturally sensitive tool where the intended meaning of items is the same or perceived similarly across cultures. To date, and due to the cultural variations discussed earlier, this may not be possible (28, 34); hence, each tool that purports to test neurocognitive functions must be adapted to the culture of the intended respondents and appropriately tested for psychometric performance.

Among lower school students (mostly aged 6 – 12 years old), neurocognitive development is rapid (35, 36), and neurocognitive tools provide crucial information on how well the student can learn, including the cognitive impairment that impedes learning (37). This information helps teachers and clinicians modify their teaching and treatment approaches to support the child's learning capacity. Early cognitive recovery and modified teaching practices to suit a child's cognitive development needs are essential interventions for cognitive deficiencies (38, 39). However, this appraisal of cognitive needs among lower school students in LMICs is foiled by a scarcity of culturally sensitive, valid, and reliable neurocognitive tools.

Overview of the Chapters

As far as researchers have developed, adapted, and standardized several neurocognitive tests among 6 – 12-year-olds (40-42), the application of these tests in new contexts may

introduce measurement and methodological biases, which requires empirically assessing the suitability of the tests considering the context of interest (43). In Chapter 2, we investigate the current neuropsychological tools used globally among 6 – 12-year-olds and their psychometric outcomes. As a narrative review, the chapter outlines the geographical diversity in tools for this age group, the procedures used to adapt or develop them, and their psychometric outcomes. We synthesize the results and offer proposals for cultural adaptation and generating psychometric information of neuropsychological tools.

For Chapters 3 and 4, we identified a neurocognitive battery that assessed the neurocognitive domains recommended for evaluation among patients by the Neurocognitive Work Group (44) in the fifth edition of the Diagnostics and Statistical Manual of Mental Disorders (DSM-V) (45). The battery needed to be useful in assessing cognitive functioning among Kenyan children suffering from HIV and stunting. The Computerized Battery for Neuropsychological Evaluation of Children (BENCI) covers almost all domains recommended by DSM-V, namely: executive functions, perceptual-motor, complex attention, language, learning, and memory (45). The current version of the BENCI does not include social cognition, which should be integrated into future versions of BENCI to cover all domains recommended in the DSM-V.

Chapter 3 contributes to the literature on much-needed neurocognitive assessments for LIMCs by translating, adapting, and providing data on the validity and reliability of the English version of the BENCI among lower school children in Kenya. With a case-control study design, we studied differences in performance between children infected with HIV and those not infected with HIV and the test-retest reliability, internal consistency, validity, and factorial structure of the BENCI. In testing for the tool's convergent validity, we related scores on the subtests of BENCI to scores of subtests measuring the same domains in the Kilifi toolkit (46). Moreover, we used Confirmatory Factor Analysis (CFA) to evaluate the fit of the BENCI data

to a multi-dimensional model of executive functioning and considered whether the BENCI exhibited measurement invariance between children living with or without HIV. With measurement invariance, we can check whether the subtests are loaded similarly onto the latent factors and whether lower school children living with and without HIV can be meaningfully compared (47).

In Chapter 4, we focused on stunting because it would be an important target for intervention given its effects on cognitive functioning, school performance, and eventual earning potential (48). Stunting and HIV often co-occur in children (14) and affect their cognitive functioning (6). However, the extent to which HIV influences stunting in developing cognitive impairment is not well known. In LMICs where stunting and HIV remain prevalent and often persist into middle childhood, having a neurocognitive battery adapted to the prevailing culture, with psychometrically sound properties and implemented in regular intervention programs, could aid in knowing and addressing the true burden of cognitive impairment among lower school children. In Chapter 4, we applied structural equation modelling to the same data as in Chapter 3 to predict cognitive functioning by HIV status, age, and gender, and we studied whether stunting mediated these effects.

In Chapter 5, we summarize the results from Chapters 2-4, discuss findings and future directions, and make recommendations for (1) researchers to adapt neurocognitive tests for generating valid and reliable psychometric information on neurocognitive performance; (2) clinicians to verify the clinical utility of these neurocognitive tools, and (3) policymakers to integrate routine cognitive assessment and management in holistic HIV care.

This thesis offers information on neurocognitive tools commonly used among lower school students, the psychometric properties of these neurocognitive tools, validates the BENCI among a large sample of Kenyan children, and offers insights into how HIV and stunting affect

neurocognitive performance, to inform us about potential interventions and remedial efforts to improve these children's neurocognitive functioning.

CHAPTER TWO

Assessing Neuropsychological Functions in Middle Childhood: A Narrative Review of Measures and their Psychometric Properties across Contexts

Revised Version of Published Article: Maina, R., van de Vijver, F., Abubakar, A., Perez-Garcia, M., Kumar, M. Assessing Neuropsychological Functions in Middle Childhood: A Narrative Review of Measures and Their Psychometric Properties Across Context. Journal of Paediatric Neuropsychology 7, 113–138 (2021). https://doi.org/10.1007/s40817-021-00096-9

ABSTRACT

There are many tools to assess neuropsychological functioning among children aged 6-12 years. However, most of these tools have been developed in High-Income Countries (HICs). These tools are often adapted to avoid or minimize bias in assessment in other cultural contexts. In selecting subtests to adapt before using the entire neuropsychological battery, researchers would benefit from having a summary of the available tools and how easily they can be used in different contexts. The aims of this narrative review were to identify neuropsychological tools commonly used among 6–12-year-olds and to summarize the psychometric properties of these tools, especially emphasizing their usage across diverse cultural contexts. We searched peer-reviewed articles in PubMed, PsycINFO, and Web of Science and published 1997-2017 for studies using neuropsychological or neurocognitive assessments or tools among children aged 6 to 12 years. A hundred and forty-five papers out of 306 reported on psychometric properties of different tools, including the Behaviour Rating Inventory of Executive Function - BRIEF (Count=6), Visual-Motor Integration - VMI (Count=6), The Test of Memory Malingering - TOMM (Count=6), Medical Symptom Validity Test - MSVT (Count=6) and Continuous Performance Tests - CPT (Count=6). Forty-six percent of the papers reported studies conducted in the United States. Most studies were based in High-Income Countries, which further highlights the need to validate these tools for use in Lower-and Middle-Income

Countries (LMICs). Psychometric checks were adequate for most tools measuring executive functioning, such as BRIEF, although tools such as CPT measuring complex attention showed mixed findings related to psychometric quality. Moreover, we found that many studies addressed certain aspects of validity and/or reliability while leaving out others, thus a comprehensive picture is lacking. To use a tool in a specific context, it is important to know its validity and reliability, which is not always reported in publications. We propose further studies to thoroughly investigate and report the psychometric properties of neuropsychological tools, especially in LMICs.

Key Words: Child neuropsychological assessments and tools, psychometrics, continuous performance, executive functioning, sensitivity, and specificity norms.

INTRODUCTION

The ages 6 – 12 are known as the *'ages of reason'* in Piagetian theories of cognitive development (49). Children aged 6 – 7 years are likely to start developing reasoning abilities related to a concrete operational level of cognitive development where they can form complex representations and solve complex problems (49). For example, a child at this age can understand that a parent can be a disciplinarian and at the same time be a provider, while a teacher can also be a parent at their own home and hence be a disciplinarian and provider to their own children. As these cognitive abilities develop, the formal operations level of cognitive development starts at ages 10 – 12 years (49). This is where the children can form generalizations across different instances and have abstract reasoning abilities. They can combine several shapes to form an overall pattern.

Performance on these cognitive abilities is founded on the physiological growth of the brain in terms of neurons whose plasticity or formation is affected by environmental factors. Performance is measured adequately by valid and reliable neurocognitive tools. This narrative review aims to assess the psychometric adequacy of these tools. This is particularly relevant for children aged 6 – 12 years whose literature on psychometric properties of cognitive tools is marred by mixed findings (21, 50) that make it hard to find one tool for a specific cognitive function whose validity and reliability indicators are suitable for assessing the functionality of a child in a given context (51). Children aged 6 – 12 years are just starting school, and their ability to learn is embedded in cognitive functions such as those related to memory formation, problem-solving, flexibility, and judgment (52, 53). Functions such as cognitive flexibility among these children have been found to be related to school performance (52). Culture-sensitive tools can be used to identify learning problems and inform instruction plans improving performance or treatments that rehabilitate cognitive deficits. Tools for children aged 6 – 12 years are diverse and show mixed findings on their validity and reliability indicators

(21, 50, 54, 55). Cultural diversity calls for the development or adaptation of tools that are appropriate for the cultural context. This narrative review aims to summarize findings on the psychometric properties of cognitive tools used among children aged 6 – 12 in various contexts.

NEUROPSYCHOLOGICAL TOOLS

Neuropsychological tools are measures used to assess the brain-behaviour relationship (56). Neuropsychological tools refer broadly to all tools that measure psychological functions behind a brain related injury/condition (e.g., traumatic brain injury) and cognitive functions, while neurocognitive tools refer to tools that measure only cognitive functioning. Executive function, memory, visuomotor coordination, processing speed, language, and attention are basic cognitive domains measured using neuropsychological tools (40). Intrusive tools such as the spinal tap were used before the advent of neuropsychological tools, which have, over the years, evolved from paper-based tools to computerized ones. Neuropsychological tools have not only made assessing cognitive functions less intrusive, but they have also become more comprehensive and easier to administer over the years, with some tools needing no training to administer and score. This enables diagnosing neurocognitive disorders and monitoring dysfunction progression and recovery, thereby better informing (remedial) interventions.

Good neuropsychological tools must be standardized, reliable, and valid. A tool is valid when it measures what it purports to measure, and it is reliable when it accurately measures what it is supposed to measure (57). A tool is said to have sensitivity when it can identify those with disease and to have specificity when it can identify those without disease (58). Testing of validity and reliability of a tool is construed in different forms. Construct validity is supported when correlations between tools align with hypothesized correlations between constructs (59). Discriminant and convergent validity are used to establish construct validity. Discriminant validity is established when two tools that are supposed to measure different phenomena

demonstrate this difference. Convergent validity, which is also referred to as concurrent validity in this review, is established whenever two tools that are supposed to measure the same construct show this similarity. Factor analysis also establishes construct validity by showing whether a cluster of items or subtests that are supposed to be caused by the target constructs indeed covary accordingly. Ecological validity is supported by correlations between tool results and measurements used in everyday practice (60). Earlier reviews of neuropsychological/cognitive tools either considered tools relevant to specific diseases or age groups with only partial relevance to early schoolers (61-64). The tool-specific reviews documented psychometric properties and cultural relevance of different neuropsychological/cognitive tools (62). The current review covers a wide range of neuropsychological and neurocognitive tools and focuses on early schoolers.

STUDY OBJECTIVE

This narrative review looks at developed and adapted neuropsychological tools in papers published between 1997 and 2017 specifically for children aged 6 – 12 years. A narrative review is recommended for a critical discussion of knowledge on a topic of interest with the aim of collating and summarizing study findings on the topic as well as identifying research gaps (65). The aims of this review are to identify and summarize commonly used neuropsychological tools for 6-12-year-olds globally and to document their psychometric properties across different contexts. Specifically, the review aims at answering the following research questions:

1. Which neuropsychological/cognitive tools are commonly used among 6 – 12-year-olds?
2. Which cultural adaptations have been made to these tools?
3. What is the reliability, validity, sensitivity, and specificity of these tools?

METHODS

We identified studies conducted between 1997 – 2017 through a thorough search of PubMed, PsycINFO, and Web of Science using the keywords (i) neuropsychological or neurocognitive with (ii) assessment or tools or tests. The search strategy is detailed in Appendix 2.2.

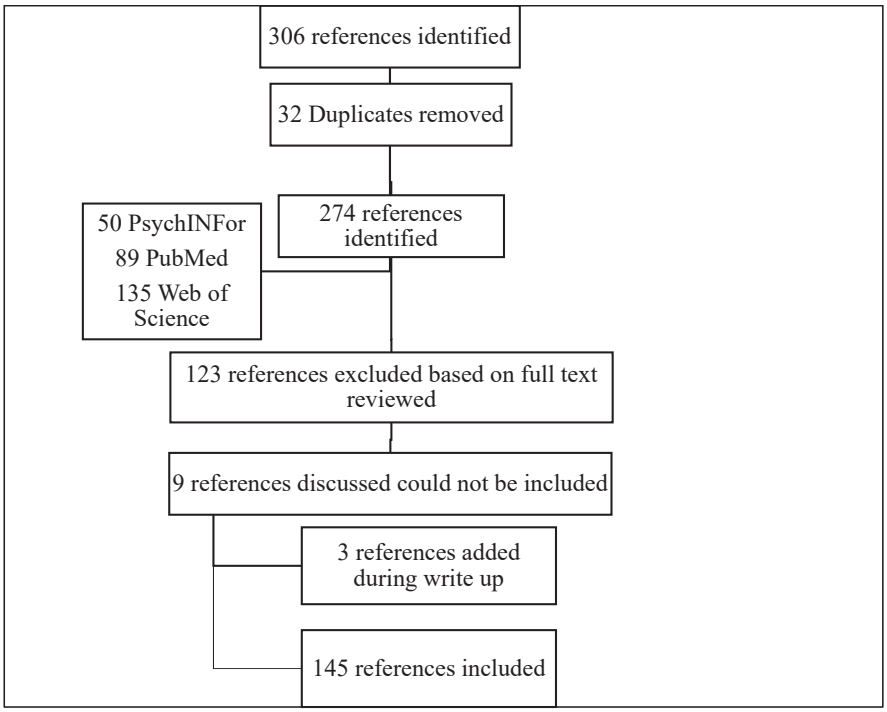
Following this search, we included original studies that examined any information on the adaptation and development of neuropsychological tools among children aged 6 – 12 years globally. RM examined each study using the abstract and title against the exclusion and inclusion criteria and determined whether it should be included in the review. The inclusion criteria were the use of neuropsychological tools, children 6 - 12 years, and English peer-reviewed journal articles published between 1997 – 2017. We also included studies that partially covered the age criteria. Exclusion criteria: studies only including neurophysiological tools, full text missing, non-English publications. We extracted information concerning the type of neuropsychological tool, cognitive domain measured as per DSM-V (executive functions, motor and perceptual-motor, complex attention, language, learning and memory), cognitive domains not recognised in DSM-V (arithmetic, cognitive reserve, intelligence/ intellectual ability, social cognition & skills, representational competence and academic achievement), study country, and type of adaptation and psychometric information reported. The classification of cognitive domains according to the Diagnostic Statistical Manual of Mental Disorders version Five (DSM-V) (44) is shown in Table 2.1 though we included other domains mentioned in the studies and yet to be recognised in DSM-V. RM developed a coding system for the studies that focused on three themes: 1. The neuropsychological/cognitive tool used among 6–12-year-olds, 2. The cultural adaptations made to the tools, and 3. The reliability, validity, sensitivity, and specificity of the tools. This a priori framework suited our line-by-line coding of the findings. When the form of construct validity was not specified as

either discriminant or convergent, we identified it as just construct validity in our review. Where correlations are significant in most or some of the tools, whether between or within studies, we considered this as partial support. RM developed a template of key findings per study on an online spreadsheet and shared it with other co-authors. She received feedback from FV, AA, & KM. There were 12 papers that lacked clarity in their psychometric findings where all the other authors reviewed these papers one by one. Out of these papers, three were selected on the basis that they did have results supporting the tools' validity. Figure 2.1 shows the data extraction flow chart. This information was entered into an online Excel sheet that was accessible to all the authors.

Table 2.1: Classification of Cognitive Domains and Subdomains

DSM V Cognitive Domains	**Sub-Domains**
Learning and Memory	Free recall, cued recall, recognition memory, semantic & autobiographical, long-term memory, implicit learning.
Executive Functions	Planning, decision making, working memory, responding to feedback, inhibition & flexibility.
Complex attention	Divided attention, sustained attention, processing speed and selective attention.
Motor and Perceptual Motor	Visual perception, visual-constructional, reasoning, perceptual-motor & coordination.
Language	Object naming, word finding, fluency, grammar & syntax, & receptive language.
Social Cognition & Skills	Recognition of emotions, insight & theory of mind.
Other Cognitive Domains	**Sub-Domains**
Arithmetic	Calculation, number processing, numeration, geometry, addition, subtraction, measurement, and time and money
Cognitive Reserve	Word reading & vocabulary.
Intelligence/ Intellectual Ability	Intelligence, verbal reasoning, verbal comprehension, perceptual organization & distractibility
Representational Competence	-
Academic Achievement	-

Figure 2.1: Data Extraction Flow Chart



In developing the results summary, we reported tool or subtest-level findings even when the findings for multiple tools were captured in the same publication. We also grouped the studies reporting on the same tool and gave a summary ranking or interpretation of the tool's psychometric findings. Where we found mixed findings, i.e., the tool had good validity outcomes in one study and poor in another, we gave a summary ranking of 0.5 to portray the less-than-optimal psychometric findings. This is as captured in the Table 2.2. We also explored possible explanations for such heterogeneous findings, which were characterized by differences in country setting, sample characteristics, and selective subtest evaluations within test batteries.

Table 2.2: Description of Columns in Table 2.4

Column Name	***Ranking and Interpretation***	
References	The numbering corresponds with the reference list in the OSF store supplementary file https://doi.org/10.17605/OSF.IO/TCYX9).	
Country	Alpha 3 numeric code for countries is used.	
Country level of income	1	HIC
	0	LMIC
	0.5	H&L/UMIC
Validity and reliability outcomes	1	Good
	0	Poor
	0.5	Mixed/partial
Normative data study	1	Yes
	0	No
Types of test/tool domains	1	Memory
	2	Executive functioning
	3	Complex Attention
	4	Motor & Perceptual Motor
	5	Learning
	6	Language
	7	Arithmetic
	8	Cognitive reserve
	9	Intelligence/ Intellectual ability
	10	Social cognition & skills
	11	Representational competence
	12	Academic achievement

Key: HIC – High Income Country (gross national income per capita of more than USD13,205) (66)*; LMIC – Low- and Middle-Income Country (gross national income per capita of USD 4,255 or less); UMIC – Upper-middle Income Country (gross national income per capita of USD4,256 – 13,205); H&L/UMIC – combinations of High and Low- or Upper-Income Countries.*

RESULTS

The search identified 306 potentially relevant papers. In total, 145 papers used neurocognitive or neuropsychological tools among 6–12-year-olds and met the inclusion criteria as indicated in the data extraction flow chart in Figure 2.1 (Also see the OSF link https://doi.org/10.17605/OSF.IO/TCYX9). Most of the papers used multiple tools, with a total of 142 different tools. Twenty-three tools were used in multiple studies. The majority of the

studies were conducted in clinical populations (N = 102). The cognitive domain (44) distribution of tools included 62 for executive functioning; 54 for complex attention; 38 for motor and perceptual motor; 27 for learning; 27 for language; 39 for memory, 2 for arithmetic, 5 for social cognition and skills, 2 for cognitive reserve, 2 for intelligence/intellectual ability, 1 each for representational competence and academic achievement. A tool can be categorized into different domains hence the more than 142 tools reported here. Almost half of the studies were conducted in the United States as shown in Table 2.3. The drawn samples were based on the objectives of the study and the targeted population, which often exhibited cognitive impairment. Thirty-seven papers studied an entirely healthy sample, while thirty-six studies considered a population with a healthy control, and seventy-two studies involved an entirely diseased population depending on the cognitive deficit of interest. Thirty-seven papers considered populations with Attention Deficit Hyperactive Disorder, representing the most (26%) targeted population in the studies.

Table 2.3: Country Distribution of the Extracted Studies

	Total Number of Papers	Detailed Description	N (%)
Countries	145	United States	69 (47.6)
		Canada	11 (7.6)
		Netherlands	6 (4.1)
		Brazil	5 (3.4)
		Australia, Kenya	4 each (5.5 in total)
		Finland, Italy, Spain, United Kingdom	3 each (8.3 in total)
		Taiwan, Colombia, France, Germany, Mexico, Hong Kong, Israel, Korea, Sweden, Uganda, Denmark	2 each (15.2 in total)
		Argentina, Austria, India, Belgium, Cyprus, Japan, Morocco, Portugal, Romania, Iran, Thailand, China	1 each (8.3 in total)

Adaptation Processes in the Reviewed Studies

Eleven papers reported on the development of completely new tools compared to a hundred and thirty-four that adapted and/or tested the psychometric properties of existing tools.

Developing New Tools

Eleven of the included studies (7.6%) developed new tools (67-69). For example, Nieuwenhuijzen et al. (68) developed a social information processing tool because no existing tool measured this cognitive domain. This tool involved using vignettes in combination with cartoons, pictures, and videos which depicted different social situations, and the child was required to respond to different questions like what was happening and how they would respond in a comparable situation. Scoring was developed to evaluate the responses' information processing trajectory within a linear scale. Chevignard et al. (67) also developed a novel open-ended naturalistic task termed the ecological cooking task for evaluating executive functioning. The task consisted of four new recipes and child friendly instructions that were added to two recipes utilized among adults in a similar task.

Adaptation

Adaptation of existing neurocognitive tools involved translation and making iterations to the items. For tools whose adaptation involves translation, it is crucial to ensure that the new versions do not lose the characteristics of the original tool. Because the respondent's language background tends to exert some effect on the tools, most cultural adaptations took language into account (42, 70). Some of the tools did not rely heavily on language hence only the instructions of the new versions had to be translated (41). Bilingual translators were preferred in five studies and a back-translation design was used for the translation in the five studies (40-42, 71). Where two translators would not agree on instruction or stimuli translation, a third one would be integrated as a tiebreaker.

Translation was reportedly done after permission was sought from the original authors (42, 72), though not all studies reported on whether researchers sought permission (46). This, at times, created challenges where the original authors were not willing to give permission for development of a different version, or in situations where they did, permission was partial in that the developers have restricted access to, for example, the tool's stimuli (42). Openly accessible tools are available to low resource settings for adaptation, but some tool developers may be hesitant to give full access due to potential misuse of the tools. Restricted access, such as that in the study by Siqueira et al. (42) or mandating seeking approvals before developing new versions may help curb potential misuse.

After translation of the subtests that formed the Kilifi Toolkit to Kiswahili among 8–11-year-olds in a semi-urban area in Kenya, the authors replaced certain items that were unfamiliar to the respondents with more familiar items (46). In adapting a neurobehavioral tool battery among Thai children, the authors substituted envelopes with paper as well as a hairbrush with a hair clip (70). The later substitution was interesting because of similar pronunciations to a toothbrush. The adaptation of the Child Hayling Test (CHT) among Brazilian children included the exclusive use of nouns instead of a mixture of nouns, adverbs, and adjectives that were used in the adult version of the tool (42). This was done to meet the linguistic preferences of Brazilian children. These forms of changes are integrated into the stimuli and instructions. When adapting the CHT, mental health practitioners, such as psychologists at the postgraduate level, judged whether each item was representative of the cognitive domains that the tool was supposed to measure and whether it would be easily comprehended (42).

Adaptation also involved creating alternative forms of the same tools to reduce practice effects in test-retest reliability measurements. Creating alternate forms was not always successful, like in a study among Thai children that reported low test-retest reliability in tools with alternate forms (70). Comparability of alternate forms may need to be improved to reduce

such effects. In several studies, particular subtests, as opposed to a full neuropsychological battery, were adapted (46, 73, 74).

Shorter versions of existing tools were also adapted for screening purposes (75, 76). For instance, Sadeh et al. (76) investigated the predictive power of the adapted EF screener within the Behaviour Assessment System for Children-Teacher Report (BASC) among 1,840 school going 6–11-year-olds in semi-rural areas in Midwest United States. An EF screener with strong predictive power would be useful in screening for behavioural problems early enough for prevention and intervention purposes.

Pilot studies evaluated the linguistic, semantic, and syntax complexities of the tools as part of adaptation (42, 46, 55, 71, 77) and as part of creating new tools (67). Kitsao-Wekulo et al. (46) did a pilot study for the Kilifi Toolkit to check translation comprehension, familiarity with the items, ceiling and floor effects of the modified scales, and ease of administration and scoring. Pilot studies exuded vital information such as the impact of using examples in helping children understand the instructions (71).

Evaluating Psychometric Properties of Neuropsychological/cognitive Tools

Validity and reliability estimates were evaluated for the tools in 141 papers depending on the objectives of the study in relation to the tool. Before we present the psychometric outcome of the various tools, we highlight the type of methods used in assessing validity and reliability.

Test-retest reliability was assessed using intraclass correlation (ICC) which is most useful when having more than two repeated measures, while internal consistency (the extent to which items hang together) was typically evaluated using Cronbach Alpha.

(M)ANOVA has been used as a descriptive tool in studies creating norms for tools where the effects of age and gender are evaluated (78, 79). Multiple regression analysis gives

a clearer picture of associations by removing confounding effects among other factors that influence outcomes. However, regression based on observed (sum) scores does not remove measurement errors.

Factor analyses with latent variables do handle these errors and correct the associations for effects of random measurement errors. Structural Equation Modelling (SEM) was used to accurately estimate associations. For instance, Budtz-Jorgensen et al. (78) used SEM to produce a factorial model for estimating the association between neuropsychological scores and biomarkers of prenatal mercury exposure while adjusting for measurement error, confounding factors, and missing data, among other validity concerns. Confirmatory factor analysis was used in the studies that assessed the tool's construct validity or assessed how well the factor structure fitted the item and test data (80). A model is considered to fit well when the exact fit test is non-significant, when the root mean square error of approximation (RMSEA) is less than .08, the comparative fit index (CFI) exceeds .90 and Tucker Lewis Index (TLI) exceeds .90 (80, 81). The tools' internal structure was often studied with principal component analysis or exploratory factor analysis with eigenvalues and other calculations used to evaluate the number of factors (82).

Construct validity has also been assessed by identifying group differences between diseased and healthy samples based on their cognitive performance on the tools (21). Discriminant validity, or a tool's ability to not measure constructs it is not supposed to measure, and convergent validity, or a tool's correlation with alternative measures of the same targeted construct, together support construct validity (80, 83). Convergent validity, where the level of agreement between two or more tools is evaluated, was measured using Pearson's correlation coefficient. Criterion validity has been used to further evaluate a tool's external structure when there is a "gold standard" that needs to be predicted with a tool/measure (80). In the study by

Woodward and Donders (80), the Memory Screening Index (MSI) was found to be equally as sensitive to severity of injuries as CT or MRI variables, the "gold standard."

Receiver Operating Characteristics (ROC) were used in studies to assess the sensitivity and specificity of tools i.e., the tool's scoring ability in differentiating those with cognitive impairment from those without (84). Area Under the Curves (AUCs) has also been used with ROC to assess for group differences in the diagnostic context. An AUC of .80 and above indicates good classification, which supports predictive validity. Higher sensitivity and specificity are predictive of the best cut-off points/scores according to categories when diagnosing impairment in children. However, assessment is broader and covers outcomes of continuous scores that are not separated into categories. An evaluation of the Test of Memory and Learning (TOMAL) indicated that a cut-off point of .80 indicated the best sensitivity and specificity combination (sensitivity .70, specificity .62) (84).

In a study of influences that improve or change test scores in repeated tests, practice effects were determined by calculating paired T-tests (85), while another similar study calculated percentage change and reliability change indexes to study change (50). Reliability change indexes consider measurement unreliability in assessing change.

Psychometric properties of tools across different cognitive domains

The psychometric results of different tools are outlined and organized into the cognitive domains that the tools measure. We summarize results in the main text and provide detailed information on the countries where the studies were conducted and the specific psychometric outcomes (including any specific psychometric checks and the reported statistics) in the OSF link https://doi.org/10.17605/OSF.IO/TCYX9).

In the executive function domain, The Behaviour Rating Inventory of Executive Functioning (BRIEF) had most studies offering psychometric information (N=7). The BRIEF passed validation checks, though we did not encounter any reliability studies in this review. Included studies with the Wechsler Intelligence Scale for Children Third and Fourth Edition (WISC-III and IV) reported good validity though reliability indicators varied regarding subsets under study. A study on the Digit Span subtest of the WISC-III found the tool to have low test-retest reliability when used in a clinical population of 437 children with carious lesion (tooth problems) followed up over five years.

Both the Medical Symptom Validity Test (MSVT) (N=6) and the Test of Memory Malingering (TOMM & TOMM 2) (N=5) and the Word Memory Test (WMT) (N=5) appeared commonly in the reviewed papers. Results concerning the validity, specificity, and sensitivity of the TOMM were mixed, while the other two tools showed high validity.

Under the complex attention tools, Continuous Performance Tests (CPT) and its revisions were most commonly used (N=7) followed by CANTAB (N=4). Different studies found differing psychometric information outcomes, as indicated in Table 2.4. CANTAB's construct validity was established, although its subtests, spatial working memory (SWM), showed low discriminant validity in a study among 54 children living with combined ADHD and concurrent vestibular impairment in Tehran, Iran. CANTAB's test-retest reliability was also found to be low among 64 healthy children in Scotland.

Six studies looked at the Developmental Test of Visual-Motor Integration psychometric indicators in the Motor and Perceptual Motor domain. These studies had differing findings when it came to discriminant validity and test-retest reliability. Construct validity of the Developmental Test of Visual-Motor Integration was supported, but two studies could not agree on the discriminant validity of the tool, as one reported the discriminant validity to be poor.

Cogstate Battery, WISC-IV, Differential Ability Scales (DAS), and NEPSY were the most frequently studied tools (Count $\geq$ 2) in the learning domain. The construct validity of the Cogstate tool was supported in two studies among 230 healthy children and those with cerebral malaria in Uganda and 87 healthy children in Australia. However, the two studies found its test-retest reliability to range from weak/low to strong. Similarly, the reliability of NEPSY was not clearly supported in two studies in the USA among 204 children made up of healthy children and those with neurological conditions and scholastic concerns.

In the language domain, we found many different tools featured in only one study, while the Developmental Neuropsychological Assessment (NEPSY) was used in two studies. The construct validity outcomes were good in neuropsychological batteries such as BENCI, Halstead-Reitan Neuropsychological Test Battery for Children (HRNB-C), and Luria-Nebraska Test for Children (TLN-C, in Portuguese), while internal consistency was supported for the HRNB-C and TLN-C. One study supported the discriminant validity of the NEPSY, but overall, the evidence for construct validity and test-retest reliability of this tool was mixed. The WISC-IV vocabulary test was found to have low validity among 104 children in Canada living with Epilepsy. Seashore Rhythm Test exhibited low internal consistency in a sample of 334 children in the USA living with specific learning disability (LD), severe emotional disturbance (ED), speech handicapped (SH), attention deficit hyperactivity disorder and other health impairment (OHI) and with diagnosis not specified. Most of the studies on language tools reported validity outcomes (N=21), while fewer studies in this domain reported reliability outcomes (N=12).

In the cognitive reserve domain, one study found non-significant correlations between the cognitive reserve subtest within WIAT-II and short-term (less than six months) neuropsychological tools outcomes within the paediatric population. Tools used for social

cognition were found to be valid, including interesting tools such as cartoons, pictures, and video vignettes.

Table 2.4: Tools with Psychometric Information Among 6 – 12-Year Olds

Tools	Reference	Country	Country Level of Income	Construct validity	Convergent validity	Discriminant validity	Test-retest reliability	Internal consistency	Sensitivity	Specificity	Normative Data	Types of tests domain
Distribution statistics			**1**	**80.68 %**	**65.22 %**	**66.33 %**	**52.22 %**	**73.81 %**	**71.15 %**	**74%**	**5**	**2**
Tower of Hanoi Test	1, 11	FIN, GBR, USA	1	1	NR	NR	0.5	NR	NR	NR	NR	2
Tower of London	134	ITA	1	1	NR	NR	NR	NR	NR	NR	NR	2
Storytelling performance measure of EF	3	USA	1	NR	NR	NR	NR	1	NR	NR	NR	2
Self-Ordered Pointing (SOP)	4	CAN	1	NR	NR	NR	1	NR	NR	NR	1	2
A standard Stroop (Golden Version); Sun-Moon Stroop & Fruit Stroop	4	CAN	1	NR	NR	NR	1	NR	NR	NR	1	2
Cogstate battery	7, 90	UGA, AUS	0.5	1	1	NR	1	NR	NR	NR	NR	2, 3, 4, 5
Children's Kitchen Task Assessment (CKTA)	116,8	USA	1	NR	0.5	0.5	NR	0.5	NR	NR	NR	2
Five to Fifteen parent questionnaire (FTF)	137	DNK	1	NR	0.5	0.5	NR	1	NR	NR	NR	2, 1, 4, 5, 6, 10
Wisconsin Card Sorting Test (categories, failure to maintain set, total errors)	17	USA	1	NR	NR	NR	0	NR	NR	NR	NR	2
Delis-Kaplan Executive Function System (D-KEFS) (Trail Making – visual scanning, number sequencing, motor speed, total errors, Verbal Fluency – set loss errors, repetition errors, Tower Test – rule violation/item ratio)	17	USA	1	NR	NR	NR	0	NR	NR	NR	NR	2
Children's Cooking Task (CCT)	22	AUS	1	NR	0.5	1	1	1	NR	NR	NR	2
The ecological 'cooking task'	23	FRA	1	NR	NR	1	NR	NR	NR	NR	NR	2
Trail-Making Test (TMT)	24	CHN, TWN	0	NR	NR	0.5	NR	NR	NR	NR	NR	2, 1
Digit span.	24, 118	CHN, TWN, BRA	0	0	NR	0.5	NR	NR	NR	NR	NR	2, 3
Korean Educational Development Institute-Wechsler Intelligence Scales (KEDI-WISC) (Subtests include Continuous Performance Test (CPT), Children's Color Trails Test (CCTT) and Stroop Color-Word Test (SCWT))	25	KOR	1	NR	0.5	NR	NR	NR	NR	NR	NR	2, 3, 5

Tools	Reference	Country	Country Level of Income	Construct validity	Convergent validity	Discriminant validity	Test-retest reliability	Internal consistency	Sensitivity	Specificity	Normative Data	Types of tests domain
Amsterdam Neuropsychological Tasks (ANT) subtests: - Baseline speed, Focused attention four letters, Shifting attentional set–visual (measures vigilance, inhibition, and cognitive flexibility), and Sustained attention.	27	ITA	1	NR	NR	0.5	NR	NR	0.5	0.5	NR	2, 3
Behavior Rating Inventory of Executive Functioning (BRIEF)	29, 57, 82, 84, 142	CAN, USA	1	0.5	0.5	1	NR	NR	NR	NR	NR	2
Luria-Nebraska Test for Children (TLN-C, in Portuguese)	30	USA	1	1	1	NR	NR	1	NR	NR	NR	2, 4, 6
FAS Verbal Fluency Test	32	BRA	0	0.5	NR	NR	NR	NR	NR	NR	NR	2, 3, 6
Arizona Cognitive Test Battery (ACTB)	34	USA	1	NR	NR	NR	0.5	NR	NR	NR	NR	2, 3
Batería de Evaluación Neuropsicológica Infantil (BENCI)	38	MAR	0	1	NR	1	0.5	NR	NR	NR	NR	2, 1, 3, 4, 6
The Cambridge Neuropsychological Test Automated Battery (CANTAB) - Subsets include Pattern recognition memory (PMR), Spatial recognition memory SRM. Spatial span (SSP), Stockings of Cambridge (SOC). Intra-extra dimensional set shift (IED). Reaction time (RTI). Rapid visual information processing (RVP).	133	FIN	1	1	NR	NR	NR	0.5	NR	NR	NR	2, 1, 3
n-back	40	ESP	1	1	NR	NR	NR	1	NR	NR	NR	2
Wechsler Intelligence Scale for Children v 3 (WISC–III)	134	ITA	1	1	NR	NR	NR	NR	NR	NR	NR	2
Wechsler Intelligence Scale for Children v 3 (WISC–III) Symbol Search subtest	87, 128	USA, PRT	1	NR	0.5	NR	0.5	NR	NR	NR	NR	2
Wechsler Intelligence Scale for Children v 3 (WISC–III) Coding subtest	87, 128	USA, PRT	1	NR	0.5	NR	0.5	NR	NR	NR	NR	2
Wechsler Intelligence Scale for Children v 3 (WISC–III) Digit Span subtest	128	PRT	1	NR	NR	NR	0.5	NR	NR	NR	NR	2
Wechsler Intelligence Scale for Children-Fourth Edition (WISC-IV)	17	USA	1	NR	NR	NR	0.5	NR	NR	NR	NR	2, 3
Wechsler Intelligence Scale for Children Fourth Edition (WISC-IV)- General Ability Index (GAI), Full Scale IQ (FSIQ) and	58, 53	USA, CAN	1	0.5	NR	NR	NR	NR	1	NR	NR	2, 3, 5

Tools	Reference	Country	Country Level of Income	Construct validity	Convergent validity	Discriminant validity	Test-retest reliability	Internal consistency	Sensitivity	Specificity	Normative Data	Types of tests domain
Cognitive Proficiency Index (CPI)												
Children's Category Test – Level 2 (CCT-2)	54	USA	1	1	0.5	0.5	NR	NR	0.5	NR	NR	2, 1, 5
Japanese short form of the Swanson Cognitive Processing Test. Swanson Cognitive Processing Test	55	JPN	1	NR	0.5	NR	1	NR	NR	NR	NR	2
Reynolds Intellectual Assessment Scale (RIAS) - subtests include: - Composite Intelligence Index (CIX), Nonverbal Intelligence Index (NIX) and Verbal Intelligence Index (VIX).	56	CAN	1	0.5	NR	NR	NR	NR	NR	NR	NR	2, 9
The Children's Executive Functions (CEFS)	57	USA	1	NR	0.5	NR	NR	NR	NR	NR	NR	2
Behavioral screener for the assessment of executive functions version 2 (BASC-2-EF) screener	59	CAN	1	1	NR	NR	NR	1	NR	NR	NR	2
EF scale from the Behavior Assessment System for Children-Teacher Report	59	CAN	1	1	NR	NR	NR	NR	0.5	0.5	NR	2, 3, 10
Testbatterie zur Aufmerksamkeitsprufung fur Kinder (KITAP)	61	AUT	1	NR	NR	0.5	NR	NR	NR	NR	NR	2, 3
clock test (clock drawing test, clock face test)	66	NR	NR	NR	NR	0.5	NR	NR	NR	NR	NR	2
Brief neurocognitive screener (DIVERGT) - subtests Digit Span Test, The Verbal Fluency Test, The Grooved Pegboard Test and The Trail Making Test.	68	USA	1	NR	NR	1	1	NR	0.5	0.5	NR	2, 3, 4
Korean Computerized Neurobehavioral Tests (KCNT) - subtests include Simple Reaction Time (response speed), Choice Reaction Time (psychomotor speed), Color Word Vigilance (attention), Addition (executive functions), Symbol Digit (executive functions), and Finger Tapping Speed (manual dexterity).	72	KOR	1	NR	NR	NR	0.5	NR	NR	NR	NR	2, 3, 4
Halstead-Reitan Neuropsychological Test Battery for Older Children (HRNB-C)	75	USA	1	0.5	NR	NR	NR	1	NR	NR	NR	2, 3, 4, 6

Tools	Reference	Country	Country Level of Income	Construct validity	Convergent validity	Discriminant validity	Test-retest reliability	Internal consistency	Sensitivity	Specificity	Normative Data	Types of tests domain
Halstead-Reitan Neuropsychological Test Battery for Children (HRNB-C)	76	USA	1	1	NR	NR	NR	NR	NR	NR	NR	2, 3, 4, 6
Halstead–Reitan Neuropsychological Test Battery - Trail Making Test	110	USA	1	NR	NR	1	NR	NR	1	NR	NR	2, 3
Kaufman Assessment Battery for Children, second edition (KABC-II)	6, 52, 83	UGA, KEN, IND	0	1	NR	NR	0.5	NR	0.5	NR	NR	2, 1, 3, 4
Online version of IMPACT	91	USA	1	NR	NR	NR	0.5	NR	NR	NR	NR	2, 3
Pediatric ImPACT	95	USA	1	NR	0.5	0.5	NR	NR	NR	NR	NR	2, 1. 3
The Cognitive Assessment System (CAS) - subtest Planned Codes	46	USA	1	NR	1	NR	NR	NR	NR	NR	NR	2
Immediate Post-concussion Assessment and Cognitive Testing (ImPACT)	111	USA	1	NR	NR	NR	NR	NR	NR	NR	1	2, 1, 3, 5
Omnibus test of cognitive functioning; Trail Making A (attention), Continuous Performance Task (CPT) (attention)]; Trail Making B (Executive Function); Cog Set Shifting (Executive Function), Controlled Oral Word Association Test (COWAT) (Executive Function); Digit Span (Working Memory), Spatial Span (Working Memory), and California Verbal Learning Test (CVLT) (Verbal Memory)	98	USA	1	NR	NR	1	NR	NR	NR	NR	NR	2, 6
Timo's Adventure	99	NLD	1	NR	NR	1	NR	NR	1	1	NR	2, 6
Combination of Kaufman Hand Movements Scale; The Stroop Color-Word Association Test (Stroop); The Controlled Oral Word Association Test (COWAT); Trail Making Test; Arithmetic and Digit Span subtests of the Wechsler Intelligence Scale for Children, Third Edition (WISC-III; Conners' Continuous Performance Test, (CPT)	103	USA	1	NR	NR	0	NR	NR	0.5	0.5	NR	2, 3, 4, 6

Tools	Reference	Country	Country Level of Income	Construct validity	Convergent validity	Discriminant validity	Test-retest reliability	Internal consistency	Sensitivity	Specificity	Normative Data	Types of tests domain
Neuropsychological Battery: subtests Mental Control; Target Detection Cancellation Test; Visual-Verbal Learning Curve; Rey-Osterrieth Complex Figure Test; Language Comprehension and Working Memory test; Language Fluency test; Wisconsin Card Sorting Test-Abbreviated Version (WCST-A)	105	COL	0	1	NR	0	NR	NR	0.5	0.5	NR	2, 1, 3, 4, 6
Lebby-Asbell Neurocognitive Screening Examination—Children and Adolescent versions (LANSE-C/A)	107	USA	1	NR	0.5	0	NR	0.5	NR	NR	NR	2, 1, 3, 4, 5, 6
Pediatric Attention Disorders Diagnostic Screener (PADDS)	109	USA	1	NR	1	NR	NR	NR	NR	NR	NR	2, 3
Swanson, Nolan, and Pelham Questionnaire (SNAP-IV scale)	118	BRA	0	NR	NR	0	NR	NR	NR	NR	NR	2, 1, 3
Behavioral Assessment of Dysexecutive Syndrome for Children (BADS-C) (Subtests: Playing Cards test, Water test, Key search test, Zoo map tests, Six parts test)	120, 127	FRA, HKG	1	NR	0.5	NR	NR	NR	NR	NR	NR	2
Developmental Neuropsychological Assessment (NEPSY)	131, 123	USA	1	0.5	NR	1	0.5	NR	NR	0.5	NR	2, 1, 3, 4, 5, 6
Groton Maze Learning Task (GMLT)	136	AUS	1	0.5	NR	NR	NR	NR	NR	NR	NR	2, 1
Child Hayling Test (CHT)	126	BRA	0	NR	NR	NR	NR	NR	0.5	NR	NR	2
The Corsi test	136	AUS	1	1	NR	NR	NR	NR	NR	NR	NR	2
A Maze task	125	BRA	0	NR	0.5	1	NR	NR	NR	NR	NR	2, 3, 5
California Verbal Learning Test, Children's Version CVLT-C	5, 92, 16	USA, DNK	1	1	NR	NR	NR	1	0.5	0.5	NR	1
QS4-G: Parent Questionnaire for the Developmental Evaluation of 4-Year-Old	10	ITA	1	NR	NR	NR	NR	NR	0.5	1	NR	1, 3, 4, 6
Test of Memory and Learning (TOMAL)	135	USA	1	1	0.5	1	NR	NR	0.5	0.5	NR	1, 3, 5
Word Completion Memory Test (WCMT)	114	NLD	1	NR	0.5	NR	NR	NR	NR	1	NR	1
The Test of Memory Malingering (TOMM); TOMM trial 2	79, 80, 101, 108, 114, 12	USA, NLD, DEU	1	NR	0.5	NR	NR	NR	0.5	0.5	NR	1
Medical Symptom Validity Test (MSVT)	12, 19, 20, 44, 50, 62	USA, CAN	1	NR	1	1	NR	NR	1	1	NR	1
Rey's Fifteen Item Test (FIT)	12	DEU	1	NR	0.5	NR	NR	NR	NR	NR	NR	1

Tools	Reference	Country	Country Level of Income	Construct validity	Convergent validity	Discriminant validity	Test-retest reliability	Internal consistency	Sensitivity	Specificity	Normative Data	Types of tests domain
Word Memory Test (WMT)	20, 44, 50, 114	CAN, NLD	1	NR	0.5	NR	NR	NR	NR	NR	NR	1
Nonverbal Medical Symptom (NV-MSVT)	44	CAN	1	NR	NR	NR	NR	NR	NR	0.5	NR	1
Memory Screening Index (MSI) from the WRAML (Wide Range Assessment of Memory and Learning)	144	USA	1	1	NR	1	NR	NR	1	NR	NR	1, 5
Rey's Auditory-Verbal Learning Test (AVLT)	138	NLD	1	NR	NR	NR	0.5	NR	NR	NR	NR	1, 5
Children's Memory Scale	17	USA	1	NR	NR	NR	0	NR	NR	NR	NR	1
Word List Delayed Recognition	17	USA	1	NR	NR	NR	0	NR	NR	NR	NR	1
Amsterdam Short-Term Memory (ASTM)	114	NLD	1	NR	0.5	NR	NR	NR	NR	1	NR	1
Cambridge Neuropsychological Test Battery (CANTAB)	39, 73, 119	GBR, IRN, USA	1	1	NR	0.5	0	NR	NR	NR	NR	1, 3, 11
WISC-IV Digit Span subtest	102	USA	1	NR	NR	NR	NR	NR	1	1	NR	1
Differential Ability Scales (DAS). Differential Ability Scales - Second Edition (DAS II)	42, 48, 36	USA	1	NR	NR	1	NR	NR	NR	1	NR	1, 5
CNS Vital Signs (CNSVS) - subtests: - verbal and visual memory, finger tapping, symbol digit coding, the Stroop Test, a test of shifting attention and the continuous performance test	45	USA	1	NR	0.5	1	0.5	NR	NR	NR	NR	1, 3
Kilifi Toolkit - Subtests include: - Tower Test, Self-Ordered Pointing Test, Verbal List Learning, Colored Progressive Matrices, Dots, Contingency Naming Test, Score, People Search,	64	KEN	0	NR	NR	NR	0.5	0.5	NR	NR	NR	2, 1, 3
Perceived cognitive function (PCF)	69	USA	1	NR	NR	1	NR	NR	NR	NR	NR	1, 3
Autism/Tics, AD/HD, and other Comorbidities (A&TAC) inventory	70	SWE	1	NR	NR	NR	NR	NR	NR	NR	NR	1, 3, 4, 5, 6, 10
Standardized Assessment of Concussion (SAC)	87	USA	1	NR	0.5	NR	0.5	NR	NR	NR	NR	1, 3
Ten Questions' Questionnaire (TQQ)	93	KEN	0	NR	NR	NR	0.5	NR	1	1	NR	1, 4
Parent Report Child Behavioral Checklist (CBCL)	97	USA	1	NR	NR	NR	NR	NR	1	1	NR	3
CMS Delayed Verbal Recall>Delayed Recognition memory subtests	102	USA	1	NR	NR	NR	NR	NR	1	1	NR	1, 3

Tools	Reference	Country	Country Level of Income	Construct validity	Convergent validity	Discriminant validity	Test-retest reliability	Internal consistency	Sensitivity	Specificity	Normative Data	Types of tests domain
Behavioral Assessment and Research System (BARS) (included tests of motor speed and dexterity, attention, memory, and visuospatial coordination)	117	THA	0	NR	NR	0.5	0.5	NR	NR	NR	NR	1, 3, 4
Continuous Performance Tests (CPT). MOXO-CPT. AULA CPT. Conners' Continuous Performance Test (CCPT). computerized Corner's continuous performance test (CPT) – Second Edition	9, 24, 35, 37, 103, 145, 28	ISR, CHN, TWN, USA, ESP	0.5	NR	1	0.5	0.5	NR	0.5	0.5	NR	3
Gordon Diagnostic System (GDS)	86	USA	1	1	NR	NR	NR	NR	0.5	0.5	NR	3
NIH Toolbox Pattern Comparison Processing Speed Test	18	USA		NR	0.5	0.5	NR	NR	NR	NR	NR	3
Cancellation Test	24	CHN, TWN	0	NR	NR	0.5	NR	NR	NR	NR	NR	3
Circle-Tracing Task.	24	CHN, TWN	0	NR	NR	0.5	NR	NR	NR	NR	NR	3
Continuous Attention Test for Children (CAT)	26	COL	0	NR	0	0.5	NR	NR	NR	NR	NR	3
Attentional Network Test (ANT)	40, 39, 43, 47, 129	ESP, GBR, USA, DEU, USA	1	1	1	0.5	0.5	0	NR	NR	NR	3
Wechsler Intelligence Scale for Children Freedom-from-Distractibility/Working Memory Index (FDI/WMI) and Processing Speed Index (PSI) (both subtests contribute towards FSIQ))	86	USA	1	1	NR	NR	NR	NR	NR	NR	NR	3
10 Wechsler Intelligence Scale for Children-Third Edition (WISC-III) subtests and 4 Wechsler Individual Achievement Test (WIAT) subtests	143	USA	1	NR	0.5	NR	NR	NR	NR	NR	NR	3, 4, 6
Children's Color Trails Test (CCTT), 1 2 CCTT	67, 89	CYP HKG	1	1	0.5	NR	0.5	NR	NR	NR	1	3
Test of Variables of Attention (TOVA)	77	USA	1	NR	NR	NR	NR	0.5	NR	NR	NR	3
Trail Making Test B (Trails B)	87	USA	1	NR	0.5	NR	0.5	NR	NR	NR	NR	3
Clinical virtual reality VR/ Classroom-CPT (VC) (attention)	94	ROU	1	NR	NR	1	NR	NR	NR	NR	NR	3
Go/No-Go	125	BRA	0	NR	1	0.5	NR	NR	NR	NR	NR	2, 3, 5

Tools	Reference	Country	Country Level of Income	Construct validity	Convergent validity	Discriminant validity	Test-retest reliability	Internal consistency	Sensitivity	Specificity	Normative Data	Types of tests domain
DiViSA - Discriminación Simple de Árboles/ Simple Tree Discrimination Test	122	ESP	1	NR	NR	1	NR	1	1	1	NR	3
Developmental Test of Visual-Motor Integration; Beery-Buktenica Developmental Test of Visual Motor Integration test. Beery Developmental Test of Visual-Motor Integration-Third Revision. Beery Visual-Motor Integration (VMI) Test	132, 2, 13, 17, 31	USA, FIN	1	1	1	0.5	0.5	NR	0.5	0.5	NR	4
Purdue Pegboard	13	USA	1	1	NR	NR	NR	NR	NR	NR	NR	4
Pegboard with the dominant (PegsDom) and non-dominant (PegsND) hands	128	PRT	1	NR	NR	NR	0.5	NR	NR	NR	NR	4
Matching Figures from the WRAVMA (Wide Range Assessment of Visual Motor Abilities)	128	PRT	1	NR	NR	NR	0.5	NR	NR	NR	NR	4
Visual Learning from the WRAML (Wide Range Assessment of Memory and Learning)	128	PRT	1	NR	NR	NR	0.5	NR	NR	NR	NR	4
Finger Windows from the WRAML (Wide Range Assessment of Memory and Learning)	128	PRT	1	NR	NR	NR	0.5	NR	NR	NR	NR	4
Rey-Osterrieth Complex Figure Task (RCFT)	13, 31	USA	1	NR	1	NR	NR	NR	NR	NR	NR	4
IT - Inspection time (speed of visualization measure)	60	USA	1	NR	NR	1	NR	0.5	NR	NR	NR	4
Pediatric Stroke Outcome Measure (PSOM)	63	CAN	1	0.5	NR	NR	NR	NR	NR	NR	NR	4, 6
Reality Monitoring (RM)	71	SWE	1	NR	NR	0.5	NR	NR	NR	NR	NR	4
The Bruininks-Oseretsky Test of Motor Proficiency, Second Edition (BOT-2)	81	AUS	1	NR	NR	NR	0.5	NR	NR	NR	NR	4
Bruininks-Oseretsky Test of Motor Proficiency - SF (BOTMP-SF)	130	CAN	1	0.5	1	NR	NR	NR	NR	NR	NR	4
The Movement Assessment Battery for Children (M-ABC)	130, 140	CAN, BEL	1	0.5	0.5	NR	NR	NR	NR	NR	NR	4
Touwen examination	104	NLD	1	NR	NR	NR	0.5	NR	NR	NR	NR	4
Conjunction Visual Search - CVS	113	ARG	0	1	1	NR	NR	1	NR	NR	NR	4
Assessment of Motor and Process Skills (AMPS)	124	USA	1	1	NR	NR	NR	NR	NR	NR	NR	4
Dean-Woodcock Sensory-Motor Battery (DWSMB)	141	USA	1	NR	NR	1	NR	NR	NR	NR	NR	4

Tools	Reference	Country	Country Level of Income	Construct validity	Convergent validity	Discriminant validity	Test-retest reliability	Internal consistency	Sensitivity	Specificity	Normative Data	Types of tests domain
Rorschach Performance Assessment System	88	USA	1	NR	0.5	NR	NR	NR	NR	NR	NR	4
Test of Visual Perceptual Skills – Third Edition (TVPS) (Visual Discrimination, Visual Memory, Visual Spatial Relationships).	17	USA	1	NR	NR	NR	0	NR	NR	NR	NR	4
A brief computerized test, incorporated into the Discrete Trial Trainer ©	49	USA	1	NR	0.5	NR	1	NR	1	NR	NR	5
Internet based measures:- Peabody Individual Achievement Test (PIAT); GOAL Formative Assessment in Literacy for Key Stage 3; Woodcock-Johnson III Reading Fluency Test; Language tests Listening Grammar, Figurative Language and Making Inferences; Items from National Foundation for Educational Research 5-14 Mathematics Series; General cognitive ability was measured using WISC-III-PI Multiple Choice Information (General Knowledge) and Vocabulary Multiple Choice subtests for verbal measures and for nonverbal measures WISC-III-UK Picture Completion and Raven's Standard Progressive Matrices. The Spatial Reasoning series.	51	GBR	1	NR	1	NR	NR	1	NR	NR	NR	5, 6
Expressive One-Word Picture Vocabulary Test - Revised	17	USA	1	NR	NR	NR	0.5	NR	NR	NR	NR	6
Wechsler Intelligence Scale for Children-Fourth Edition (WISC-IV) Vocabulary subtest	41	CAN	1	0.5	NR	NR	NR	NR	NR	NR	NR	6, 8
Seashore Rhythm Test (SRT)	74	USA	1	NR	NR	NR	NR	0.5	NR	NR	NR	6
Two forms of the Speech Sounds Perception Test (SSPT)	74	USA	1	NR	NR	NR	NR	0.5	NR	NR	NR	6
Aphasia Screening Test (AST)	74	USA	1	NR	NR	NR	NR	0.5	NR	NR	NR	6
Evaluación Neuropsicológica Infantil (ENI)	85	MEX	0	NR	1	NR	NR	NR	NR	NR	NR	6
Buschke Selective Reminding Test (SRT)	87	USA	1	NR	0.5	NR	0.5	NR	NR	NR	NR	6
Woodcock Reading Mastery Test	96	USA	1	NR	1	NR	NR	NR	NR	NR	NR	6
Revised Token Test (RTT)	106	MEX	0	NR	NR	1	NR	NR	NR	NR	NR	6
Zareki-R. Arithmetic subtest of WISC-III	33	BRA	0	0.5	NR	NR	NR	NR	NR	NR	1	7

Tools	Reference	Country	Country Level of Income	Construct validity	Convergent validity	Discriminant validity	Test-retest reliability	Internal consistency	Sensitivity	Specificity	Normative Data	Types of tests domain
Key Math-Revised Inventory (KM-R)	112	USA	1	1	NR	0.5	NR	NR	NR	NR	NR	7
Wechsler Individual Achievement Test-Second Edition (WIAT-II) reading subtest (measured Cognitive reserve)	41	CAN	1	0.5	NR	NR	NR	NR	NR	NR	NR	6, 8
Human figure drawings (Matching Familiar Figure Test) - two drawings were used: person and house, tree, and person	100	ISR	1	NR	NR	0.5	NR	NR	NR	NR	NR	10
Cartoons, pictures, and video vignettes	139	NLD	1	NR	NR	0.5	NR	NR	NR	NR	NR	10
Woodcock Johnson III Tests of Achievement.	17	USA	1	NR	NR	NR	0.5	NR	NR	NR	NR	12
WISC-RN (the Dutch version of the WISC-R)	115	NLD	1	0	NR	0	NR	NR	NR	1	NR	9
Bolt Board, Pegboard and Bead Threading Tests	65	KEN	0	1	NR	NR	1	NR	NR	NR	NR	4

Key: Reference numbers can be traced to the OSF store supplementary file https://doi.org/10.17605/OSF.IO/TCYX9; Country name is coded accorded to alpha-3 country codes; Country level of income is coded as 1 for High Income Country, 0.5 for High and Upper/Middle Income Country and 0 for Upper/Middle Income Country; Validity and reliability outcomes are coded as 1 for good, 0.5 for mixed findings and 0 for poor; Presence of Normative data evaluation is coded as 1 for Yes and 0 for No; Type of test/tool domain is coded as 1 for Memory, 2 for Executive functioning, 3 for Complex attention, 4 for Motor & Perceptual Motor, 5 for Learning, 6 for Language, 7 for Arithmetic, 8 for Cognitive reserve, 9 for Intelligence/ Intellectual ability, 10 for Social cognition & skills, 11 for Representational competence and 12 for Academic achievement.

Tools Tested in LMICs, including Sub-Saharan Africa (SSA)

Sub-Saharan Africa

Six studies from SSA, four in Kenya and two in Uganda, were included. In Uganda, the authors tested construct, concurrent, and convergent validity, and test–retest reliability for the computerized, self-administered Cogstate battery and the construct validity of the KABC-II (41, 86). Moderate test-retest correlations were found, while good convergent validity correlations were found with tools such as KABC-II and TOVA. In Kenya, on the other hand, internal consistency was tested for Tower Test (planning), Self-Ordered Pointing Test (SOPT; verbal/visual selective reminding), Verbal List Learning (VLL; working memory), Coloured

Progressive Matrices (CPM; reasoning), Dots (nonverbal memory), Contingency Naming Test (CNT; attention and attention shift, Score (auditory sustained and selective attention), and People Search (visual sustained and selective attention) (46). Test-retest reliability for immediate memory span and CNT was found to be below acceptable levels, while the other subtests had marginally to acceptable reliability. Internal consistency results ranged from .70 to .84. The sensitivity, specificity, and test-retest reliability of the Ten Questions Questionnaire, which measures perceptual-motor skills and memory, was also tested among 6 – 9-year-old Kenyan children (87). Test-retest reliability was found to be excellent for motor, vision, speech, and four cognition questions, while specificity and sensitivity rates were greater than 70% and 96%, respectively.

LMIC vs. HIC

In Table 2.4, the psychometric properties of the Cogstate battery and Continuous Performance Tests (CPT) were tested in both LMICs and HICs. In both settings (Uganda-LMIC and USA - HIC), the Cogstate battery showed good psychometric outcomes in terms of construct validity, convergent validity, and test-retest reliability. As for the CPT, test-retest reliability and sensitivity and specificity were only tested in a HIC (USA) and were moderate, while convergent validity (tested in Spain) was supported. There were mixed findings in CPT's discriminant validity in both HIC (Israel and USA) and LMICs (China and Taiwan).

Other Findings

Four studies sought to gather age-related norms for the neuropsychological tools. Archibald and Kerns (88) collected normative data for newly modified tools of executive function i.e., a standard Stroop (Golden Version), Sun-Moon Stroop & Fruit Stroop and modified Self-Ordered Pointing (SOP) among 7-12-year-olds in Victoria, Canada. Dos et al. (89) collected normative data for 7-12-year-old Portuguese-speaking children using the Zareki-R battery. In contrast, Konstantopoulos et al. (90) collected normative data for 7–16-year-old Cypriot

children using the Children's Colour Trails Test (CCTT). Reynolds et al. (79) provided normative data on each subtest used to calculate the composite scores of ImPACT among 10-12-year-olds in the USA. The ratio of the studies that developed norms vis a vie those that involved development and adaptation of tools in this review is substantial and worrying. This is because respondents from diverse cultures will inevitably have different scores due to underlying cultural differences. Therefore, it is not enough to translate, modify and pilot the tools and test the psychometric properties of the tools. These procedures just take care of construct, item, and methodological biases. It is equally as important to norm the scores in making them relevant to the new culture where the adapted tool will be utilised. Norms from HIC are not relevant to LMIC. However, since our review was restricted to a particular duration and types of publication, it is possible that the norms were published elsewhere and beyond our review years.

We observed common missing information on psychometric outcomes tested and reported in the papers. Most papers reported either validity or reliability outcomes of the tools, but rarely reported both outcomes. It may not have been possible to report on some psychometric outcomes, for example, where the tool did not have an item structure that allowed evaluating internal consistency. However, the scant reporting practices are questionable in instances where data were available to test for varied psychometric outcomes that speak to the quality of measurement. For example, internal consistency could be calculated in tools with dichotomous item scores, such as Tower of London, yet, for the studies covered in our review, this was not reported. Convergent validity could be evaluated between Attention Network Test (ANT) and attention tools in the Cambridge Neuropsychological Test Battery (CANTAB) (91); between memory tools such as Word Memory Test (WMT) and Medical Symptom Validity Test (MSVT) (92); and between inhibitory control tools such as Self-Ordered Pointing, Delayed Alternation/Non-alternation (DANA), Go/No Go and A standard Stroop (Golden Version) and

Sun-Moon Stroop & Fruit Stroop tests. The factorial structure could be evaluated for test batteries like the Arizona Cognitive Test Battery (ACTB). However, for small sample sizes, in some studies this is smaller than 150 respondents (80), confirmatory factor analysis may be curtailed in favour of exploratory factor analysis or principal component analysis for scaled scores. Such psychometric outcomes were not evaluated though they may have been published elsewhere in the literature that was not captured in our review. That said, different samples differ in cultural characteristics hence it is important to check for replication of the psychometric findings in different samples. The reasons behind such missing information could be justifiable if accidental but it could be problematic if the information is hidden from report to hide disappointing results from reviewers and readers (93). Common lack of information on psychometric quality of neuropsychological tools makes it difficult to determine the usefulness of tools if applied in different populations and cultural settings.

DISCUSSION

This narrative review covered studies on the adaptation and psychometric information of neuropsychological/cognitive tools among children aged 6 – 12 years that were published between 1997 and 2017. The narrative review investigated the neuropsychological tools that are commonly used, the cultural adaptations made to these tools, and the reliability and validity of these tools.

Commonly Used Tools and Psychometric Outcomes

The cognitive domains covered were exhaustive of the DSM V classification, although tools that covered executive functions, complex attention, and memory domains were the most used in the reviewed literature. The included studies of neuropsychological tools among children reported mostly on psychometric outcomes of executive functioning, where the BRIEF appeared most commonly (N=6), followed by KABC-II (N=3). Construct and

convergent validity indicators for the BRIEF showed partial-to-low correlation outcomes. Discriminant validity, however, was supported when it came to its three composite or scale scores, as well as comparison of its teacher rated to parent-rated versions. The BRIEF may have been a common tool due to the ease of administration, that is, through the parents (94). The construct validity of the KABC-II was supported in all reviewed studies, although its predictive validity and reliability findings were rated as low to moderate in several studies. KABC-II was among the few executive function tools with psychometric information in LMICs despite the complexity in administrating it (86).

Complex attention psychometric outcomes were mainly reported for the CPT (N=7) and the Attention Network Test (ANT) (N=5). The latter tool exhibited poor reliability, with only one study reporting moderate to high test-retest reliability. However, construct validity of the ANT was supported. This refers to the good construct and convergent validity, although ANT's discriminant validity was only partially supported (91, 95-97). The CPT was also commonly studied and had several versions. The reviewed studies supported the discriminant validity of the CPT and showed moderate test-retest reliability of this tool. However, findings concerning CPT's specificity and sensitivity ranged from moderate to high.

The Medical Symptom Validity Test (MSVT) (N=6), the Test of Memory Malingering (TOMM) (N=6), and the Word Memory Test (WMT) (N-4) were commonly studied under the memory domain. The WMT showed mixed results regarding validity outcomes, but its specificity was supported in two studies. This trend was not seen in the MSVT, which showed good validity and specificity outcomes, while studies of the TOMM showed mixed findings concerning validity, good performance in terms of specificity, but substandard sensitivity. In some cases, insufficient effort by participants could have affected the variability in these validity and sensitivity outcomes.

Visual-Motor Integration (VMI) was the only perceptual motor tool that was commonly used in the reviewed studies (N=6), which showed mixed findings concerning discriminant validity but did support its convergent, construct, concurrent, and criterion validity. Test-retest reliability of the VMI ranged from low to high in assorted studies while interrater reliability was ranked as high in one study. The popularity of this tool could be attributed to ease of administration (98), especially due to the age of our population of interest. It could also be due to being among the very few tools available for the perceptual-motor domain.

Studies often administered certain subtests rather than entire neuropsychological batteries. However, entire batteries were also widely studied. The tools include the Wechsler Intelligence Scale for Children (WISC), the Halstead-Reitan Neuropsychological Test Battery for Children (HRNB-C), and the Cambridge Neuropsychological Test Automated Battery (CANTAB). The HRNB-C was found to have good discriminant and construct validity, and reviewed studies supported its reliability and sensitivity. The CANTAB as well was found to have good construct validity, though internal consistency ranged from poor to high in between the subtests (99). The WISC-III and IV subsets were commonly studied with reliability findings ranging from poor to high depending on the subtest, while validity outcomes showed the same mixed outcomes.

The Cogstate battery, KABC-II (41, 55, 86), Kilifi Toolkit (46) and Ten Questions Questionnaire (87) and Bolt Board, Pegboard and Bead Threading Tests (85) were validated in Sub-Saharan Africa. Only six studies were conducted in Kenya and Uganda (41, 46, 55, 85-87), but the number of tools covered in these studies is nearly exhaustive of the cognitive domains identified as vital in DSM V. Executive functions covered include planning, working memory, and reasoning; complex attention subdomains covered include attention and attention shift/ selective attention; memory subdomains include non-verbal memory; while perceptual motor sub-domains include visuomotor coordination and visuospatial perception.

Findings with respect to the validity and reliability findings of the tools were not reported for all settings. Notably, discriminant analysis was commonly reported, but none of the studies conducted in Sub-Saharan Africa reported on this form of validity. This selective testing of validity arguably reflects authors' preferences for what is relevant to them and what is easily obtained (51). Authors tend to choose the type of validity to be tested based on the purposes for which they would like the tools to be used. If they want to see whether the tool can measure attention in the same way as other validated attention tools, they will choose to do convergent validity testing. However distinct the types of validity are, a tool cannot be assumed to work well unless it shows evidence of reliability, correlation with variables that it is expected to correlate with, and lack of correlation with variables that it is not expected to correlate as well as evidence that the tool items reflect the cultural construct (100). If a tool cannot correlate with itself, it may not correlate with another measuring the same construct due to its own poor internal consistency/reliability. In most of the studies reported in this review, both validity and tests for reliability could have been reported, but often papers reported only one of these crucial psychometric properties. Reporting both outcomes should be the norm since low reliability may suppress validity hence reporting on either validity or reliability may not give a complete picture of the tool's psychometric quality. Among the tools reviewed, The Developmental Test of Visual-Motor Integration was the only tool with multiple studies reporting on its reliability as well as discriminant, convergent, and construct validity. Reporting all relevant psychometric quality indicators should become standard practice among researchers before assuming that a tool works well. Educators and clinicians should check on these properties before integrating the tools into practice. Interpretation, use, and relevance across diverse cultural settings should be the norm.

Cultural Adaptations

Adaptation processes took different dimensions each dependent on the objectives of the studies. Recommendations for cognitive tools adaptation consist of translation, piloting, and tool modification (101). The adaptation processes captured in this review involved changes to the tools in terms of language and items while the objectives of the study at times necessitated just the testing of different psychometric properties of full batteries or their subsets. The reviewed studies partially tapped into the recommended adaptation procedures. It is beyond the objectives of this review to make recommendations on appropriate adaptation of cognitive tools in diverse cultural contexts. However, some of the adapted tools resulted in cognitive tools with good results concerning validity and reliability, even though mixed findings remain common and call for more psychometric research in various settings. Tools such as the Behavioural Assessment and Research System (BARS) had test-retest validity ranging from low to high depending on the subtest; the Brazilian Child Hayling Test had high content validity but low specificity; the Behaviour Assessment System for Children-Teacher Report was found to show good reliability, good construct validity, but its predictive validity was found to be weak and partially supported; while the Kilifi toolkit was found to have moderate internal consistency, low to moderate test-retest reliability, and partially supported predictive validity (42, 46, 70, 76). The variability in psychometric indicators could result from many factors including differences in test population, differences in individual task scores that may affect reliability or also the adapted tool items do not reflect the cultural construct (102).

Implications for Domains Well Covered

A total of seventy-seven and seventy-five of the studies tested the psychometric properties of tools that measure executive function and complex attention, respectively. The executive function domain has been extensively covered in studies among pre-schoolers and

children in early school years, despite the development of this domain starting at around 3 – 5 years and its maturity only appearing in adolescence (103). This trend presupposes that these studies are inclined to find out the developmental trend rather than whether or not the function has reached maturity. In addition to this, the tools such as the Batería de Evaluación Neuropsicológica Infantil (BENCI), the Developmental Neuropsychological Assessment (NEPSY), and subtests of the Behavioural Assessment of Dysexecutive Syndrome for Children (BADS-C) monitor executive dysfunction progression and recovery. This interest in executive function development means many tools are likely to be developed and their psychometric characteristics tested for measurement of this domain compared to other domains. In addition, the interest may lead to the development of different versions of the same tools in different settings.

Key Gaps and Areas for Intervention

For most of the tools used in the United States, psychometric properties are well documented, yet in many other settings diverse cultural practices and settings might create a different orientation to cognitive functioning that warrants more research. In the United States, processing speed of information is valued in education, which may underpin the quality of information which is inadvertently valued among Hispanics (56).

Only four studies reported on the development of normative data, with other studies reporting on the decision to change the tool to improve validity. Though there is still a debate on which option to pick before integrating a tool in a certain setting, it is interesting to note that the researchers are hesitant, or resource constrained to develop normative data. Tool results need to be interpreted with regard to the general population, as clinical data may not cover the full range of possible scores. Moreover, normative data can tell whether or not a child's functioning score is well within that of the general population in reference to age (104). The

general population here is culture specific where norms from one context cannot be assumed to fit the cultural traits of another. There may be differences in values such as support for accuracy over speed or vice versa which may render a population impaired when using foreign norms, yet these are the upheld cultural values that are seen as healthy. Moreover, since cultural traits tend to change due to globalization, the norm references scores may need to be re-evaluated and updated as needed. These are among cross-cultural effects that make norming a vital component of tools adaptation. However, normative data studies are difficult to conduct as several methods of data collection need to be integrated to obtain an ethnically diverse sample that is truly representative of the general population (105). To achieve adequate norms, the studies need a large sample size that is representative of the raw scores distribution against age and schooling among other variables that represent variability in the population (106). Conducting such studies may prove expensive and require expertise that may not be readily available especially in low resource setting. Despite the challenges, it is not appropriate to utilize norms from other contexts. Researchers and practitioners may need to think of innovative ways to raise funds and collaborate with experts to conduct norm studies just like in the NIH Toolbox norming study (106).

Strengths and Limitations

We considered a narrative review as the best form of interrogating the research questions because it critically examines and discusses the knowledge of interest. That said, there are other types of reviews, such as systematic reviews of specific tools, that require much more work but could yield a more comprehensive picture on the psychometric quality of specific tools in particular contexts. This is like the review of the Raven's test in Sub-Saharan Africa (107). Our review concentrated on studies done between 1997 and 2017; hence studies falling off this timeline were not integrated. Further, the search was limited to the databases in PubMed, Web of Science, and PsycINFO, yet there are other databases that could have

generated more information. However, the common duplication of results in our search suggested we did reach a good level of saturation in finding relevant studies. The terms used for the search were limited to "neuropsychological" or "neurocognitive" and "assessment" or "test," and so using of additional terms like "cognition" could have yielded a broader sample. However, during screening, we also tried other search terms, but they resulted in a roughly similar set of studies. Moreover, these search criteria resulted in many studies being screened in comparison to other earlier reviews that covered a narrower set of studies. In addition, narrative reviews are said to be subjective, and their search criteria may not have explicit specifications (65). Finally, we used an a priori framework to code the findings as per the research question. This form of coding may preclude some findings that do not fit the framework. The search did not include studies with non-English reported findings due to a lack of resources for hiring translators. However, earlier reviews that did not include non-English publications did not show large systematic bias (108, 109). The search did not include data published in tool manuals that were not published in research journals.

Conclusion

Our narrative review indicates that more needs to be done in cultural adaptation and generating psychometric information of neuropsychological tools. A lot of psychometric information was missing from the papers, which is particularly problematic if psychometric problems are underreported for tools used across diverse cultural contexts. There is a need to extensively generate psychometric information for all cognitive domains in the DSM-V; adapt and test psychometric properties of tools in diverse settings; integrate diverse validity and reliability analysis; and courageously do normative data studies and report all relevant psychometric outcomes.

Our review offers a summarised write-up of neuropsychological tools used in the literature among 6 – 12-year-olds, highlighted psychometric properties of a range of tools in

several cognitive domains and settings, but also showed that many tools should be studied further for their psychometric quality, particularly in ensuring that they function well in diverse settings, particularly in LMICs that face major challenges in ensuring that children develop well in the neuropsychological and neurocognitive domains.

Chapter Two Appendices

Appendix 2.1: List of Abbreviations

ANT - Attention Network Test.

BADS-C - Behavioural Assessment of Dysexecutive Syndrome for Children
BENCI - Batería de Evaluación Neuropsicológica Infantil.

BRIEF - Behaviour Rating Inventory of Executive Functioning.

CANTAB - Cambridge Neuropsychological Test Automated Battery.

CANTAB - Cambridge Neuropsychological Test Battery.

CNT - Contingency Naming Test.

CPM - Coloured Progressive Matrices.

CPT - Continuous Performance Tests.

DAS - Differential Ability Scales.

DSM V – Diagnostic and Statistical Manual of Mental Disorders, 5th Edition.

HRNB-C - Halstead-Reitan Neuropsychological Test Battery for Children.

KABC - Kaufman Assessment Battery for Children.

LMIC – Lower and Middle-Income Countries.

MSVT - Medical Symptom Validity Test

NEPSY - Developmental Neuropsychological Assessment.

SOPT - Self-Ordered Pointing Test.

SSA - Sub-Saharan Africa.

SWM - Spatial Working Memory.

TOMM - Test of Memory Malingering.

TOVA - Test of Variables of Attention.

VLL - Verbal List Learning.

WIAT - Wechsler Individual Achievement Test.

WISC -Wechsler Intelligence Scales.

WMT - Word Memory Test.

Appendix 2.2: Search Strings

Here is an example of a search string used in PubMed Search *((((((((((neuropsychological) OR neurocognitive)) AND ((assessment) OR test))) AND ((((child) NOT infant) NOT adolescent) NOT preschool)) AND full text[sb] AND ("1997/01/01"[PDat] : "2017/12/31"[PDat]) AND Humans[Mesh] AND English[lang] AND child[MeSH:noexp])) NOT (neurophysiological AND full text[sb] AND ("1997/01/01"[PDat] : "2017/12/31"[PDat]*

) AND Humans[Mesh] AND English[lang] AND child[MeSH:noexp])) AND full text[sb] AND ("1997/01/01"[PDat] : "2017/12/31"[PDat]) AND Humans[Mesh] AND English[lang] AND child[MeSH:noexp])) AND ((((reliability AND full text[sb] AND ("1997/01/01"[PDat] : "2017/12/31"[PDat]) AND Humans[Mesh] AND English[lang] AND child[MeSH:noexp])) OR (validity AND full text[sb] AND ("1997/01/01"[PDat] : "2017/12/31"[PDat]) AND Humans[Mesh] AND English[lang] AND child[MeSH:noexp])) AND full text[sb] AND ("1997/01/01"[PDat] : "2017/12/31"[PDat]) AND Humans[Mesh] AND English[lang] AND child[MeSH:noexp]) Filters: Full text; Publication date from 1997/01/01 to 2017/12/31; Humans; English; Child: 6-12 years.

Here is a search string used in Web of Science: *(ALL=(((((((cognitive) OR neuropsychological) OR neurocognitive)) AND (((test) OR tool) OR measure))) AND ((((child) NOT preschool) NOT infant) NOT adolescent)) NOT ALL=(neurophysiological) AND ALL=(English) AND ALL=(Reliability) AND ALL=(Validity)) AND (PY==("1997" OR "1998" OR "1999" OR "2000" OR "2001" OR "2002" OR "2003" OR "2004" OR "2005" OR "2006" OR "2007" OR "2008" OR "2009" OR "2010" OR "2011" OR "2012" OR "2013" OR "2014" OR "2015" OR "2016" OR "2017") AND DT==("ARTICLE")) AND (OA==("OPEN ACCESS"))*

Here is a search string used in PsycINFO: *(neuropsychological or neurocognitive) AND (assessment or test) NOT (infant or preschool or adolescents) NOT neurophysiological AND (validity or reliability) Limiters - Linked Full Text; Publication Year: 1997-2017; English; Age Groups: School Age (6-12 yrs.); Population Group: Human*

CHAPTER THREE

Psychometric Evaluation of the Computerized Battery for Neuropsychological Evaluation of Children (BENCI) among School Aged Children in the Context of HIV in an Urban Kenyan Setting

Published Article: Maina, R., He, J., Abubakar, A., Perez-Garcia, M., Kumar, Jelte M. Wicherts. Psychometric evaluation of the computerized battery for neuropsychological evaluation of children (BENCI) among school aged children in the context of HIV in an urban Kenyan setting. BMC Psychiatry 23, 373 (2023). https://doi.org/10.1186/s12888-023-04880-z

Abstract

Introduction. Culturally validated neurocognitive measures for children in Low- and Middle-Income Countries are important in the timely and correct identification of neurocognitive impairments. Such measures can inform development of interventions for children exposed to additional vulnerabilities like HIV infection. The Battery for Neuropsychological Evaluation of Children (BENCI) is an openly available, computerized neuropsychological battery specifically developed to evaluate neurocognitive impairment. This study adapted the BENCI and evaluated its reliability and validity in Kenya.

Methodology. The BENCI was adapted using translation and back-translation from Spanish to English. The psychometric properties were evaluated in a case-control study of 328 children (aged 6 – 14 years) living with HIV and 260 children not living with HIV in Kenya. We assessed reliability, factor structure, and measurement invariance with respect to HIV. Additionally, we examined convergent validity of the BENCI using tests from the Kilifi Toolkit.

Results. Internal consistencies ($0.49 < \alpha < 0.97$) and test-retest reliabilities (-.34 to .81) were sufficient-to-good for most of the subtests. Convergent validity was supported by significant correlations between the BENCI's Verbal memory and Kilifi's Verbal List Learning ($r = .41$), the BENCI's Visual memory and Kilifi's Verbal List Learning ($r = .32$) and the BENCI's Planning total time test and Kilifi's Tower Test ($r = -.21$) and the BENCI's Abstract Reasoning

test and Kilifi's Raven's Progressive Matrix (r = .21). The BENCI subtests highlighted meaningful differences between children living with HIV and those not living with HIV. After some minor adaptions, a confirmatory four-factor model consisting of flexibility, fluency, reasoning and working memory fitted well (χ^2 =135.57, DF = 51, N = 604, p < .001, RMSEA = .052, CFI = .944, TLI =.914) and was partially scalar invariant between HIV positive and negative groups.

Conclusion. The English version of the BENCI formally translated for use in Kenya can be further adapted and integrated in clinical and research settings as a valid and reliable cognitive test battery.

Key Words: Cognitive tests, Validity, Reliability, The BENCI, Kenya, School aged children, HIV.

Introduction

Human Immunodeficiency Virus (HIV) is a neurotropic virus that can infect the nerve cells (110). Widespread access to antiretroviral drugs (ARVs) has reduced the severity of HIV related brain diseases (111). However, even when children are on ARVs and virologically suppressed, they may continue to manifest neurocognitive impairments (112-114). The monitoring of neurocognitive performance among children with HIV should be included in a comprehensive HIV management plan (115, 116). However, in sub-Saharan Africa (SSA) the lack of adequately standardized neurocognitive tools that are easy to implement at a relatively low cost inhibits the implementation of recommended neurocognitive monitoring among HIV-positive children. To address this gap in health care, it is important to identify and validate neurocognitive measures that can be easily implemented in health care settings within the African setting. Given how limited the resources are in many of these settings, neurocognitive tools for use in SSA need to be open-access and relatively easy to administer so that they can be implemented by paraprofessionals or professionals with limited training. These tools should also be engaging to the children.

In recent years, there has been a proliferation of computerized neurocognitive tools which are relatively easy to implement, yet many of these tools have largely been developed and tested in high-income countries (115). They include the NIH toolbox, Conner's Continuous Performance Test, Attentional Network Task (ANT), CNS Vital Signs and Paediatric Immediate Post Concussion Assessment and Cognitive Testing (Paediatric ImPACT) (79, 95, 117-120). Due to potential measurement biases that may arise from adopting test from one context to another, it is crucial that these new promising tests are thoroughly evaluated in the SSA context (111, 121-123). Here, we study the psychometric properties and potential utility of the computerized Battery for Neuropsychological Evaluation of Children (The BENCI)

which covers several neuropsychological domains and was originally developed in Spanish for Ecuadorian children. The BENCI measures the seven cognitive domains with the following subtests: Simple Reaction Time, Visual-motor, Continuous Performance, Verbal Memory, Visual Memory, Verbal Comprehension Images, Verbal Comprehension Figures, Phonetic Fluency, Working Memory, Abstract Reasoning, Semantic Fluency, Go/NO-GO, Spatial Stroop, Alternate Visual-motor, and Planning-Attraction Park tests (124). See Table 3.1 for their specific domains and administration. The fact that the BENCI is openly available and computerized makes it relatively easy to access and administer. It is also enjoyable for children (40), hence curtailing for loss of interest and distraction, which may result in low completion rates, missing responses, and erroneous responses.

Since the BENCI is a promising tool with its psychometric properties already documented in Morocco among 7-, 9-, and 11-year-old children in schools, its adaptation and implementation in Kenya among children living with HIV and children who are HIV negative can expand our school-age children toolbox and provide clinics with rigorously validated measures (40). Data from Moroccan children supported a factorial structure of executive functioning with inhibition, flexibility, fluency, reasoning, and verbal memory in the Arabic version of the BENCI our study (40). In deciding the executive function tests to include in the factorial model, the previous study acknowledged the lack of a theoretical model that could explain the battery's structure. Hence, we opted to use Diamond model functions (125) of executive functioning to create our model. We included verbal tests as indicators of executive function because tests of verbal memory (126, 127) have been associated with executive function outcomes with up to 55-60% shared variance (126). However, factorial structure and measurement invariance with respect to HIV status has yet to be evaluated in a similar LMIC region. Measurement invariance evaluates whether the subtests are loaded similarly onto the latent factors and whether groups based on, e.g., educational attainment, health status, ethnicity

and age can be meaningfully compared (47). Since the language of instruction in the Kenyan schools is English, (128) we choose to adapt an English version of the BENCI. Moreover, computerized assessment is rare in Kenya, and this study with the computerized BENCI is an important first step to assess the feasibility of reliably evaluating neurocognitive functions using computerized measures in the Sub-Saharan context. To conduct a comprehensive evaluation of the BENCI, we carried out the following:

1. Adapted the BENCI in a culturally appropriate adaptation format and user-centered testing
2. Evaluated its internal consistency and test-retest reliability
3. Examined the associations between the results of the BENCI (a computerized test) and those of a paper-and pencil standardized test
4. Evaluated differences in performance and measurement properties among children who are living with HIV versus those who are not living with it.

Methodology

Participants and Settings

A total of 604 (311 females, 291 males and two with missing gender information) children from Nairobi participated in the study. Nairobi is the capital city of Kenya with an 87.1% literacy level and the language of instruction in the schools is English (128). We recruited two samples from different study sites. One group of children was sampled from a children HIV outpatient programme. The programme, implemented in seven resource poor settings in Nairobi, included children living with HIV of different ethnic backgrounds who receive home-based care. The sample of children not living with HIV was drawn from three primary schools in Nairobi. The schools were chosen on the basis of their similarity to most schools in Kenya with regards to the mode of education at that time which was the 8.4.4 system with the examining body under the Ministry of Education being the Kenya National

Examination Council (129). These children come from diverse socio-economic settings with most of them from middle-class families. We chose this to rule out the impact of sharp socioeconomic status differences. The study sample size computation was based on data from an earlier study in Africa that found the means on the KABC – 2 to differ between HIV-infected (N = 93) and uninfected (N = 106) by $\mu_1 = 184.7\ (sd = 63.72)$ and $\mu_2 = 200.6\ (sd = 68.72)$, respectively, yielding a Cohen's d of 16.1/66.3 = 0.24 (2). Together with an alpha level of 5% and a power of 80%, these resulted in a total sample size of 544 respondents, thus the target sample size was 272 children living without HIV and 272 children living with HIV, respectively. We slightly oversampled to address any potential loss of data due to missingness.

Measures

The BENCI: The existing BENCI test was first developed in Ecuador and offers norms for children aged 6 – 17 years in Ecuador, 7, 9 and 11 years in Morocco and 6 - 8 years in Palestine (40, 124) . The test can be administered within 75 minutes with one 10-minute break in between the 14 neuropsychological tests. On average, however, the administration takes around 90 minutes. The test can be administered by skilled psychologists with additional training specific to BENCI.

Paper and Pencil Measures: To test convergent validity of the BENCI, we used paper and pencil tests that are internationally accepted and standardized and have previously been adapted and validated in a rural Kenyan community (46). This so-called Kilifi Toolkit covers executive functioning, memory, and attention and can be administered within 120 minutes. The neurocognitive tests have good psychometric properties with *split- half reliability between .70 and .84 while internal consistency is* $\geq$ *.70* among 7 – 11-year-old children in Kenya (46). Table 3.1 lists tests in Kilifi toolkit and the BENCI. As part of our study, we also measured age, gender, height, and weight.

The BENCI Adaptation process

The adaptation process was guided by the translation and adaptation guidelines of the International Test Commission (130). We obtained authorization to adapt the original BENCI test and the original test developers including MPG who also had an advisory role in test adaptation. Since the original BENCI was in Spanish, the translation was the first stage of adaptation where one bilingual researcher translated it from Spanish to English and another native English speaker checked the English translation for linguistic and semantic consistency. Clinical psychologists in Kenya, in discussions with other professionals in Spain, evaluated the tools' structure and appropriateness against the tool's original markers in terms of sentence structure and familiarity of images in the Kenyan context. This work was complemented by a pilot study involving 5 females and 3 males with a median age of 13 years to check the appropriateness of the items, pictures, and instructions. The pilot study involved administering all the sub tests within the BENCI and later interviewing each child individually on how they experienced the tests.

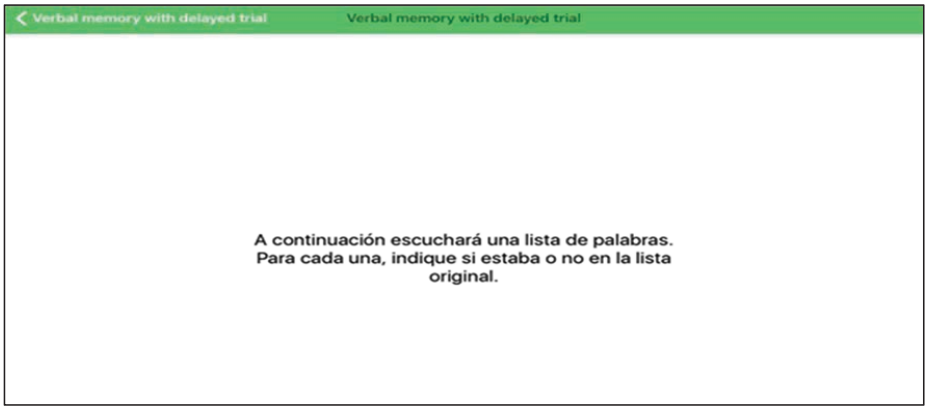
In terms of the BENCI administration, some children expressed that the sustained attention test was too lengthy which lowered their enthusiasm for doing the rest of the tests. This was discussed with the study team and changes were made to place the sustained attention test right before the 10 minutes' break. Children tended to touch the screen with their fingers playfully even when not responding and this resulted in unintended responses especially in the Visual Memory and Verbal Memory with Delayed Trial test. Hence, BENCI administrators were instructed to caution the children against moving their hands on the screen if they did not have any intention to respond.

In language, some English words in the instructions of some BENCI subtests were unclear to some young children. An example is the word 'figures' which was changed to

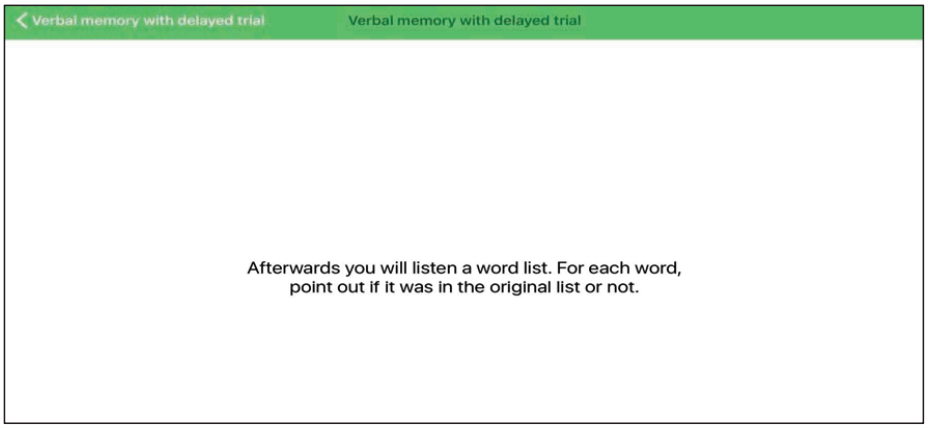
'shape' as Kenyan children are more familiar with the latter than the former. Some instructions were not clear enough; hence recommendations were made to ensure that children understood what to do when a certain stimulus appeared, especially in the verbal comprehension subtest. In the Continuous performance test, instructions on pressing screen right after letter X appeared after letter A were not clear. We therefore agreed that we would draw a letter A followed by letter X to help in indicating when the screen should be pressed. Several instructions were changed to simpler English. Young children had a better understanding of the test requirements when additional information was given in Kiswahili – the national language of Kenya.

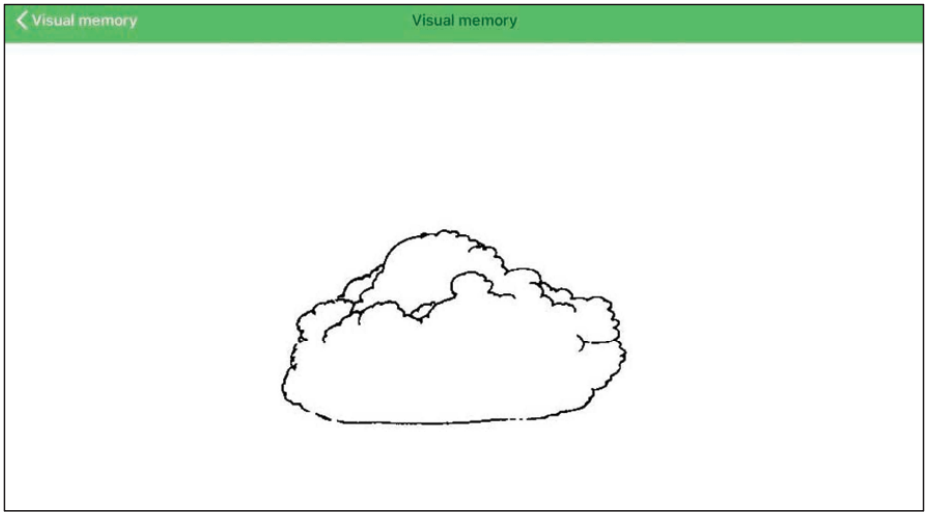
Cultural adaptations were also made to images in the verbal comprehension test, as young children did not recognize some animals like the difference between a squirrel and a rabbit, while some animals had some striking resemblance to animals familiar to the Kenyan children. Images within the visual memory subtest, which could not be recognized by children, were also changed, or scoring changed to include the interpretation that was familiar to the children. For example, some children could not differentiate between cloud and bush as the images were similar so both answers were integrated as the correct answers in the scoring guide. See Figure 3.1 for the pictorial presentations on the changes made in the BENCI and appendix 3.1 for the other observations and recommendations made during the pilot phase.

Figure 3.1: Translation and Cultural Adaptations Made in BENCI

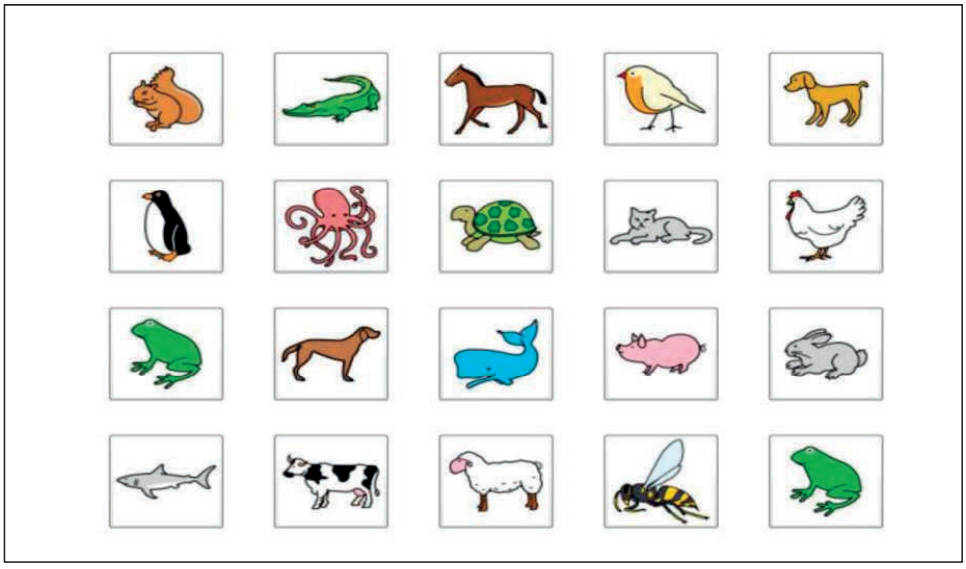


Translation from Spanish (shown above) to English (shown below)





This figure was recognised as “bush” by some children and “cloud” by others.



Verbal Comprehension test: Could not tell the difference between a squirrel and rabbit. Incorporated a teaching guide for training section before the test.

Procedure

In the clinics, a database of children aged 6 – 14 years old was generated and the children were informed to come to the clinic on a certain day of the week when the programme arranged for some fun activities to take place. Most of the time the assessment day fell on a weekend and on the same day as the children were scheduled for their clinical appointments. On the scheduled day, the children and their parents were randomly identified and individually informed about the study with voluntary participation of the children being requested. We included children aged 6-14 years[1] that are HIV-positive and not having any comorbid conditions as reported in their medical reports. We did not include children with comorbid and/or severe medical conditions associated with being HIV-positive as indicated in their medical reports, as well as children who did not meet the age criteria. In the school setting, the children were randomly selected from their classrooms, which ranged from Grades 1 to 5. In this population, we included children aged between 6 – 14 years old and not having any medical condition as reported by the school and the students themselves. Children who did not meet these conditions were excluded from the study. The institutions provided a room where the neurocognitive assessments could be carried out. Relevant subtests in Kilifi toolkit (see Table 3.1) were administered with paper and pencil by a trained interviewer (46). For test-retest reliability, 38 HIV negative children (21 females) were re-assessed 2 months after the initial assessment.

[1] The original plan was to include children aged 8 – 11 years but we included also 6, 7 and 12 – 14 year olds (N = 52, 9.5%) because they were in the same grade.

Table 3.1: BENCI and Kilifi Toolkit Tests

BENCI (90 minutes)			**KILIFI TOOLKIT (120 MINUTES)**		
Domain	**Sub-test**	**Outcome Measures**	**Domain**	**Sub-test**	**Outcome Measures**
Processing Speed	Simple Reaction Time Test (a plus sign of the screen prompts the child to press a key on the keyboard fast)	Mean RT & Median RT	-	-	
Visual-motor Coordination	Visual-motor test (involves connection of elements/number in a given sequence)	TT & Total Errors			
Sustained Attention	Continuous Performance Test (respondent presses any key every time the required stimulus appears)	Hits/CA, EO, EC, Mean RT & Median RT.	Visual Sustained and Selective Attention Auditory Sustained and Selective Attention	People Search (A stimulus sheet comprising complete and incomplete stick figures is presented. The subject is required to cross out only complete figures, as quickly as possible) Digit span as we could not find the tape. The child is instructed to repeat a series of numbers (with increasing numbers of digits) forward. Each correct response is worth one point; with a maximum of 14 points for each sub-score series	TT, RT, Errors of Omission (EO) and Errors of Commission (EC) TT and Highest Score

BENCI (90 minutes)			**KILIFI TOOLKIT (120 MINUTES)**		
Domain	**Sub-test**	**Outcome Measures**	**Domain**	**Sub-test**	**Outcome Measures**
Memory	Verbal memory test (child listens to some words then repeats the ones remembered) Verbal memory delayed recall test (the series of words said are repeated after 20 minutes) Verbal Memory Essay of Recognition test (words are read out loud and respondents identifies those that were in the previous list) Visual memory (series of images are presented after which respondents verbalizes those remembered) Visual Memory delayed Essay (the images remembered are said out loud after 20 min) Visual Memory Essay of Recognition (respondent identifies if images presented were in previous list)	Hits/CA, P & I Hits/CA, P & I Hits/CA & Errors Hits/CA, EC & EO Hits/CA, EC & EO Hits/CA, EC & EO	Memory	Working Memory: Verbal List Learning – VLL (Two lists of 15 items are read out to the child as a shopping list. The first is presented five times and the second only once) Subtests within include: - Verbal Memory Test Free Recall Trial Test Short Delay Free Recall Trial Short Delay Cued Recall Trial Long Delay Free Recall Trial Long Delay Cued Recall Trial Long Delay Recognition Trial	Intrusions (I), Perseverations (P), CA and TT
Language	Verbal Comprehension Images Test (respondent matches images to given conditions) Verbal Comprehension Figures (respondent matches geographic shapes to given conditions) Phonetic Fluency (a letter is presented, and respondents verbalizes all words that start with the letter given.)	Hits/CA & Errors Hits/CA & Errors Hits/CA, I & P	-	-	

Executive Functioning	Working Memory (a list of colour and numbers are said, and respondent repeats the numbers then the colours) Abstract Reasoning (respondent completes a logical series by selecting the right element) Semantic Fluency (a category is given, and respondents says the elements known in that category) Inhibition: Go/NO-GO (respondents identifies distinguishing factor between two elements and later identify the distinguishing element) Flexibility: Spatial Stroop (respondent matches arrow directions to arrow labels) (Two components of spatial Stroop - attention shifting task measures flexibility while proper spatial Stroop task measures inhibition) Flexibility: Alternate Visual motor (is flexibility measure that involves two distinct series in which the respondent should connect alternatively) Planning: Attraction Park (respondent chooses a number of attractions according to money in hand with each attraction chosen expiring after a given period)	Hits/CA Hits/CA Hits/CA, I & P EC, Hits/CA & Mean RT Median RT, EC, EO TT & Total Errors Planning Time, TT, Rule 1, Rule 2, fairground amusements & different fairground amusements/CA	Executive Functioning	Self-Ordered Pointing Test - SOPT (Selection of pictures displayed in varying positions on separate sheets in sets of 6, 8, 10, and 12. As each page is turned the subject is required to identify all members of the set, but to point to each item of the set only once. Touching a picture more than once is considered an error). Raven progressive matrices: Reasoning: Coloured Progressive Matrices – CPM (Three sets with 12 matrices made of abstract patterns. The subject is asked to complete the matrix by placing one of a choice of four patterns in the empty space) Attention and attention shift: Contingency Naming Test – CNT (The child is taught a series of rules to name nine drawings displayed in a single series. Each drawing consists of a large outer coloured shape and a smaller inner coloured shape. Each drawing is named according to the shape or colour of one of its two shapes. The rules taught for selecting the name of the item become more	Time Taken (TT), Reaction Time (RT) and Correct Answers (CA)

BENCI (90 minutes)			KILIFI TOOLKIT (120 MINUTES)		
Domain	**Sub-test**	**Outcome Measures**	**Domain**	**Sub-test**	**Outcome Measures**
				complex over four trials) Planning: Tower Test/ Tower of London (Three coloured wooden balls are moved between three pegs to match a goal position. Time and number of moves required are recorded**).**	

RT – Reaction Time, TT – Total Time; CA – Correct Answers; EO – Errors of Omission; EC – Errors of Commission; I – Intrusions; P – Perseverations;

Analyses

Data from the BENCI was automatically captured in the tablet as programmed in the original Spanish version and exported to an Excel sheet. The Kilifi Toolkit data were input into Excel sheets and codes/matching identifications were realigned to ensure correct matching with similar cases in the file with BENCI data. We double-checked the age, gender, and clinic/school groupings to ensure the correct ID matching. Analyses were run in SPSS version 20 and AMOS version 22. We used Alpha = .05 as the nominal significance level.

Data were cleaned by first having a visual inspection of a scatter plot and statistical evaluation of each of the subtest scores for outliers. Data with influential outliers were then evaluated through a three-step process to identify if certain scores should be deleted. First, we checked the residuals of the regression of age on the subtests where cases with high standardized residual value, low effect size, and low p-value were noted. Second, we evaluated cases with *z* scores beyond z=|2| for possible deletion. Third, we conducted a case-by-case check to evaluate whether a certain score would be expected given other subtest scores from these participants. For instance, we discarded scores on the Verbal Memory Immediate hits and

Continuous Performance hits subtest that had z-scores below -3 and whose z- scores were not expected for the age groups we were looking at. Through this process, we decided whether certain scores should remain as they are or identify them as missing. We then carried out a missing data pattern analysis where Little's MCAR test statistic was significant (χ^2 = 2455.2, DF = 1725, $p < .001$), highlighting that scores were not missing completely at random. However, a check on whether the missingness was significantly related to age, HIV status, and date of data collection, uncovered no significant relationship with the missingness pattern in subtest scores. We could, therefore, not identify what the missingness was related to.

Internal consistency in terms of Cronbach's Alpha (KR-20 for dichotomous items) was determined for all seven tests for which item-level data were available. We opted for Cronbach's Alpha because it is widely used in testing the internal consistency of the items within a test that reflects the degree to which items covary positively. The test-retest reliability was analysed using ICC and Pearson's correlations. We then checked whether the performance within the BENCI subtests aligned with developmental models' expectation of growth in cognitive performance as children grew in age. Convergent validity was analysed using Pearson correlation where scores of the BENCI subtests were correlated with the raw scores of corresponding subtests in Kilifi Toolkit. We hypothesized that tests measuring the same cognitive domain would correlate positively. We compared differences between HIV-positive and HIV-negative groups with t-tests and considered possible floor and ceiling effects by checking histograms and outliers by calculating skewness for each subtest.

We run a confirmatory factor analysis in AMOS to assess the construct validity the BENCI using a model of Executive Function proposed by Diamond, in which executive function comprises reasoning, inhibition, flexibility, fluency, and working memory cognitive functions (125). The model fit was evaluated with the Chi-square tests, Root Mean Square Error of Approximation (RMSEA), Comparative Fit Index (CFI), and the Tucker-Lewis Index (TLI).

A model is considered a good fit if the value of RMSEA is below .06, and CFI and TLI above .90, respectively.

Results

The two test batteries were administered among 274 children living with HIV and 330 children without HIV with a mean age of 9.48 (SD = 1.31), of which roughly half were male. Table 3.2 summarizes the demographics of participants in the two groups. The second assessment of the BENCI among 38 Children not living with HIV consists of 21 females, with a mean age of 9.18 (*SD* = 1.21).

Table 3.2: Socio-Demographic Information

Variables		**HIV Negative N (%)**	**HIV Positive N (%)**
Gender	Male	163 (49.40)	148 (54.00)
	Female	166 (50.30)	125 (45.60)
	Missing	1 (0.30)	1 (0.40)
Age in months (Mean ± SD)		117.2 ± 16.24	119.40 ± 14.63
Age in Years (Mean ± SD)		9.41 ± 1.37	9.56 ± 1.24
Nutrition	Weight in kg (Mean ± SD)	34.98 ± 7.12	32.27 ± 5.85
	Height in cm (Mean ± SD)	136.34 ± 8.00	133.02 ± 8.11

Scale Attenuation Effects

Using correlational and descriptive statistics including histograms, we evaluated attenuation patterns in the BENCI tests. Eight of the BENCI subtests exhibited ceiling and floor effects that tend to suppress correlations and reliabilities. Specifically, on Verbal Comprehension Figures, 30% (N = 181) of the sample scored the highest possible score of 8 hits, while on Verbal Comprehension Images hits, 51% (N = 308) of the sample scored the highest possible score of 8 hits. Other subtests with ceiling effects included *Continuous Performance hits, Go No Go hits, Working Memory hits,* and *Spatial Stroop.* Both *Verbal*

Memory Recognition 13.4% (N = 16) and *Visual Memory Recognition* 16.7% (N = 20) showed some ceiling effects meaning that the number of participants having the highest scores was almost equal to those with average scores. At the same time, floor effects were evident on the *Planning Time of First Option* and *Spatial Stroop errors scores*. *Semantic Fluency hits 13.4% (N=16), Phonetic Fluency hits 16% (N = 19), Verbal Memory Delayed hits* 16.8% (N = 20), and *Planning time total 33.9% (N = 38)* showed some floor effects. This meant that the number of participants having the lowest scores was almost as equal to those with average scores. The floor and ceiling effects highlighted that these subtests psychometric functioning could be improved by adding easier and more difficult items, respectively, in any future revisions of the BENCI. The remaining BENCI subtests showed no such attenuation effects.

Internal Consistency of the BENCI

We computed the Cronbach's Alphas (KR-20s) for seven of the subtests with dichotomous item scores. The internal consistency of the BENCI subtests varied from poor to excellent reliability. As shown in Table 3.3, the Language Comprehension tests, Verbal Comprehension Images, and Figures, had the fewest items (N = 8) and Cronbach Alpha $0.49 < \alpha < 0.68$ which was the lowest among the other BENCI subtests. Low Cronbach Alphas tend to suppress correlations, but most of the BENCI subtests had high Alphas. The Abstract reasoning, Planning, Go No Go, Spatial Stroop, and Processing Speed tests correlated well with themselves ($0.75 < \alpha < 0.97$ or alpha range from .75 to .97) hence showing that there was little random measurement error.

Possibly due to the ceiling effects being less severe because of lower mean scores, we found Verbal Comprehension Figures and Images tests to show higher internal consistencies among children living with HIV ($0.57 < \alpha < 0.68$) than among children not living with HIV ($0.35 < \alpha < 0.56$), whose scores were more affected by the ceiling effect. In the Abstract

reasoning, Planning, Go No Go, Spatial Stroop, and Processing Speed sub-tests the items had acceptable and excellent ($0.76 < \alpha < 0.97$, or alpha range from .76 to .97) internal consistency showing that the tests are reliable for both children living with HIV and those not living with HIV, as shown in Table 3.3. The Alphas in the latter tests were higher in the lower-scoring sample of children living with HIV than in children not living with HIV due to less severe attenuation effects in the former group.

Table 3.3: BENCI Items Internal Consistency

BENCI Subtests	No. of Items	Skewness	Overall Cronbach's Alpha	HIV Negative Cronbach's Alpha	HIV Positive Cronbach's Alpha
Verbal comprehension images Total Time	8	3.113	.689	.519	.682
Verbal comprehension images Hits	8	-1.302	0.592	0.386	0.602
Verbal comprehension figures Total Time	8	1.926	0.613	0.56	0.609
Verbal comprehension figures Hits	8	-.737	0.496	0.349	0.571
Abstract reasoning Hits	25	.019	0.832	0.813	0.781
Abstract reasoning Total Time	25	.851	0.890	0.904	0.875
Go No Go Total Hits	101	-.745	0.870	0.824	0.895
Go No Go Total Time	101	1.136	0.879	0.872	0.864
Planning Total time	12	.895	0.753	0.760	0.744
Spatial Stroop Hits	90	-.933	0.973	0.966	0.975
Spatial Stroop Time	90	-1.701	0.950	0.924	0.959
Processing speed Reaction Time	50	1.598	0.832	0.832	0.822

Tests Retest Reliability of the BENCI

Table 3.4 presents the Intraclass Correlation (ICC) of the test and retest scores of the BENCI and the Pearson correlations between the repeated measurements among the 38 children not living with HIV. The Intraclass correlation for specific tests ranged from -.34 to .81. The coefficients were rather high in Sustained Attention RT, Immediate Visual Memory, and Alternate Visual-motor Coordination (*ICC range from .74 to .81, r =.68 - .62).* Moderate correlations were found in Immediate Verbal Memory, Delayed Visual Memory, and Visual Recognition Memory *(ICC range from .52 to .58, r =.39 - .38).* Test-retest reliability was poor for Go/No-Go (RT), Sustained Attention CA, and Reasoning *(ICC range from .14 to -.34, r =.08 - -.15).*

The test-retest reliability results showed that most of the tests were consistent on the two occasions (2 months in between t1 and t2). With clear significant gains in performance as

expected by increasing test familiarity and maturation for fifteen out of nineteen subtests, except for Sustained Attention CPT, Verbal Recognition Memory (CA), Reasoning (CA), and Go/No-Go (RT) that showed no clear improvements in mean performance.

Table 3.4: Reliability test-retest of the BENCI battery

Test (N=38)	First Visit Mean (SD)	Second Visit Mean (SD)	ICC	CI 95%	Pearson correlation
Visual-motor Coordination (TT)	73772.32 (34587.11)	54539.74 (27326.12)	**.66****	.35 - .82	.51**
Alternate Visual-motor Coordination (TT)	75473.65 (32581.90)	50385.71 (23478.43)	**.74****	.49 - .87	**.62****
Sustained Attention CPT (CA)	49.06 (12.06)	51.70 (6.26)	.13	-.74 - .57	.08
Sustained Attention CPT (RT)	626.96 (191.41)	618.57 (196.13)	**.81****	.62 - .90	**.68****
Immediate Verbal Memory (CA)	5.19 (2.60)	6.26 (3.38)	.58*	.12 - .77	.39*
Delayed Verbal Memory (CA)	5.30 (3.01)	5.58 (3.00)	**.71****	.44 - .85	.55**
Verbal Recognition Memory (CA)	18.89 (3.49)	20.00 (3.08)	.41	-.15 - .70	.26
Immediate Visual Memory (CA)	5.76 (2.60)	6.30 (3.29)	**.75****	.51 - .87	**.61****
Delayed Visual Memory (CA)	5.35 (3.22)	6.47 (3.29)	.55*	.13 - .77	.38*
Visual Recognition Memory (CA)	44.30 (6.08)	45.47 (4.11)	.52*	.07 - .75	.38*
Comprehension of Images (CA)	7.53 (0.97)	7.78 (0.42)	.49*	-.01 - .74	.45*
Working Memory (CA)	11.58 (5.69)	13.76 (4.92)	**.71****	.43 - .85	.55**
Reasoning (CA)	13.89 (4.11)	15.74 (4.39)	-.34	-1.63 - .32	-.15
Semantic Fluency (CA)	8.00 (3.01)	6.84 (3.58)	**.64***	.30 - .81	.48*
Phonetic Fluency (CA)	4.89 (2.48)	5.68 (2.83)	.48*	-.00 - .73	.32
Go/No-Go (CA)	0.87 (0.14)	0.84 (0.16)	.43*	-11 - .71	.27
Go/No-Go (RT)	0.64 (0.08)	0.66 (0.11)	.14	-.71 - .56	.08
Selective Attention (RT)	575.21 (148.28)	573.16 (153.80)	**.66****	.32 - .83	.49*
Planning FO (RT)	5047.03 (5865.45)	2600.13 (3168.40)	.43*	-.11 - .71	.32*

TT – Total Time; RT – Reaction Time; CA – Correct Answers; CPT – Continuous Performance Test; FO – First Option; **. Correlation is significant at the 0.01 level (2-tailed); *. Correlation is significant at the 0.05 level (2-tailed).

Convergent Validity

Table 3.5 presents the correlations between corresponding BENCI and Kilifi toolkit tests. The attention, memory, inhibition/planning, reasoning, and flexibility tests in the BENCI and Kilifi were expected to correlate. However, some of these tests did not correlate as expected due attenuation effects, while others correlated as expected despite the attenuation effects.

In domains of reasoning, several inhibitions, and a few memory-related tests in the BENCI were positively correlated with tests in Kilifi toolkit, supporting convergent validity across these domains. The BENCI's Working Memory test was expected to correlate with Kilifi's Self-Ordered Pointing Test (SOPT) because they both measure working memory. However, the BENCI Working Memory test did not have a significant correlation with Kilifi's working memory test, Self-Ordered Pointing Test (SOPT). This could be because the BENCI Working Memory test showed ceiling effects and might have been too easy for most test takers.

Kilifi's Verbal List Learning Test and Nonverbal Selective Reminding Memory test were expected to correlate with the BENCI's Verbal Memory and Visual Memory tests because they all measure memory. However, none of the BENCI's memory tests had a significant correlation with Kilifi's Nonverbal Selective Reminding Memory Test (NVSRT). Moreover, the BENCI's Verbal Memory Recognition and Visual Memory Recognition tests had no significant correlation to any of Kilifi's memory tests. This outcome could be because the BENCI's Verbal Memory Recognition and Visual Memory Recognition tests had some ceiling effects while Kilifi's NVSRT had floor effects. However, the BENCI's Verbal Memory Immediate hits had a significant correlation with Kilifi's Verbal List Learning's (VLL) Immediate Memory Span (r = .37), Level of Learning (r = .40) and Total correct answers (r = .41). In addition, the BENCI's Verbal Memory Delayed Trial was also significantly correlated with Kilifi's Verbal List Learning's Immediate Memory Span (r = .21). Moreover, the

BENCI's Visual Memory Immediate hits had a significant correlation with Kilifi's Verbal List Learning's (VLL) Immediate Memory Span (r = .23), Level of Learning (r = .34) and Total correct answers (r = .32). In addition, BENCI's Visual Memory Delayed Trial was also significantly correlated with Kilifi's Verbal List Learning's (VLL) Level of Learning (r = .23) and Total correct answers (r = .25). The significance was found despite the BENCI's Verbal Memory Delayed showing some floor effects. The rest of the memory tests in the BENCI and Kilifi had no ceiling or floor effects. The correlation between Kilifi's Verbal List Learning's (VLL) Level of Learning and Total correct answers and the BENCI's Reasoning test was not expected. As expected, the BENCI Abstract Reasoning Test significantly correlated with Kilifi's Raven's Progressive Matrix (RPM) (r = .21). Both reasoning tests had no attenuation effects.

Kilifi's People Search test and FNRT test were expected to correlate with BENCI's Continuous Performance test and Spatial Stroop Attention test because they all measure attention. Among the attention tests, the BENCI sustained attention test, Continuous Performance hits and reaction time test, did not have a significant correlation with Kilifi's visual sustained and selective attention - People Search test (r = -.10; r = .12), as well as auditory sustained and selective attention test - Forward Digit Span total score (r = -.14; r = .07). People Search test had floor effects while Continuous Performance hits had ceiling effects. Moreover, the BENCI tests that contain an attention component, Reasoning (r = -.37) and Working Memory (r = .19) were also significantly correlated to Kilifi's People Search. Kilifi's People Search and its correlation with the BENCI's Reasoning and Working Memory tests was unexpected as these BENCI tests are not primarily meant to measure attention.

BENCI's Spatial Stroop was expected to correlate with Kilifi's Contingency Naming test (CNT) because they both measure flexibility. However, the Spatial Stroop test, had no

significant correlation with the Contingency Naming test (CNT) (r =.03). The Spatial Stroop test showed ceiling effects while CNT had no attenuation effects.

Kilifi's Tower Test was expected to correlate with the BENCI's planning test because they both measure inhibition. This is indeed the case, as the BENCI Planning Total Time test had a significant association with Kilifi's Tower test (r = -.21). However, BENCI's Planning Time of First Option test had no significant association with Kilifi's Tower test (r = -.11). These results should be interpreted cautiously because the BENCI's Planning Total Time test had some floor effects while the Planning Time of First Option had floor effects indicating that items were relatively difficult for our test takers.

Overall, in the reasoning domain, much convergence between the BENCI and Kilifi Toolkit was supported, whereas in the memory and inhibition domains there was only partial convergence. Subtests in the flexibility, attention, and working memory domains showed little convergent validity with the Kilifi mostly because of attenuation effects.

Table 3.5: Convergent Validity of the BENCI Battery

	Kilifi Toolkit Tests										
		People Search test	**Digit Span test**	**Contingency Naming test**	**Self-Ordered Pointing Test**	**Verbal List Learning (VLL) test – Total CA**	**Nonverbal Selective Reminding Memory Test (NVSRT)**	**Tower Test**	**Ravens Progressive Matrices test**	**VLL Immediate Memory Span**	**VLL Level of Learning**
BENCI Tests	*Domains*	*Visual Sustained and Selective Attention*	*Auditory Sustained and Selective Attention*	*EF: Flexibility - Attention and attention shift*	*EF: Working Memory*	*Memory*	*Non-Verbal Memory*	*EF: Inhibition - Planning*	*EF: Reasoning*	*Memory*	*Memory*
Sustained Attention CPT (CA)	*Sustained Attention*	-0.103	0.053	-0.157	0.093	.303**	0.089	0.056	.288**	0.056	.266**
Sustained Attention CPT (RT)	*Sustained Attention*	0.123	-0.024	0.130	-0.043	-0.110	0.123	0.066	-0.151	0.066	-0.062
Working Memory (CA)	*EF: Working Memory*	.194*	-0.124	-0.047	0.004	.276**	0.085	-0.143	0.049	-0.143	.297**

BENCI Tests	Kilifi Toolkit Tests										
		People Search test	**Digit Span test**	**Contingency Naming test**	**Self-Ordered Pointing Test**	**Verbal List Learning (VLL) test – Total CA**	**Nonverbal Selective Reminding Memory Test (NVSRT)**	**Tower Test**	**Ravens Progressive Matrices test**	**VLL Immediate Memory Span**	**VLL Level of Learning**
	Domains	*Visual Sustained and Selective Attention*	*Auditory Sustained and Selective Attention*	*EF: Flexibility - Attention and attention shift*	*EF: Working Memory*	*Memory*	*Non-Verbal Memory*	*EF: Inhibition - Planning*	*EF: Reasoning*	*Memory*	*Memory*
Verbal memory (CA)	*Memory*	-0.006	0.005	-0.030	0.014	.414**	-0.181	-0.165	.346**	-0.165	.372**
Verbal Memory Delayed (CA)	*Memory*	-0.038	0.043	-0.089	0.116	0.193	-0.102	-0.171	0.162	-0.171	.212*
Verbal Memory Recognition (CA)	*Memory*	-0.010	-0.025	-0.001	0.068	-0.076	0.012	-.186*	-0.073	-.186*	-0.085
Planning Total Time	*EF: Inhibition - Planning*	-0.083	-0.176	.279**	-0.011	-0.030	-.405**	-.209*	-0.010	-.209*	0.004
Planning Time FO	*EF: Inhibition - Planning*	-0.169	-0.128	.219*	-0.070	-0.060	-.310**	-0.113	0.047	-0.113	-0.028
Reasoning (CA)	*EF: Reasoning*	-.367**	0.119	0.042	0.000	.424**	-.279**	-0.087	.206*	-0.087	.380**
Visual Memory Immediate (CA)	*Memory*	0.041	0.012	0.010	-0.081	.322**	-0.112	0.060	0.119	0.060	.234*
Visual Memory Delayed (CA)	*Memory*	0.038	-0.033	-0.007	0.064	.252*	-0.088	-0.179	.261*	-0.179	0.220
Visual Memory Recognition (CA)	*Memory*	-0.109	0.215	-0.057	0.092	0.129	-0.022	0.008	0.078	0.008	0.047
Spatial Stroop Flexibility	*EF: Flexibility*	-0.142	0.086	0.027	0.084	.414**	-.202*	-0.037	.327**	-0.037	.361**

TT – Total Time; RT – Reaction Time; CA – Correct Answers; CPT – Continuous Performance Test; FO – First Option; **. Correlation is significant at the 0.01 level (2-tailed); *. Correlation is significant at the 0.05 level (2-tailed).

The BENCI Functionality in Age and HIV Groups

As can be seen in Table 3.6, children not living with HIV outperformed those living with HIV on all BENCI tests. However, the mean group difference was significant in all subtests except Continuous Performance Test hits and reaction time, Go No Go hits, Verbal Memory Recognition hits, Processing Speed median reaction time, and Planning total time.

We checked whether the performance within the BENCI subtests aligned with developmental models' expectation of growth in cognitive performance as children aged, and report Pearson correlations between age in years and the BENCI subtest performance for the children living with HIV- and those not living with HIV separately in Table 3.7. We hypothesized that children not living with HIV would significantly outperform those living with HIV. Among the children living with HIV, there were significant associations in the expected direction between age and Verbal Comprehension Images hits, Verbal Memory hits, Verbal Memory Recognition hits, planning total time, Planning Time of First Option, Abstract Reasoning hits, Visual Memory Immediate hits, Visual Memory Recognition hits and Spatial Stroop omission errors. Among children not living with HIV, there was a significant association between age and Continuous Performance reaction time, Processing Speed reaction time, Verbal Memory hits, Abstract Reasoning hits, and Visual Memory Delayed hits. The lack of significant correlations between some cognitive indicators and age could be because of attenuation effects, but might also relate to sampling issues (e.g., older participants appearing in the sample because of delayed development and the repeating of grades in school).

Table 3.6: Mean Group Differences in BENCI Subtests Responses

		N	Mean	Std. Deviation	Std. Error Mean	Mean Difference	Significance (2-tailed)
Verbal Comprehension Figures Hits	HIV negative	317	7.030	.990	.056	.498	.000
	HIV positive	258	6.530	1.171	.073		
Verbal Comprehension Images Hits	HIV negative	318	7.540	.743	.042	.760	.000
	HIV positive	259	6.780	1.220	.076		
Continuous performance Hits	HIV negative	322	47.329	12.756	.712	1.928	.092
	HIV positive	264	45.401	14.891	.916		
Continuous performance RT Median	HIV negative	318	585.997	149.796	8.400	20.093	.070
	HIV positive	260	565.904	107.263	6.652		
Go No Go Hits	HIV negative	315	42.333	8.007	.451	.287	.703
	HIV positive	258	42.047	9.968	.621		
Go No Go Mean RT	HIV negative	315	.825	.007	.000	-.002	.020
	HIV positive	252	.827	.009	.001		
Processing Speed Median Reaction Time	HIV negative	313	584.931	146.541	8.283	20.6036	.054
	HIV positive	267	564.328	102.179	6.253		
Phonetic Fluency Hits	HIV negative	319	5.100	2.729	.153	1.658	.000
	HIV positive	271	3.440	2.546	.155		
Semantic Fluency Hits	HIV negative	321	7.310	3.295	.184	1.909	.000
	HIV positive	271	5.410	3.440	.209		
Working Memory Hits	HIV negative	319	10.510	6.060	.339	1.955	.000
	HIV positive	270	8.560	6.314	.384		
Verbal Memory Hits	HIV negative	264	6.240	2.249	.138	.747	.000
	HIV positive	206	5.500	2.040	.142		
Verbal Memory Hits Delayed	HIV negative	317	4.670	3.192	.179	1.125	.000
	HIV positive	270	3.540	2.706	.165		
Verbal Memory Hits Recognition	HIV negative	317	18.310	3.533	.198	.517	.089
	HIV positive	270	17.790	3.810	.232		
Planning Total Time	HIV negative	303	18178.770	12791.017	734.825	148.873	.891
	HIV positive	259	18029.900	12776.036	793.864		
Planning Time of First Option	HIV negative	304	3404.050	3027.831	173.658	-567.889	.034
	HIV positive	259	3971.940	3316.305	206.065		
Abstract Reasoning Hits	HIV negative	319	14.870	4.835	.271	4.261	.000
	HIV positive	269	10.610	4.775	.291		
Visual Motor Total time	HIV negative	320	70080.740	28797.814	1609.847	-18559.339	.000
	HIV positive	254	88640.080	35451.218	2224.407		
	HIV negative	315	74940.830	36611.539	2062.827	-22239.897	.000

		N	Mean	Std. Deviation	Std. Error Mean	Mean Difference	Significance (2-tailed)
Alternative Visual-Motor Total Time	HIV positive	255	97180.730	45435.071	2845.254		
Visual Memory Immediate Hits	HIV negative	278	6.140	2.295	.138	1.277	.000
	HIV positive	207	4.860	2.030	.141		
Visual Memory Delayed Hits	HIV negative	273	6.150	2.459	.149	1.289	.000
	HIV positive	201	4.860	1.990	.140		
Visual Memory Recognition Hits	HIV negative	314	44.580	5.122	.289	1.527	.003
	HIV positive	264	43.050	6.975	.429		
Spatial Stroop Hits	HIV negative	328	66.500	21.956	1.212	7.270	.000
	HIV positive	270	59.230	23.769	1.447		
Spatial Stroop Omission Errors	HIV negative	328	9.050	10.469	.578	-5.962	.000
	HIV positive	270	15.010	16.667	1.014		
Spatial Stroop Commission Errors	HIV negative	328	11.710	15.242	.842	-2.719	.032
	HIV positive	270	14.430	15.571	.948		
Spatial Stroop Mean Time	HIV negative	328	979.504	223.457	12.338	-68.133	.000
	HIV positive	270	1047.638	249.922	15.210		

Table 3.7: Age Correlations in BENCI Subtests Responses

	HIV Positive			HIV Negative		
	Pearson Correlation	Sig. (2-tailed)	N	Pearson Correlation	Sig. (2-tailed)	N
Age in years			252			292
Verbal comprehension Images Hits	.188**	.004	239	.038	.520	283
Continuous Performance Hits	-.050	.435	244	-.004	.945	288
Continuous Performance RT Median	-.054	.409	240	-.168**	.004	285
Go No Go Total Hits	.074	.253	239	-.023	.698	281
Go No Go Mean RT	-.113	.085	233	.048	.422	281
Processing Speed Median Reaction Time	-.123	.053	247	-.156**	.009	279
Phonetic Fluency Hits	-.063	.321	250	.034	.571	284
Semantic Fluency Hits	.025	.689	250	.006	.921	286

	HIV Positive			HIV Negative		
	Pearson Correlation	**Sig. (2-tailed)**	**N**	**Pearson Correlation**	**Sig. (2-tailed)**	**N**
Working Memory Hits	-.046	.468	249	-.100	.094	284
Verbal Memory Hits	.156*	.032	190	.255**	.000	234
Verbal Memory Hits Delayed	.048	.450	249	-.006	.921	282
Verbal Memory Hits recognition	.151*	.017	249	.034	.564	282
Planning Total Time	.128*	.049	238	-.015	.803	267
Planning Time of First Option	.237**	.000	240	.006	.922	269
Abstract Reasoning Hits	.156*	.014	248	.210**	.000	284
Visual Motor Total Time	-.126	.055	233	-.083	.161	285
Alternative Visual Motor Total Time	.105	.108	237	-.006	.914	281
Visual Memory Immediate Hits	.160*	.027	191	.082	.196	248
Visual Memory Delayed Hits	.096	.195	185	.230**	.000	243
Visual Memory Recognition Hits	.150*	.019	245	.037	.541	279
Spatial Stroop Hits	.106	.095	249	.034	.561	292
Spatial Stroop Omission Errors	-.152*	.017	249	-.082	.163	292
Spatial Stroop Commission Errors	-.013	.838	249	.003	.964	292
Spatial Stroop Mean Time	-.017	.788	249	-.072	.218	292

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

Confirmatory Factor Analyses

We tested the construct validity of Executive Functioning as proposed by Diamond for normal development (125). According to his model, the subtests that measure inhibition, flexibility, reasoning, memory, and fluency together constitute executive functioning (125). These are tests that evaluate the ability to make decisions, exercise self-control, pay attention, be creative, solve problems, and plan towards having good health and success in life. These are considered core functions in the brain hence the name executive functions. We fitted a confirmatory factor analysis model previously fitted successfully in the Arabic version of the BENCI and sought to adjust the model slightly to improve fit if necessary.

A second-order model with Executive Functioning as a second-order latent factor and five first-order latent factors (i.e., Fluency, Reasoning, Memory, Inhibition and Flexibility) measured by the specific the BENCI subtests (Figure 3.2) was specified and tested with the pooled sample including missingness handled by Full Information Maximum Likelihood. The model fit indexes suggested a good fitting model (χ2 (100, N = 604) = 245.55, p < 001, RMSEA = .049, CFI = .908, TLI = .875). However, this model had several issues. First, the Fluency factor was estimated to have a negative residual variance that we fixed at zero. Second, in this revised model, the Verbal memory factor also yielded an estimate negative residual variance that we treated similarly by fixing it at zero. Third, in the third model, the residual variance of the Alternate Visual-motor total time also needed to be fixed to zero. Next, we considered modification indices and found that the model could be improved if we included a covariance between the residuals of Reasoning and Flexibility and between the residuals of Semantic Fluency correct answers and Verbal Memory Recognition correct answers. This further modified model showed an acceptable fit (χ^2 (101, N = 604) = 205.73, p < .001, RMSEA = .041, CFI = .934, TLI =.911). Figure 3.3 presents the standardized factor loadings. An inspection of this model showed that not all indicators of Inhibition (Go No Go RT = λ -.46;

Go No Go CA = λ .74) had significant loadings on their respective factor, indicating that these specific tests did not measure Inhibition as intended (Figure 3.3). It also showed that the latent factor of Inhibition did not load on the Executive Functioning factor. Therefore, we removed the Inhibition factor together with its indicators and tested a second-order factor with only four factors. This model fitted well (χ^2 (51, N = 604) = 135.57, p < .001, RMSEA = .052, CFI = .944, TLI =.914). Figure 3.4 presents the factor loadings of this model. Therefore, the five components of Executive Functioning as validated before did not all show up in the Kenyan sample, while Executive functioning comprised of fluency, reasoning, verbal memory, and flexibility was found to fit well in the Kenyan sample. The final model with four factors each measuring executive functioning supports the construct validity for the BENCI battery, despite Heywood cases on the Alternative Visual-motor subtest.

AMOS treats missing data using full information maximum likelihood, which is considered a robust method for treating missing data. However, we checked whether model fit would be affected when using a dataset with no missing data. On running the model with no missing data, the model fit was excellent (χ^2 (51, N = 327) = 64.07, p > .05, RMSEA = .028, CFI = .968, TLI =.958). This shows that the BENCI does have good construct validity though some changes in some test items and instructions are needed in future revisions of some subtests.

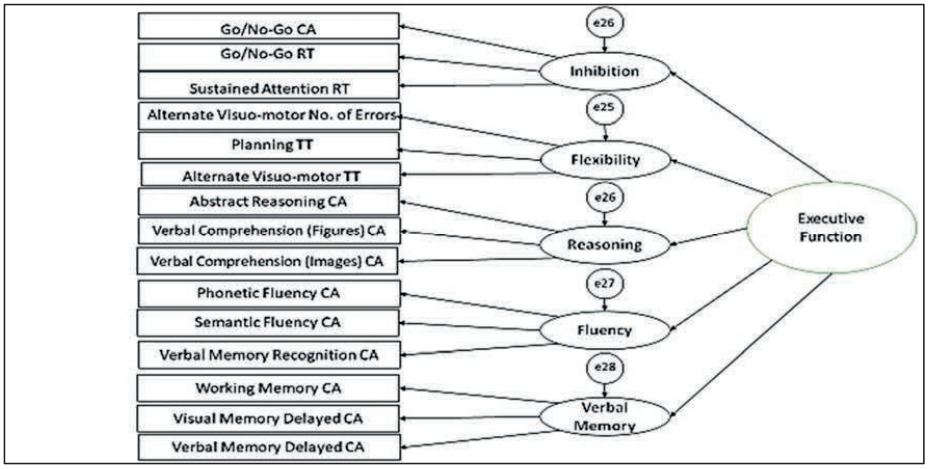


Figure 3.2: Five Factor Executive Function Model ($\chi 2$ (100, N = 604) = 245.55, p < 001, RMSEA = .049, CFI = .908, TLI = .875)

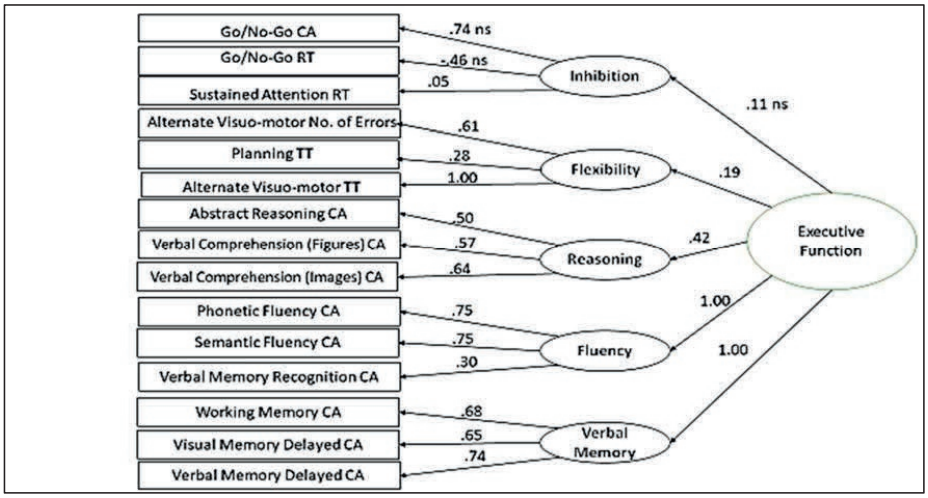


Figure 3.3: Five Factor Executive Function Model ($\chi 2$ (101, N = 604) = 205.73, p < .001, RMSEA = .041, CFI = .934, TLI =.911) ns – not significant

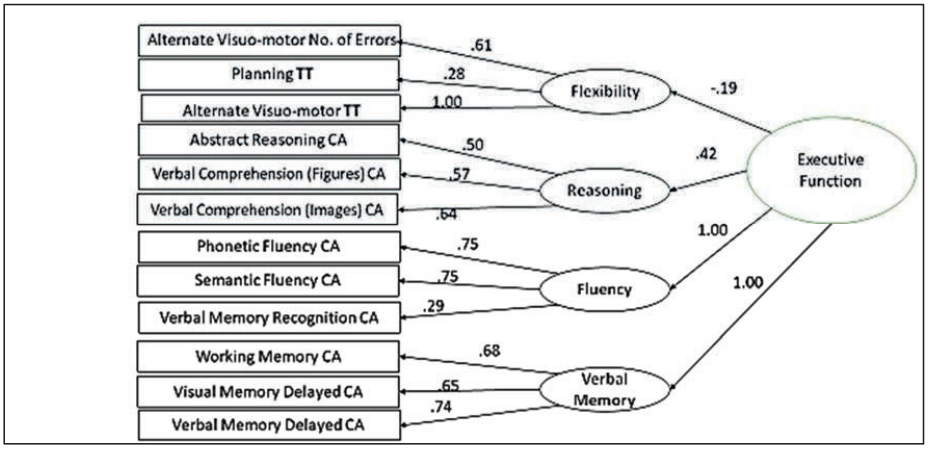


Figure 3.4: Four Factor Executive Function Model ($\chi 2$ (51, n = 604) = 135.57, p < .001, RMSEA = .052, CFI = .944, TLI =.914)

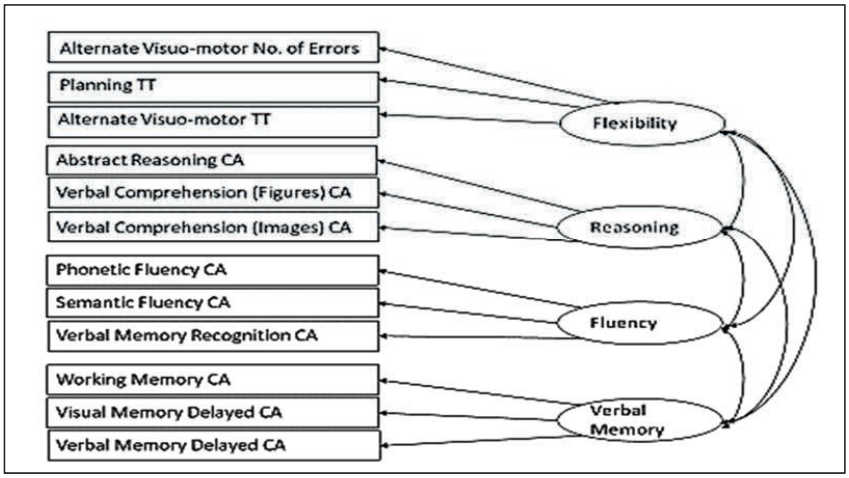


Figure 3.5: Four Factor First Order Model ($\chi 2$ (47, n = 604) = 107.76, p < 001, RMSEA = .046, CFI = .960, TLI = .933)

Measurement Invariance

We set out to test whether the BENCI behaves the same way across the HIV-positive (N = 274) and HIV-negative groups (N =330) using measurement invariance testing with multi-group confirmatory factor analysis. We used the factor model that was identified as having an excellent fit using the pooled sample as the basis and modified it to have only the four

correlated first-order factors (i.e., Fluency, Reasoning, Memory, and Flexibility, each of them had their observed indicators) but no second-order factor (which is not required for testing measurement invariance). The model fit was excellent (χ^2 (47, n = 604) = 107.76, p < 001, RMSEA = .046, CFI = .960, TLI = .933) as shown in Figure 3.5.

We first tested for configural invariance where all factor loading, item intercepts and residual parameters were freely estimated. The model fit indexes suggested a well-fitting model (χ^2 (94, N = 604) = 175.09, p < .001, RMSEA = .038, CFI = .941, TLI = .902). The factor loadings of all the indicators in both groups were significant.

We then specified a model for metric invariance where all the factor loadings were restrained to be the same across the two groups and all the other parameters were freely estimated. This model had a good fit (χ^2 (102, N = 604) = 198.35, p < .001, RMSEA = .040, CFI = .930, TLI = .893). On comparing the configural to the metric invariance model, we found that there was no statistically significant difference between the chi-square values, suggesting that the metric invariance was supported ($\Delta\chi^2$ =23.26, DF = 8, p = .003). This meant that the factor loadings were invariant and the indicator items across groups have the same associations with the latent constructs. Differences in other fit indexes also showed that the metric invariance was tenable (ΔCFI from configural to metric model < 0.01).

A scalar invariance model was then specified where the item intercepts and factor loadings were restrained to be the same across groups, while the latent mean of the latent factors in the HIV-positive group was released (with an aim to check latent mean differences in flexibility, fluency, verbal memory, and reasoning). This model had a poorer fit compared to the metric invariance model (χ^2 (110, N = 604) = 245.12, p < .001, RMSEA = .045, CFI = .901, TLI = .860). On comparing this scalar invariance model to the metric invariance model, there was a worsening fit due to constraints on the intercepts; this was due to a statistically significant

difference between the chi-square values of the scalar invariance and metric invariance model ($\Delta\chi^2$ =46.77, DF= 8, p <.001). The CFI difference also showed that the scalar invariance was not holding across all subtests (ΔCFI = 0.029). This indicates that some intercepts were not invariant and that these subtests are uniformly biased.

Using modification indices, we then specified a partial scalar invariance model where we constrained one intercept for each indicator at a time and tested whether this restraint resulted in a significant chi-square difference. For items for Verbal Comprehension (figures) CA and Visual Memory Delayed CA, the tests showed significant chi-square difference hence we freely estimated these two intercepts across groups while holding the rest of the intercepts and factor loadings to be the same across groups. This partially invariant model fitted well (χ^2 (108, N = 604) = 218.38, p < .001, RMSEA = .041, CFI = .920, TLI = .884). The fit for the partial scalar invariance was better than the strict scalar invariance, and the difference between the chi square values between this model and the metric invariance model shows that partial scalar invariance fits reasonably well ($\Delta\ \chi^2$= 20.03, DF = 6, p >.001). The CFI difference also showed that the partial scalar invariance was tenable (ΔCFI 0.010).

To summarize the series of measurement invariance tests, we conclude that metric invariance is achieved indicating that factor loadings of the BENCI are comparable across the HIV-positive and HIV-negative samples, and we can compare the association of the BENCI with other invariant constructs across the two groups, but not the mean comparisons of Verbal Comprehension (figures) CA and Visual Memory Delayed CA. These subtests are not well-calibrated. A partially scalar invariant model fitted the data reasonably well meaning you could compare mean difference for most of the subtests with caution for Verbal Comprehension Figures CA and Visual Memory Delayed CA.

Discussion

This study aimed to validate the BENCI battery in Kenya with children living with HIV and those not living with HIV and contribute to a toolset of evaluation tests for primary school students in Kenya and other similar settings. There were four main analyses to address internal consistency, test-retest reliability, convergent validity, and construct validity among 6 to 14-year-olds. The adaptation of the English version of the BENCI resulted in a battery with good test-retest and validity checks. We discuss each finding and its implications in detail.

Reliability

Some subtests were found to have floor effects due to having too many difficult items while others had ceiling effects due to having too many easy items. Too few and easy items resulted in ceiling effects for the language tests. The BENCI's subtests showed poor to excellent internal consistency with most subtests showing higher alpha values for the HIV-positive group than the school sample. This was likely caused by smaller attenuation effects in the subtests with ceiling effects or the HIV-positive group showed more variation in true scores leading to higher Alphas as seen in the N-back working memory test (131). The internal consistencies in our study were similar albeit slightly lower than those found in the Moroccan sample, possibly because the level of difficulty of the test items suited the younger cohort in the Moroccan sample better than in our data. This points to the need to develop age-appropriate norms and to add items with age-suitable difficulties in future revisions.

Our results for the BENCI test-retest reliability were fairly similar to a previous study conducted in Morocco (40). The Arabic adaptation of the same tool reported Intraclass correlation to range from -.23 to .81, similar to our study (40). However, the poor test-retest reliability of the reasoning test could be due to the relatively long-time interval between the two assessments in our study as a longer interval may create changes in the construct (132). It is possible that the respondents were thinking about the test items more often than before the

first administration (132). The latter is more likely with children who have high mental imagery skills meaning they are likely to think about the test items quite often and grow familiar with them and forthwith give different responses in the second assessment (133). A child may respond substantially different in a language test whose retest is one year compared to verbal memory because their language ability has improved well past their memory ability. Studies on cognitive tests have had a re-test time interval of 15 to 60 days though there were recommendations for within a 14-day lapse of time especially for tests such as visual memory which would lose reliability over longer durations (40, 41, 132, 134, 135). However, some studies have shown that for verbal memory and visual motor speed tests, the test-retest reliability with a one-year time-lapse remains stable while for language tests a recommendation for not less than 14 days has been made (135, 136). The mixed results in our study suggest that the test domains and time lapses play a role here (41). Our test-retest results in attention tests are also similar to those of other studies that show higher reliability in attention speed tests compared to accuracy tests (137). Tests that call for speed over accuracy have been found to have high reliability than those that call for accuracy over speed (137).

Convergent Validity

BENCI attention tests do not correlate with Kilifi's People Search and Forward Digit Span as expected, but they showed convergence with tests that had attention components. Studies have cited the tendency of attention tests to confound with other cognitive functions (138-140). In our study, similar administration processes between tests with attentional components could have contributed to convergence as seen in the BENCI's Working Memory test with Kilifi's People Search test. These two are attentional control tasks as they call for a response to correct stimuli during incorrect stimulus thereby inhibiting a response. Correlations between attention tests have been found to support convergent validity with a range from low-to-moderate. Speed measures have higher significant correlations compared to accuracy

attention measures (137, 141). In our study, however, the BENCI attention accuracy tests showed moderate convergent validity while attention speed measures showed weak convergent validity. Poor convergent validity between some attention tests has been documented in other studies (138). In the memory domain, BENCI's working memory and Kilifi's people search correlated well, a finding that has also been found in other studies comparing working memory tests to attention tests.

BENCI's Visual Memory test showed a weak correlation with Kilifi Toolkit's Nonverbal Selective Reminding Memory Test (NVRST). The administration is similar between these two tests. An explanation for this could be found in studies showing the impact of familiarity with the tools on scoring. In our study, the NVRST test involved memorizing the shape formed by a set of 8 dots and then replicating the shape by placing a marble on a set of dots. In the BENCI version, the child was supposed to memorize several images and then correctly point them out when shown amidst a set of other pictures; a task that would involve other cognitive functions such as visual-motor coordination. Pointing out pictures is a familiar learning concept in the Kenyan context. This is because among the methods used in teaching pre-schoolers is by pointing out images and encouraging the children to read and memorize their names. The administration was fairly similar but their scores in terms of correct answers were not highly correlated. Probably other psychological processes are involved in the BENCI subtest that are not in the Kilifi subtest. There are some studies that have found a similar lack of correlation between tests. In a study done in Zambia, a non-verbal test called draw-a-person was locally adapted and the two tests, the original and adapted one, were compared and found to not be correlated (142). However, when the ratings were done by adults and correlated to educational outcomes, the two tests had significant correlations. Further research can explore similar comparisons between uncorrelated tests to find out if other psychological processes are involved. Such an evaluation could be similar to the one conducted in the Zambian study. This

is in trying to find out whether the BENCI visual memory test expectations do truly reflect the cultural indicators for non-verbal memory. However, the NVRST in the form of Children's Memory Scale (CMS) dot location subtest has also been found not to have significant correlations with the Lebby-Asbell Neurocognitive Screening Examination (LANSE) visual memory test (140). In addition, NVRST administration involves visual-motor coordination and other cognitive functions in addition to memory.

Computerized assessments are preferred due to ease of administration and scoring as well as precision (41). However, Kenyan children are not very used to computerized assessments and a lack of familiarity may introduce variance in test scores that are not related to the construct being measured. Some of the factors that have been known to introduce construct irrelevant variance with computerized assessments include proficiency with the computer-based tests, ease of interaction with the platform, speediness of the tests and test-taker's anxiety (143). Some administration processes, such as tasks calling for inhibitory control, within the tablet may affect some domains more than others (144). The lack of familiarity and some administration processes associated with tablet-based testing could affect convergence validity when compared to some paper-based tests. However, there are some studies that have shown no significant differences in test performance between tests using computer-based platforms and those using paper-based ones meaning that variation in convergence may apply to some tests more than others (144). To reduce variation in some of these tests, studies have suggested several approaches including reducing the difficulty level of computer-based tests as well as clarifying the relationship between tasks and the expected test takers performance (143, 145). It is however, beyond the objectives of this study to investigate approaches that would have worked best in reducing validity variance between the BENCI and Kilifi toolkit. These are next level questions to consider.

The lack of convergence in some tests may also be contributed to by lack of a common construct between some of the BENCI and Kilifi toolkit tests. Since the latter is the gold standard, comparing it to a test that that does not capture the same constructs may give us erroneous findings. Differences in correlations between measures have been found to increase when comparisons are made to alternate measures with low convergence validity (145). Improvements and adaptations of some of the BENCI tests may improve convergence with the Kilifi toolkit tests.

The BENCI Functionality in Age and HIV Groups

The BENCI highlighted clear mean differences between the HIV-positive and HIV-negative groups. Just as indicated in the BENCI results, tests can have mean differences but the score differences between the groups may not be significantly different as seen in the scores for correct answers in the inhibition test and time taken in the planning test. An earlier study showed that certain tests like inhibition and planning can have the ability to differentiate healthy from unhealthy populations but the difference in scoring within the tests may not be significantly different (146). However, the BENCI did affirm what other studies have found that children living with disease score lower than children living without disease in tests of working memory, inhibition, memory, and planning among other cognitive functions (140, 146-148). Moreover, taking more time when doing a test has been associated with taking more mental effort to achieve a desired outcome, in this case a correct response, entails a healthy approach to inhibitory tasks (125). Better performance in correct answers is denoted by higher scores while in reaction time it is denoted by lower scores. Therefore, for children having high reaction time, performance will be regarded as poor. Overall, this is true when the dependent variable is time but not when it is accuracy. For example, higher reaction time is worse than a lower one in Selective Attention, Sustained Attention and Go/No-Go tests. These findings add

to the body of literature on the significance of testing for cognitive deficiencies among unhealthy children.

Construct validity

The planning test did not have significant loading on the inhibition factor in the pooled sample and subsequently, this factor did not load well onto executive function. This has not been the case in another Sub-Saharan African study that supported the construct validity of a planning test (86). Inhibitory control has been found to be higher in children within settings that emphasize obedience and self-control such as East Asian countries and been found to be lower, to a point where there are no significant age differences, among children in developing countries and communities (149). The study also reported cross-national differences in inhibition, shifting, and updating. We would then expect the children in this study to have the BENCI inhibition tests to load onto executive function just like other western adapted tests have done in a sub-Saharan setting. This is more so since inhibition tends to develop rapidly among younger children hence, we would not expect a lack of this cognitive function among 6- to 14-year-olds even though inhibitory control tends to mature at adolescence (125, 150). However, studies looking into whether maturity of inhibitory control affects how well the function can load into an executive function model may clarify the results we found in this study. Observations of the school and home executive function stimulation activities give a broader picture of the activities emphasized and how they encourage inhibitory control development. These observations could be integrated in further research with the BENCI. Flexibility on the other hand builds developmentally onto inhibition and loads well on executive function. This finding does not reflect the arguments pointed out earlier on inhibition. Inhibition is a first-order component that appears around 6-8 years and flexibility is a second-order component that appears later in development (151). Since flexibility loaded well onto executive function, the lack of significant loadings in the inhibition construct could potentially

be because of the lack of culturally aligned items in the inhibition tests or a problem with instructions. The findings in the construct validity indicators call for a developmental approach when interpreting scores and the need to norm the BENCI for age groups.

The BENCI also showed support for metric and partial scalar invariance as opposed to strict scalar invariance. This means that the BENCI items are loaded onto the latent factors similarly across groups, hence can be compared across the groups. The same applies to items per subtest. However, comparability of means between the latent factors was not supported in its entirety meaning that we cannot compare the means of fluency, flexibility, verbal memory, and reasoning across the groups. We can choose to create separate norms for HIV+ and HIV- groups since the tests behave differently in the two groups, but this will not give us an opportunity to compare performance. One of the options that can enable performance comparison is to create norms with the healthy and optimally functioning group, but caution should be integrated when norming for Verbal Comprehension Figures and Visual Memory Delayed tests. We may underestimate or overestimate between-group abilities due to miscalibration of the tests and the results may be marred with measurement bias. This means that we may not have true between-groups construct differences due to other construct irrelevant variables causing differences in test scores. In this case, we may choose to correct for intercept differences during norming by estimating their effect sizes and relating this to effects on the norm scores (152). As an alternative, we can choose to carry out a study on why the two tests are biased and correct for any item level (attenuation effects in Verbal Comprehension Figure). We are yet to come across a study that investigates measurement invariance of a neurocognitive tool in Kenya and its regions. Children studies that we have come across are based in high income countries and cannot be compared to our setting due to different group dynamics and cultural dynamics that underlie cognitive performance and developed test items (47).

Limitations

In this study, one drawback was that the results could only be generalized in a community setting and not a clinical one. We could not find comparison tests for some domains due to the limited availability of validated tools within the Kenyan culture.

The study also noted that some subtests had floor and ceiling effects, which compromised the interpretation of other findings. In this case, any results pertaining to the subtests having ceiling and floor effects should be interpreted with caution. Moreover, further studies may revise the tests by perhaps adding more items to the tests with ceiling effects and decreasing the difficulty of the items in the tests with floor effects so as to match the difficulty to ability level and reduce attenuation effects. In addition, age-appropriate norms for the subtests should be considered.

The methods used to capture reaction time and total time may not have been completely accurate because the paper-pencil tests used a stopwatch that is prone to administration errors while the iPad-based tests used an internally configured watch. In the paper-based tests, errors may be integrated when timing is not stopped immediately a task is completed or when an administrator gives more time for task completion than would be required. These can create systematic or random measurement errors where the latter could suppress correlations. This may have been the case in convergent validity where random measurement could have suppressed some correlations. Nevertheless, the possibility of errors in paper-based tools is another reason to prefer automated computerized tests with internalized and consistent timing across participants.

Conclusion

The Spanish version of the BENCI was successfully adapted to English, and its psychometric checks showed that it had good convergent validity in reasoning and some memory and inhibition tests. However, further research is needed to fully understand the non-

verbal memory, working memory and flexibility tests from a convergent validity view. The BENCI was also found to have good discriminant validity with only a few tests not showing a significant difference between the case and control populations. Construct validity showed good goodness of fit indicators though the inhibition did not load onto executive function as expected. Future language adaptations can consider Kiswahili translations which is Kenya's national language.

HIV is a known risk factor for poor neurocognitive outcomes due to its negative impact of CNS and exposure to a host of negative psychosocial factors. We therefore hypothesized that children living with HIV would perform worse than those who are uninfected. Confirming our hypothesis, children living with HIV performed significantly worse than those who were uninfected, thus showing that the BENCI is sensitive to a well-documented biological risk factor.

Chapter Three Appendixes

Appendix 3.1: Supplementary Table for Pilot Study BENCI Observations, Respondents Feedback and Researchers Recommendations

Subtest	Observations by Two Testers	Recommendations by the Research Team
Verbal Comprehension - Figures	a. The instructions do not prepare some children well for the test. The instructions orient the child to tapping either to the right or left of an image. However, the test requires the child to tap a figure with certain traits. Probably the wording should be changed to exude this. b. Some respondents tend to tap on 2-3 figures rapidly before the next instructions. The instructions should therefore indicate that only one figure should be tapped on per instruction. c. I did not have a problem on my end regarding the instructions. The instructions are simple enough to follow once they see the task itself.	a. Maybe in the instructions instead of saying 'figures' we use the word 'shapes' or 'images' cause that is what we are taught they are called in school. b. Provide new instructions adequate for Kenyan children "Follow the instructions indicated below. If the instructions ask you to tap at the right or the left of the images, you will have to tap at your right or your left. Press the START button to launch the test" c. We can use the training section of the test to succinctly orient the children to the test.
Working Memory	a. Clear instructions and intense training needed as younger children tend to apply the same instructions as the ones in previous tests precluding the grouping bit. Also inform the child that she/he needs to wait for the microphone. b. I feel that part of the instructions that was difficult for them to understand was that they are meant to repeat either colour first only and then the numbers next because they just repeated what they heard.	a. We have to fix the order, for instance, first number and then colours. This is important for the standardization of the test. b. Make the instructions clear and use the training option to evaluate whether the child has understood the instructions.

Subtest	Observations by Two Testers		Recommendations by the Research Team
Planning	**Visual Memory**	**Verbal Memory with Delayed Trial**	
a. Two children found it really hard to reason out what was needed. I will explore further to find out an easier way of doing this. Probably simpler English would do. b. When reading out the instructions to them (in Kiswahili), I used examples that they knew. Instead of 'theme park,' I talked of 'Uhuru Park' which most are familiar with, and instead of 'tokens' I use the word 'shillings.	a. Screen is very sensitive to any tapping so child's hands should be off the screen if not responding. b. Some of the images that respondents had difficulty naming or telling apart include Cloud/bush and hair/head c. There are instructions still in Spanish. d. I did not experience a challenge with this task. e. The instructions of 'Yes/No' is not in English.	a. Have the children keep their hands off the screen unless they are responding. Double tapping the screen or idly passing their fingers on the screen may interfere with preferred responses. b. The instructions in between are not in English. c. The instructions of 'Yes/No' is not in English.	a. Advice the children to keep their hands off the screen if not responding. b. Translation "Next, you will hear a list of words. You have to tap "yes" if that word was in the list of word that you listen before or tap "no" if not."
a. 'Shillings' is more comprehensible than 'Tokens.' Also call it a 'play park.'	a. Advice the children to keep their hands off the screen if not responding. b. Images with two possible names, for example hair/head. In this case, we only have to specify in the manual that hair or head are correct answers for Kenyan children. c. Translation "Next, you will see several images. You have to tap "yes" if that image was in the list of images that you saw before or tap "no" if not."		

Subtest	Observations by Two Testers	Recommendations by the Research Team
Continuous Performance	a. Two children found the test too long and had to be encouraged to finish. One gave up and did other tests after which she was encouraged to finish. b. I timed this test, and it took 9 minutes. The test is configured to have 3 blocks.	a. I realize that is a very boring test, but it should be so. The main objective is to sustain attention in a low demanding task. Children with attention problems can sustain attention during the first minutes but after that, they fail the task and produce more errors. However, if you feel that it is too long, we could reduce the duration. In my opinion, it should be 8-10 minutes at least. Nine minutes is quite acceptable even for 6 years old. b. If reducing the duration will compromise the test, then we can maintain the configuration and schedule to have a break right before this test.
Alternate Visual Motor	a. The instructions indicate that there will be numbers and letters, but the test has numbers only with some of the numbers inside squares and others inside circles. The previous test before upgrading to the app version had test as per instructions (letters and numbers). b. Even in Kiswahili these instructions are unclear, but the training helped a little bit.	a. Use the settings to reconfigure this to animals and fruits instead. Children can start with the animals first then the fruits. b. Let us do more training.
Abstract reasoning	a. Translation needs to be done *'complet a la serie.'*	a. Translation "Complete the pattern"

Subtest	Observations by Two Testers	Recommendations by the Research Team
Go-No-Go	a. The instructions are simple enough for one's understanding. However, the sound that comes as a result of tapping the elements confuses the children such that when the sound, they are supposed to be watching out for comes, they are unable to hear it. I understand, however, that it is part of the test. A lot of training needs to be done so that the child gets to familiarize themselves with the sound. b. I gave all my instructions to the children in Kiswahili because they were all from poverty area and I realized that the English instructions would not be understood easily especially those tasks with a lot of wording. E.g., go-no-go, working memory, visual and verbal memory with delayed trial, planning.	a. Even as we are translating the instructions, we need to make them simple enough for them to understand and ensure that the task is also understood.
Semantic & Phonetic Fluency a. The training in phonetic fluency and semantic fluency uses examples that are also used in the test. Is this appropriate?		a. This is to be fixed. The training set should be different from the testing set.

CHAPTER FOUR

The Effects of Height-for-age and HIV on Cognitive Development of School-Aged Children in Nairobi, Kenya: A Structural Equation Modelling Analysis

Published Article: Maina, R., He, J., Abubakar, A., Perez-Garcia, M., Kumar, Jelte M. Wicherts. The Effects of Height-for-age and HIV on Cognitive Development of School-Aged Children in Nairobi, Kenya: A Structural Equation Modelling Analysis. Front. Public Health - Sec. Children and Health. Volume 11 - 2023 | Doi: 10.3389/fpubh.2023.1171851

Abstract

Background. Empirical evidence indicates that both HIV infection and stunting impede cognitive functions of school-going children. However, there is less evidence on how these two risk factors amplify each other's negative effects. This study aimed to examine the direct effects of stunting on cognitive outcomes and the extent to which stunting (partially) mediates the effects of HIV, age, and gender on cognitive outcomes.

Methodology. We applied structural equation modelling to cross-sectional data from 328 children living with HIV and 260 children living without HIV aged 6-14 years from Nairobi, Kenya to test the mediating effect of stunting and predictive effects of HIV, age, and gender on cognitive latent variables flexibility, fluency, reasoning, and verbal memory.

Results. The model predicting the cognitive outcomes fitted well (RMSEA = .041, CFI = 0.966, χ^2 = 154.29, DF=77, p < .001). Height-for-age (a continuous indicator of stunting) predicted fluency (β = .14) and reasoning (β = .16). HIV predicted height-for-age (β = -.24) and showed direct effects on reasoning (β = -.66), fluency (β = -.34), flexibility (β = .26), and verbal memory (β = -.22), highlighting that the effect of HIV on cognitive variables was partly mediated by height-for-age.

Conclusion. In this study, we found evidence that stunting partly explains the effects of HIV on cognitive outcomes. The model suggests there is urgency to develop targeted preventative and rehabilitative nutritional interventions for school children with HIV as part of a comprehensive set of interventions to improve cognitive functioning in this high-risk group of children. Being infected or having been born to a mother who is HIV positive poses a risk to normal child development.

Key Words: Stunting, Mediation, HIV, Lower school students, Executive functioning, Reasoning, Flexibility, Lower & Middle-Income Countries.

Introduction

Stunting (a height-for-age Z score of below -2 SD) (153) affects more than 149.2 million children worldwide and is associated with cognitive impairment (48, 154, 155) linked to poor academic performance (156, 157). Children who are stunted are at risk of underperforming in school and consequently dropping out (158). Over time, decreased years of education may result in low intelligence or cognitive ability (159). These may further contribute to long-term effects of reduced income and increased poverty (153, 158, 160, 161). Indeed, Hoddinott and colleagues (161) found that stunting at two years was associated with increased probability of poverty in adulthood. Stunting has also been associated with increased mortality, morbidity, and a vicious cycle of stunting between mothers and children if left unaddressed (48, 160). This cycle is characterized by stunted mothers who tend to have a higher probability of lower age at first birth and multiple births (161) leading to increased nutritional demands on the mother (48); if not met, may lead to undernutrition in children (162). Moreover, mothers with a history of stunting, are likely to have short stature/adult height (158) which is linked to obstetric complications during birth and having children with small gestational age (SGA) (158, 160). SGA has been associated with up to 20% of stunting in children under the age of 5 years (160). This cyclical disability effects of stunting have attracted worldwide attention with underlying factors such as poverty and hunger forming part of the amelioration efforts in the sustainable development goals (163). These two factors have been incorporated as targets of intervention given the proven association between stunting and poverty (158, 161), and hunger resulting in deficient diets that do not meet the nutritional standards needed to prevent stunting (162). These primarily nutrition-specific interventions have achieved population-wide traction and success in reducing stunting. For example, countries have already put in place measures to curtail stunting that have borne fruits, with Asian countries showing a stunting decrease from 49% to 28% between 1990 and 2010. However, in Africa stunting has

remained stagnant at around 40% (48). To achieve results similar to those achieved in Asia, nutrition-specific interventions have been primarily advocated assuming that they will reverse the effects of stunting (160). However, an increase in height does not necessarily mean that the child's cognitive function is restored and working according to age. Nutrition may increase a child's height but not necessarily ameliorate cognitive impairment post stunting because other factors may also impair cognitive functioning in a child with stunting. Specifically, other factors such as HIV infection, poverty, and poor health may affect cognitive outcomes, particularly in low and middle-income countries (LMICs) (5). Children with short stature (having a height that is well below that of other children of the same age and sex) may exhibit poor or delayed cognitive development for various reasons. For example, stunting is highly prevalent among children with HIV (28.6% in Kenya) (164) and children who are both HIV positive and stunted could have worse cognitive outcomes. Investigating causal mechanisms between stunting and cognitive performance may inform the alignment of stunting interventions to programmatic goals for the comprehensive management of HIV for school-going children.

HIV, Stunting, and Cognitive Development

Given normal cognitive development, children's cognitive functioning develops because of environmental factors and brain myelination among other neurological mechanisms and other factors involved in cognitive development (165, 166). HIV has been found to negatively impact cognitive function (167), and stunting (168) may partially explain this link. HIV is neurotropic meaning it directly affects the central nervous system (CNS) which may lead to cognitive impairment (169). Indirectly, HIV infection puts children at risk of undernutrition through inadequate and imperfect absorption of food, opportunistic infections, some HIV drugs, and other aetiological factors (170).Chronic undernutrition manifests itself as stunting. Children who are stunted have been found to perform poorly in receptive

vocabulary and numerical ability compared to children who are not stunted, whereas children infected with HIV perform poorly in receptive and expressive language and attention compared to those without HIV (10, 167). Children with HIV are also found to have poorer cognitive performance in draw-a-person task and digit span, (22) and working memory and executive functioning (171) though some studies have not found any difference in general cognitive function (6). The few earlier studies on both HIV and stunting among school-age children or lower school students used only a partial set of cognitive functions (172, 173). Stunting in children has been found to predict performance in reasoning, memory, language, executive functions, and motor ability (172, 173) while HIV predicts performance in nonverbal cognitive abilities, executive function, processing speed, memory, planning, reasoning, working memory, and visual-spatial abilities (2, 22, 23). These cognitive functions fall short of the recommended assessment domains (45, 174) of memory, language, attention, perceptual-motor, executive function, and social cognition (45). Therefore, more research using a broad battery of tests could shed light on how both HIV and stunting affect cognitive development.

We study the predictive effects (in relation to our model) of stunting and HIV on cognitive outcomes, while also considering age (175-177) and gender (165, 178-181) as relevant factors in predicting both stunting and cognitive outcomes. Age is central to stunting because the definition of stunting includes height-for-age ratio (153). Moreover, cognitive performance normally increases with age right from birth (166, 173, 182), although the developmental trajectories might vary over cognitive functions (151, 178) and might differ between children (183) for a host of reasons such as nutrition, exposure to HIV, parental education, and parental income (5, 22, 172, 173). Gender is relevant to our understanding of the effects of HIV and stunting on cognitive development because gender differences that vary in strength and direction have previously been found in relatively healthy children's populations (165, 178, 179).

In this study, we investigate the mediation effects of stunting (as measured by height for age) of the link between HIV, age, and gender on cognitive functions recommended for assessment in the diagnostic statistical manual of mental disorders version five (DSM V) among school-age children from Kenya. The DSM V recommended classification of neurocognitive domains was preferred in this study due to the domains' consistent with available knowledge on aetiology of neurocognitive disorders and their impaired cognitive functions and with assessment criteria developed by experts (44, 125). We also prefer DSM V criteria because we hope that the findings can inform an integrated approach to clinical management of children with HIV. We hypothesized that stunting would partially mediate the effects of HIV on cognitive outcomes among 6-14-year-olds. The study used the Computerised Battery for Neuropsychological Evaluation of Children (BENCI) (184) – a cognitive battery that has good validity and reliability for diverse cultures including low-income settings and that can measure the cognitive functions recommended in DSM V. The outcome of this study may shed light on which cognitive domains are most impacted by stunting within a population infected with HIV to inform future interventions for improving cognitive functioning.

Methods

Design and Setting

We evaluated the effects of stunting and HIV status on cognitive functions in a cross-sectional case control study among 6 to 14-year-olds within an HIV programme and three public schools. This study was part of a larger study that validated the Computerised Battery for Neuropsychological Evaluation of Children (BENCI) in Kenya (169).

The HIV uninfected sample was taken from three public primary schools in a middle-class urban setting. The schools follow the Kenyan government structured curriculum where children aged 6 years are in grade 1. The case sample was taken from a HIV programme in a

middle-class urban setting. The programme provides a community-based intervention to address medical, social, and economic needs of HIV positive children and their families. Both study settings are in Kenya's capital city, Nairobi. Nairobi's population is above the national poverty average (36.1%) and also above the national severe stunting average (11.4%) (185). Nairobi's food consumption relies heavily on food production from other regions within the country and its inhabitants spend more on food than those in rural regions with the major food category being cereals (186).

Ethics Approval

The study received ethics approval from Tilburg University's School of Humanities Research Ethics Committee (REC# 2017/25) and the Kenyatta National Hospital/ University of Nairobi Ethical Review Committee (P556/07/2016). Additional approvals were sought from the County Government of Nairobi. Heads of the study sites authorized the study, while the caregivers gave informed consent, and the children gave assent after a careful explanation of what the study entailed.

Study Sample Characteristics

Children who met the eligibility criteria were recruited into the study. The inclusion criteria were all children aged 6-14 years and, for the control group, not having any medical condition as reported by the school and the students themselves, while for the experimental group, not having comorbid conditions as reported by caregivers and children themselves. We excluded children with comorbid and/or severe medical conditions associated with being HIV-positive as indicated in their medical reports. Children were recruited from four clinics within the HIV programme and three public primary schools. In the clinics, the staff helped in generating a database of children who met the inclusion criteria, and we aligned our recruitment process to their next hospital visit, which was also a play day for the children and fell on a

weekend. On attending their scheduled clinic appointment, the parents of potential participants were randomly identified and informed about the study while being requested to sign up for the study. In the school setting, the same procedure was undertaken though the teachers here helped in randomly selecting the students who met inclusion criteria. Language wise, the Kenyan government obligates parents to send all children to school. The language of instruction in the schools is English, although the children prefer to use Kiswahili, Kenya's national language, in their daily communication.

Data Collection Procedure

The data collection was conducted by clinical psychologists. Once consent was given, the children were immediately shown to a room in which data collectors designated them to a table with an iPad. After anthropometric measures were taken, children completed the cognitive assessment using the BENCI on the iPads, which took around 90 to 120 minutes. There was a 10-minute break between the BENCI subtests which was scheduled right after the sustained attention subtest which can be tedious for children.

Measures

The computerised BENCI has been adapted and validated for use among Kenyan children aged 6-14 years in urban settings (169, 184) and has seventeen tests that measure the following: processing speed, motor coordination, attention (sustained and selective), memory (verbal and visual), language (comprehension and production), and executive function (updating / monitoring, inhibition / impulsivity, flexibility, working memory, planning) our study (40). A detailed description of the tests, their administration, and scoring has been written about elsewhere (169, 184). Using the same data as reported here, BENCI has been found to have good test-retest reliability for most subtests and sufficient internal consistencies ranging from .50 to .97 (169). A four-factor model consisting of flexibility, fluency, reasoning, and

verbal memory fitted the data well (RMSEA = .052, CFI = .944, TLI =.914) and showed metric (RMSEA = .040, CFI = .930, TLI = .893) and partial scalar measurement invariance (RMSEA = .041, CFI = .920, TLI = .884) between the HIV positive and negative groups (169). The tool has shown convergent validity with reasoning, memory, and inhibition tests from a local test battery called the Kilifi Toolkit (169). The BENCI was ideal for this study because it integrates tests that measure neurocognitive indicators recommended in DSM V and its good psychometric properties in our setting. We collected socio-demographic information about age, gender, weight, and height. Age was determined from the year of birth and calculated in terms of complete years and months since birth. Weight in kilograms was measured by a body scale as per the WHO protocol (187), and height in meters was measured by a tape measure.

Age was measured as complete years while gender was measured as either male or female.

Analyses

Stunting was calculated using the height-for-age z score (HAZ) based on 5- 19-year-olds' WHO Child Growth Reference standards where age was calculated in months (188). A WHO developed syntax was used to compute height-for-age (188). Children who were not stunted were defined as having a HAZ of $\geq$ -2.0 SD, those moderately stunted scored <−2.0 SD to > −3.0 SD, while those that were severely stunted scored $\leq$ -3.0 SD (188). The analyses used a continuous variable defined as height-for-age z score to measure stunting, with lower z values indicating more stunting.

We used maximum likelihood estimation in AMOS (189) to fit a structural equation model (SEM) as depicted in Figure 4.1. We used SEM as opposed to a multivariate path analysis as SEM allowed us to estimate a well-fitting complex model featuring latent cognitive variables underlying subtest scores (190). The model tests the mediating effect of height-for-age and

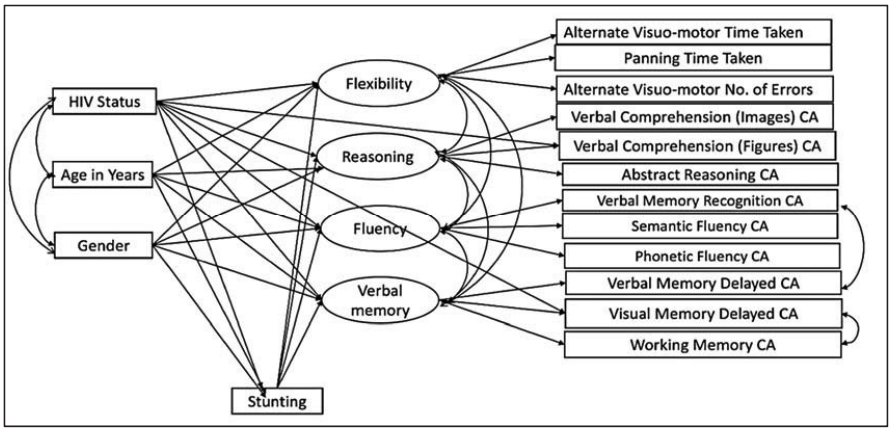
includes all direct effects of HIV, age, and gender on the cognitive latent variables. The measurement model for the BENCI had been previously confirmed in another study using the same data (169). We adapted the model a bit due to the partial scalar invariance findings. The adaptation involved additional direct paths from HIV to Verbal Comprehension Figures CA and Visual Memory Delayed CA to accommodate the intercept differences identified in the earlier validation study.

Our previous paper on adapting and validating the BENCI in Kenyan children outlines the data cleaning process, including decisions in dealing with problematic data (169). Also see appendix 4.4. The missing data pattern was not completely at random (Little's MCAR test $\chi 2 = 2455.2$, DF = 1725, $p < .001$) but was not significantly related to factors that may have produced a missing pattern (169). Little's MCAR test is also sensitive to non-normality, which might also play a role in the missing data pattern. However, we used data imputation in AMOS to check for modification indexes and calculate bootstrapped indirect effects. The modification indexes were used to check whether adding some paths would improve the model through a method of forward selection. Without overfitting the model too much, we decided ad hoc to add two residual covariance based on improper estimates of negative residual variances and modification indices. Residuals of Verbal Memory Recognition and Verbal Memory Delay were positively correlated, arguably due to the use of the same items across these indicators. The other residual covariance between Visual Memory Delay and Working Memory was unexpected but implemented to improve model fit. No further adjustments in the model as shown in Figure 4.1 were made. Bootstrapping based on 1000 samples was performed to determine the significance of the direct, indirect, and total effects as well as their standard errors. We also fitted a model in which effects of HIV, age, and gender were fully mediated by stunting and ran a specification search model (Figure 4.2) in AMOS using the model in Figure 4.1 to assess the robustness of the results. We compared the fit of the models to assess

mediation by stunting. Model fit was evaluated using goodness of fit indicators where an excellent fitting model would have a non-significant Chi-square test, Tucker Lewis Index (TLI) $\geq$ 0.95, Comparative Fit Index (CFI) $\geq$ 0.95, and Root Mean Square Error of Approximation (RMSEA) $\leq$ 0.08 (191). The term 'predictive effect' is used in this study to refer the hypothesized direction or the arrows within the model and finding prediction does not preclude that other factors have a role in causality.

We used a sample size of 604 as calculated in our paper putting forth psychometric validity of BENCI[2] (169).

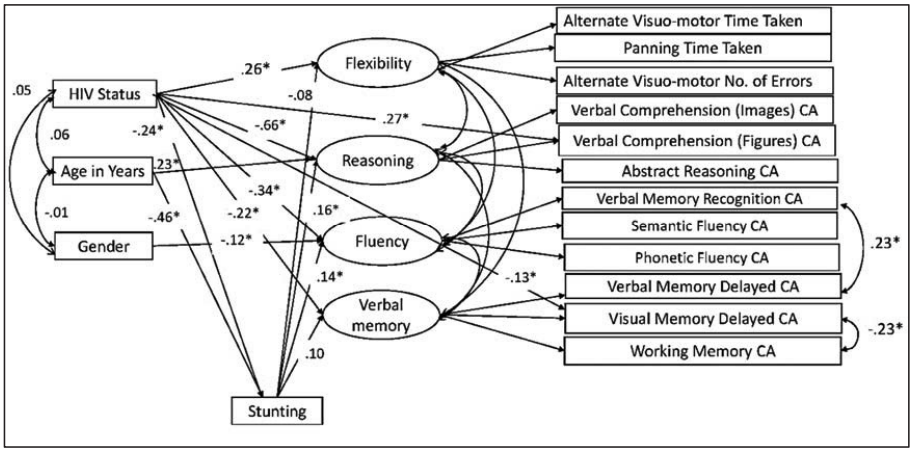
Figure 4.1: The BENCI Measurement Model with Adapted Partial Scalar Invariance and Modification Index Paths.



CA – Correct Answers

[2] Using the sample in the validity paper (169) we determined the power of our model given the sample size of 604, RMSEA of .08 to assess misfit (if a model does not fit, RMSEA > .08) and a model with 77 degrees of freedom (df). Using the function SemPower.PostHoc in R (192), a sample size of N = 604 is associated with a power larger than > 99.9 % to reject a wrong model with DF = 77 with an amount of misspecification corresponding to RMSEA = .08 using Alpha = .05 (192).

Figure 4.2: Height-for-age Mediation Model: Significant Paths in Specification Search



CA – Correct Answers; * P ≤ .05.

Results

Socio-demographic Results

The total sample mean age was 9.48 (SD = 1.31) and the mean stunting (HAZ) was -.44 (SD = 1.38). The prevalence of stunting in the HIV-positive sample was 17.9% while in the HIV-negative sample was 3.9%. The mean height-for-age in males was -0.42 (SD = 1.30) while in females it was -0.47 (SD = 1.45). Females who were HIV positive had more stunting (mean -0.94, SD = 1.51) than their HIV-negative counterparts and both HIV-positive and negative males[3]. The details of the sociodemographic indicators are presented in Table 4.1.

[3] Females who were HIV positive had more stunting (mean -0.94, SD = 1.51) than females who were HIV negative (Mean -0.02, SD = 1.24), males who were HIV positive (Mean -0.68, SD = 1.23) and males who were HIV negative (Mean -0.23, SD = 1.32) (F (1, 503) = 3.89, p = 0.049).

Table 4.1: Socio-demographic Indicators of the Study Population

Variables		**HIV Uninfected N (%)**	**HIV Infected N (%)**
Gender	Female	163 (49.4)	148 (54.0)
	Male	166 (50.3)	125 (45.6)
	Missing	1 (0.3)	1 (0.4)
Age in months (Mean ± SD)		117.2 ± 16.24	119.40 ± 14.63
Age in Years (Mean ± SD)		9.41 ± 1.37	9.56 ± 1.24
Nutrition	Weight in kg (Mean ± SD)	34.98 ± 7.12	32.27 ± 5.85
	Height in cm (Mean ± SD)	136.34 ± 8.00	133.02 ± 8.11
	Height -for -age z score (Mean ± SD)	-.13 (1.28)	-0.82 (1.39)
	Not stunted (≥ -1.9 SD)	265 (80.3)	180 (65.7)
	Moderately Stunted (- 2.9 to-2.0 SD)	11 (3.3)	41 (15)
	Severely Stunted (≤ -3.0 SD)	2 (0.6)	8 (2.9)
	Missing	52 (15.8)	45 (16.4)

Z score indicators are as provided by the WHO Child Growth Reference standards for 5 – 19-year-olds (188).

Height-for-age Mediation Model

We tested a full model (Figure 4.1) in which HIV status, age, and gender predicted the four cognitive executive functioning factors, and stunting acted as (partial) mediator of these predictions. This model showed good fit in terms of RMSEA = .041, CFI = 0.966, and TLI = 0.947, while the exact fit formally rejected the model ($\chi^2 = 154.29$, DF = 77, N = 604, $p < .001$), probably because of sensitivity to minor (distributional) violations and the relatively large sample size. The standardized effects and their level of significance results are presented in Table 4.2 and their standard errors are given in the Appendix 4.2. We also ran a full mediation

model without direct paths from age, gender, and HIV status on the cognitive latent variables, but this model showed poor fit (RMSEA = .074, CFI = .869, TLI = .824, $\chi 2$ = 384.22, DF = 89, N = 604, $p < .001$), highlighting that stunting does not fully mediate the effects of HIV on cognitive outcomes. A specification search yielded a more parsimonious model (Figure 4.2) that fitted well (RMSEA = .038, CFI = .967, TLI = .953, χ^2 = 158.73, DF = 84, N = 604, $p <$ .001) and corroborated our proposed model albeit without the significant paths.

Table 4.2: Height-for-age Model Standardized Effects

Bootstrapped Estimates	**Standardized Indirect Effects**				**Standardized Direct Effects**				**Standardized Total Effects**			
	HIV Status	**Age in years**	**Gender**	**Height-for-age**	**HIV Status**	**Age in years**	**Gender**	**Height-for-age**	**HIV Status**	**Age in years**	**Gender**	**Height-for-age**
Height-for-age					-0.242*	-0.462*	-0.004	-	-0.242*	-0.462*	-0.004	-
Flexibility	0.020	0.038	0.000	-	0.255*	-0.013	0.051	-0.082	0.275*	0.024	0.051	-0.082
Verbal Memory	-0.024*	-0.046	0.000	-	-0.212*	0.041	-0.016	0.100	-0.236*	-0.005	-0.016	0.100
Fluency	-0.033*	-0.063*	0.000	-	-0.338*	0.075	-0.133*	0.136*	-0.371*	0.012	-0.134*	0.136*
Reasoning	-0.038**	-0.072*	-0.001	-	-0.655*	0.245*	-0.021	0.157*	-0.692*	0.173*	-0.021	0.157*
Alternative Visual Motor Number of Errors	0.167**	0.015	0.031	-0.050	-	-	-	-	0.167*	0.015	0.031	-0.050

Bootstrapped Estimates	Standardized Indirect Effects				Standardized Direct Effects				Standardized Total Effects			
	HIV Status	Age in years	Gender	Height-for-age	HIV Status	Age in years	Gender	Height-for-age	HIV Status	Age in years	Gender	Height-for-age
Planning Time Taken	0.079*	0.007	0.015	-0.023	-	-	-	-	0.079*	0.007	0.015	-0.023*
Verbal Memory Delayed Hits	-0.174*	-0.004	-0.012	0.074	-	-	-	-	-0.174*	-0.004	-0.012	0.074
Visual Memory Delayed Hits	-0.147*	-0.003	-0.010	0.062	-0.126*	-	-	-	-0.272*	-0.003	-0.010	0.062
Working Memory Hits	-0.174*	-0.004	-0.012	0.074	-	-	-	-	-0.174*	-0.004	-0.012	0.074
Verbal Memory Recognition Hits	-0.068*	0.002	-0.025*	0.025*	-	-	-	-	-0.068*	0.002	-0.025*	0.025*
Semantic Fluency Hits	-0.281*	0.009	-0.101*	0.103*	-	-	-	-	-0.281*	0.009	-0.101*	0.103*
Phonetic Fluency Hits	-0.282*	0.009	-0.102*	0.103*	-	-	-	-	-0.282*	0.009	-0.101*	0.104*
Verbal Comprehension Images Hits	-0.391*	0.098*	-0.012	0.088*	-	-	-	-	-0.391*	0.098*	-0.012	0.088*
Verbal Comprehension Figures Hits	-0.523*	0.131*	-0.017	0.118*	0.268*	-	-	-	-0.255*	0.131*	-0.017	0.118*
Abstract Reasoning Hits	-0.394*	0.098*	-0.012	0.089*	-	-	-	-	-0.394*	0.098*	-0.012	0.089*

Bootstrapped Estimates	Standardized Indirect Effects				Standardized Direct Effects				Standardized Total Effects			
	HIV Status	Age in years	Gender	Height-for-age	HIV Status	Age in years	Gender	Height-for-age	HIV Status	Age in years	Gender	Height-for-age
Alternate Visual-motor Time Taken	0.275*	0.024	0.051	-0.082	-	-	-	-	0.275*	0.024	0.051	-0.082

* P ≤ .05; ** ≤ .001; Hits-Correct Answers.

Direct Effects

As expected by its effects on poor nutrition and cognitive development, HIV infection (coded as 1 = HIV+) had a significant direct effect on height-for-age (Z score with lower scores, more stunting) (β = -.242, p < .002) and on all cognitive latent variables (reasoning, fluency, verbal memory, and flexibility). HIV also had a direct effect on Verbal Comprehension Figures Hits and Visual Memory Delayed Hits reflective of the uniform measurement bias we described earlier (169). Age significantly predicted height-for-age (β = -.462, p = .004). Also, age showed a direct effect on reasoning (β = .245, p < .001), but we found little evidence of direct age effects on fluency, verbal memory, and flexibility.

There was no gender difference in height-for-age (β = .00, p = .927). While males averaged higher fluency scores than females (β = -.133, p = .005), gender did not significantly predict performance in flexibility, verbal memory, and reasoning. As expected from earlier works on the negative impact of stunting on cognitive outcomes, height-for-age predicted both fluency (β = .136, p = .008) and reasoning (β = .157, p = 002).

Mediation Effects

As shown in Table 4.2 and the Appendix 4.2, we found three significant indirect effects due to stunting between HIV grouping and verbal memory (β = -.024, SE = 0.013, p = .047),

fluency (β = -.033, SE = 0.013, p = .005), and reasoning (β = -.038, SE = 0.013, p = .001). An additional analysis captured in Appendix 4.5 suggested that the non-significant indirect path for flexibility (β = .020, SE = .012, p = .100) could be due to low power.

Given the failure of gender as a variable to predict height-for-age, height-for-age did not mediate the relationship between gender and any of the cognitive latent variables. In addition, an exploratory analysis captured in the Appendix 4.6 highlighted that females who were HIV positive showed more severe stunting, but adding the interaction between gender and HIV status rendered the direct path of gender on fluency non-significant. We deliberate on this finding in the Appendix 4.6 on interaction effect.

Height-for-age mediated the relationship of age with fluency (β = -.063, SE = 0.023, p = .009) and reasoning (β = -.072, SE = 0.023, p = .002). Total effects are reported in Table 4.2, while their standard errors are reported in appendix 4.2. Figure 4.2 reports the results of the specification search. Sensitivity analyses that checked for specification errors in the model are reported in the Appendix 4.3.

Discussion

We studied the mediating effects of stunting and the predictive effects of HIV, age, and gender on cognitive outcomes in a sample of 604 Kenyan children, and found that fluency, verbal memory, and reasoning are functions that may need to be targeted for intervention in children who were stunted and HIV positive. Next, we discuss these findings in detail.

Height-for-age Effects

Similar effects of height-for-age on language were found in a recent study in Kenya (172). Our study confirms this earlier study in showing the persistent nature of cognitive impairment among children who are stunted. Children who are stunted when aged 2 and who later recovered from stunting, remain underperforming on cognitive tests aged 5 compared to

children who were never stunted (182). Two cross-sectional studies among children older than 5 years have found that children with better HAZ have better performance in the cognitive tests (172, 173) and found that height-for-age mediates the prediction with age of reasoning, memory, language, executive functions, and motor ability.

Our finding that HIV directly contributes to stunting has also been found in other studies within Sub-Saharan Africa (173, 193, 194). Children living with HIV infection and those exposed to HIV yet uninfected have a higher prevalence of stunting than children who are neither infected nor exposed to HIV. Hence, being infected or having been born to a mother who is HIV positive poses a risk to normal child development (22, 110, 176). Moreover, children who are stunted and living with HIV or exposed to HIV are likely to have persistent stunting as they age (176).

Our findings of stunting increasing with age have been found in other studies (172) and are consistent with the notion that stunting often persists over age. A study looking at changes in height-for-age among children living with HIV and started ART around 8 years found that stunting reached its peak at 13 years for boys and 12 years for girls (180). After this age, stunting declined though more slowly in boys (at 13 years 50%, 15 years 48% and 18 years 31% stunted) than in girls (at 13 years 35%, 15 years 25% and 18 years 15% stunted) (180). Though there is a dearth of studies and indeed consensus on the exact age when HIV most directly leads to stunting, there are variable suggestions such as a study that showed male and females do not differ by age and stunting at ART initiation when aged around 8 years (180). However, stunting z scores start to dip as early as the first year of life among children living with HIV than those exposed but not infected (177). Another explanation for the strong effect is how stunting is calculated, i.e., as height relative to age z score.

HIV Effects

We found children without HIV to outperform their HIV infected counterparts in all domains of cognitive functioning, with up to 44% of reasoning performance variation due to HIV. Our findings are consistent with those of earlier studies showing that children living with HIV score poorly in tests of nonverbal cognitive abilities, executive function, processing speed, memory, planning, reasoning, working memory, and visual-spatial abilities (2, 22, 23), especially in advanced stages of the disease (2). Suboptimal cognitive functioning significantly impedes the wellbeing of children. For instance, adherence requires memory capabilities for learning new information, encoding, storing, and retrieving it when required (195). Similarly, for teenagers negotiating for healthy lifestyles, reasoning becomes an important asset. Deficits in these cognitive domains caused by HIV thus hinder psychosocial, learning processes including wading through routine functions and activities of daily life.

Mediation by Stunting

An earlier study in Kenya found that stunting mediated the effect of age and years in school on executive function, language, and motor skills, but not on verbal memory (172). These results are consistent with the current study findings related to language comprehension.

There is a dearth of evidence in form of comparative studies for such mediation findings among lower school students/ school-age children. Among younger and older cohorts, age of stunting onset and gender impact cognitive development among children with HIV (180) and without HIV infection (177). A study that followed up children from birth till 5 years found significant lower cognitive scores among those with early stunting onset (1-6 months) compared to those who were never stunted (at 60 months). The effect of stunting on cognitive performance, however, was no longer significant among those with late stunting onset (7-24 months after birth) although this might have been due to low power (175). We would therefore expect indirect effects of age and HIV on cognition among lower school students/ school-age children due to persistent stunting.

The statistically non-significant findings of the indirect effect of HIV on flexibility via height-for-age in our study could be attributed to low power. However, we note that there are additional underlying factors that determine good cognitive functioning among children who are stunted, such as lack of parental stimulation and few learning opportunities, which could contribute to cognitive deficiencies (110, 182). Indeed, parental stimulation among children who are stunted has been seen to improve performance in language and IQ tests (196), and such factors warrant more research in the future.

Age Differences

The prediction of reasoning based on age was expected as reasoning increases with age and height-for-age reflects a history of stunted growth. However, other studies among a community sample have not found age differences related to reasoning and memory in 6–8-year-olds though the narrow age range could lower correlations with age (173). Such age differences in cognitive functions are expected because some functions such as inhibition develop earlier and rapidly more than others that appear later on in development (151). In other cases, late school onset and repeating a grade may create spurious age differences in cognitive performance (197) or obscure aging effects. Repeating children either improve academic achievement (198) or experience a decline in cognitive performance (199, 200) depending on how long they were retained though repeating may also reflect an existing low cognitive ability (201) amongst other persistent psychosocial and academic challenges.

The age-wise trend for cognitive performance should be steep but we found non-significant age differences in some of the cognitive indicators. Aging effects on cognitive performance could have been obscured by other risk factors that were not included and controlled for in our analysis. The few risk factors may have underpowered the findings resulting in non-significant correlations with age.

Gender Differences

Gender has been found to be associated with risk of stunting among children aged below 5 years where having female gender was protective against stunting (175). Though our study did not reflect the same findings, the direction of our outcome is seen in other studies where girls were found to be more stunted than boys (although Intiful, Abdulai (202) found this difference to be non-significant).

Males in our study outperformed the females in fluency function, but no other gender gaps emerged. Earlier studies documented that males outperformed females in other cognitive functions such as visual-spatial ability though females have better scores than males in memory (203). Gender differences in cognitive function have been linked to school achievement with females performing better in languages though some studies have not found any differences in some subjects (181, 204). Whereas such outcomes may bring up confusion on which gender is need of a certain cognitive intervention, such findings should be interpreted with caution because studies have shown age related sex differences in cognitive maturation (178).

A study on underlying factors in gender differences may contribute to giving boys and girls equal opportunities in development may it be in improved school performance and increased earning potential.

Limitations

Our study interrogated a few independent variables while additional socio-economic factors and other confounding factors could affect cognitive functioning alongside stunting, HIV, age, and gender. Including additional factors such as poverty caregiver socioeconomic status, children schooling and related factors, and children's familiarity with technology such as iPads in future studies might shed further light on the mechanisms causing lower cognitive functioning in populations infected by HIV. There is a dearth of studies evaluating the interplay

between technology familiarity and cognition in children living with HIV and stunting. However, technological tools have been associated with cognitive development depending on exposure and pre-existing cognitive deficits (205). Moreover, though the Kenyan government obligates all parents to send children to school, our findings on the level of cognitive performance could be confounded by factors such as absentia and repeating grades among other factors. Our study did not control for such educational factors.

Our study used a cross-sectional design that is less able to uncover when and how effects emerge. A longitudinal study would point out the exact point where the severity of HIV strongly predicts cognitive deficiency in interaction with other determinants of stunting.

Our cross-sectional study design and study assessments do not allow us to uncover cognitive development trends within the children, or their ability to cope with early functional deficits (206). In addition, in cross-sectional studies, we cannot see whether the older children at an earlier time- point differ from the younger children in our sample. Of note, is that even longitudinal studies may miss this learning/coping confounding effect. Therefore, in situations where a child may be seen as underperforming, for example, in reasoning, they may have developed alternate ways of making sense of their environment such as through memorization. Indeed, children of the same age group have been found to have different patterns of developing reasoning functions (183). We may also not adequately explain differences in cognitive functioning of children of the same age who are brought up in different cognitively stimulating environments. With age, it is important to consider differences between following up the same cohort over time (177, 180, 200) and studying at one time-point (178, 182). Another limitation encompasses the cohort we used. These study findings and implications were drawn from a community sample and school factors such as student-to-teacher ratio and resources available in public vs private schools may not have been matched to the sample. A hospital sample may present different findings hence the implications should not be overgeneralised.

Using longitudinal case-control designs, future studies could consider trends in different cognitive functions as factors of the environment they had grown in, compensatory mechanisms for deficits and neurological mechanisms. Whereas our study takes a cross-sectional approach with few predictors, it is equally important in reviewing paediatric HIV programmes and setting up stunting and cognitive interventions.

Conclusion

As strides are made to mitigate and better manage HIV in children while reducing new infections, addressing stunting as well as its cognitive effects remain crucial, especially with the added burden of HIV (48). Stunting appears to play a role in the effects of HIV on cognitive domains. Our results point to the importance of integrating interventions that target reasoning, fluency, and verbal memory cognitive functions among children suffering from HIV infection and stunting. Nutrition programmes looking into reversing the effects of HIV on cognitive outcomes among lower school children in LMIC can tailor interventions targeting stunting. This is by targeting reasoning, fluency, and verbal memory and a wider set of cognitive functions that may need to be rehabilitated based on future research findings.

There is a dearth in comparative studies for such mediation findings among lower school students. Among younger and older cohorts, age of stunting onset and gender impact cognitive development among children with HIV (180) and without HIV infection (177). A study that followed up children from birth till 5 years found significantly lower cognitive scores among those with early stunting onset (1 – 6 months) compared to those who were never stunted (at 60 months). The effect of stunting on cognitive performance, however, was no longer significant among those with late stunting onset (7-24 months after birth) although this might have been due to low power (175). We would therefore expect indirect effects of age and HIV on cognition among lower school students due to persistent stunting.

The insignificance of the indirect effect of HIV on flexibility via height for age could be attributed to low power. Also, other factors other than stunting could have affected cognitive performance. Our study interrogated a few independent variables while there are socio-economic factors and other confounding factors that would affect cognitive functioning where stunting is mediating the relationship with HIV, age, and gender. Including additional factors such as poverty and caregiver socioeconomic status in future studies might shed further light on the mechanisms causing lower cognitive functioning in populations affected by HIV.

Age Differences

The prediction of reasoning based on age was expected as reasoning increases with age and height for age reflects a history of stunted growth. However, other studies among a community sample have not found age differences related to reasoning and memory in 6–8-year-olds though the narrow age range could lower correlations with age (173). Such age differences in cognitive functions are expected because some functions such as inhibition develop earlier and rapidly more than others that appear later on in development (151). In other cases, late school onset and repeating a grade may create spurious age differences in cognitive performance (197) or obscure aging effects. Repeating children either improve academic achievement (198) or experience a decline in cognitive performance (199, 200) depending on how long they were retained though repeating may also reflect an existing low cognitive ability (201).

The age-wise trend for cognitive performance should be steep but we found non-significant age differences in some of the cognitive indicators. The effect of age on cognitive performance could have been obscured by other risk factors that were not included and controlled for in the analysis.

The interventions aimed to improve cognitive functioning of children living with HIV and stunting, should target reasoning, fluency, and verbal memory and a wider set of cognitive functions that may need to be targeted based on future research findings.

Chapter Four Appendixes

Appendix 4.1: Specification Search Stunting Mediation Models

Model	Name	Params	Df	C	C – df	AIC 0	BCC 0	BIC 0	C/df	P	RMSEA	CFI 1	CFI 2
132	Unconstrained	52	84	158.73	74.73	0.00	0.00	11.65	1.89	0.00	0.04	0.97	0.96142
142	Unconstrained	53	83	157.00	74.00	0.27	0.33	16.32	1.89	0.00	0.04	0.97	0.97
143	Unconstrained	53	83	157.25	74.25	0.52	0.58	16.57	1.89	0.00	0.04	0.97	0.96
112	Unconstrained	50	86	163.34	77.34	0.61	0.49	3.45	1.90	0.00	0.04	0.97	0.96
152	Unconstrained	54	82	155.52	73.52	0.79	0.90	21.24	1.90	0.00	0.04	0.97	0.97
122	Unconstrained	51	85	161.66	76.66	0.93	0.87	8.17	1.90	0.00	0.04	0.97	0.96
123	Unconstrained	51	85	161.86	76.86	1.13	1.07	8.37	1.90	0.00	0.04	0.97	0.96
133	Unconstrained	52	84	159.95	75.95	1.22	1.22	12.87	1.90	0.00	0.04	0.97	0.96
124	Unconstrained	51	85	162.14	77.14	1.41	1.35	8.65	1.91	0.00	0.04	0.97	0.96
134	Unconstrained	52	84	160.18	76.18	1.44	1.44	13.09	1.91	0.00	0.04	0.97	0.96

Appendix 4.2: Height-for-age Model Standardized Errors

Standard Errors	Standardized Indirect Effects				Standardized Direct Effects				Standardized Total Effects			
	HIV Status	Age in years	Gender	Height-for-age	HIV Status	Age in years	Gender	Height-for-age	HIV Status	Age in years	Gender	Height-for-age
Height-for-age	-	-	-	-	0.034	0.032	0.035	-	0.034	0.032	0.035	-
Flexibility	0.012	0.023	0.003	-	0.039	0.045	0.038	0.049	0.036	0.038	0.038	0.049
Verbal Memory	0.013	0.024	0.004	-	0.048	0.055	0.045	0.051	0.048	0.047	0.045	0.051
Fluency	0.013	0.023	0.005	-	0.045	0.053	0.045	0.049	0.043	0.048	0.045	0.049
Reasoning	0.013	0.023	0.006	-	0.042	0.048	0.042	0.047	0.040	0.042	0.042	0.047
Alternative Visual Motor Number of Errors	0.024	0.023	0.023	0.030	-	-	-	-	0.024	0.023	0.023	0.030
Planning Time Taken	0.014	0.011	0.011	0.015	-	-	-	-	0.014	0.011	0.011	0.015

Standard Errors	Standardized Effects		Indirect		Standardized Direct Effects				Standardized Total Effects			
Verbal Memory Delayed Hits	0.036	0.035	0.033	0.038	-	-	-	-	0.036	0.035	0.033	0.038
Visual Memory Delayed Hits	0.031	0.030	0.028	0.032	0.035	-	-	-	0.036	0.03	0.028	0.032
Working Memory Hits	0.037	0.035	0.033	0.038	-	-	-	-	0.037	0.035	0.033	0.038
Verbal Memory Recognition Hits	0.020	0.009	0.011	0.012	-	-	-	-	0.020	0.009	0.011	0.012
Semantic Fluency Hits	0.034	0.036	0.034	0.037	-	-	-	-	0.034	0.036	0.034	0.037
Phonetic Fluency Hits	0.035	0.036	0.034	0.038	-	-	-	-	0.035	0.036	0.034	0.038
Verbal Comprehension Images Hits	0.029	0.024	0.024	0.027	-	-	-	-	0.029	0.024	0.024	0.027
Verbal Comprehension Figures Hits	0.092	0.030	0.032	0.039	0.097	-	-	-	0.037	0.03	0.032	0.039
Abstract Reasoning Hits	0.033	0.026	0.024	0.028	-	-	-	-	0.033	0.026	0.024	0.028
Alternate Visual-motor Time Taken	0.036	0.038	0.038	0.049	-	-	-	-	0.036	0.038	0.038	0.049

Appendix 4.3: Sensitivity Analysis

A model specification search using AMOS resulted in the same as model shown in figure 4.2 after 2,097,152 models were tried. A model specification search is recommended for detection and correction of specification errors so that the initial theory implied model can reflect the true population model cognizant to the study variables (189). The resulting models and their fit indicators are indicated in Appendix 4.1. The best model goodness of fit indicators was excellent with RMSEA = .038, CFI = .967, TLI = .953, $\chi 2$ (84, n = 604) = 158.731, $p < .001$. The model showed the specification search did not reveal missing links.

Appendix 4.4: Data Cleaning

With weight and height measurements, there were 43 respondents with default entries. These were converted to missing data. We then checked for outliers through a scatter plot and statistical evaluation of weight and height. Before deleting the outliers, we checked for the residuals of the regression of age on the measurements where we noted measurements with high standardized residual value, low effect size, and low p-value (169). We then checked the *z* scores that were beyond z=2 and also triangulated the scores against what would be expected in other participants. As per this evaluation, we did not discard any more weight and height entries.

Appendix 4.5: Flexibility Power Analysis

With flexibility, though there is a direct significant path between HIV and flexibility (p = .001), when height-for-age mediates this relationship, the path becomes non-significant (p =.100) though the power of this mediation is 0.5. In calculating this power (207), we used N = 604, path HIV to stunting β = -.24, path stunting to Flexibility β = -.08, path Flexibility to HIV β = .26 and alpha = .05 (207). This suggests that height-for-age does not fully mediate the relationship between HIV and flexibility though there is a 50% chance that we missed the indirect path if one exists in the population. Therefore, power could have been an issue.

Appendix 4.6: Interaction Effect

Our study found no significant gender effects on stunting yet other studies have found stunting to be significantly higher in males compared to females (208, 209). We therefore set out to find out whether the relation between HIV and stunting was different for genders. We reran the path model again, but this time added an interaction dummy variable for HIV and gender. The results showed a direct effect of the interaction variable on height-for-age (β = -.173, < .05) showing that a child that was female and HIV positive was most likely to be stunted. The only path that changed in the full model after adding the interaction between gender and HIV was that the gender effect on fluency which became non-significant (β = -.054, p = .362).

With flexibility, though there is a direct significant path between HIV and flexibility (p = .001), when height-for-age mediates this relationship, the path becomes non -significant (p =.100) though the power of this mediation is 0.5. In calculating this power (207), we used N = 604, path HIV to stunting β = -.24, path stunting to Flexibility β = -.08, path Flexibility to HIV β = .26 and alpha = .05 (207). This suggests that height-for-age does not fully mediate the relationship between HIV and flexibility though there is a 50% chance that we missed the indirect path if one exists in the population. Therefore, power could have been an issue.

CHAPTER FIVE: GENERAL DISCUSSION

Summary of Results

Main Chapter Conclusions

The general goals of this thesis were to review the neurocognitive tools used with lower school students and to adapt and validate the Computerized Battery for Neuropsychological Evaluation of Children (BENCI) in a cohort of Kenyan children living with HIV and those living without HIV. Moreover, we used the BENCI to evaluate the role of stunting in cognitive outcomes among children living with HIV. These investigations would help with evaluating the true burden of neurocognitive impairment among 6 – 12-year-old children in Lower- and Middle-Income Countries (LMICs), such as Kenya. This thesis is organized into five chapters. In the first chapter, we introduced subsequent chapters and discussed gaps in knowledge on neurocognitive tools used in middle childhood and the need to conduct subsequent studies to enable valid neurocognitive assessments that monitor development and inform future interventions.

In Chapter 2, we presented a narrative review of neurocognitive tools commonly used in research in middle childhood globally that considered their psychometric properties across diverse cultural contexts. Specifically, we documented where these tools have been developed and/or tested and evaluated the adequacy of psychometric outcomes, such as the reliability and validity of the tools, across different contexts. We found numerous neurocognitive tools that have been developed and tested among 6 - 12-year-olds, with most tools having been developed and validated in high-income countries. The tested and developed tools measured certain domains of neurocognitive outcomes, such as executive function, while leaving out other domains that are considered relevant according to the Diagnostic and Statistical Manual of

Mental Disorders, Fifth Edition (DSM-V). In addition, we found that several standardized tools for the different domains were confined to one geographical setting or were insufficiently studied for psychometric properties. In the reviewed literature, tests of executive functioning were overrepresented compared to tests in other domains. Finally, we observed several shortcomings in the evidence on the development and adaptation of neurocognitive tools for 6 – 12-year-olds in LMICs. These include mixed findings concerning test-retest reliability and construct validity, and lack of norming data/standardization samples. A key gap in the literature is the scarcity of adequately adapted, validated, and standardized tools for Sub-Saharan Africa.

In Chapter 3, we adapted and validated the Computerized Battery for Neuropsychological Evaluation of Children (BENCI) among lower-school students in Kenya. We preferred this tool because it covers all domains recommended by the DSM-V, except social cognition. We translated the tool from Spanish to English, piloted the adapted tool among Kenyan children, and finally tested the psychometric properties of the adapted tool in a large sample of Kenyan children aged 6-12 living with and without HIV. The adaptation process involved having bilingual researchers translate the tool to English, checking the linguistic and semantic consistency, and evaluating both the tool's structure (subtests, instructions, and order of administration) and its appropriateness against original markers (e.g., the number of blocks per subtest, number of trials per block, percentage of target stimuli, presentation time). In comparing the performance of 328 children living with HIV against 260 children not living with HIV in a case-control study, we found good internal consistency and test-retest reliability for most subtests. We studied the convergent validity of the BENCI using locally validated paper and pencil tools and observed good correlations between BENCI's memory, reasoning, and inhibition domains and the local tools from the Kilifi toolkit (46). A confirmatory four-factor model consisting of reasoning, flexibility, verbal memory, and fluency fitted the data well and showed metric and partial scalar invariance with respect to the

HIV-positive and HIV-negative groups. In summary, even though some subtests could be improved in future work, we found the adapted English version of the BENCI to be a valid and reliable neurocognitive tool for use among Kenyan children.

In Chapter 4, we studied the potential impact of HIV and stunting on neurocognitive performance on the BENCI. Specifically, we applied structural equation modelling to the same data as in Chapter 3 to predict performance on reasoning, flexibility, verbal memory, and fluency of Kenyan children by HIV, gender, age, and stunting. We also tested whether stunting mediates the effects of HIV, gender, and age on reasoning, flexibility, verbal memory, and fluency. We found evidence that HIV directly impacted stunting and all cognitive outcomes, while age was only directly predictive of stunting and reasoning. Stunting was found to (partially) mediate the effects of HIV on reasoning, verbal memory, and fluency, and stunting was found to mediate the effects of age on fluency and reasoning. We observed that children with a double burden of HIV and stunting have poorer cognitive performance compared to those with either HIV only or stunting only. We did not observe any direct effects of gender on stunting but found some evidence that female children living with HIV were most likely to be stunted.

General Conclusions

Like many other LMICs settings, Kenya has a shortage of adequately standardized neurocognitive tools with robust psychometric properties that can be used to inform the needs of vulnerable children such as those living with HIV. Culturally tailored and validated tests would enable us to know the true burden of cognitive impairment in LMICs and the associated risk factors, but our review of cognitive tests used among 6 – 12-year-olds revealed important gaps characterized by few domains and inadequate psychometric information in research on neurocognitive tools in LIMCs. Therefore, we successfully adapted the BENCI for Kenyan children and found evidence that children with a double burden of HIV and stunting have

poorer cognitive performance compared to those with either HIV only or stunting only. We also identified reasoning, verbal memory, and fluency as targets for intervention among children with HIV and stunting.

General Discussion and Future Directions

Research Implications

We managed to adapt a test developed in a high-income setting and made it suitable for use in a low-income setting, i.e., Kenya, while largely retaining its robust psychometric properties. However, our results align with earlier findings (210) showing that even sub-scales claimed to be "culture free" and carefully adapted may still be affected by culture. Notably, we found the factor structure of the BENCI to differ from that of the Arabic version of the BENCI (40) among Moroccan children. In the latter sample, inhibition, flexibility, fluency, reasoning, and verbal memory factors reflected the higher-order executive function factor well (40). However, in the Kenyan data, the subtests for inhibition did not reflect the executive function construct as well as they did in the Moroccan sample. Although we did not formally test for measurement invariance between the Moroccan and Kenyan data, our finding that the factor structure in the Arabic version did not emerge in the data from the Kenyan cohort already highlights that configural invariance (i.e., the first step towards stricter variants of measurement invariance) between the versions did not emerge. Other tests that have also been found to measure differently across cultural groups include the Children's Colour Trails Test and Raven's Progressive Matrices (107, 210, 211). A study comparing performance in the Children's Colour Trails Test among Moroccan and North American children showed that the former was much slower in completing the test than the latter (210). The test measures mental processing speed, but there could be cultural underpinnings that promote accuracy over speed in ways that differ by population. If so, we expect failures of measurement invariance, creating biased outcomes. Cultural underpinnings such as language have also been found to influence cognitive

performance in non-verbal tests using coloured stimuli in other societies (195). This stresses the need to carefully adapt and validate neurocognitive tools within the respondents' culture. Developing a culture-fair test that measures equivalently across cultural groups is challenging and might not be possible given the many cultural variations that play a role in neurocognitive measurement.

Despite the challenges in creating culture-fair cognitive tests, researchers have produced methods to validate cross-cultural comparisons of neurocognitive functioning (47, 212). With increasing globalization, there is a need for neurocognitive tests that are validated and standardized for the populations of interest. Development and adaptation of many subtests to suit cultural diversity may render cross-cultural comparison of cognitive outcomes problematic. However, measurement invariance (the ability of a subtest or test item to function the same across groups) has been proposed as a solution to this challenge (47, 212). If neurocognitive tests are non-invariant (i.e., do not have functional equivalence), the cognitive outcomes may be biased across groups based on gender, age, education levels, and other factors, or some of the test items or subtests may not be functioning well in some of the groups (47). In the case of BENCI with respect to HIV groups, we found support for metric and partial scalar invariance, specifically because the Verbal Comprehension (figures) CA and Visual Memory Delayed CA were not well-calibrated (169). The two subtests did not function the same way across the HIV groupings. Such an outcome may necessitate discounting these subtests when computing reasoning and verbal memory domain scores or further study to identify the source of the bias and recalibrate the subtests in future revisions. It would be good to conduct further studies of measurement invariance of the BENCI across other types of groupings based on culture, gender, age, and educational levels either to corroborate invariance or to identify the sources of bias that could improve (cross-cultural) mean comparisons and lead to the creation of (sub)tests and tools that function equivalently across (cultural) groups.

This would facilitate cross-cultural comparisons of cognitive outcomes, particularly mean differences between subtests, a venture that has been untenable in other studies (212). Evaluating measurement invariance should be customary in cross-cultural comparisons and should be studied further across relevant groupings.

To accurately reflect differences in latent abilities, the difficulty of items in neurocognitive tests should also match the level of ability in the sample. However, some BENCI subtests in our study showed ceiling effects meaning that many children found them too easy, while other subtests showed floor effects meaning that many children found them too difficult. This implies that the items in the subtest should either be revised to suit the test-takers' ability levels or that more difficult or easy items should be added to these subtests. Such revisions are needed in the future to accurately assess the true burden of cognitive impairment or performance levels in the targeted population of lower-school Kenyan children with and without a disease burden.

Apart from such test improvements, there is a need to develop normative data for the standardization of scores on the BENCI. Normative data helps in interpreting test scores in relation to normal cognitive functioning in the general population. Earlier versions of the BENCI have test norms for children aged 6 – 17 years in Ecuador, 7, 9, and 11 years in Morocco, and 6 - 8 years in Palestine (40, 124). We are yet to develop similar norms in Kenya. Our narrative review found no studies reporting on the development of normative data in any LMIC, although our review only covered studies published in journals between 1997 and 2017, so other studies may have collected normative data in these settings. Given that commercial tests are not preferred in these low-resource settings, it is imperative that the collection of normative data on well-validated neurocognitive tools is appropriately funded and conducted in collaboration with experts (106).

It would be important to replicate our study in other LMICs where the number of adapted and validated neurocognitive tests among 6 – 12-year-olds remains small (213). Many settings in LMICs have minimal resources, so free, non-invasive, and easy-to-use tests would be preferred. BENCI is a free-of-charge tool that only requires purchasing iPads and personnel costs.

We carried out our empirical investigation within a cross-sectional design, which limited any findings related to cognitive developmental changes over time and environmental factors that may have triggered the changes. Together with the sampling of participants based on grades rather than age groups, our study design may have precluded finding age differences in fluency, verbal memory, and flexibility performance. Future longitudinal studies may uncover other mediating, moderating, and confounding factors such as socioeconomic status, maternal education, child schooling, and negative prenatal and perinatal adverse events (3, 4) that may provide more insight into developmental trajectories and uncover underlying factors that impact cognitive functioning among lower school students.

Cultural norms and expectations affect behaviours and ways of thinking that eventually influence cognitive performance (195). Debate (214) continues on whether or not there is a critical point or age at which a child's cognition is most responsive to cultural input. Some discussion frames it as a matter of processes (physiological maturation, heritability, intensity of cultural/environmental exposure including cultural diffusion, rehabilitation and innovation and other factors) and not a certain point in time (214). The complexity of neurocognitive development is also highlighted in the mutualism model, which states that cognitive processes mutually influence each other during cognitive development (215). Such models could be expanded to include environmental effects on development (216, 217) and to consider cross-cultural differences. As developing children respond to the varying cultural and environmental inputs, certain neural pathways are opened while others are closed, reopened, and redeveloped.

Future studies can investigate such developmental changes in diverse cultural contexts to outline the extent to which these cultural and environmental features are likely to alter cognitive performance. For this effort, well-validated and culturally appropriate neurocognitive assessment batteries like the BENCI are essential. It is also clear that such necessary research requires funding and collaboration between experts, practitioners, and people with lived experience of neurocognitive deficits, and/or with knowledge of the children, cultural context, and psychometrics in neurocognitive assessments.

Clinical Implications

Our findings on the psychometric robustness of the English version of the BENCI showed that the battery can be relied upon to evaluate cognitive functions but with a few reservations. This includes the miscalibration of the Verbal Comprehension (figures) CA and Visual Memory Delayed CA subtests which we found in our test of invariance between HIV-positive and HIV-negative groups. This is relevant when the BENCI is used in clinical practice.

The goal of a clinician using a neurocognitive tool is to make an optimal clinical evaluation of the type of cognitive impairment and the severity of the impairment. This, in turn, will guide therapeutic interventions that are optimally aligned to the child. In as much as the evaluation also considers clinical history during patient intake, 'neurocognitive tools aid in making an objective severity evaluation and "close to" real/true performance evaluation for the condition. With poorly calibrated tools, the evaluation of performance may be underestimated or overestimated (218). The remaining BENCI subtests were well-calibrated, showing that their performance scores are "close to" a true reflection of the underlying cognitive function in our cohort. However, it is good to consider the discriminant validity of the tool in relation to other traits or factors that the tool should not inadvertently measure, such as anxiety or depression. For instance, if a child's slow response rate is due to anxiety or depression, discriminant

validity would not be supported. Future research on the BENCI should further study discriminant validity in relation to other traits and improve its accuracy in diagnosing symptomatic cognitive impairment.

Though not explicitly studied in this study, eliminating administrator bias is also important in ensuring valid and reliable results. There have been reports of a lack of consistency and uniformity in neurocognitive test administration and results which could not be attributed to the test itself (219). This could be attributed to a lack of training in administration and score interpretation. Electronic tools, such as our iPad-based BENCI, may help reduce such bias due to the training features integrated before each subtest, streamlined timings, and programmed scoring, which minimizes administrators' interference during the assessment. Also, administration by professionals with a background in neurocognitive assessment may help minimize bias. In low-resource settings such as Sub-Saharan Africa, expertise and electronic tools may not be readily available hence the need to train other healthcare workers and paraprofessionals on tenets of neurocognitive assessment and impairment.

Policy Implications

Without information on the true burden of cognitive impairment, its correlates, and implications on general well-being, the optimal institutional and country-specific policies remain unclear. The BENCI is a culturally sensitive tool for assessing cognitive impairment among children in Kenya and is useful in gathering information on the true burden of cognitive impairment among children with HIV and/or stunting. The tool is, therefore, helpful in informing government policies and institutional strategies and interventions in protecting the cognitive functioning of these vulnerable children as part of holistic HIV care. In Kenya, we have policy guidelines for addressing stunting as part of HIV care (220). The care may also protect the cognitive functioning of at-risk children, but cognitive functioning is neither

explicitly mentioned in the guidelines as an indicator of well-being nor included in routine assessments and interventions. Routine screening may help identify children at risk of neurocognitive impairment, thereby signalling the need for early interventions. The latter is important in light of other studies reporting on the persistent nature of cognitive impairment despite stunting interventions (182). In totality, our study findings identify that good cognitive functioning needs to be a vital component of the governments and other institutions' efforts to ameliorate the effects of HIV on children's development.

Personal Reflection in Relation to the Thesis

While working as a clinical psychologist in an HIV clinic in 2016 – 2019, I encountered children living with HIV who, or whose parents, complained of difficulty learning in school. An example was a child being slow compared to other children in copying the teacher's notes from the blackboard, which disallowed the child to copy the notes in time before the teacher erased them to make room for new notes. The child reported good adherence to medication, showed no opportunistic infection, and was virally suppressed. Further clinical investigations to identify the problem would call for intrusive investigations, including Cerebrospinal Fluid (CSF) analysis, which was unaffordable to the patient's caregivers. The HIV clinic offered free comprehensive HIV services, but the services did not include neurocognitive investigations unrelated to common HIV opportunistic infections. Investigations into cognitive impairment were not provided in the clinic. This child could have a problem related to processing speed or attention, which non-invasive paper-and-pencil cognitive tools could assess, but we did not have such assessments adapted to our setting. In such cases, the child's complaints would go unaddressed from a clinical standpoint. A different arrangement with the teacher would instead be considered to accommodate the child's learning process, although factors such as costs and class size shrouded any benefits. A teacher may have found it difficult to give special consideration to one child in a class of many children, and the caregivers would not be able to

afford to move a child to an expensive school that had fewer children per class. The lack of equal learning opportunities for children with HIV and cognitive impairment puts them at risk of underperforming in school, dropping out, and earning low wages. Neurocognitive deficits associated with HIV infection among children have been reported in planning, reasoning, attention, visual processing, and memory (2, 22, 23), which impede the well-being of the children. Learning is supported by, among other things, healthy cognitive functions, which formed the basis of our investigation. I want to support these children in accessing affordable and valid neurocognitive assessments. This has been made possible through this thesis. BENCI is a freely accessible iPad-based neurocognitive battery that has been standardized and found to have valid and reliable subtests for use among lower-school students in Kenya.

As part of my future work, I want to improve the BENCI subtests that showed attenuation effects and did conform to measurement invariance, develop a test to measure social cognition to cover all domains of assessment recommended by the DSM-V, and create Kenyan norms for the BENCI. I am also hoping to complement this work by integrating cognitive rehabilitation among the 6 – 12-year-olds who are stunted and evaluating the cognitive rehabilitation outcomes using the BENCI. Monitoring and improving cognitive deficits among persistently stunted children who are also living with HIV will allow them to learn and eventually succeed in life just like their non-stunted classmates who are not living with HIV.

REFERENCES

1. Walker SP, Wachs TD, Grantham-McGregor S, Black MM, Nelson CA, Huffman SL, et al. Inequality in early childhood: Risk and protective factors for early child development. Lancet. 2011;378(9799):1325-38.
2. Ruel T, Boivin M, Boal H, Bangirana P, Charlebois E, Havlir D, et al. Neurocognitive and motor deficits in HIV-infected Ugandan children with high CD4 cell counts. Clinical Infectious Diseases. 2012;54(7):1001-9.
3. Firth J, Torous J, Stubbs B, Firth JA, Steiner GZ, Smith L, et al. The "online brain": how the internet may be changing our cognition. World Psychiatry. 2019;18(2):119-29.
4. Fogarty L, Creanza N, Feldman MW. The life history of learning: Demographic structure changes cultural outcomes. PLOS Computational Biology. 2019;15(4):e1006821.
5. Musindo O, Krabbendam L, Mutahi J, García MP, Bangirana P, Kumar M. Neurocognitive deficits and socioeconomic risk factors among children and adolescents living with HIV in Sub-Saharan Africa: A systematic review. Child and Adolescent Psychiatry and Mental Health. 2022;16(1):31.
6. Kandawasvika GQ, Kuona P, Chandiwana P, Masanganise M, Gumbo FZ, Mapingure MP, et al. The burden and predictors of cognitive impairment among 6- to 8-year-old children infected and uninfected with HIV from Harare, Zimbabwe: A cross-sectional study. Child Neuropsychol. 2015;21(1):106-20.
7. Roze E, Reijneveld SA, Stewart RE, Bos AF. Multi-domain cognitive impairments at school age in very preterm-born children compared to term-born peers. BMC Pediatrics. 2021;21(1):169.
8. Bücker J, Kapczinski F, Post R, Ceresér KM, Szobot C, Yatham LN, et al. Cognitive impairment in school-aged children with early trauma. Compr Psychiatry. 2012;53(6):758-64.
9. Iloh KK, Emodi IJ, Ibeziako NS, Ikefuna AN, Ubesie AC, Iloh ON, et al. Neurocognitive function of school-aged HIV-infected children in Enugu, Nigeria. Journal of Tropical Pediatrics. 2017;63(6):425-30.
10. Woldehanna T, Behrman JR, Araya MW. The effect of early childhood stunting on children's cognitive achievements: Evidence from young lives Ethiopia. Ethiop J Health Dev. 2017;31(2):75-84.
11. Khan DSA, Das JK, Zareen S, Lassi ZS, Salman A, Raashid M, et al. Nutritional status and dietary intake of school-age children and early adolescents: Systematic review in a developing country and lessons for the global perspective. Frontiers in Nutrition. 2022;8.
12. Bogale TY, Bala ET, Tadesse M, Asamoah BO. Prevalence and associated factors for stunting among 6–12 years old school age children from rural community of Humbo district, Southern Ethiopia. BMC Public Health. 2018;18(1):653.
13. UNICEF. HIV estimates for children dashboard: UNICEF data 2022 [Accessed on January 2, 2023]. Available from: https://data.unicef.org/resources/hiv-estimates-for-children-dashboard/
14. Okafor C, Fadupin G, Oladokun R. Nutritional status and virological outcomes of children HIV positive attending anti-retroviral clinic at University College Hospital, Ibadan. Food and Nutrition Sciences. 2021;12:1088-97.
15. Sabanathan S, Wills B, Gladstone M. Child development assessment tools in low-income and middle-income countries: How can we use them more appropriately? Arch Dis Child. 2015;100(5):482-8.
16. McCoy DC, Sudfeld CR, Bellinger DC, Muhihi A, Ashery G, Weary TE, et al. Development and validation of an early childhood development scale for use in low-resourced settings. Population Health Metrics. 2017;15(1):3.
17. Muthukrishna M, Bell AV, Henrich J, Curtin CM, Gedranovich A, McInerney J, et al. Beyond western, educated, industrial, rich, and democratic (WEIRD) psychology: Measuring and mapping scales of cultural and psychological distance. Psychological Science. 2020;31(6):678-701.
18. Nielsen M, Haun D, Kärtner J, Legare CH. The persistent sampling bias in developmental psychology: A call to action. Journal of Experimental Child Psychology. 2017;162:31-8.

19. Kusi-Mensah K, Nuamah ND, Wemakor S, Agorinya J, Seidu R, Martyn-Dickens C, et al. Assessment tools for executive function and adaptive function following brain pathology among children in developing country contexts: A scoping review of current tools. Neuropsychol Rev. 2022;32(3):459-82.
20. Nsereko N, Palmer S, Basa V. Advancing psychometrics in Uganda's institutions of higher learning with emphasis on structural equation modeling (SEM). Nkumba International Research Journal (NIRJ). 2022;2(1):7-21.
21. Spironello C, Hay J, Missiuna C, Faught BE, Cairney J. Concurrent and construct validation of the short form of the Bruininks-Oseretsky Test of Motor Proficiency and the Movement-ABC when administered under field conditions: Implications for screening. Child: Care, Health and Development. 2010;36(4):499-507.
22. Sherr L, Hensels IS, Tomlinson M, Skeen S, Macedo A. Cognitive and physical development in HIV-positive children in South Africa and Malawi: A community-based follow-up comparison study. Child Care Health Dev. 2018;44(1):89-98.
23. Cohen S, ter Stege JA, Geurtsen GJ, Scherpbier HJ, Kuijpers TW, Reiss P, et al. Poorer cognitive performance in perinatally HIV-infected children versus healthy socioeconomically matched controls. Clinical Infectious Diseases. 2014;60(7):1111-9.
24. Jewsbury PA, Bowden SC, Duff K. The Cattell–Horn–Carroll model of cognition for clinical assessment. Journal of Psychoeducational Assessment. 2017;35(6):547-67.
25. Schneider W, McGrew K. The Cattell-Horn-Carroll model of intelligence. In: Dawn PF, Patti LH, editors. Contemporary intellectual assessment: Theories, tests, and issues. 3 ed. New York: Guilford Press; 2012. p. 99-144.
26. Heyes C. Culture. Curr Biol. 2020;30(20):R1246-r50.
27. Legare CH, Dale MT, Kim SY, Deák GO. Cultural variation in cognitive flexibility reveals diversity in the development of executive functions. Scientific Reports. 2018;8(1):16326.
28. Gonthier C. Cross-cultural differences in visuo-spatial processing and the culture-fairness of visuo-spatial intelligence tests: an integrative review and a model for matrices tasks. Cognitive Research: Principles and Implications. 2022;7(1):11.
29. Uchiyama R, Spicer R, Muthukrishna M. Cultural evolution of genetic heritability. Behavioral and Brain Sciences. 2021:1-147.
30. Guida A, Megreya AM, Lavielle-Guida M, Noël Y, Mathy F, van Dijck J-P, et al. Spatialization in working memory is related to literacy and reading direction: Culture “literarily” directs our thoughts. Cognition. 2018;175:96-100.
31. Boyd R. A different kind of animal: How culture transformed our species. New Jersey Princeton University Press; 2017.
32. Kendal RL, Boogert NJ, Rendell L, Laland KN, Webster M, Jones PL. Social learning strategies: Bridge-building between fields. Trends in Cognitive Sciences. 2018;22(7):651-65.
33. Huang C-M, Doole R, Wu CW, Huang H-W, Chao Y-P. Culture-related and individual differences in regional brain volumes: A cross-cultural voxel-based morphometry study. Frontiers in Human Neuroscience. 2019;13.
34. Bathje GJ, Feiss C. Culture-fair tests. In: Teo T, editor. Encyclopedia of Critical Psychology. New York, NY: Springer; 2014. p. 358-61.
35. Korkman M, Kemp S, Kirk U. Effects of age on neurocognitive measures of children ages 5 to 12: A cross-sectional study on 800 children from the United States. Developmental neuropsychology. 2001;20:331-54.
36. Korkman M, Lahti-Nuuttila P, Laasonen M, Kemp SL, Holdnack J. Neurocognitive development in 5- to 16-year-old North American children: A cross-sectional study. Child Neuropsychology. 2013;19(5):516-39.
37. Noble KG, Tottenham N, Casey BJ. Neuroscience perspectives on disparities in school readiness and cognitive achievement. Future Child. 2005;15(1):71-89.

38. National Research Council (US), Institute of Medicine (US) Committee on Integrating the Science of Early Childhood Development. Promoting healthy development through intervention. In: Shonkoff J, Phillips D, editors. Neurons to Neighborhoods: The Science of Early Childhood Development. Washington DC: National Academies Press (US); 2000. p. 612.
39. Committee on the Science of Children Birth to Age 8 Deepening and Broadening the Foundation for Success, Board on Children Youth and Families, Institute of Medicine, National Research Council. Child development and early learning. In: Allen L, Kelly B, editors. Transforming the Workforce for Children Birth Through Age 8: A Unifying Foundation. 4. Washington DC: National Academies Press (US); 2015. p. 707.
40. Fasfous, Peralta-Ramirez MI, Perez-Marfil MN, Cruz-Quintana F, Catena-Martinez A, Perez-Garcia M. Reliability and validity of the Arabic version of the computerized Battery for Neuropsychological Evaluation of Children (BENCI). Child Neuropsychol. 2015;21(2):210-24.
41. Bangirana P, Sikorskii A, Giordani B, Nakasujja N, Boivin MJ. Validation of the CogState battery for rapid neurocognitive assessment in Ugandan school age children. Child and Adolescent Psychiatry and Mental Health. 2015;9.
42. Siqueira LS, Gonçalves HA, Hübner LC, Fonseca RP. Development of the Brazilian version of the Child Hayling Test. Trends Psychiatry Psychother. 2016;38(3):164-74.
43. Fernandez A, Abe J. Bias in cross-cultural neuropsychological testing: problems and possible solutions. Culture & Brain. 2018;6.
44. Sachdev P, Blacker D, Blazer D, Ganguli M, Jeste D, Paulsen J, et al. Classifying neurocognitive disorders: The DSM-5 approach. Nature Reviews Neurology. 2014;10:634-42.
45. American Psychiatric Association (APA). Diagnostic and statistical manual of mental disorders (5th ed.). Washington, DC: American Psychiatric Association; 2013. 991 p.
46. Kitsao-Wekulo PK, Holding PA, Taylor HG, Abubakar A, Connolly K. Neuropsychological Testing in a Rural African School-Age Population: Evaluating Contributions to Variability in Test Performance. Assessment. 2013;20(6):776-84.
47. Wicherts JM. The importance of measurement invariance in neurocognitive ability testing. Clin Neuropsychol. 2016;30(7):1006-16.
48. Prendergast AJ, Humphrey JH. The stunting syndrome in developing countries. Paediatr Int Child Health. 2014;34(4):250-65.
49. Malik F, Marwaha R. Cognitive Development. Treasure Island (FL): StatPearls Publishing LLC; 2023.
50. Llorente AM, Voigt RG, Williams J, Frailey JK, Satz P, D'Elia LF. Children's Color Trails Test 1 2: Test-Retest Reliability and Factorial Validity. Clinical Neuropsychologist. 2009;23(4):645-60.
51. Hubley AM, Zumbo BD. A Dialectic on Validity: Where We Have Been and Where We Are Going. The Journal of General Psychology. 1996;123(3):207-15.
52. Stad FE, Wiedl KH, Vogelaar B, Bakker M, Resing WCM. The role of cognitive flexibility in young children's potential for learning under dynamic testing conditions. European Journal of Psychology of Education. 2019;34(1):123-46.
53. Chen X, Chen H, Li D, Wang L. Early childhood behavioral inhibition and social and school adjustment in chinese children: a 5-year longitudinal study. Child Dev. 2009;80(6):1692-704.
54. Ahonniska J, Ahonen T, Aro T, Tolvanen A, Lyytinen H. Repeated assessment of the Tower of Hanoi test: Reliability and age effects. Assessment. 2000;7(3):297-310.
55. Holding PA, Taylor HG, Kazungu SD, Mkala T, Gona J, Mwamuye B, et al. Assessing cognitive outcomes in a rural African population: development of a neuropsychological battery in Kilifi District, Kenya. Journal of the International Neuropsychological Society : JINS. 2004;10(2):246-60.
56. Casaletto KB, Heaton RK. Neuropsychological Assessment: Past and Future. J Int Neuropsychol Soc. 2017;23(9-10):778-90.
57. Kelley TL. Interpretation of educational measurements. Oxford: World Book Co.; 1927.
58. Parikh R, Mathai A, Parikh S, Chandra Sekhar G, Thomas R. Understanding and using sensitivity, specificity and predictive values. Indian journal of ophthalmology. 2008;56(1):45-50.

59. Teglasi H, Nebbergall AJ, Newman D. Construct validity and case validity in assessment. Psychol Assess. 2012;24(2):464-75.
60. Andrade C. Internal, external, and ecological validity in research design, conduct, and evaluation. Indian J Psychol Med. 2018;40(5):498-9.
61. Williams ME, Sando L, Soles TG. Cognitive Tests in Early Childhood:Psychometric and Cultural Considerations. Journal of Psychoeducational Assessment. 2014;32(5):455-76.
62. Bradley-Johnson S. Cognitive Assessment for the Youngest Children: A Critical Review of Tests. Journal of Psychoeducational Assessment. 2001;19(1):19-44.
63. Stadskleiv K. Cognitive functioning in children with cerebral palsy. Developmental Medicine & Child Neurology. 2020;62(3):283-9.
64. Ezeamama AE, Bustinduy AL, Nkwata AK, Martinez L, Pabalan N, Boivin MJ, et al. Cognitive deficits and educational loss in children with schistosome infection—A systematic review and meta-analysis. PLOS Neglected Tropical Diseases. 2018;12(1):e0005524.
65. Ferrari R. Writing narrative style literature reviews. Medical Writing. 2015;24(4):230-5.
66. Hamadeh N, Rompaey VC, Metreau E, Eapen SG. World Bank Blogs [Internet]2022. [cited 2023]. Available from: https://blogs.worldbank.org/opendata/new-world-bank-country-classifications-income-level-2022-2023.
67. Chevignard MP, Servant V, Mariller A, Abada G, Pradat-Diehl P, Laurent-Vannier A. Assessment of executive functioning in children after TBI with a naturalistic open-ended task: a pilot study. Dev Neurorehabil. 2009;12(2):76-91.
68. van Nieuwenhuijzen M, Vriens A, Scheepmaker M, Smit M, Porton E. The development of a diagnostic instrument to measure social information processing in children with mild to borderline intellectual disabilities. Res Dev Disabil. 2011;32(1):358-70.
69. Rocke K, Hays P, Edwards D, Berg C. Development of a performance assessment of executive function: the Children's Kitchen Task Assessment. Am J Occup Ther. 2008;62(5):528-37.
70. Rohitrattana J, Siriwong W, Suittiwan P, Robson MG, Strickland PO, Rohlman DS, et al. Adaptation of a neurobehavioral test battery for Thai children. Rocz Panstw Zakl Hig. 2014;65(3):205-12.
71. Hwang Y, Hosokawa T, Swanson HL, Ishizaka I, Kifune N, Ohira D, et al. A Japanese short form of the Swanson Cognitive Processing Test to measure working memory: reliability, validity, and differences in scores between primary school children of the United States and Japan. Psychol Rep. 2006;99(1):27-38.
72. Rienstra A, Spaan PEJ, Schmand B. Validation of symptom validity tests using a “Child-model” of adult cognitive impairments. Archives of Clinical Neuropsychology. 2010;25(5):371-82.
73. Thomas E, Maruff P, Paul J, Reeve R. Spatial sequence memory and spatial error monitoring in the Groton Maze Learning Task (GMLT): A validation study of GMLT sub-measures in healthy children. Child Neuropsychology. 2016;22(7):837-52.
74. Reitan RM, Wolfson D. The Trail Making Test as an initial screening procedure for neuropsychological impairment in older children. Archives of Clinical Neuropsychology. 2004;19(2):281-8.
75. Karr JE, Garcia-Barrera MA. The assessment of executive functions using the BASC-2. Psychol Assess. 2017;29(9):1182-7.
76. Sadeh SS, Burns MK, Sullivan AL. Examining an executive function rating scale as a predictor of achievement in children at risk for behavior problems. Sch Psychol Q. 2012;27(4):236-46.
77. Richard's MM, Introzzi I, Zamora E, Vernucci S. Analysis of internal and external validity criteria for a computerized visual search task: A pilot study. Appl Neuropsychol Child. 2017;6(2):110-9.
78. Budtz-Jorgensen E, Keiding N, Grandjean P, Weihe P. Estimation of health effects of prenatal methylmercury exposure using structural equation models. Environ Health. 2002;1(1):2.

79. Reynolds E, Fazio VC, Sandel N, Schatz P, Henry LC. Cognitive Development and the Immediate Postconcussion Assessment and Cognitive Testing: A Case for Separate Norms in Preadolescents. Appl Neuropsychol Child. 2016;5(4):283-93.
80. Woodward H, Donders J. The performance of children with traumatic head injury on the Wide Range Assessment of Memory and Learningâ€"Screening. Applied Neuropsychology. 1998;5(3):113-9.
81. Rose SA, Feldman JF, Jankowski JJ, Van Rossem R. Basic information processing abilities at 11 years account for deficits in IQ associated with preterm birth. Intelligence. 2011;39(4):198-209.
82. Stinnett TA, Oehler-Stinnett J, Fuqua DR, Palmer LS. Examination of the underlying structure of the NEPSY: A developmental neuropsychological assessment. Journal of Psychoeducational Assessment. 2002;20(1):66-82.
83. Campbell D, Fiske D. Convergent and discriminant validation by the multitrait-multimethod matrix. Psychological bulletin. 1959;56 2:81-105.
84. Thaler NS, Allen DN, McMurray JC, Mayfield J. Sensitivity of the test of memory and learning to attention and memory deficits in children with ADHD. Clin Neuropsychol. 2010;24(2):246-64.
85. Kitsao-Wekulo PK, Holding PA, Taylor HG, Kvalsvig JD, Connolly KJ. Determinants of variability in motor performance in middle childhood: a cross-sectional study of balance and motor co-ordination skills. BMC Psychology. 2013;1(1):29-.
86. Bangirana P, Seggane-Musisi., Allebeck P, Giordani B, John C, Opoka O, et al. A Preliminary Examination of the Construct Validity of the KABC-II in Ugandan Children with a History of Cerebral Malaria. African Health Sciences. 2009;9(3).
87. Mung'ala-Odera V, Meehan R, Njuguna P, Mturi N, Alcock K, Carter JA, et al. Validity and reliability of the 'Ten Questions' questionnaire for detecting moderate to severe neurological impairment in children aged 6-9 years in rural Kenya. Neuroepidemiology. 2004;23(1-2):67-72.
88. Archibald SJ, Kerns KA. Identification and description of new tests of executive functioning in children. Child Neuropsychology. 1999;5(2):115-29.
89. Dos SF, Da Silva PA, Ribeiro FS, Dias AL, Frigério MC, Dellatolas G, et al. Number processing and calculation in Brazilian children aged 7-12 years. Span J Psychol. 2012;15(2):513-25.
90. Konstantopoulos K, Vogazianos P, Thodi C, Nikopoulou-Smyrni P. A normative study of the Children's Color Trails Test (CCTT) in the Cypriot population. Child Neuropsychol. 2015;21(6):751-8.
91. Fisher A, Boyle JM, Paton JY, Tomporowski P, Watson C, McColl JH, et al. Effects of a physical education intervention on cognitive function in young children: randomized controlled pilot study. BMC Pediatr. 2011;11:97.
92. Carone DA. Young child with severe brain volume loss easily passes the word memory test and medical symptom validity test: Implications for mild TBI. The Clinical Neuropsychologist. 2014;28(1):146-62.
93. Flake JK, Fried EI. Measurement schmeasurement: Questionable measurement practices and how to avoid them. Advances in Methods and Practices in Psychological Science. 2020;3(4):456-65.
94. Vriezen ER, Pigott SE. The relationship between parental report on the BRIEF and performance-based measures of executive function in children with moderate to severe traumatic brain injury. Child Neuropsychol. 2002;8(4):296-303.
95. Forns J, Esnaola M, Lopez-Vicente M, Suades-Gonzalez E, Alvarez-Pedrerol M, Julvez J, et al. The n-back Test and the Attentional Network Task as Measures of Child Neuropsychological Development in Epidemiological Studies. Neuropsychology. 2014;28(4):519-29.
96. Gold AB, Ewing-Cobbs L, Cirino P, Fuchs LS, Stuebing KK, Fletcher JM. Cognitive and behavioral attention in children with math difficulties. Child Neuropsychol. 2013;19(4):420-37.
97. Sobin C, Kiley-Brabeck K, Daniels S, Blundell M, Anyane-Yeboa K, Karayiorgou M. Networks of attention in children with the 22q11 deletion syndrome. Developmental Neuropsychology. 2004;26(2):611-26.
98. Ahonniska J, Ahonen T, Aro T, Tolvanen A, Lyytinen H. Practice effects on visuomotor and problem-solving tests by children. Perceptual and Motor Skills. 2001;92(2):479-94.

99. Syvaoja HJ, Tammelin TH, Ahonen T, Rasanen P, Tolvanen A, Kankaanpaa A, et al. Internal consistency and stability of the CANTAB neuropsychological test battery in children. Psychol Assess. 2015;27(2):698-709.
100. Chiang IA, Jhangiani RS, Price PC. Research Methods in Psychology: Reliability and Validity of Measurement. 2 ed. Canada: BC Campus; 2015 (October, 13).
101. Malda M, Vijver FJRVD, Transler C, Sukumar P, Srinivasan K, Rao K. Adapting a cognitive test for a different culture: An illustration of qualitative procedures. Psychology Science Quarterly. 2008;50(4):451-68.
102. Cooper SR, Gonthier C, Barch DM, Braver TS. The Role of Psychometrics in Individual Differences Research in Cognition: A Case Study of the AX-CPT. Frontiers in psychology. 2017;8:1482-.
103. Best JR, Miller PH. A Developmental Perspective on Executive Function. Child Dev. 2010;81(6).
104. Ellingsen KM. Standardized Assessment of Cognitive Development: Instruments and Issues. Early Childhood Assessment in School and Clinical Child Psychology2016. p. 25-49.
105. Nolte MT, Shauver MJ, Chung KC. Analysis of four recruitment methods for obtaining normative data through a Web-based questionnaire: a pilot study. Hand (N Y). 2015;10(3):529-34.
106. Beaumont JL, Havlik R, Cook KF, Hays RD, Wallner-Allen K, Korper SP, et al. Norming plans for the NIH Toolbox. Neurology. 2013;80(11 Supplement 3):S87-S92.
107. Wicherts J, Dolan C, Carlson J, Maas H. Raven's test performance of sub-Saharan Africans: Average performance, psychometric properties, and the Flynn Effect. Learning and Individual Differences. 2010;20:135-51.
108. Nussbaumer-Streit B, Klerings I, Dobrescu AI, Persad E, Stevens A, Garritty C, et al. Excluding non-English publications from evidence-syntheses did not change conclusions: a meta-epidemiological study. Journal of Clinical Epidemiology. 2020;118:42-54.
109. Morrison A, Polisena J, Husereau D, Moulton K, Clark M, Fiander M, et al. The effect of english-language restriction on systematic review-based meta-analyses: A systematic review of empirical studies. International journal of technology assessment in health care. 2012;28:138-44.
110. Wedderburn CJ, Evans C, Yeung S, Gibb DM, Donald KA, Prendergast AJ. Growth and Neurodevelopment of HIV-Exposed Uninfected Children: a Conceptual Framework. Current HIV/AIDS reports. 2019;16(6):501-13.
111. Kammerer B, Isquith PK, Lundy S. Approaches to Assessment of Very Young Children in Africa in the Context of HIV. Neuropsychology of Children in Africa2013. p. 17-36.
112. Musindo O, Bangirana P, Kigamwa P, Okoth R, Kumar M. Neurocognitive functioning of HIV positive children attending the comprehensive care clinic at Kenyatta national hospital: exploring neurocognitive deficits and psychosocial risk factors. AIDS Care. 2018;30(5):618-22.
113. Eckard AR, Rosebush JC, O'Riordan MA, Graves CC, Alexander A, Grover AK, et al. Neurocognitive dysfunction in HIV-infected youth: investigating the relationship with immune activation. Antiviral Therapy. 2017;22(8):669-80.
114. Boivin MJ. African Multi-Site 2-Year Neuropsychological Study of School-Age Children Perinatally Infected, Exposed, and Unexposed to Human Immunodeficiency Virus. Clinical infectious diseases : an official publication of the Infectious Diseases Society of America. 2019.
115. Sherr L, Croome N, Parra Castaneda K, Bradshaw K, Herrero Romero R. Developmental challenges in HIV infected children—An updated systematic review. Children and Youth Services Review. 2014;45:74-89.
116. WHO. Global Health Sector Strategy on HIV 2016-2021: Towards Ending AIDS. Geneva: World Health Organisation; 2016. p. 57.
117. Carlozzi NE, Beaumont JL, Tulsky DS, Gershon RC. The NIH Toolbox Pattern Comparison Processing Speed Test: Normative Data. Archives of Clinical Neuropsychology. 2015;30(5):359-68.
118. Gualtieri CT, Johnson LG. Reliability and validity of a computerized neurocognitive test battery, CNS Vital Signs. Archives of Clinical Neuropsychology. 2006;21(7):623-43.

119. Hahn E, Thi MTT, Hahn C, Kuehl LK, Ruehl C, Neuhaus AH, et al. Test retest reliability of Attention Network Test measures in schizophrenia. Schizophrenia Research. 2011;133(1-3):218-22.
120. Rosa VO, Schmitz M, Moreira-Maia CR, Wagner F, Londero I, Bassotto CF, et al. Computerized cognitive training in children and adolescents with attention deficit/hyperactivity disorder as add-on treatment to stimulants: feasibility study and protocol description. Trends Psychiatry Psychother. 2017;39(2):65-76.
121. Sternberg RJ, Nokes C, Geissler PW, Prince R, Okatcha F, Bundy DA, et al. The relationship between academic and practical intelligence: a case study in Kenya. Intelligence. 2001;29(5):401-18.
122. Fernald LCH, Engle P, Kariger P, Raikes A, World Bank eLibrary - York U. Examining Early Child Development in Low-Income Countries : a Toolkit for the Assessment of Children in the First Five Years of Life. Washington, D.C.: The World Bank; 2009. Available from: http://elibrary.worldbank.org/doi/book/10.1596/28107.
123. Rosselli M, Ardila A. The impact of culture and education on non-verbal neuropsychological measurements: a critical review. Brain Cogn. 2003;52(3):326-33.
124. Burneo-Garces C, Cruz-Quintana F, Perez-Garcia M, Fernandez-Alcantara M, Fasfous A, Perez-Marfil MN. Interaction between Socioeconomic Status and Cognitive Development in Children Aged 7, 9, and 11 Years: A Cross-Sectional Study. Dev Neuropsychol. 2019;44(1):1-16.
125. Diamond A. Executive Functions. Annual Review of Psychology. 2013;64(1):135-68.
126. Duff K, Schoenberg MR, Scott JG, Adams RL. The relationship between executive functioning and verbal and visual learning and memory. Archives of Clinical Neuropsychology. 2005;20(1):111-22.
127. Chang YL, Jacobson MW, Fennema-Notestine C, Hagler DJ, Jr., Jennings RG, Dale AM, et al. Level of executive function influences verbal memory in amnestic mild cognitive impairment and predicts prefrontal and posterior cingulate thickness. Cereb Cortex. 2010;20(6):1305-13.
128. Kenya National Bureau of Statistics. Kenya National Adult Literacy Survey Report. Nairobi; 2007.
129. Clark N. Education in Kenya 2015 [Available from: https://wenr.wes.org/2015/06/education-kenya.
130. International Test Commission. The ITC Guidelines for Translating and Adapting Tests. 2017.
131. Thomas ML, Patt VM, Bismark A, Sprock J, Tarasenko M, Light GA, et al. Evidence of systematic attenuation in the measurement of cognitive deficits in schizophrenia. J Abnorm Psychol. 2017;126(3):312-24.
132. Streiner DL, Norman GR, Cairney J. Health Measurement Scales: A Practical Guide to Their Development and Use: Oxford University Press; 2015.
133. Boerma IE, Mol SE, Jolles J. Reading Pictures for Story Comprehension Requires Mental Imagery Skills. Front Psychol. 2016;7:1630.
134. Schatz P, Ferris CS. One-Month Test–Retest Reliability of the ImPACT Test Battery. Archives of Clinical Neuropsychology. 2013;28(5):499-504.
135. Moser RS, Schatz P, Grosner E, Kollias K. One year test-retest reliability of neurocognitive baseline scores in 10- to 12-year olds. Appl Neuropsychol Child. 2017;6(2):166-71.
136. Pearson. Retesting Time Advice for Clinical Assessments 2020 [Available from: https://support.pearson.com/usclinical/s/article/Clinical-Customer-Support-Test-Retest-Minimum-Time-Advice.
137. Fernández-Marcos T, de la Fuente C, Santacreu J. Test–retest reliability and convergent validity of attention measures. Applied Neuropsychology: Adult. 2018;25(5):464-72.
138. Christensen KM, Joschko M. Construct validity of the continuous attention test for children. Clin Neuropsychol. 2001;15(2):203-9.
139. Chernoff MC, Laughton B, Ratswana M, Familiar I, Fairlie L, Vhembo T, et al. Validity of Neuropsychological Testing in Young African Children Affected by HIV. J Pediatr Infect Dis. 2018;13(3):185-201.

140. Raiker JS, Manning E, Herrington B, May AC, Haynes S, Graves PE, et al. Brief neurocognitive screening in youth with brain tumours: A preliminary investigation of the Lebby-Asbell Neurocognitive Screening Examination (LANSE). Brain Injury. 2015;29(10):1192-8.
141. Shaked D, Faulkner LMD, Tolle K, Wendell CR, Waldstein SR, Spencer RJ. Reliability and validity of the Conners' Continuous Performance Test. Applied Neuropsychology: Adult. 2019:1-10.
142. Serpell R, Jere-Folotiya J. Basic Education for Children with Special Needs in Zambia. Psychology and Developing Societies. 2011;23(2):211-45.
143. Huff KL, Sireci SG. Validity Issues in Computer-Based Testing. Educational Measurement: Issues and Practice. 2005;20(3):16-25.
144. Hassler Hallstedt M, Ghaderi A. Tablets instead of paper-based tests for young children? Comparability between paper and tablet versions of the mathematical Heidelberger Rechen Test 1-4. Educational Assessment. 2018;23(3):195-210.
145. Carlson KD, Herdman AO. Understanding the Impact of Convergent Validity on Research Results. Organizational Research Methods. 2010;15(1):17-32.
146. Davidson F, Cherry K, Corkum P. Validating the Behavior Rating Inventory of Executive Functioning for Children With ADHD and Their Typically Developing Peers. Appl Neuropsychol Child. 2016;5(2):127-37.
147. Berg C, Edwards DF, King A. Executive function performance on the children's kitchen task assessment ith children with sickle cell disease and matched controls. Child Neuropsychology. 2012;18(5):432-48.
148. Berger I, Slobodin O, Cassuto H. Usefulness and Validity of Continuous Performance Tests in the Diagnosis of Attention-Deficit Hyperactivity Disorder Children. Arch Clin Neuropsychol. 2017;32(1):81-93.
149. Schirmbeck K, Rao N, Maehler C. Similarities and differences across countries in the development of executive functions in children: A systematic review. Infant and Child Development. 2020;29(1):e2164.
150. Röthlisberger M, Neuenschwander R, Cimeli P, Roebers CM. Executive Functions in 5- to 8-Year Olds: Developmental Changes and Relationship to Academic Achievement. Journal of Educational and Developmental Psychology. 2013;3(2).
151. Buttelmann F, Karbach J. Development and Plasticity of Cognitive Flexibility in Early and Middle Childhood. Front Psychol. 2017;8:1040.
152. Wicherts JM, Dolan CV. Measurement invariance in confirmatory factor analysis: An illustration using IQ test performance of minorities. Educational Measurement: Issues and Practice. 2010;29(3):39-47.
153. United Nations Children's Fund, World Health Organization, The World Bank. UNICEF-WHO-The World Bank: Joint child malnutrition estimates, levels and trends. New York, Geneva and Washington, DC: UNICEF, WHO and The World Bank; 2021.
154. Crookston BT, Dearden KA, Alder SC, Porucznik CA, Stanford JB, Merrill RM, et al. Impact of early and concurrent stunting on cognition. Maternal & child nutrition. 2011;7(4):397-409.
155. Ekholuenetale M, Barrow A, Ekholuenetale CE, Tudeme G. Impact of stunting on early childhood cognitive development in Benin: evidence from Demographic and Health Survey. Egyptian Pediatric Association Gazette. 2020;68(1):31.
156. Cortés Pascual A, Moyano Muñoz N, Quílez Robres A. The relationship between executive functions and academic performance in primary education: Review and meta-analysis. Front Psychol. 2019;10:1582.
157. Pellicano E, Kenny L, Brede J, Klaric E, Lichwa H, McMillin R. Executive function predicts school readiness in autistic and typical preschool children. Cognitive Development. 2017;43:1-13.
158. Victora CG, Adair L, Fall C, Hallal PC, Martorell R, Richter L, et al. Maternal and child undernutrition: consequences for adult health and human capital. The Lancet. 2008;371(9609):340-57.

159. Ritchie SJ, Tucker-Drob EM. How much does education improve intelligence? A meta-analysis. Psychol Sci. 2018;29(8):1358-69.
160. Leroy JL, Frongillo EA. Perspective: What does stunting really mean? A critical review of the evidence. Advances in Nutrition. 2019;10(2):196-204.
161. Hoddinott J, Behrman JR, Maluccio JA, Melgar P, Quisumbing AR, Ramirez-Zea M, et al. Adult consequences of growth failure in early childhood. Am J Clin Nutr. 2013;98(5):1170-8.
162. Black RE, Allen LH, Bhutta ZA, Caulfield LE, de Onis M, Ezzati M, et al. Maternal and child undernutrition: global and regional exposures and health consequences. The Lancet. 2008;371(9608):243-60.
163. UN General Assembly. Transforming our world : The 2030 agenda for sustainable development Geneva2015 [Available from: https://www.refworld.org/docid/57b6e3e44.html.
164. Abate BB, Aragie TG, Tesfaw G. Magnitude of underweight, wasting and stunting among HIV positive children in East Africa: A systematic review and meta-analysis. PLOS ONE. 2020;15(9):e0238403.
165. Ardila A, Rosselli M, Matute E, Inozemtseva O. Gender Differences in Cognitive Development. Developmental Psychology. 2011;47(4):984-90.
166. Jäncke L. Sex/gender differences in cognition, neurophysiology, and neuroanatomy. F1000Research. 2018;7.
167. Wedderburn CJ, Yeung S, Rehman AM, Stadler JAM, Nhapi RT, Barnett W, et al. Neurodevelopment of HIV-exposed uninfected children in South Africa: outcomes from an observational birth cohort study. The Lancet Child & Adolescent Health. 2019;3(11):803-13.
168. McDonald CM, Manji KP, Kupka R, Bellinger DC, Spiegelman D, Kisenge R, et al. Stunting and wasting are associated with poorer psychomotor and mental development in HIV-exposed Tanzanian infants. The Journal of nutrition. 2013;143(2):204-14.
169. Maina R, He J, Abubakar A, Perez-Garcia M, Kumar M, Wicherts J. Psychometric evaluation of the computerized battery for neuropsychological evaluation of children (BENCI) among school aged children in the context of HIV in an urban Kenyan setting. BMC Psychiatry. 2023;23(373).
170. Vangal KS, Rajneesh T. Malnutrition in HIV/AIDS - aetiopathogenesis. In: Nancy D, editor. Nutrition and HIV/AIDS. Rijeka: IntechOpen; 2020.
171. Phillips N, Amos T, Kuo C, Hoare J, Ipser J, Thomas KG, et al. HIV-Associated cognitive impairment in perinatally infected children: A meta-analysis. Pediatrics. 2016;138(5).
172. Kitsao-Wekulo P, Holding P, Taylor HG, Abubakar A, Kvalsvig J, Connolly K. Nutrition as an important mediator of the impact of background variables on outcome in middle childhood. Frontiers in Human Neuroscience. 2013;7(713).
173. Ajayi OR, Matthews GB, Taylor M, Kvalsvig JD, Davidson L, Kauchali S, et al. Structural Equation Modeling of the Effects of Family, Preschool, and Stunting on the Cognitive Development of School Children. Frontiers in Nutrition. 2017;4(17).
174. Lezak MD, Howieson DB, Loring DW. Neuropsychological assessment. 4 ed. New York: Oxford University Press; 2004.
175. Alam MA, Richard SA, Fahim SM, Mahfuz M, Nahar B, Das S, et al. Impact of early-onset persistent stunting on cognitive development at 5 years of age: Results from a multi-country cohort study. PLoS One. 2020;15(1):e0227839.
176. Neary J, Langat A, Singa B, Kinuthia J, Itindi J, Nyaboe E, et al. Higher prevalence of stunting and poor growth outcomes in HIV-exposed uninfected than HIV-unexposed infants in Kenya. AIDS. 2022;36(4):605-10.
177. McHenry MS, Apondi E, Ayaya SO, Yang Z, Li W, Tu W, et al. Growth of young HIV-infected and HIV-exposed children in western Kenya: A retrospective chart review. PLOS ONE. 2019;14(12):e0224295.
178. Gur RC, Richard J, Calkins ME, Chiavacci R, Hansen JA, Bilker WB, et al. Age Group and Sex Differences in Performance on a Computerized Neurocognitive Battery in Children Age 8-21. Neuropsychology. 2012;26(2):251-65.

179. Ramful A, Lowrie T. Spatial Visualisation and Cognitive Style: How Do Gender Differences Play Out? Mathematics Education Research Group of Australasia. 2015:508-15.
180. Jesson J, Schomaker M, Malasteste K, Wati DK, Kariminia A, Sylla M, et al. **Stunting and growth velocity of adolescents with perinatally acquired HIV: differential evolution for males and females. A multiregional analysis from the IeDEA global paediatric collaboration**. J Int AIDS Soc. 2019;22(11):e25412.
181. Weis M, Heikamp T, Trommsdorff G. Gender differences in school achievement: The role of self-regulation. Frontiers in Psychology. 2013;4.
182. Casale D, Desmond C. Recovery from stunting and cognitive outcomes in young children: evidence from the South African Birth to Twenty Cohort Study. Journal of Developmental Origins of Health and Disease. 2016;7(2):163-71.
183. Lazonder AW, Janssen N, Gijlers H, Walraven A. Patterns of development in children's scientific reasoning: Results from a three-year longitudinal study. Journal of Cognition and Development. 2021;22(1):108-24.
184. Maina RW, Abubakar A, Perez-Garcia M, Van De Vijver FJR, Kumar M. Standardization of the Computerized Battery for Neuropsychological Evaluation of Children (BENCI) in an urban setting, in Kenya: a study protocol. BMC Res Notes. 2019;12(799).
185. Statistics KNBo. Basic report on well-being in Kenya: based on the 2015/16 Kenya Integrated Household Budget Survey (KIHBS). Kenya National Bureau of Statistics (KNBS); 2018. Report No.: 978-9966-102-02-7.
186. Ministry of Planning and National Development. Basic Report on Well-being in Kenya Based on Kenya Integrated Household Budget Survey- 2005/06. 2007. Report No.: 9966-767-08-8.
187. World Health Organisation. Training course on child growth assessment. Geneva: WHO,; 2008.
188. World Health Organisation. Growth Reference Data for 5-19 Years 2013 [cited 2020 6th December]. Available from: www.who.int/growthref/en/.
189. Schumacker RE. Teacher's Corner: Conducting Specification Searches With Amos. Structural Equation Modeling: A Multidisciplinary Journal. 2006;13(1):118-29.
190. Tarka P. An overview of structural equation modeling: its beginnings, historical development, usefulness and controversies in the social sciences. Qual Quant. 2018;52(1):313-54.
191. Hu L, Bentler PM. Cutoff Criteria for Fit Indexes in Covariance Structure Analysis: Conventional Criteria Versus New Alternatives. Structural Equation Modeling. 1999;6:1-55.
192. Moshagen M, Erdfelder E. A New Strategy for Testing Structural Equation Models. Structural Equation Modeling: A Multidisciplinary Journal. 2016;23(1):54-60.
193. Stanaway JD, Afshin A, Gakidou E, Lim SS, Abate D, Abate KH, et al. Global, regional, and national comparative risk assessment of 84 behavioural, environmental and occupational, and metabolic risks or clusters of risks for 195 countries and territories, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. The Lancet. 2018;392(10159):1923-94.
194. Nigussie J, Girma B, Molla A, Mareg M, Mihretu E. Under-nutrition and associated factors among children infected with human immunodeficiency virus in sub-Saharan Africa: a systematic review and meta-analysis. Arch Public Health. 2022;80(1):19.
195. Wang Q. The cultural foundation of human memory. Annual Review of Psychology. 2021;72(1):151-79.
196. Walker SP, Chang SM, Powell CA, Grantham-McGregor SM. Effects of early childhood psychosocial stimulation and nutritional supplementation on cognition and education in growth-stunted Jamaican children: prospective cohort study. Lancet. 2005;366(9499):1804-7.
197. Chen Q. Impacts of Late School Entry on Children's Cognitive Development in Rural Northwestern China—Does Preprimary Education Matter? Asia & the Pacific Policy Studies. 2017;4(3):586-601.

198. Aduda PO, Kodero HMN, Sichari M. The Effect of Class Repetition on the Academic Performance of Pupils in Lower Primary Schools in Homa-Bay Sub-County. International Journal For Research In Educational Studies. 2019;5(9):37-54.
199. Hong G, Raudenbush SW. Effects of Kindergarten Retention Policy on Children's Cognitive Growth in Reading and Mathematics. Educational Evaluation and Policy Analysis. 2005;27(3):205-24.
200. Wu W, West SG, Hughes JN. Effect of Retention in First Grade on Children's Achievement Trajectories Over 4 Years: A Piecewise Growth Analysis Using Propensity Score Matching. J Educ Psychol. 2008;100(4):727-40.
201. Glick P, Sahn DE. Early academic performance, grade repetition, and school attainment in Senegal: A panel data analysis. The World Bank Economic Review. 2010;24(1):93-120.
202. Intiful FD, Abdulai H, Nyarko R, Tette E, Asante M. Malnutrition in HIV infected children on antiretroviral drugs in a cohort of Ghanaian children. Heliyon. 2021;7(12):e08636.
203. Halpern DF. Sex differences in cognitive abilities. 4 ed: Taylor & Francis; 2013.
204. Scheiber C, Reynolds M, Hajovsky D, Kaufman A. Gender differences in achievement in a large, nationally representative sample of children and adolescents. Psychology in the Schools. 2015;52.
205. Vedechkina M, Borgonovi F. A review of evidence on the role of digital technology in shaping attention and cognitive control in children. Frontiers in Psychology. 2021;12.
206. Grammer JK, Coffman JL, Ornstein PA, Morrison FJ. Change over time: Conducting longitudinal studies of children's cognitive development. Journal of Cognition and Development. 2013;14(4):515-28.
207. Kenny DA. MedPower: An interactive tool for the estimation of power in tests of mediation [Computer software] 2017 [Available from: https://davidakenny.shinyapps.io/MedPower/.
208. Khan T, Khan REA, Raza MA. Gender Analysis of Malnutrition: A Case Study of School-Going Children in Bahawalpur. Asian Development Policy Review. 2015;3(2):29-48.
209. Astatkie A. Dynamics of stunting from childhood to youthhood in Ethiopia: Evidence from the Young Lives panel data. PLOS ONE. 2020;15(2):e0229011.
210. Fasfous A, Puente A, Perez-Marfil M, Cruz-quintana F, Peralta-ramirez I, Perez-Garcia M. Is the color trails culture free ? Archives of clinical neuropsychology : the official journal of the National Academy of Neuropsychologists. 2013;28(7):743-9.
211. Dutton E, Becker D, Osman HA, Bakhiet SF, Essa YAS, Ali HAA, et al. The Raven's test performance of South Sudanese samples: A validation of criticisms of the utility of Raven's among Sub-Saharan Africans. Personality and Individual Differences. 2018;128:122-6.
212. Holding P, Anum A, van de Vijver FJR, Vokhiwa M, Bugase N, Hossen T, et al. Can we measure cognitive constructs consistently within and across cultures? Evidence from a test battery in Bangladesh, Ghana, and Tanzania. Appl Neuropsychol Child. 2018;7(1):1-13.
213. Rachel M, Fons VDVJR, Amina A, Perez-Garcia M, Manasi K. Assessing neuropsychological functions in middle childhood: A narrative review of measures and their psychometric properties across context. Journal of Pediatric Neuropsychology. 2021;7(3):113-38.
214. Werker JF, Hensch TK. Critical periods in speech perception: New directions. Annual Review of Psychology. 2015;66(1):173-96.
215. Maas H, Dolan C, Grasman R, Wicherts J, Huizenga Hm, Raijmakers M. A dynamical model of general intelligence: The positive manifold of intelligence by mutualism. Psychological Review. 2006;113:842-61.
216. Van Der Maas HLJ, Kan KJ, Marsman M, Stevenson CE. Network models for cognitive development and intelligence. J Intell. 2017;5(2).
217. de Bruijn AGM, Meijer A, Königs M, Oosterlaan J, Smith J, Hartman E. The mediating role of neurocognitive functions in the relation between physical competencies and academic achievement of primary school children. Psychology of Sport and Exercise. 2023;66:102390.
218. Lindhiem O, Petersen IT, Mentch LK, Youngstrom EA. The importance of calibration in clinical psychology. Assessment. 2020;27(4):840-54.

219. Nasreddine ZS. MoCA test mandatory training and certification: What is the purpose? Journal of the American Geriatrics Society. 2020;68(2):444-5.
220. Ministry of Health-Kenya. Kenyan national guidelines on nutrition and HIV. Kenya: Ministry of Health; 2014.

INDEX

SAMENVATTING VAN RESULTATEN

De algemene doelstellingen van dit proefschrift waren het beoordelen van de neurocognitieve meetinstrumenten die worden gebruikt bij basisschoolleerlingen en het aanpassen en valideren van de Computerized Battery for Neuropsychological Evaluation of Children (BENCI) in een cohort van Keniaanse kinderen met en zonder HIV. Bovendien hebben we de BENCI gebruikt om de rol van groeiachterstand in de cognitieve uitkomsten van kinderen met HIV te evalueren. Deze onderzoeken dragen bij aan de meting van neurocognitieve vaardigheden van 6-12-jarige kinderen in lagere- en middeninkomenslanden (LMICs), zoals Kenia, en kinderen die lijden aan groeiachterstanden en HIV. Dit proefschrift is onderverdeeld in vijf hoofdstukken. In het eerste hoofdstuk introduceren we de volgende hoofdstukken en bespreken we hiaten in de kennis over neurocognitieve meetinstrumenten die in de leeftijdsgroepen 6-12 jaar worden gebruikt en de noodzaak om vervolgonderzoek uit te voeren om geldige neurocognitieve beoordelingen mogelijk te maken die de ontwikkeling monitoren en toekomstige interventies informeren.

Hoofdstuk 2 presenteert een narratieve overzichtsstudie van neurocognitieve meetinstrumenten die vaak worden gebruikt in onderzoek in de middenkindertijd wereldwijd, waarbij rekening wordt gehouden met hun psychometrische eigenschappen in diverse culturele contexten. In het bijzonder hebben we gedocumenteerd waar deze instrumenten zijn ontwikkeld en/of getest, en hebben we de geschiktheid van psychometrische uitkomsten, zoals de betrouwbaarheid en validiteit van de instrumenten, in verschillende contexten geëvalueerd. We hebben talloze neurocognitieve meetinstrumenten gevonden die zijn ontwikkeld en getest onder 6-12-jarigen, waarbij de meeste meetinstrumenten zijn ontwikkeld en gevalideerd in welvarende landen. De geteste en ontwikkelde instrumenten maten bepaalde domeinen van neurocognitieve prestaties, zoals de executieve functies, terwijl andere domeinen buiten

beschouwing werden gelaten die als relevant worden beschouwd volgens de Diagnostic and Statistical Manual of Mental Disorders, vijfde editie (DSM-V). Bovendien ontdekten we dat verschillende gestandaardiseerde instrumenten voor de verschillende domeinen beperkt waren tot één geografische setting of onvoldoende bestudeerd waren op psychometrische eigenschappen. In de beoordeelde literatuur waren tests van het executief functioneren oververtegenwoordigd in vergelijking met tests op andere domeinen. Ten slotte hebben we verschillende tekortkomingen waargenomen in het literatuur over de ontwikkeling en aanpassing van neurocognitieve hulpmiddelen voor 6-12-jarigen in LMICs. Deze omvatten gemengde bevindingen met betrekking tot test-hertestbetrouwbaarheid en constructvaliditeit, en een gebrek aan normgegevens. Een belangrijk hiaat in de literatuur is de schaarste aan adequaat aangepaste, gevalideerde en gestandaardiseerde instrumenten voor Sub-Sahara Afrika.

In de studie gerapporteerd in Hoofdstuk 3 hebben we de Computerised Battery for Neuropsychological Evaluation of Children (BENCI) onder basisschoolleerlingen in Kenia aangepast en gevalideerd. We gaven de voorkeur aan dit meetinstrument omdat het alle domeinen bestrijkt die door de DSM-V worden aanbevolen, met uitzondering van sociale cognitie. We vertaalden de tool van het Spaans naar het Engels, testten de aangepaste tool onder Keniaanse kinderen en testten uiteindelijk de psychometrische eigenschappen van de aangepaste tool bij een grote steekproef van Keniaanse kinderen in de leeftijd van 6 tot 12 jaar die met en zonder HIV leefden. Het adaptatiesproces hield in dat tweetalige onderzoekers het instrument naar het Engels vertaalden, de taalkundige en semantische consistentie moesten controleren en zowel de structuur van het instrument (subtests, instructies en volgorde van afname) als de geschiktheid ervan ten opzichte van de oorspronkelijke instructies (bijvoorbeeld het aantal blokken per subtest, aantal pogingen per blok, percentage doelstimuli, presentatietijd). Bij het vergelijken van de prestaties van 328 kinderen met HIV en die van 260

kinderen die niet met HIV leefden, vonden we voor de meeste subtests een goede interne consistentie en test-hertestbetrouwbaarheid. We hebben de convergente validiteit van de BENCI bestudeerd met behulp van lokaal gevalideerde papier- en potloodinstrumenten en hebben goede correlaties gevonden tussen de geheugen-, redeneer- en inhibitie-domeinen van BENCI en de lokale instrumenten uit de Kilifi-toolkit (46). Een confirmatief vierfactorenmodel bestaande uit redenering, flexibiliteit, verbaal geheugen en vloeiendheid paste goed bij de gegevens en vertoonde metrische en gedeeltelijke scalaire meetinvariantie met betrekking tot de HIV-positieve en HIV-negatieve groepen. Samenvattend: hoewel sommige subtests in toekomstig werk verbeterd zouden kunnen worden, vonden we dat de aangepaste Engelse versie van de BENCI een valide en betrouwbaar neurocognitief meetinstrument is voor gebruik bij Keniaanse kinderen.

In Hoofdstuk 4 bestudeerden we de potentiële impact van HIV en groeiachterstand op de neurocognitieve prestaties op de BENCI. Concreet hebben we structurele vergelijkingsmodellen gepast op dezelfde gegevens als in hoofdstuk 3 om de prestaties op het gebied van redeneren, flexibiliteit, verbaal geheugen en spreekvaardigheid van Keniaanse kinderen op basis van HIV, geslacht, leeftijd en groeiachterstand te voorspellen. We hebben ook getest of groeiachterstand de effecten van HIV, geslacht en leeftijd op redeneren, flexibiliteit, verbaal geheugen en vloeiendheid medieert. We vonden dat HIV een directe invloed heeft op groeiachterstand en alle cognitieve uitkomsten, terwijl leeftijd alleen een directe voorspellende waarde heeft voor groeiachterstand en redenering. Er werd vastgesteld dat groeiachterstand de effecten van HIV op het redeneren, het verbale geheugen en de spreekvaardigheid (gedeeltelijk) mediëert, en dat groeiachterstand de effecten van leeftijd op de spreekvaardigheid en redenering medieert. We vonden dat kinderen met een dubbele last van HIV en groeiachterstand slechtere cognitieve prestaties hebben vergeleken met kinderen met alleen HIV of alleen groeiachterstand. We hebben geen directe effecten van geslacht op

groeiachterstand waargenomen, maar we hebben wel aanwijzingen gevonden dat vrouwelijke kinderen met HIV het meest waarschijnlijk lijden aan een groeiachterstand.

Algemene conclusies

Net als veel andere LMIC-omgevingen heeft Kenia een tekort aan adequaat gestandaardiseerde neurocognitieve meetinstrumenten met robuuste psychometrische eigenschappen die kunnen worden gebruikt om helderheid te schaffen over de behoeften van kwetsbare kinderen, zoals kinderen met HIV. Cultureel op maat gemaakte en gevalideerde tests zouden ons in staat stellen de werkelijke last van cognitieve stoornissen in LMIC's en de bijbehorende risicofactoren te kennen, maar onze literatuurstudie van cognitieve tests die onder 6- tot 12-jarigen worden gebruikt, bracht belangrijke hiaten aan het licht die worden gekenmerkt door weinig domeinen en inadequate psychometrische informatie in onderzoek naar neurocognitieve hulpmiddelen in LIMC's. Daarom hebben we de BENCI met succes aangepast voor Keniaanse kinderen en vonden we dat kinderen met een dubbele last van HIV en groeiachterstand slechtere cognitieve prestaties hebben vergeleken met kinderen met alleen HIV of alleen groeiachterstand. We hebben ook redenering, verbaal geheugen en vloeiendheid geïdentificeerd als doelwitten voor interventie bij kinderen met HIV en een groeiachterstand.

My research to date, as captured in this book, indicates process outcomes of adaptation of an iPad based neurocognitive battery - BENCI. These outcomes show the viability of adapting psychometrically sound neurocognitive tests for LMICs e.g. Kenya where the scourge of stunting and HIV may impair cognitive functioning and inadvertently put children at risk of underperforming in school, dropping out, and long-term effects of reduced income and increased poverty. This novel work in my country results from interaction with children living with HIV reporting difficulty learning at the same pace as their healthy classmates.

Clinically, they couldn't afford intrusive neurocognitive lab tests and we did not have culturally validated tests. In school, their parents couldn't afford placement in schools with fewer children where teachers could give their children extra attention in learning. Assessing for neurocognitive impairment is an important first step towards cognitive recovery. With proper culturally adapted and psychometrically sound open access cognitive tools such as BENCI, cognitive deficiencies can be assessed, and educational and treatment approaches can be tailored to promote equity in education.

In relation to my research, children living with HIV and stunting have deficits in executive functioning which can now be assessed and the children given an opportunity to learn like their schoolmates without explicit vulnerabilities. By promoting accessibility and striving for educational equity, BENCI offers a promising solution to unlocking the potential of children in resource-poor settings. Together, let us bridge the gap and create education equity for all!

Rachel Maina is a Clinical Psychologists whose interests lie around cross-cultural psychology techniques for integration of 1. Culturally aligned neurocognitive measures in LMICs and 2. Culturally aligned cognitive recovery and support in LMICs to address inequalities related to cognitive disparities.