Please cite the Published Version

Additional Information: This is an open access article which first appeared in Wellcome Open Research

Data Access Statement: Open Science Framework: Adaptation of the Kaufman Assessment Battery for Children—2nd edition (KABC-II) to assess school-age neurodevelopment in rural Zimbabwe. https://doi.org/10.17605/OSF.IO/T6GXB18.

This project contains the following underlying data:

- Extended data & results KABC-II.pdf (supplementary material)

- KABC_II_modification.xlsx (raw spreadsheet data of test scores)

Data are available under the terms of the Creative Commons Zero "No rights reserved" data waiver (CC0 1.0 Public domain dedication).

Check for updates

RESEARCH ARTICLE

REVISED

Piloting the adaptation of the Kaufman Assessment Battery for Children—2nd edition (KABC-II) to assess school-age neurodevelopment in rural Zimbabwe [version 2; peer review: 2 approved]

Joseph D. Piper [1,2], Clever Mazhanga[2], Gloria Mapako[2], Idah Mapurisa[2], Tsitsi Mashedze[2], Eunice Munyama[2], Marian Mwapaura[2], Dzivaidzo Chidhanguro[2], Grace Gerema[2], Naume V. Tavengwa[2], Robert Ntozini [2], Lisa F. Langhaug [2], Melanie Smuk [1], Tamsen Rochat[3], Alan Kaufman[4], Nadeen Kaufman[4], Melissa Gladstone [5], Elizabeth Allen[5], Andrew J. Prendergast[1]

[1]Department of Genomics and Child Health, Queen Mary University of London, London, E1 4AT, UK
[2]Zvitambo Institute for Maternal and Child Health Research, Mabelreign, Harare, Zimbabwe
[3]Department of Psychology, University of the Witwatersrand, Johannesburg, 2000, South Africa
[4]Department of Medicine, Yale University, New Haven, Connecticut, 06520-8081, USA
[5]Department of Life Course and Medical Sciences, University of Liverpool, Liverpool, L69 7ZX, UK

Abstract

Background

Neurodevelopment assessment tools for low-resource settings are urgently needed. However, most available tools were developed in high-income settings and may lack cross-cultural validity.

Methods

We piloted and adapted two subtests within the planning domain of the Kaufman Assessment Battery for Children-2nd edition (KABC-II) for use in rural Zimbabwean children aged 7years. After initial assessments of face validity, we created 4 substitutions for the story completion subtest and 7 additions for the pattern reasoning subtest through a co-design process with fieldworkers and child development experts. To assess how successful the changes were, T-tests adjusting

Open Peer Review

Approval Status ✓ ✓

	1	2
version 2 (revision) 30 May 2024	✓ view ↑	✓ view ↑
version 1 09 Nov 2022	? view	? view

1. **Supriya Bhavnani** [iD], Sangath, Bardez, India

2. **Katherine Brown** [iD], Great Ormond Street Hospital for Children National Health Service Foundation Trust, London, UK

Any reports and responses or comments on the

for unequal variances were used to compare scores between the original and adapted versions of the same subtest. ANOVA and pairwise analysis was performed to compare the performance of KABC-II subtests across domains. Intraclass correlation coefficient was calculated to explore the variability between domains.

Results

Initial test scores on the planning domain were significantly lower than the other three domains of learning, sequential memory and simultaneous reasoning (P<0.001) in 50 children (mean age 7.6(SD 0.2) years). Modified subtests were administered to another 20 children (mean age 7.6(SD 0.2) years), who showed story completion scores that were 0.7 marks higher (95% CI 0.0, 1.4; P=0.05) and pattern reasoning scores 1.8 marks higher (95% CI 0.5, 3.2; P=0.01). Overall, the planning domain mean score increased from 8.1 (SD 2.9) to 10.6 (SD 3.4). The intra class correlation coefficient between all four KABC-II domains was initially 0.43 (95% CI 0.13, 0.64) and after modification was 0.69 (95% CI 0.37, 0.87), suggesting an increase in the construct validity.

Conclusions

The KABC-II planning domain was successfully adapted to improve cross-cultural validity. Construct validity was enhanced, based on increased inter-correlations among scales. The process of co-design to modify tests for new settings may be beneficial for other commonly used neurodevelopmental tools.

Keywords
Child development, Sub-Saharan Africa, School-age, cognition

Corresponding author: Joseph D. Piper (j.piper@qmul.ac.uk)

Author roles: Piper JD: Conceptualization, Data Curation, Formal Analysis, Funding Acquisition, Investigation, Methodology, Project Administration, Software, Validation, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Mazhanga C**: Conceptualization, Data Curation, Methodology, Project Administration, Supervision; **Mapako G**: Conceptualization, Data Curation, Methodology, Project Administration; **Mapurisa I**: Conceptualization, Data Curation, Methodology, Project Administration; **Mashedze T**: Conceptualization, Data Curation, Investigation, Methodology, Project Administration; **Munyama E**: Conceptualization, Data Curation, Investigation, Methodology, Project Administration; **Mwapaura M**: Conceptualization, Data Curation, Formal Analysis, Investigation, Methodology, Project Administration, Visualization; **Chidhanguro D**: Investigation, Methodology, Project Administration; **Gerema G**: Data Curation, Methodology; **Tavengwa NV**: Conceptualization, Methodology, Project Administration, Supervision, Writing – Original Draft Preparation; **Ntozini R**: Data Curation, Formal Analysis, Methodology, Resources, Software, Supervision, Validation, Writing – Original Draft Preparation, Writing – Review & Editing; **Langhaug LF**: Conceptualization, Methodology, Project Administration, Supervision; **Smuk M**: Formal Analysis, Software, Validation, Writing – Original Draft Preparation, Writing – Review & Editing; **Rochat T**: Conceptualization, Methodology, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Kaufman A**: Conceptualization, Methodology, Writing – Original Draft Preparation, Writing – Review & Editing; **Kaufman N**: Conceptualization, Methodology, Validation, Writing – Original Draft Preparation, Writing – Review & Editing; **Gladstone M**: Conceptualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Allen E**: Investigation, Validation, Writing – Original Draft Preparation, Writing – Review & Editing; **Prendergast AJ**: Conceptualization, Formal Analysis, Funding Acquisition, Investigation, Methodology, Project Administration, Resources, Supervision, Validation, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing

How to cite this article: Piper JD, Mazhanga C, Mapako G *et al.* **Piloting the adaptation of the Kaufman Assessment Battery for Children—2 nd edition (KABC-II) to assess school-age neurodevelopment in rural Zimbabwe [version 2; peer review: 2 approved]** Wellcome Open Research 2024, **7**:274 https://doi.org/10.12688/wellcomeopenres.17902.2

First published: 09 Nov 2022, **7**:274 https://doi.org/10.12688/wellcomeopenres.17902.1

Introduction

Over 250 million children worldwide are at risk of poor cognitive development[1]. Neurodevelopmental assessment requires tools that are adaptable to low- and middle-income countries (LMIC) settings. While some locally developed tools are available[2], many widely used tools have been developed and validated in high-resource settings, including Wechsler's Intelligence Scales[3], the Bayley Scales for Infant and Toddler Development[4] and the second edition of the Kaufman Assessment Battery for Children (KABC-II)[5]. These tools incorporate tests and questions that are relevant to the settings where they were developed and validated but may require a lengthy adaptation process for cross-cultural use.

One of the advantages of KABC-II is that it has a dual theoretical framework, with up to 18 individual tests (called subtests) that use either the Cattell-Horn-Carroll (CHC) psychometric model or Luria's cognitive processing model. Luria's approach is often used in LMICs because it measures cognitive processing by focusing on novel tests and puzzles not generally seen in schools, removing subtests that rely on acquired knowledge. This helps to correct for variable school enrolment and exposure. The KABC-II can be condensed to eight subtests and the raw subtest scores are then scaled, based on the participant's age. The scaled results of two subtests are added together to each of the four Luria domains of cognitive processing as follows (Figure 1)[6]: Number Recall and Word Order subtests give the Sequential memory domain; Story Completion and Pattern Reasoning subtests provide the Planning domain; Atlantis and Atlantis Delayed subtests constitute the Learning domain; and Rover and Triangles subtests provide the Simultaneous logic domain. The subtest scaled scores can also be combined to provide the mental processing index (MPI) as an overall measure of cognitive function[6].
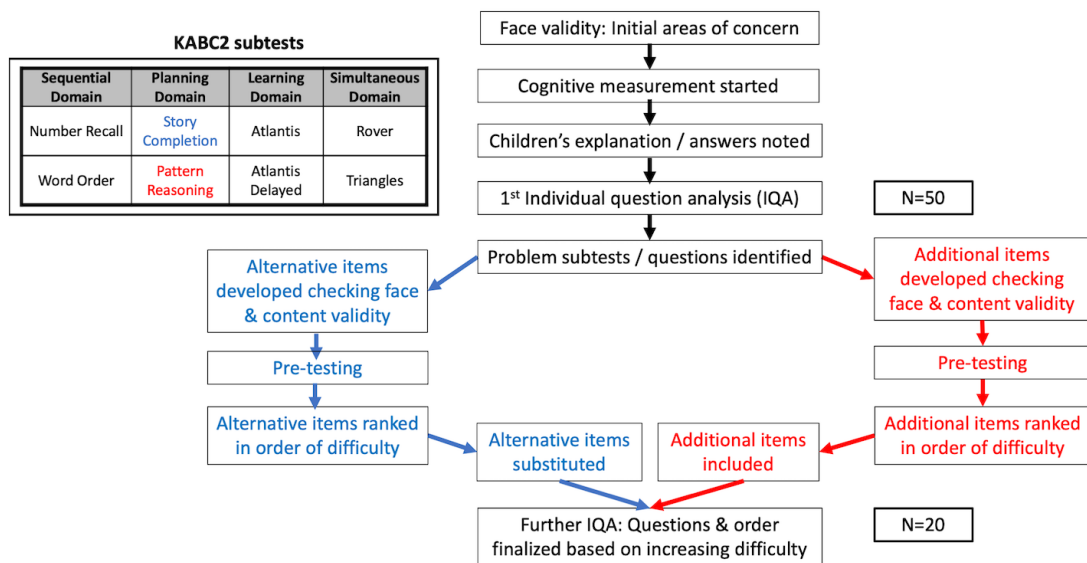

Figure 1. Stepwise process of monitoring and adaptation used for cognitive measurement. Inset: Structure of the KABC-II showing domains and subtests. KABC-II, Kaufman Assessment Battery for Children-2nd edition.

The KABC-II was originally developed and validated using a large sample of children in the USA[5]. It has since been widely used across Africa[7], demonstrating robust factor analysis in Uganda[8] and psychometric validity and reliability in rural South Africa[6,9]. Recently, improved KABC-II monitoring and quality assurance has been demonstrated using regular video review[10] across multiple countries and languages in Africa, including with a Shona translation in Zimbabwe[10,11]. Using KABC-II, a significant effect on cognition was detected following a nutrition intervention in South African children aged 6–11 years old on two of the subtests[12], whilst in Ethiopia, 5-year-old children with poorer growth had worse KABC-II scores than those with good growth[13]. In Burkina Faso, five of the KABC-II subtests identified poorer scores in children with stunting[14]. However, these studies[12–14] did not include the Story Completion subtest and so did not calculate a mental processing index (MPI). For studies that did calculate a MPI using the Luria model[10,11] the planning domain did score lower than other domains, but no major concerns with validity were noted by the authors. To optimally measure cognitive processing, a certain level of baseline understanding of the subtest and individual items should be achieved. Although it has been widely used across sub-Saharan Africa, there has been little documented pre-testing and piloting exploring cross-cultural validity within these individual subtests. Here we report two separate methods for piloting the adaptation of KABC-II cognitive subtests following concerns raised in their face validity during its use in rural Zimbabwe. The adaptations from this pilot study will then be applied for measuring school-aged children in rural Zimbabwe who were previously in The Sanitation Hygiene Infant Nutrition Efficacy (SHINE) cluster randomised study[15].

Methods
Participants: Study site and data collection
A pilot study to assess school-age children's growth, physical and cognitive development was conducted in Zvamabande (rural) and Makusha (peri-urban) regions of Shurugwi district in Midlands Province, Zimbabwe between 3rd September 2020 and 3rd December 2020[16]. Inclusion criteria were children aged 7 years old who were resident in either of these two regions, identified by the Community healthworkers (CHWs) and with a primary caregiver available who was able to consent. The site for this pilot study included where the SHINE cluster randomised trial was previously conducted, so any children who were born into the SHINE trial were excluded. From the children initially identified by CHWs, 80 children (mean age 7.6 years, SD 0.2) were randomly selected by computer in Harare, thus minimising selection bias (80 was a convenient sample size to achieve during the three months of piloting, and enabled the team to test a variety of other measures of cognitive and physical function in the pilot study detailed elsewhere[16]). After community sensitisation, a specific sensitisation visit to the family was conducted by the CHWs using a community sensitisation sheet in Shona or Ndebele (the local languages). If the family expressed an interest in participating, a mutually convenient date was arranged. Caregivers then gave written informed consent (including the right to withdraw at any time) and children gave written assent for participation. Ethical approval for the pilot study was obtained from the Medical Research

Council of Zimbabwe on 6th April 2020 (MRCZ/A/1675). There was an amendment for this pilot study that was approved on 31st July 2020. There were no recorded harms from this study. The main SHINE follow-up study was approved by the Medical Research Council of Zimbabwe (MRCZ), approval number MRCZ/A/1675 on 8th February 2021. Measurements and analyses for this main follow-up study are ongoing and due to finish in 2023. Further details of the SHINE Follow-up study are available as a preprint[17] and on Open Science Framework.

Cognitive tests comprised the KABC-II, a school achievement test measuring literacy and numeracy, a finger-tapping task to measure fine motor skills, the caregiver-reported Strength and Difficulties Questionnaire, and a child socioemotional score, as previously described[16]. The KABC-II was selected as part of the test battery because previous factor analysis had shown it was applicable in a similar setting in rural South Africa[6]. KABC-II assessments were undertaken at the household (can be found as Supplementary Figure 1 in *Underlying data*[18]). The KABC-II measurements were administered in a tent pitched close to the household, where the caregiver and child could see each other at all times.

The KABC-II
The KABC-II measures the processing and cognitive abilities of children aged 3–18 years old. It is directly administered to one child at a time and the answers are recorded by an individual fieldworker, who has undergone extensive training, typically by a psychologist. Each subtest starts with a standardised explanation to introduce the concepts required and then sample and training questions follow. This provides the child with a standardised training for each subtest. Starting points for several subtests are based on the child's age, with the option of dropping back to earlier starting points if an older child gets initial questions wrong. Later questions then increase in difficulty and the child stops after a discontinue rule is met, usually after getting a certain number of sequential questions wrong. Responses are recorded on a custom answer sheet to provide raw totals for each subtest, which are converted to scaled scores based on the child's age, with younger children scoring comparatively higher for the same raw total. The eight subtests take approximately two hours to administer (including breaks).

Online training for the KABC-II with feedback was provided from expert trainers via zoom in Uganda, Malawi and South Africa, combined with local training within Zimbabwe in Shona. The KABC-II was administered by fieldworkers who were trained primary care nurses. They were monitored by the study paediatrician (JP) and project lead (CM). A standardisation exercise was performed during this pilot study which showed good inter-rate reliability, where each primary care nurse measured one child using the KABC-II and the other three marked independently.

KABC-II adaptation
After the first 50 children were assessed, we noted that children were scoring relatively poorly on the planning domain of the KABC-II due to low scores in both story completion and pattern reasoning subtests. This observation had previously

been noted in rural South Africa[6]. We therefore decided to adapt these two subtests using a stepwise process, based on initial concerns regarding face validity, and then subsequent monitoring as cognitive measurement continued (Figure 1). This monitoring included noting the children's explanation of answers, and then analysing the proportion of correct answers for each individual question (individual question analysis, IQA) in both subtests for the first 50 children. This identified certain questions where unfamiliarity with the question or concept was causing poor performance. From these initial concerns of face validity noted by the data collection team, alternative and additional items were designed by those researchers based in Zimbabwe (JP, CM, GM, IM, TM, MM, DC and NVT). A diverse board of international experts were then consulted (MG, TF, AK and NK) including the original developers of the KABC-II (AK, NK). Therefore, for story completion, alternative items were developed, pre-tested and then ranked in order of difficulty before substituting problematic questions (Figure 1). For pattern reasoning, additional items were developed, pre-tested and then ranked in order of difficulty before being included (Figure 1). After these subtests were modified, custom scoring sheets for the raw scores were also developed, but the scaling of scores based on age was unchanged. The performance of the remaining 20 children on the modified subtests was evaluated using IQA to determine the effect of the modifications on the subtests. Permission for adaptation and translation was obtained from Pearson. We hypothesized that (a) scores would increase significantly on the two modified planning subtests, thereby providing evidence that the changes made to enhance cultural fairness were successful; and (b) correlations between domains of the KABC-II would increase after the modifications, suggesting an improvement in construct validity of the adapted test.

Data analysis

Mean scores of the subtests for the 50 children before modification were compared to 20 children who performed the adapted subtests by using independent sample T-tests with unequal variance. One-way repeated measures analysis of variance (ANOVA) was also performed for the first 50 children to determine if the difference amongst domains was significant, with a post-hoc Tukey's test to describe the pairwise differences between individual domains. This was similarly performed for the last 20 children. Finally, the intraclass correlation coefficient for absolute agreement using a two-way mixed effects model was calculated between domains before and after modification. Data were analysed using Microsoft Excel (RRID:SCR_016137), IBM SPSS Statistics v27 (RRID:SCR_016479) and Stata v15.0 (RRID:SCR_012763). An alternative freely available software that could perform this analysis would be R-project.

Reflexive statement

JP is a white, male, married paediatric doctor with a foundation course in art, a medical degree from Oxford University, and diploma in tropical medicine from Liverpool University. He previously worked on body composition within the SHINE trial in a neighbouring district (Chirumanzu) in

2016. He has worked closely with many collaborators on this project (CM, IM, TM, NVT, BM, MG, AP) since 2016, which helped contribute to the co-design of the adaptations.

Results

A total of 157 eligible children aged 7 years old were identified by CHWs, from whom 80 were randomly selected remotely by the study statistician. Among these 80 children, two families declined to be measured and three children had the wrong age on documentation, therefore five random replacements were made. Overall, 80 children (39 girls; 49%) were enrolled and underwent assessments between September 3rd and December 4th, 2020. This was part of a broader pilot study examining growth, physical and cognitive function, which is detailed elsewhere[16], including baseline characteristics. Data from 10 of the 80 child assessments were used for standardisation between fieldworkers and were therefore excluded from the KABC-II adaptation process. These 10 children were all chosen during one week of standardisation measurements. Results from 70 children (35 girls; 50%) were used in the analysis of the modifications. All children successfully performed the tasks with standard instructions as recommended by the KABC-II protocol.

Results[18] from the first 50 children showed that scores were significantly lower in the planning domain (mean score 8.1 (SD 2.9)), compared to the other domains (sequential 12.6 (SD 2.8), learning 13.5 (SD 3.1), simultaneous 11.6 (SD 3.5); P<0.001) (Figure 2). A post-hoc Tukey's test revealed that marks were significantly lower for planning compared to learning with a mean difference of 5.4 marks, (95% CI 3.8, 7.0; P<0.001), simultaneous (3.5 marks; 95% CI 1.9, 5.1; P<0.001) and sequential domains (4.6 marks; 95% CI 3.0, 6.2; P<0.001). Comparing between the other domains, the simultaneous score was less compared to learning (1.9; 95% CI 0.3, 3.5; P=0.01) but there were no other significant differences for sequential compared to learning (0.9; 95% CI -0.7, 2.5; P=0.5) or sequential compared to simultaneous (1.0; 95% CI -0.6, 2.6; P=0.34).
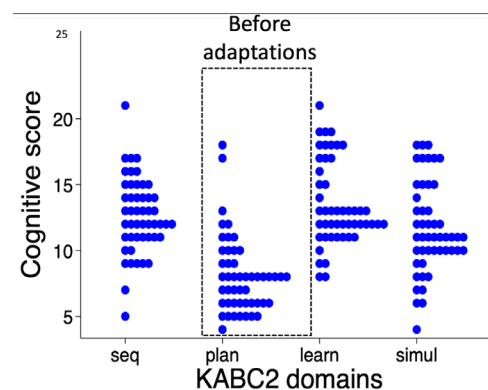

Figure 2. Scores for the first 50 children showed reduced scoring in the Planning domain, comprised of Story Completion and Pattern Reasoning subtests. KABC-II, Kaufman Assessment Battery for Children-2nd edition.

The planning domain comprises two subtests: first, a story completion task, in which children have to pick the missing picture(s) from a selection of pictures, and then align them in the correct sequence to complete a picture-based story; and second, a pattern reasoning task, in which children have to select the correct printed shape or image to fit within a repeating pattern (can be found as Supplementary Figure 2–Figure 3 in *Underlying data*[18])

Story completion

The story completion task contains 18 questions, although this rural Zimbabwean population of 7-year-olds did not progress past item 12. There is one sample question and three training questions, with a discontinue rule once children answer three sequential questions incorrectly. Individual question analysis identified challenges with problematic items, which were observed to be 4, 6, 8 and 9 (Figure 3a) (see extended data and results for further explanation). When asked to explain their choices, it was apparent that children did not sufficiently understand the picture stories on these items to be able to complete them logically or consistently. For example, for item 4, the picture story was of a birthday cake with candles being lit, blown out and then shared. In a rural Zimbabwean context, children did not recognise the small candles on the cake and thought they were flowers, or realised that there was fire on the cake, but did not realise it was arising from birthday candles. They did not understand the sequence to blow the candles out and also did not pick consistent alternatives.

Alternative picture stories to replace the problematic items 4, 6, 8 and 9 of the story completion subtest were designed based on more familiar and locally relevant stories (*e.g.*, washing, cooking or riding a bicycle) that could be easily described in pictures. A sequence of photographs was taken locally for each of the picture-based stories to provide an appropriate Zimbabwean context. These pictures were designed to match a story sequence that was similar to the KABC-II item they were replacing. They were then printed and pre-tested in a small group of eight children (can be found as Supplementary Figure 2 in *Underlying data*[18]). However, with feedback from pretesting, the alternative stories varied in difficulty in a different way from the original sequence. For example, the new picture story of falling off a bicycle appeared easier to complete (can be found as Supplementary Figure 2 in *Underlying data*[18]), although it was designed to replace item 9 (a picture story of a tightrope walker). We therefore re-arranged the sequence of questions so that the picture stories were administered in order of increasing difficulty, defined as the proportion of children who got them correct (Figure 3b). The proportion correct on each story completion item was calculated, both for the original story completion questions and the alternative story completion questions. Among the 20 children assessed using the modified story completion, their mean score was significantly higher (mean 4.8; 95% CI 4.2, 5.4) than among the initial 50 children assessed using only the original story completion task (mean 4.1; 95% CI 3.8, 4.4); mean difference 0.7 (95% CI 0.0, 1.4; P=0.05; Figure 3c).

Pattern reasoning

Individual question analysis on the pattern reasoning subtest showed that few children were getting correct answers after item 4 (Figure 4a). Although questions after item 4 had reasonable face validity, it seemed they were not being answered with
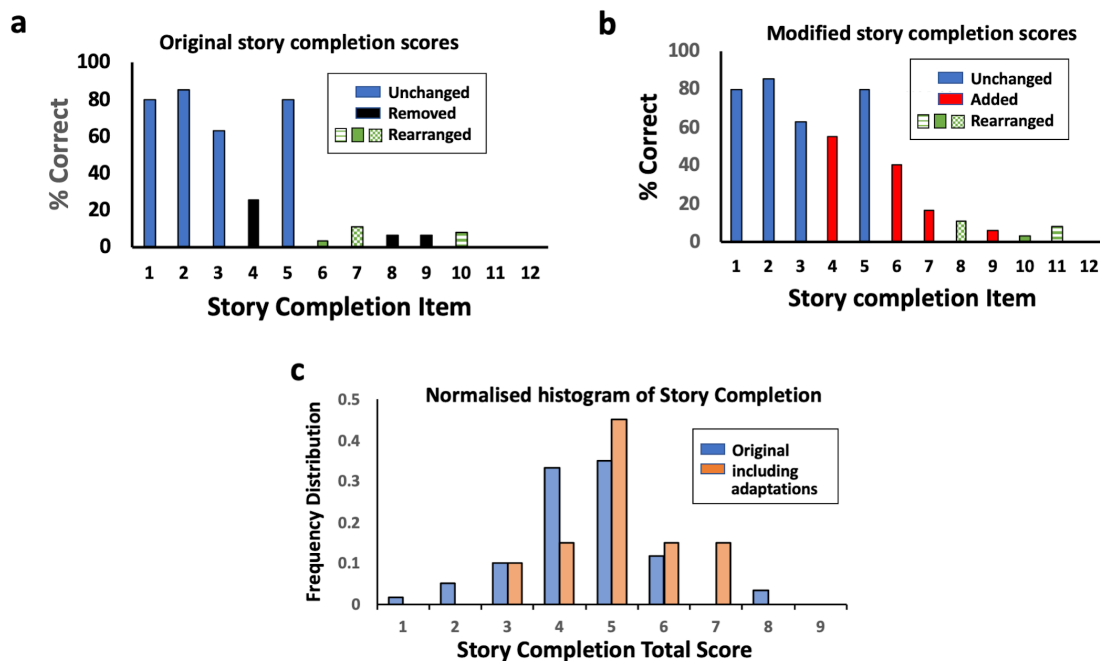

Figure 3. Story completion adaptation. (**a**) Individual question analysis showed items 4, 6, 8 and 9 were problematic items and scored poorly (see extended data and results). (**b**) Adaptation of story completion showing substitutions and rearrangement of items in order of increasing difficulty. (**c**) Normalised histogram of total scores for original and modified story completion.
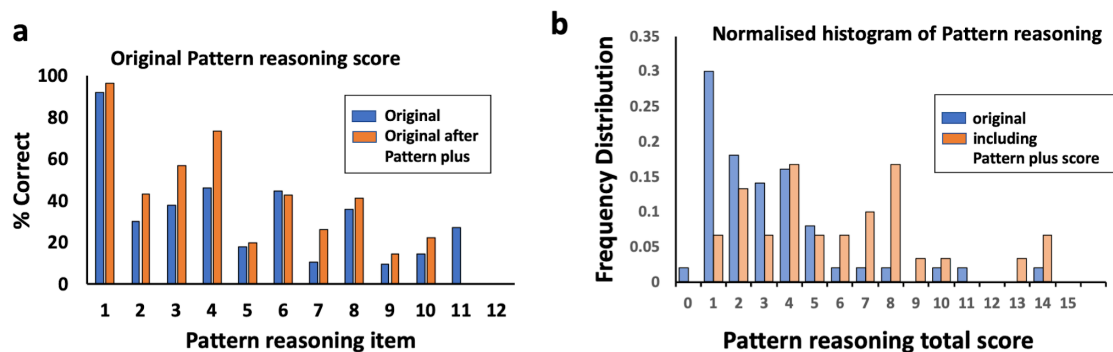
Figure 4. Pattern reasoning substitution. (**a**) Scores for original KABC-II pattern reasoning items, before and after Pattern Plus training. (**b**) Normalised histogram of total scores for original and modified pattern reasoning, including the scoring from the Pattern Plus training questions. KABC-II, Kaufman Assessment Battery for Children-2nd edition.

sufficient understanding to accurately represent the child's problem solving ability. Discussion with community members and primary school teachers highlighted that these children had not previously seen puzzles with alternating patterns. The KABC-II only has a single training question that demonstrates alternating patterns (item 2). It became apparent this was insufficient for children to grasp the concept consistently before answering further questions, and that they needed more examples. Therefore instead of any substitutions, additional pattern reasoning items were developed to supplement the original items. It was decided that all these additional items would be training questions, so that the child would receive feedback to explain what the correct answer was if they got the item wrong. A total of 12 additional questions using alternating patterns were developed and then pre-tested. From these 12, the 7 most appropriate questions were chosen for inclusion as items in the final test battery, based on face validity and feedback from the pre-testing (all supplemental items included can be found as Supplementary Figure 3 *in Underlying data*[18]). The selected pattern questions included shapes as well as contextually appropriate pictures of goats and chickens (can be found as Supplementary Figure 3 *in Underlying data*[18]). Each item had a training point about learning individual colours, orientation or number of objects within the pattern. All seven additional pattern questions were included as training questions for every child: Therefore they were included in the scoring, with explanations given if the child got the question wrong. We termed this modified pattern reasoning "Pattern Plus" because it provided additional exposure to alternating patterns prior to the remainder of the pattern reasoning subtest. As these were additional questions, all of the original KABC-II pattern reasoning questions were included with no substitutions made. We developed routine explanations for the Pattern Plus questions to standardise the training given to the child.

For the 20 children who had these additional training questions for pattern reasoning, we first examined whether their scores had improved for the original KABC-II questions only (Figure 4a). Their scores (mean 4.6 marks; 95% CI 3.6, 5.5) were not significantly higher than among the 50 children

assessed using the original pattern completion task without Pattern Plus training (4.0 marks; 95% CI 3.3, 4.6); mean difference 0.6 (95% CI -0.5, 1.7; P=0.3, Figure 4a). However, when the scoring from the Pattern Plus training questions was also included, the mean score significantly increased to 5.8 (95% CI 4.6, 7.0); mean difference compared to original test 1.8 (95% CI 0.5, 3.2; P=0.01) (Figure 4b).

Total scores in the planning domain among the 20 children who performed the modified story completion and pattern reasoning subtests showed a significant improvement compared to the 50 children assessed using the original KABC-II tests (mean score 10.6 (95% CI 9.0, 12.2) *versus* 8.1 (95% CI 7.2, 8.9; P=0.01), respectively; mean difference 2.5 (95% CI 0.8, 4.3; P=0.01) (Figure 5).

When comparing domains for the last 20 children, scores still remained lower in the modified planning domain (mean score 10.6 (SD 3.4)), compared to the other domains (learning 14.3 (SD 4.0), simultaneous 12.7 (SD 3.8), sequential 13.2 (SD 3.7); P=0.001). However, a post-hoc Tukey's test revealed that marks were significantly lower only for planning compared to learning (mean difference 3.7 marks; 95% CI 0.7, 6.8), but not for simultaneous (2.1 marks; 95% CI -0.5, 5.2) or sequential scales (2.6 marks; 95% CI -0.5, 5.7). Comparisons between the other domains showed no other significant differences for these 20 children. The intraclass correlation coefficient between domains for the first 50 children using the unmodified planning domain was 0.43 (95% CI 0.13, 0.64). For the last 20 children using the modified planning domain it was 0.69 (95% CI 0.37, 0.87).

Discussion

There is a recognised need for context-specific tools to measure neurodevelopment in LMIC settings[19,20]. Many existing proprietary tools were developed in high-income settings and may not be immediately transferable to an LMIC context. We used the KABC-II to evaluate school-age cognition in rural Zimbabwe and successfully adapted two subtests using complementary methods: substitution, addition and rearrangement of items for the story completion task, and addition of further training
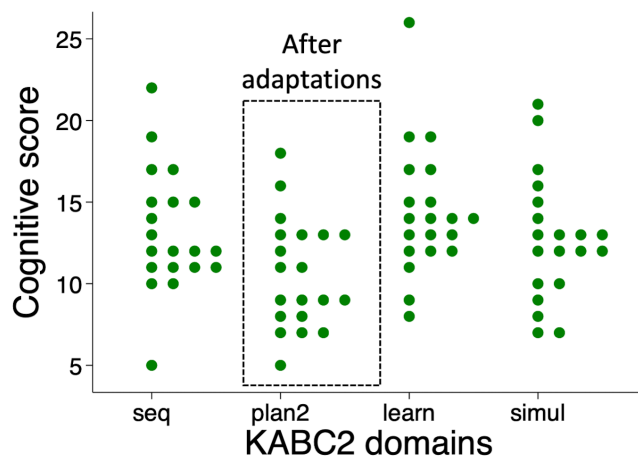
Figure 5. Scores for the last 20 children showed improvement in the Planning domain after modifications to the Story Completion and Pattern Reasoning subtests. KABC-II, Kaufman Assessment Battery for Children-2nd edition.

questions for the pattern reasoning task. Both adaptations increased the scores that children achieved on these subtests, improving their overall performance in the planning domain, which was previously noted to be reduced compared to other domains in rural South Africa[9].

As previously mentioned, results from the first 50 children showed that scores were significantly lower in the planning domain (mean score 8.1 (SD 2.9)), compared to others (sequential 12.6 (SD 2.8), learning 13.5 (SD 3.1), simultaneous 11.6 (SD 3.5); P<0.001). After modification, the planning domain mean (SD) of 10.6 (3.4) increased but was still lower compared to the others (learning 14.3 (4.0), simultaneous 12.7 (3.8) and sequential 13.2 (3.7); P=0.001). However, individual marks for subtests or domains were not always reported by other studies[12–14]. For the study in rural South Africa in 376 children aged 9–12 years[9] comparable mean (SD) domain scores of learning 14.5 (5.6), Sequential 17 (2.7), simultaneous 12.5 (2.8) and planning 10.1 (2.6); P<0.001) were measured. This study made a small change in story completion (item 9) so that the child could score the point if they washed the frying pan first as well as for the last card[9] but otherwise did not comment on any concerns with validity. The IMPAACT study[10] measured similar subtests and domains within the KABC-II and the planning domain appeared lowest in models adjusted for age and clinical site, although it was not possible to calculate the significance of the difference from the data available[10]. Again the authors did not raise any concerns of validity[10]. Therefore, there appears to be some consistency in the planning domain scoring lower, although these other studies did not report exploring the validity of story completion or pattern reasoning.

The planning domain is linked to executive function, so these adaptations may improve the predictive power of the

KABC-II at age 7 years old and beyond. This is important because executive functions are higher level cognitive processes needed for self-control and decision-making. They are central to healthy behaviours and development in children and adolescents[21]. The development of executive function is highly sensitive to positive exposures such as high-quality education, and to negative exposures such as high adversity, poverty and lower-quality education[22]. Deficits in executive functions have been associated with a wide range of negative outcomes including behavioural and mental health problems. hars from this study.

To our knowledge, individual subtests within the KABC-II have not been previously examined and adapted in Africa. Many studies have used all subtests mentioned here across Africa[7] including Uganda[8] and rural South Africa[6,9], sometimes with video monitoring[10,11]. An alternative approach has been to select individual subtests which do not provide an overall score such as the mental processing index,[12–14] For example in South Africa, difference in nutrition were observed to associate with the Rover and Atlantis subtests only[12]. In Ethiopia, certain subtests were selected based on the local context and combined differently, with the story completion subtest being excluded. Both mothers and children were assessed using number recall, word order and hand movement for short-term memory subtests. Visual processing assessment was assessed by combining triangles, pattern reasoning and conceptual thinking[13]. Scores between mothers and children were correlated and mother's education was significantly correlated with pattern reasoning in particular. Scores on number recall, hand movements, triangles were reported higher in the non-stunted and normal-weight groups compared with the stunted or underweight groups. Conceptual thinking scores were reported lower in underweight children, and word order scores were lower in stunted children[13]. In Burkina Faso, stunted children were shown to perform significantly worse in memory (measured by atlantis and number recall subtests) and spatial abilities (triangles, conceptual thinking and face recognition subtests[14] but no overall measure of cognition divided into specific domains (such as is possible with the MPI) was made. It is possible that these studies may have had concerns of how culturally appropriate story completion was, but this was not documented.

The increase in mean scores for both subtests (15% increase in story completion, 30% increase in pattern reasoning) suggested they became more culturally appropriate. The intraclass correlation coefficient (ICC) between domains also increased with the adaptations, suggesting variability between domains may have reduced with the modifications. The increase in ICC suggests a potential improvement in construct validity of the planning domain, with higher intercorrelations between planning and the other domains. This is supported by the confidence intervals where the ICC before modification of 0.43 (95% CI 0.13, 0.64) excludes the ICC after modification of 0.69 (95% CI 0.37, 0.87). Note, however, that the pre-modification value of 0.43 is included in the post-modification confidence interval of 0.37 to 0.87; therefore, inferences about construct validity are tentative and require cross-validation.

The outcomes of this study show the value of careful monitoring and refinement of tools for use in new settings with different contexts from their original development. Adaptation is an important process in ensuring that neurodevelopmental tests are cross-culturally applicable. Other studies have previously shown improvements in test performance with culturally relevant adaptations, for example using wire to model patterns in Zambia instead of drawing[23]. Previously, nine different subtests from the KABC first edition (KABC) were adapted for use amongst 5–6 year olds in rural Kenya[24]. Alterations included translation, substitution of materials including more culturally relevant pictures, as well as altering the task structures to improve cross-cultural validity. It has been proposed that adaptation of individual tests should include monitoring the distribution of scores to identify floor and ceiling effects, undertaking test and re-test reliability to ensure stable measurement, ensuring high inter-rater agreement, and comparing internal consistency with similar measures[19]. Simultaneous measurement of contributing factors such as socioeconomic status may also demonstrate associations with variables that are expected to be related to the adapted tests (convergent validity). Repeated testing could also demonstrate increasing scores with age[19]. Another alternative is locally developed tools such as the Malawi Developmental Assessment Tool (MDAT)[2] or Kilifi Development Inventory[25], which can then establish culturally appropriate norms. However, using these in different contexts may also require similar processes of monitoring participants' answers and subsequent adaptation.

This study has several strengths. The process of checking face validity of items and then monitoring individual responses to questions empowered fieldworkers and the local community to monitor children's answers and suggest adaptations, leading to a more culturally inclusive tool through co-design. The IQA also provided a way to confirm or refute initial concerns regarding face validity: for example, in the story completion task, a story about cooking a fried egg was immediately flagged and eventually replaced because rural Zimbabwean households boil or scramble eggs but do not fry them (Item 8). Similarly, item 4 (blowing candles on a cake) was poorly understood as candles are rarely used in this setting to celebrate candles and replaced. Item 9 describing a high-wire artist falling in a circus had poor face validity because very few children had seen images of a circus or acrobats so this was replaced. The concept behind item 6 of building a model hut or 'kitchen', was well understood but the picture of the sticks scattered randomly was rarely picked correctly as the first picture. This was hypothesised that a building's foundations and initial low walls are often first observed as in place early in construction in rural areas. The conceptual understanding was observed in most children and hence this item was included but moved to item 10 given its level of difficulty. IQA also highlighted that a series of pictures of a person blowing up balloons was well understood by children, even though the type of balloons were not commonly seen. Further details are available in extended data and results.

Although scores improved, they still remained lower than for other test domains, although after modification this difference was significant only when compared to the learning domain. It is plausible that children will continue to score lower in the planning domain despite adaptation, due to cultural factors such as reduced exposure to these types of puzzle[9].

The study also has several limitations. Tukey's pairwise comparison test between domains does include an adjustment for multiple testing, but the results of comparing multiple subtests should be interpreted with some caution due to the increased risk of chance errors. For our population of 7-year-old children, no child progressed beyond item 12 on the story completion task, so our adaptation did not modify later items. Therefore, for studies using the KABC-II in older children in rural sub-Saharan Africa, we would recommend a similar monitoring phase for later items of story completion to highlight any problematic questions and then undertake pretesting of any alterations. The number of children trying each new story completion item varied, as these items became available at different timepoints. The use of photographs in story completion as an alternative to illustration may have changed some cognitive processing of the task, so ideally imaging software should be used to convert these items to illustrations. The order of the Pattern Plus sequence was decided based on increasingly complex alternating patterns, partly informed by the observation that patterns with pictures appeared more challenging than shapes. All Pattern Plus questions were asked as training questions to all the 20 children with explanations given. For young children, the addition of seven pattern plus questions may be too many, and so similar adaptations may use fewer and simpler patterns. For older children, it is hoped the concept of alternating patterns with seven examples was sufficient, but this would need to be verified. The process of developing, printing and pre-testing alternative or additional items took time so that more children were tested before the adaptations (n=50) compared to after (n=20). Finally, test re-test reliability was not measured because it was not possible to revisit the children due to the rural locations of the measurement, although this step is recommended to demonstrate test stability[19].

The adaptations in the KABC-II performed in children within this pilot study were made in preparation for following up the SHINE trial: The SHINE trial was a 2×2 factorial trial in rural Zimbabwe that was ethically approved by the Medical Research Council of Zimbabwe (MRCZ/A/1675) and registered in 2013 (NCT01824940). This trial randomized children to lipid-based nutrient supplements (LNS) and/or a comprehensive household water, sanitation and hygiene WASH intervention[26–29]. Mothers were enrolled in early pregnancy, with detailed data collection on home, maternal, birth and early-life factors. The adaptations from this pilot study detailed here are being applied to children born into the SHINE trial who are aged 7 years old. The SHINE follow-up study (PACTR number PACTR202201828512110) aims to measure approximately 1,300 children. Further details including the SHINE follow-up CRF's and protocol are also available at Open Science Framework. Further analysis of the SHINE follow-up dataset may include monitoring for floor and ceiling effects within these KABC-II adaptations, checking inter-rater

agreement and comparing internal consistency with similar measurements. Convergent validity demonstrating expected associations with socioeconomic status and caregiver education may also be explored as future steps.

In conclusion, two subtests of the KABC-II were successfully adapted for use in rural Zimbabwe, drawing on both local and international expertise. This helped to increase scoring on the planning domain of cognitive processing, so that performance became more comparable to other domains in the KABC-II. The process of reviewing the face validity of items, together with monitoring of children's responses both qualitatively and by individual question analysis helped to identify items for support. Substitution, rearrangement and addition of items can improve cross-cultural validity of cognition tools, working in collaboration with the original developers, local community and participants.

Data availability

Underlying data

Open Science Framework: Adaptation of the Kaufman Assessment Battery for Children—2nd edition (KABC-II) to assess school-age neurodevelopment in rural Zimbabwe. https://doi.org/10.17605/OSF.IO/T6GXB[18].

This project contains the following underlying data:
- Extended data & results KABC-II.pdf (supplementary material)
- KABC_II_modification.xlsx (raw spreadsheet data of test scores)

Data are available under the terms of the Creative Commons Zero "No rights reserved" data waiver (CC0 1.0 Public domain dedication).

References

1. Black MM, Walker SP, Fernald LCH, et al.: **Early childhood development coming of age: science through the life course.** *Lancet.* 2017; **389**(10064): 77–90.
   PubMed Abstract | **Publisher Full Text** | **Free Full Text**

2. Gladstone M, Lancaster GA, Umar E, et al.: **The Malawi Developmental Assessment Tool (MDAT): the creation, validation, and reliability of a tool to assess child development in rural African settings.** *PLoS Med.* 2010; **7**(5): e1000273.
   PubMed Abstract | **Publisher Full Text** | **Free Full Text**

3. Wechsler D: **Wechsler intelligence scale for children.** Pearson; 2014.
   Reference Source

4. Bayley N: **Bayley scales of infant and toddler development.** Pearson; 2005.
   Reference Source

5. Kaufmann A, Kaufmann N: **Kaufman Assessment Battery for Children 2nd edition.** In: *PsychCorp*. ed.: Pearson; 2004.

6. Mitchell JM, Tomlinson M, Bland RM, et al.: **Confirmatory factor analysis of the Kaufman assessment battery in a sample of primary school-aged children in rural South Africa.** *S Afr J Psychol.* 2018; **48**(4): 434–52.
   Publisher Full Text

7. Semrud-Clikeman M, Romero RAA, Prado EL, et al.: **Selecting measures for the neurodevelopmental assessment of children in Low- and Middle-Income Countries.** *Child Neuropsychol.* 2017; **23**(7): 761–802.
   PubMed Abstract | **Publisher Full Text** | **Free Full Text**

8. Bangirana P, Seggane M, Allebeck P, et al.: **A preliminary examination of the construct validity of the KABC-II in Ugandan children with a history of cerebral malaria.** *Afr Health Sci.* 2009; **9**(3): 186–92.
   PubMed Abstract | **Free Full Text**

9. Mitchell JM, Tomlinson M, Rochat DT: **Psychometric evaluation of the Kaufman Assessment Battery for Children, second edition (KABC-II) in rural South Africa.** Stellenbosch; 2015.
   Reference Source

10. Ruisenor-Escudero H, Familiar I, Nyakato M, et al.: **Building capacity in neurodevelopment assessment of children in sub-Saharan Africa: a quality assurance model to implement standardized neurodevelopment testing.** *Child Neuropsychol.* 2019; **25**(4): 466–81.
    PubMed Abstract | **Publisher Full Text** | **Free Full Text**

11. Boivin MJ, Barlow-Mosha L, Chernoff MC, et al.: **Neuropsychological performance in African children with HIV enrolled in a multisite antiretroviral clinical trial.** *AIDS.* 2018; **32**(2): 189–204.
    PubMed Abstract | **Publisher Full Text** | **Free Full Text**

12. Taljaard C, Covic NM, van Graan AE, et al.: **Effects of a multi-micronutrient-fortified beverage, with and without sugar, on growth and cognition in South African schoolchildren: a randomised, double-blind, controlled intervention.** *Br J Nutr.* 2013; **110**(12): 2271–84.
    PubMed Abstract | **Publisher Full Text**

13. Bogale A, Stoecker BJ, Kennedy T, et al.: **Nutritional status and cognitive**

performance of mother-child pairs in Sidama, Southern Ethiopia. *Matern Child Nutr.* 2013; **9**(2): 274–84.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

14. Sanou AS, Diallo AH, Holding P, et al.: **Association between stunting and neuro-psychological outcomes among children in Burkina Faso, West Africa.** *Child Adolesc Psychiatry Ment Health.* 2018; **12**: 30.
    PubMed Abstract | **Publisher Full Text** | **Free Full Text**

15. Piper JD, Mazhanga C, Mwapaura M, et al.: **The Sanitation Hygiene Infant Nutrition Efficacy (SHINE) trial: protocol for school-age follow-up [version 1; peer review: 2 approved].** *Wellcome Open Res.* 2023; **8**: 306.
    PubMed Abstract | **Publisher Full Text** | **Free Full Text**

16. Piper JD, Mazhanga C, Mapako G, et al.: **Characterising school-age health and function in rural Zimbabwe using the SAHARAN toolbox.** *medRxiv.* 2021; 2021.09.22.21263533.
    Publisher Full Text

17. Piper JD, Mazhanga C, Mwapaura M, et al.: **The Sanitation Hygiene Infant Nutrition Efficacy (SHINE) trial: protocol for school-age follow-up.** *medRxiv.* 2022; 2022.08.17.22278247.
    Publisher Full Text

18. Piper JD, Ntozini R, C M, et al.: **Adaptation of the Kaufman Assessment Battery for Children—2nd edition (KABC-II) to assess school-age neurodevelopment in rural Zimbabwe.** Open Science Framework. [Dataset]. 2022.

19. Prado EL, Hartini S, Rahmawati A, et al.: **Test selection, adaptation, and evaluation: a systematic approach to assess nutritional influences on child development in developing countries.** *Br J Educ Psychol.* 2010; **80**(Pt 1): 31–53.
    PubMed Abstract | **Publisher Full Text**

20. 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.
    PubMed Abstract | **Publisher Full Text** | **Free Full Text**

21. Obradović J, Willoughby MT: **Studying executive function skills in young children in Low- and Middle-Income Countries: progress and directions.** *Child Dev Perspect.* 2019; **13**(4): 227–34.
    Publisher Full Text

22. Haft SL, Hoeft F: **Poverty's impact on children's Executive Functions: global considerations.** *New Dir Child Adolesc Dev.* 2017; **2017**(158): 69–79.
    PubMed Abstract | **Publisher Full Text** | **Free Full Text**

23. Serpell R: **How specific are perceptual skills? a cross-cultural study of pattern reproduction.** *Br J Psychol.* 1979; **70**(3): 365–80.
    PubMed Abstract | **Publisher Full Text**

24. Holding PA, Taylor HG, Kazungu SD, et al.: **Assessing cognitive outcomes in a rural African population: development of a neuropsychological battery in Kilifi District, Kenya.** *J Int Neuropsychol Soc.* 2004; **10**(2): 246–60.
    PubMed Abstract | **Publisher Full Text**

25. Abubakar A, Holding P, van Baar A, et al.: **Monitoring psychomotor**

Development in a resource-limited setting: an evaluation of the Kilifi developmental Inventory. *Ann Trop Paediatr.* 2008; **28**(3): 217–26.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

26.   Humphrey JH, Mbuya MNN, Ntozini R, *et al.*: **Independent and combined effects of improved water, sanitation, and hygiene, and improved complementary feeding, on child stunting and anaemia in rural Zimbabwe: a cluster-randomised trial.** *Lancet Glob Health.* 2019; **7**(1): e132–e47.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

27.   Prendergast AJ, Chasekwa B, Evans C, *et al.*: **Independent and combined effects of improved water, sanitation, and hygiene, and improved complementary feeding, on stunting and anaemia among HIV-exposed children in rural Zimbabwe: a cluster-randomised controlled trial.** *Lancet*

Child Adolesc Health. 2019; **3**(2): 77–90.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

28.   Ntozini R, Chandna J, Evans C, *et al.*: **Early child development in children who are HIV-exposed uninfected compared to children who are HIV-unexposed: observational sub-study of a cluster-randomized trial in rural Zimbabwe.** *J Int AIDS Soc.* 2020; **23**(5): e25456.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

29.   Chandna J, Ntozini R, Evans C, *et al.*: **Effects of improved complementary feeding and improved water, sanitation and hygiene on early child development among HIV-exposed children: substudy of a cluster randomised trial in rural Zimbabwe.** *BMJ Glob Health.* 2020; **5**(1): e001718.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Open Peer Review

Current Peer Review Status: ✓ ✓

Version 2

Reviewer Report 20 June 2024

https://doi.org/10.21956/wellcomeopenres.24725.r85265

✓   **Katherine Brown** (iD)

Great Ormond Street Hospital for Children National Health Service Foundation Trust, London, UK

Thank you. I am now happy to approve the revised article without reservations. BW Kate Brown.

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: I undertake health services research related to children with critical illness and with heart disease using a range of methods (data science, qualitative and health care evaluation). My interests include health care access for children with critical illness, evaluation of treatments for critically ill children, longer-term impacts of critical illness and heart diseases (neuro developmental outcomes and quality of life) and measuring experience / outcomes from the perspective of child or parent

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Reviewer Report 10 June 2024

https://doi.org/10.21956/wellcomeopenres.24725.r85266

✓   **Supriya Bhavnani** (iD)

Child Development Group, Sangath, Bardez, Goa, India

My suggestions have been appropriately addressed.

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Early Childhood Development, Cognitive assessments

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Version 1

Reviewer Report 10 October 2023

https://doi.org/10.21956/wellcomeopenres.19837.r67312

? **Katherine Brown** (iD)

Great Ormond Street Hospital for Children National Health Service Foundation Trust, London, UK

Katherine Brown, from the Centre for Outcomes in Children's Critical Illness and Cardiovascular Disease, University College London.

'Piloting the adaptation of the Kaufman Assessment Battery for Children—2nd edition (KABC-II) to assess school-age neurodevelopment in rural Zimbabwe' is an important, carefully planned and executed piece of work. The authors are addressing an important issue, which is that access to suitable valid and culturally appropriate measures to assess child development is not always as good as it should be. Although many measures for the assessment of child development exist, they are dominated by tools developed in high income settings, often also measures that require a very high level of training and sometimes special equipment in order for them to be used. Studies of this type are challenging to undertake and the authors have sought to carefully address an important gap within their study which has led to adaptations of the KABC-2.

Comments and suggestions:

Please make it clearer in the introduction whether the referenced studies 12-14 identified similar problems with the same subscales (story completion and pattern reasoning)?: these statements in introduction are not specific enough.

The references to the Shine study in the introduction and methods are a bit confusing. For example, in the methods the authors state that participants in Shine were excluded, but then go on to refer to the Shine follow study when discussing ethical approvals. Please clarify the relationship between this study and Shine?

In methods, please provide a bit more detail on who undertook the assessments, their training and any additional validation or additional person review of tester activities that were undertaken.

For clarity, it might be helpful to use a specific section of text on the measure and a separate section on the administration of the measure.

In the results section, please provide a bit more specific detail about the items that were problematic in the story completion section and pattern reasoning, which item numbers and a complete list is needed.

The pattern reasoning section is harder to follow than the story completion section even with reference to the supplementary materials. The rationale / reason for the substitutions is to use contextually appropriate goats and chickens, can the authors be clearer were all replaced or some, and how was this selection made? In supplementary materials pattern reasoning account, I was not able to determine exactly which patterns were original versus added. I did not find accompanying supplementary text detailing the exact reasons / issues / rationale related to the pattern reasoning section (whereas this is presented more clearly with better signposting for the story completion section again in supp materials).

In discussion please could the authors expand on what if any further validation work is needed for the adapted measure. It is not uncommon for test retest to be infeasible for a this type of study in a high income setting. At the same time concurrent validity / construct validity of the adapted measure could be assessed in a future study?

Is the work clearly and accurately presented and does it cite the current literature?
Yes

Is the study design appropriate and is the work technically sound?
Yes

Are sufficient details of methods and analysis provided to allow replication by others?
Partly

If applicable, is the statistical analysis and its interpretation appropriate?
Yes

Are all the source data underlying the results available to ensure full reproducibility?
Yes

Are the conclusions drawn adequately supported by the results?
Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: I undertake health services research related to children with critical illness and with heart disease using a range of methods (data science, qualitative and health care evaluation). My interests include health care access for children with critical illness, evaluation of treatments for critically ill children, longer-term impacts of critical illness and heart diseases (neuro developmental outcomes and quality of life) and measuring experience / outcomes from the

perspective of child or parent

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Author Response 12 May 2024

Joseph Piper

Thank you for the constructive feedback. Please find responses in bold below.

Reviewer Comment:

The pattern reasoning section is harder to follow than the story completion section even with reference to the supplementary materials. The rationale / reason for the substitutions is to use contextually appropriate goats and chickens, can the authors be clearer were all replaced or some, and how was this selection made? In supplementary materials pattern reasoning account, I was not able to determine exactly which patterns were original versus added. I did not find accompanying supplementary text detailing the exact reasons / issues / rationale related to the pattern reasoning section (whereas this is presented more clearly with better signposting for the story completion section again in supp materials).

Author Response:

We apologise for the confusion. We have tried to clarify that for pattern reasoning, only further additions were made and no substitutions occurred. All of the patterns in supplementary materials were added. For copyright purposes, we have not included any of the original patterns. We have added further explanation for pattern reasoning.

Reviewer Comment:

In discussion please could the authors expand on what if any further validation work is needed for the adapted measure. It is not uncommon for test retest to be infeasible for a this type of study in a high income setting. At the same time concurrent validity / construct validity of the adapted measure could be assessed in a future study?

Author Response:

Thank you for this point. This had been stated in limitations, but is now made clearer with plans for further verification within the SHINE follow-up dataset.

Competing Interests: No competing interests were disclosed.

Reviewer Report 20 December 2022

❓　　**Supriya Bhavnani** iD

Child Development Group, Sangath, Bardez, Goa, India

Having reviewed "Piloting the adaptation of the Kaufman Assessment Battery for Children—2nd edition (KABC-II) to assess school-age neurodevelopment in rural Zimbabwe", I have found it to be an extremely interesting adaptation of the KABC-II. It is a detailed study describing the adaptation of the test, the data-driven rationale for choosing which items to adapt and piloting the adapted version on children has shown how their adaptations have improved the scale. Overall I have found this to be a well-written article. My comments are listed below –

Abstract:
1. In methods sentence 1, consider moving "*both within the planning domain*" to the beginning of the sentence instead of the end, i.e. after "*We piloted and adapted two subtests (pattern reasoning and story completion)*". It will make for a less interrupted read.

2. Can the number of items substituted and added into each subsets of the test be mentioned somewhere in the abstract since this is one of the main results of the paper i.e. what all had to be done to the battery to make it appropriate for their settings. This could be in abstract methods or even in results.

3. Include age range or mean (sd) of the children in the methods

4. I would move the numbers of children for both phases of the pilot into the methods instead of the results in abstract

Introduction:
1. Expansion of LMIC is missing the word "countries"

2. Within the literature citing KABC-II use in Africa, it is unclear what subtests have worked and what haven't in these studies in Burkina Faso, Ethiopia and South Africa. Are they also the planning domain? While I understand that adaptation and piloting for these settings have not been systematically described, could these studies be brought back into the discussion to compare how these findings compare to how KABC-II use has been described in other African settings?

Methods:
1. As in abstract, I suggest you mention either age range of participants of the mean and sd

2. The SHINE study appears abruptly in the methods in the sentence "Any children who were born into the SHINE cluster randomised trial were excluded." I suggest adding in the previous sentence that the site was the same one as that in which the SHINE study was conducted.

3. In the sentence "*After community sensitisation, a specific sensitisation visit to the family **was***

conducted by the CHWs using a community sensitisation sheet in Shona or Ndebele (the local languages)." The word "was" is missing

4. The actual process of creating the modified items for either story completion or pattern reasoning is not detailed. The word "co-design" is mentioned in some places in the manuscript but it is not clear in methods how the individual items were identified, who vetted their face validity, was Pearson involved in this process at all? Was there a committee of experts established who reviewed these items? Details of this process will enable its replication for this or indeed other developmental assessment tools in other studies

Results:
1. The score on the KABC-II domains before adaptation are reported in the first section. Could these be brought back into the discussion to compare them against other African studies which might have been conducted on children of similar ages?

2. Story completion section – specific items identified to be challenging mentioned here aren't clearly different from other higher difficulty items in the graph e.g. 7, 10, 11, 12. I think the explanation is there in supplementary methods, can that please be referenced at this point so the reader knows that there is more rationale given by the authors albeit buried in supplementary data

3. Story completion section – Could the discussion comment on "why" some items of the story completion task might not have been recognizable by the children? This would help highlight the socio-cultural reasons for which the adaptation needed to be done

4. Story completion section – In the sentence "*A sequence of photographs was taken for each of the picture-based stories.*" Please clarify that these were locally taken i.e. were African in context

5. Story completion section – In the text, please clarify where figures 3b and 3c have been referenced and not only call them figure 3.

6. Pattern reasoning section – in the sentence "*Their scores (mean 4.6 marks; 95% CI 3.6, 5.5) were not significantly higher than among the 50 children assessed using the original pattern completion task without training (4.0 marks; 95% CI 3.3, 4.6); mean difference 0.6 (95% CI -0.5, 1.7; P=0.3, Figure 4a).*" mention that this training was Pattern Plus Training as is mentioned in the legend of Figure 4

Discussion:
1. Please note a few suggestions for improvement within the relevant sections above

2. The sentence "*There were no recorded harms from this study.*" Appears out of the blue. Could it be moved to methods where the ethics are being described?

3. In the section with the sentence "*For young children, the addition of seven pattern plus questions may be too many, and so similar adaptations may use fewer and simpler patterns.*" Could a similar comment be made on how these adaptations may or may not be sufficient for older children as well given the age range of KABC-II is 3-18 years?

4. In the sentence "*Finally, test re-test reliability was not measured because it was not possible to*

revisit the children due to the rural locations of the measurement, although this step is recommended to demonstrate test stability18." The authors rightly comment on how not all forms of validation of the tool have been possible. I feel this needs to be expanded to include that other validations have also not been done yet, of the types that the authors themselves have mentioned in the paragraph before strengths i.e. "*It has been proposed that adaptation of individual tests should include monitoring the distribution of scores to identify floor and ceiling effects, undertaking test and re-test reliability to ensure stable measurement, ensuring high inter-rater agreement, and comparing internal consistency with similar measures18. Simultaneous measurement of contributing factors such as socioeconomic status may also demonstrate associations with variables that are expected to be related to the adapted tests (convergent validity). Repeated testing could also demonstrate increasing scores with age18. Another alternative is locally developed tools such as the Malawi Developmental Assessment Tool (MDAT)2 or Kilifi Development Inventory24, which can then establish culturally appropriate norms. However, using these in different contexts may also require similar processes of monitoring participants' answers and subsequent adaptation.*" – I imagine some of these will emerge from the SHINE follow up study and this can be mentioned as next steps of this work

Figure 1: Please include sample sizes in the flowchart at "*cognitive measurement started*", "*pre-testing*" and I imagine another box would need to be added for the N=20. I feel this would make the methods flowchart more comprehensive

Figure 2 & 5: I wonder if it is possible to show the planning domain data as a histogram for the two conditions of before and after adaptations. It would be useful to be able to compare the results directly without looking at 2 separate figures. Plotting percentage of participants on the y axis would allow for comparison despite the different sample size.

Is the work clearly and accurately presented and does it cite the current literature?
Yes

Is the study design appropriate and is the work technically sound?
Yes

Are sufficient details of methods and analysis provided to allow replication by others?
Partly

If applicable, is the statistical analysis and its interpretation appropriate?
Yes

Are all the source data underlying the results available to ensure full reproducibility?
Yes

Are the conclusions drawn adequately supported by the results?
Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Early Childhood Development, Cognitive assessments

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Author Response 12 May 2024

Joseph Piper

Thank you for the constructive feedback. Please find individual responses as below:

Reviewer Comment:

Having reviewed "Piloting the adaptation of the Kaufman Assessment Battery for Children—2nd edition (KABC-II) to assess school-age neurodevelopment in rural Zimbabwe", I have found it to be an extremely interesting adaptation of the KABC-II. It is a detailed study describing the adaptation of the test, the data-driven rationale for choosing which items to adapt and piloting the adapted version on children has shown how their adaptations have improved the scale. Overall I have found this to be a well-written article. My comments are listed below –

Abstract:
1. In methods sentence 1, consider moving "*both within the planning domain*" to the beginning of the sentence instead of the end, i.e. after "*We piloted and adapted two subtests (pattern reasoning and story completion)*". It will make for a less interrupted read.

   Author Response: **Thank you: this has been corrected.**
1. Can the number of items substituted and added into each subsets of the test be mentioned somewhere in the abstract since this is one of the main results of the paper i.e. what all had to be done to the battery to make it appropriate for their settings. This could be in abstract methods or even in results.

   Author Response: **Thank you: This is included in abstract results**
1. Include age range or mean (sd) of the children in the methods

   Author Response: **This has been included in results so that it is clear to readers there is no difference in mean or SD in age for those before or after the modifications.**

2. I would move the numbers of children for both phases of the pilot into the methods instead of the results in abstract

   Author Response: **We have kept the numbers of children in the results so the difference between initial and later numbers of children in pilot is clear for the reader.**

Reviewer Comment:

Introduction:
    1. Expansion of LMIC is missing the word "countries"

Author Response: **Thank you this has been included**

Reviewer Comment:
    1. Within the literature citing KABC-II use in Africa, it is unclear what subtests have worked and what haven't in these studies in Burkina Faso, Ethiopia and South Africa. Are they also the planning domain? While I understand that adaptation and piloting for these settings have not been systematically described, could these studies be brought back into the discussion to compare how these findings compare to how KABC-II use has been described in other African settings?

Author Response: **To our knowledge, examination of the suitability of KABC-II subtests has not been previously examined. We have included the studies that used all the tests to get an MPI along with those who were quoted as selecting individual subtests now in the discussion.**

Reviewer Comment:

Methods:
    1. As in abstract, I suggest you mention either age range of participants of the mean and sd

Author Response: **Thank you for this suggestion which has been included.**

    1. The SHINE study appears abruptly in the methods in the sentence "Any children who were born into the SHINE cluster randomised trial were excluded." I suggest adding in the previous sentence that the site was the same one as that in which the SHINE study was conducted.

Author Response: **Thank you for this suggestion which has been included.**

    1. In the sentence "*After community sensitisation, a specific sensitisation visit to the family **was** conducted by the CHWs using a community sensitisation sheet in Shona or Ndebele (the local languages).*" The word "was" is missing

Author Response: **Thank you for this correction which has been done.**

    1. The actual process of creating the modified items for either story completion or pattern reasoning is not detailed. The word "co-design" is mentioned in some places in the manuscript but it is not clear in methods how the individual items were identified, who vetted their face validity, was Pearson involved in this process at all? Was there a committee of experts established who reviewed these items? Details of this process will enable its replication for this or indeed other developmental assessment tools in other studies

Author Response: **We have clarified the process. Pearson were not involved apart from giving appropriate approvals, but professors Alan and Nadeen Kaufman, Tamsen Rochat and Melissa Gladstone provided appropriate input.**

Reviewer Comment:

Results:

1. The score on the KABC-II domains before adaptation are reported in the first section. Could these be brought back into the discussion to compare them against other African studies which might have been conducted on children of similar ages?

Author Response: **Thank you for this point. Many African studies did not use the domains as specified in this paper. However, the study closest in demographics (Mitchell et al.) did and these values are now quoted with an ANOVA calculated that shows Planning is also significantly reduced.**

1. Story completion section – specific items identified to be challenging mentioned here aren't clearly different from other higher difficulty items in the graph e.g. 7, 10, 11, 12. I think the explanation is there in supplementary methods, can that please be referenced at this point so the reader knows that there is more rationale given by the authors albeit buried in supplementary data

Author Response: **Thank you for this: this has now be referenced to extended data and results in both the text and figure to assist the reader.**

1. Story completion section – Could the discussion comment on "why" some items of the story completion task might not have been recognizable by the children? This would help highlight the socio-cultural reasons for which the adaptation needed to be done

Author Response: **This has also been extended in the discussion section**

1. Story completion section – In the sentence "*A sequence of photographs was taken for each of the picture-based stories.*" Please clarify that these were locally taken i.e. were African in context

Author Response: **Thank you this has been clarified**

1. Story completion section – In the text, please clarify where figures 3b and 3c have been referenced and not only call them figure 3.

Author Response: **Thank you this has been clarified**

1. Pattern reasoning section – in the sentence "*Their scores (mean 4.6 marks; 95% CI 3.6, 5.5) were not significantly higher than among the 50 children assessed using the original pattern completion task without training (4.0 marks; 95% CI 3.3, 4.6); mean difference 0.6 (95% CI -0.5, 1.7; P=0.3, Figure 4a).*" mention that this training was Pattern Plus Training as is mentioned in the legend of Figure 4

Author Response: **Thank you this has been included Discussion:**
1. Please note a few suggestions for improvement within the relevant sections above

2. The sentence "*There were no recorded harms from this study.*" Appears out of the blue. Could it be moved to methods where the ethics are being described?

Author Response: **Thank you this has been added into methods**

1. In the section with the sentence "*For young children, the addition of seven pattern plus questions may be too many, and so similar adaptations may use fewer and simpler*

patterns." Could a similar comment be made on how these adaptations may or may not be sufficient for older children as well given the age range of KABC-II is 3-18 years?

Author Response: **Thank you, a comment on older children has been entered.**

1. In the sentence "*Finally, test re-test reliability was not measured because it was not possible to revisit the children due to the rural locations of the measurement, although this step is recommended to demonstrate test stability18.*" The authors rightly comment on how not all forms of validation of the tool have been possible. I feel this needs to be expanded to include that other validations have also not been done yet, of the types that the authors themselves have mentioned in the paragraph before strengths i.e. "*It has been proposed that adaptation of individual tests should include monitoring the distribution of scores to identify floor and ceiling effects, undertaking test and re-test reliability to ensure stable measurement, ensuring high inter-rater agreement, and comparing internal consistency with similar measures18. Simultaneous measurement of contributing factors such as socioeconomic status may also demonstrate associations with variables that are expected to be related to the adapted tests (convergent validity). Repeated testing could also demonstrate increasing scores with age18. Another alternative is locally developed tools such as the Malawi Developmental Assessment Tool (MDAT)2 or Kilifi Development Inventory24, which can then establish culturally appropriate norms. However, using these in different contexts may also require similar processes of monitoring participants' answers and subsequent adaptation.*" – I imagine some of these will emerge from the SHINE follow up study and this can be mentioned as next steps of this work

Author Response: **Thank you this has been included.**

Reviewer Comment:

Figure 1: Please include sample sizes in the flowchart at "*cognitive measurement started*", "*pre-testing*" and I imagine another box would need to be added for the N=20. I feel this would make the methods flowchart more comprehensive.

Author Response: **Thank you this has been included.**

Figure 2 & 5: I wonder if it is possible to show the planning domain data as a histogram for the two conditions of before and after adaptations. It would be useful to be able to compare the results directly without looking at 2 separate figures. Plotting percentage of participants on the y axis would allow for comparison despite the different sample size.

Author Response: **Thank you for this suggestion, however the histograms presented are based on individual scores and we feel this is clearer for the reader given the small numbers used in this study.**

Competing Interests: No competing interests were disclosed.

Comments on this article

Version 1

Author Response 12 May 2024

Joseph Piper

Please note the following changes:

1) **Update text:** Abstract methods updated to describe 4 substitutions for pattern reasoning and 7 additions for story completion. Abstract results updated to describe mean and standard deviation (SD) for children before and after adaptations.

2) Introduction edited to correct LMIC with countries. In Introduction, Figure 1 to include the numbers of 50 and 20 showing stages in the pilot. Further detail of previous studies using KABC-II subtests added.

3) Methods clarified that the site was where the SHINE study was previously conducted. Mean and standard deviation of the children given. Additional word 'was' added before conducted for describing the initial sensitisation visit to the family. Line added of no recorded harms.

4) For results, additional subsection of KABC-II added and training of fieldworkers described. In KABC-II adaptation section, additional details of item design and consultation with experts described.

5) For results section, further clarification of identifying problematic items for story completion was added, including an explanation that these items were replaced with pictures photographed with an appropriate Zimbabwean context. For pattern reasoning, further explanation was provided that items after item 4 required further examples and explanation. Hence additional questions were added with a specific learning point about colour or orientation. Addition of 'Pattern Plus' describing children using the adapted patterning. Section added comparing the scores before and after the pilot as well as describing scores from other studies that used the KABC-II total or MPI, also showing reduced planning section. Section added describing how other studies in Africa used individual subtests and showed some plausible associations, but did not use the mental processing index (MPI). For those that did use the MPI, similar raw scores were observed. Future work within the SHINE trial outlined including suggestions for examining a larger dataset.

Competing Interests: No competing interests were disclosed.