Journal of Psychology in Africa, 2025
Vol. 35, No. 4, 431–439, http://doi.org/10.32604/jpa.2025.070057

WPPSI-III in Sudan: Validity, reliability, and confirmatory factor analysis in khartoum kindergarten and primary schools

Rwaa Omer Ali Ahmed[1], Salaheldin Farah Attallah Bakhiet[2,*], Ayman Mohamed Taha Abdelaziz Ahmed[1], FadlAlMawla AbdulRadi[1] and Ismael Salamah Albursan[3]

[1]*Department of Educational Psychology, Faculty of Education, Khartoum University, Khartoum, P.O. Box 102, Sudan*
[2]*Department of Special Education, King Saud University, Riyadh, 11433, Saudi Arabia*
[3]*Department of Psychology, College of Education, King Saud University, Riyadh, 11433, Saudi Arabia*

**Correspondence: Salaheldin Farah Attallah Bakhiet,* bakhiet@ksu.edu.sa
Received: 19 March 2025; Accepted: 22 June 2025; Published: 17 August 2025

Abstract: The study aims to determine the validity and reliability of the Wechsler Preschool and Primary Scale of Intelligence–Third Edition (WPPSI-III) scores in a sample of kindergarten and lower primary pupils from Khartoum State, Sudan. It also aims to examine whether test's factor structure in this sample replicated that of the original WPPSI-III. The study sample consisted of 384 kindergarten and primary school children in Khartoum State (females = 50% mean age = 4.14, SD = 1.37), selected using stratified random sampling across its seven localities: Khartoum, Jebel Awliya, Khartoum Bahri, East Nile, Omdurman, Ombada, Karari. For concurrent validation, the children additionally completed the Goodenough Draw-a-Man Test, and the Colored Progressive Matrices. WPPSI-III scores demonstrated high internal consistency across the subtest items. Confirmatory factor analysis indicators for total, verbal, and performance intelligence were all excellent. The scale also showed weak to strong score stability ranging from 0.25 (weak) to 0.88 (strong) based on the Spearman-Brown equation, 0.25 to 0.75 based on the Guttman split-half method. The Cronbach's alpha coefficient scores ranged from 0.54 to 0.93. The WPPSI-III and Goodenough Draw-a-Man Test scores concurrent validity scores were poor (0.05) to modest (0.31), and while those with the Colored Progressive Matrices test were poor (r = 0.04–0.18). These findings provide evidence to suggest that the WPPSI-III is appropriate for research use with kindergarten and lower primary school students in Khartoum State, Sudan.

Keywords: intelligence tests; validation; reliability; primary schools; kindergartens; confirmatory factor analysis; cognitive abilities; WPPSI-III; Sudan

Introduction

Intelligence scales are widely used to evaluate children's cognitive abilities and to identify their educational needs. The Wechsler Preschool and Primary Scale of Intelligence, Third Edition (WPPSI-III), is one of the most widely used scales globally for assessing young children's intelligence. However, its cross-cultural applicability remains an open question. The WPPSI-III measures various cognitive domains, including thinking, problem-solving, reasoning, and decision-making skills. There is evidence suggesting that WPPSI-III scores predict children's behavioral, emotional, and developmental difficulties (Renaud et al., 2022), including attention deficit hyperactivity disorder and developmental coordination disorders (Jascenoka et al., 2015). However, concerns have been raised regarding the psychometric properties of the WPPSI-III scores and their applicability in low- and middle-income nations, given the test's crucial role in evaluating cognitive abilities. This study aims to assess the validity and reliability of the Wechsler Preschool and Primary Scale of Intelligence–Third Edition (WPPSI-III) scores among kindergarten and lower primary school children in Khartoum State, Sudan.

WPPSI-III history and description. The WPPSI-III was standardized on a large sample of 2100 children that was nationally representative of the population of the United States in terms of gender, race, geographic region, and socioeconomic status (Wechsler, 2002). The standardization study demonstrated strong reliability, including internal consistency and test-retest reliability, as well as excellent validity-encompassing construct, convergent, and discriminant validity. Additionally, the WPPSI-III showed significant correlations with other established measures of intelligence. The factor structure of the test supported the two primary indices: verbal intelligence (VIQ) and performance intelligence (PIQ), as well as the full-scale intelligence (FSIQ).

Findings from the Netherlands indicated excellent convergent validity when compared with a Dutch measure of intelligence (Karino et al., 2011). Similarly, a Spanish adaption of the WPPSI-III (Corral et al., 2015) revealed outstanding psychometric properties. Few studies have validated the WPPSI-III in cross-cultural, non-western settings.

Cross-cultural studies. Of the cross-cultural WPPSI-III validations is the study by Karakas and Yalin (2003). The aimed to standardize and adapt the WPPSI-III for use in Turkey, taking into account the country's environmental and cultural context. To ensure comprehensive national representation, the study included a sample of 1540 children aged 2.5 to 7, selected from various geographical regions across Turkey. The results indicated high levels of test-retest reliability (REPORT THE RANGE OF INTERNAL COSISTENCY SCORES HERE), and internal consistency (REPORT THE RANGE OF INTERNAL COSISTENCY SCORES HERE), in the Turkish context, Figueiredo and Nascimento (2007) conducted a study to adapt and validate the WPPSI-III in the Brazilian

context, while Chen and Zhu (2008) validated the WPPSI-III in the Chinese context, observing high structural, convergent, and discriminant validity, as well as high internal consistency and test-retest reliability of scores (STATE THE RANGE HERE). Studies conducted in Pakistan (Rasheed et al., 2018) and Jordan (Al-Mustafa & Al-Rawsan, 2013) yielded analogous results.

African studies. Of the WPPSI-III studies implemented in Africa, the study by Holding et al. (2019) is a notable exception. Holding et al. (2019) adapted the WPPSI-III in six countries across multiple locations and continents, including South Africa. The factor structure of the scale was analyzed, and the "fluid reasoning" dimension—comprising block design, matrix reasoning, and picture completion—was validated. The results provided support for the validity of this dimension across cultural settings, suggesting that these subtests could be used in research to explore relationships between variables. However, they are not recommended for use in mean comparisons or for clinical purposes due to the variability in scores across locations. This study highlights the importance of adapting cognitive assessment instruments to diverse cultural contexts and validating them before they are used in clinical or research settings. Bornman et al. (2021) adapted and validated the WPPSI-IV.

The sudanese context. Research on the standardization and psychometric properties of the Wechsler scales has been very limited in Sudan, focusing mainly on the WISC (Al-Hussein, 2008, 2005; Hussein, 1988; Khaleefa, 1987) and the Wechsler Adult Intelligence Scale (WAIS). To date, there have been no studies aimed at standardizing or validating the WPPSI scales—neither the current version nor its two previous versions—in Sudan. This is also true for most of the African continent, despite the WPPSI having been applied in nations such as South Africa and Tanzania (Ruan-Iu et al., 2020), where insufficient data have been published on its psychometric properties.

In conclusion, there is a clear need for studies on the WPPSI-III in Sudan, particularly among younger children. Such research is essential to determine the validity and reliability of scores among kindergarten and primary school students. Findings would guide the appropriate use of the WPPSI-III for making cognitive and educational decisions at the kindergarten and lower primary school levels.

Study goals. This study aimed to determine the validity and reliability of the Wechsler Preschool and Primary Scale of Intelligence–Third Edition (WPPSI-III) scores among children in kindergarten and lower primary school grades in Khartoum State, Sudan. Specifically, the study aimed to address the following questions:

1. What is the internal consistency of the WPPSI-III subtest scores?

2. What are the indicators of the confirmatory factorial validity of the WPPSI-III scores?

3. What is the concurrent validity of the WPPSI-III scores, as measured by their correlations with the Goodenough Draw-a-Man Test scores, and the Colored Progressive Matrices Test scores?

Method

Participants and setting

The study sample consisted of 384 kindergarten children (n = 192) and primary school children (n = 192) in Khartoum State (females = 50%, mean age 4.14, SD = 1.37 Years), selected using stratified random sampling across its seven localities: Khartoum, Jebel Awliya, Khartoum Bahri, East Nile, Omdurman, Ombada, Karari. The sample size was determined using Stephen Thompson's formula via computer software below:

$$n = \frac{N \times p\,(1-p)}{\left[\left[N - 1 \times \left(d^2 \div z^2\right)\right] + p\,(1-p)\right]}$$

The study sample was also distributed according to the chronological age of the children and divided based on the WPPSI-III age classification, into two groups: the first age group (2 years 6 months to 3 years 11 months) and the second age group (4 years 0 months to 7 years 3 months)

Measures

Criterion measure. Wechsler Preschool and Primary Scale of Intelligence—Third Edition (WPPSI-III) (WPPSI-III) The WPPSI-III scale consists of 14 subtests: Design Block: (20 items), Information: (34 items), Reasoning Matrix (39 items), Vocabulary (25 items), Concept Picture (38 items), Search Symbol (optional) (42 items), Verbal Reasoning Word: (28 items), Coding (optional (59 geometric shapes), Comprehension (22 items), Completion Picture (32 items), Similarities (24 items), Vocabulary Receptive (optional) (38 Items), Assembly Object: (14 items), Naming Picture (optional) (30 items) (Hadi and Murad, 2014, pp. 228–230).

Concurrent validity measures

Colored progressive matrices

The Sudanese version (Al-Khatib et al., 2006; Al-Khatib et al., 2021), consisting of 36 geometric shapes divided into 3 groups, was used to verify the concurrent validity of the WPPSI-III scale.

Goodenough–Harris draw-A-man test

The Sudanese version of the Goodenough–Harris Draw-A-Man Test, standardized by Badri (1997a, 1997b), was applied to verify the concurrent validity of the WPPSI-III scale.

Procedure

Ethical approval for the study was obtained from the Faculty of Education Research Board subcommittee on Research Ethics, University of Khartoum, ensuring that all procedures complied with established ethical standards for research involving human participants. Prior to data collection, informed consent was obtained from the parents or legal guardians of all participating children, who were provided with detailed information about the purpose, procedures, potential risks, and benefits of the study. Additionally, age-appropriate assent was obtained from the children themselves, ensuring that participation was

Table 1. Pearson correlation between the dimensions and indicators scale (WPPSI-III) and the Draw-A-Man Test and colored matrices

Goodenough Draw-a-Man Test		Colored Progressive Matrices		Intelligence
Value (r)	Sample size	Value (r)	Sample size	
0.27**	384	0.11	163	Verbal intelligence
0.23**	384	0.13	163	Practical intelligence
0.28**	384	0.14	163	Processing speed
0.28**	384	0.08	163	Linguistic complex
0.29**	384	0.14	163	Total Intelligence

Note. (**) Function at 0.01.

voluntary and that the children understood their right to withdraw at any time without any negative consequences. All data were treated with strict confidentiality and used solely for research purposes.

Data analysis

We utilized SPSS (Version 29) and Mplus (Version 8) for data analysis. The collected data were initially screened and cleaned using SPSS Version 29 to check for completeness, normality, and any outliers. Descriptive statistics such as means, standard deviations, and frequencies were computed to summarize the demographic characteristics of the participants and the distribution of test scores. Reliability analyses, including Cronbach's alpha, were conducted to assess the internal consistency of the WPPSI-III subtests.

Confirmatory factor analysis (CFA) was performed using Mplus Version 8 to examine the factorial validity of the WPPSI-III scores. Various model fit indices, including the Comparative Fit Index (CFI), Tucker-Lewis Index (TLI), Root Mean Square Error of Approximation (RMSEA), and Standardized Root Mean Square Residual (SRMR), were evaluated to determine the adequacy of the hypothesized factor structure. Additionally, correlation analyses were conducted to assess the concurrent validity of the WPPSI-III scores with external measures, such as the Sudanese geometric shapes test.

Results

Descriptive statistics

Tables 1 and 2 present the correlation coefficients between the WPPSI-III subtest scores and the Goodenough Draw-a-Man Test scores, which range from 0.05 to 0.31 ($p < 0.05$). In contrast, the correlation coefficients between the WPPSI-III, and the Colored Progressive Matrices scores range from 0.04 to 0.18 ($p > 0.05$). These results indicate that the WPPSI-III demonstrates concurrent validity with the Goodenough Draw-a-Man Test, but not with the Colored Progressive Matrices test.

WPPSI-III items intercorrelations. The Pearson correlation coefficient test was applied between the WPPSI-III items and subtests t. The results are as follows:

Cube design

The results of the item intercorrelations for the Block Design subtest, which consists of 20 items, revealed that items 17, 19, and 20 were not significantly correlated with the total test score. In contrast, all other items showed significant positive correlations with the total score of the Block Design subtest, with significance levels ranging between $p < 0.05$ and $p < 0.01$.

Information

The results of the item intercorrelations for the Information subtest, which comprises 34 items, showed that items 1 and 2 were not significantly correlated with the total test score. Items 3, 4, 5, 6, 7, 9, 11, and 17 exhibited weak correlations, while item 10 was negatively correlated with the total score. All remaining items demonstrated significant positive correlations with the total score of the Information subtest, with significance levels ranging between $p < 0.05$ and $p < 0.01$.

Matrices

The results of the item intercorrelations for the Matrix Reasoning subtest, which consists of 44 items, indicated that items 1 through 6 were not significantly correlated with the total test score. Items 7, 10, 11, 12, and 13 showed weak or non-significant correlations, while item 8 was negatively correlated with the total score. All remaining items demonstrated significant positive correlations with the total score of the Matrix Reasoning subtest, with significance levels ranging between $p < 0.05$ and $p < 0.01$.

Verbal vocabulary

The results of the item intercorrelations for the Verbal Vocabulary subtest, which consists of 25 items, revealed that items 1 through 6 were not significantly correlated with the total test score. Items 8, 12, 13, 14, 16, 17, 18, 19, 24, and 25 showed weak correlations, while item 7 was negatively correlated with the total score. All remaining items demonstrated significant positive correlations with the total score of the Verbal Vocabulary subtest, with significance levels ranging between $p < 0.05$ and $p < 0.01$.

Image concepts

The results of the item intercorrelations for the Image Concepts subtest, which consists of 38 items, showed that items 1 through 7, as well as items 27, 29, 30, 31, 32, 33, 34, 35, 36, 37, and 38, were not significantly correlated with the total test score. Items 8, 11, 12, 23, and 28 showed weak correlations. All remaining items demonstrated significant positive correlations with the total score of the

Table 2. Pearson correlation between the subtests of the WPPSI-III and the Draw-a-Man test and Colored Matrices test

The number	Colored Progressive Matrices		Goodenough Draw-a-Man Test		Subtests
	Value (r)	Sample size)	Value (r)	Sample size	
1	0.27 ∗∗	384	0.08	163	Cube design
2	0.31 ∗∗	384	0.05	163	Information
3	0.08	384	0.15	163	Matrices
4	0.25 ∗∗	384	0.18∗	163	Verbal vocabulary
5	0.21 ∗∗	384	0.11	163	Image Concepts
6	0.16 ∗∗	384	0.13	163	Search for symbols
7	0.28 ∗∗	384	0.13	163	Word inference
8	0.14∗	384	0.06	163	Coding
9	0.21 ∗∗	384	0.11	163	Understanding
10	0.19 ∗∗	384	0.13	163	Complete pictures
11	0.05	384	0.05	163	Similarities
12	0.28 ∗∗	384	0.04	163	Sensory vocabulary
13	0.24 ∗∗	384	0.09	163	Collect Things
14	0.30 ∗∗	384	0.15	163	Image naming

Note. (*) Function at 0.05; (**) Function at 0.01.

Image Concepts subtest, with significance levels ranging between $p < 0.05$ and $p < 0.01$.

Searching for symbols

The results of the item intercorrelations for the Symbol Search subtest, which consists of 42 items, indicated that items 2 through 6 and items 40, 41, and 42 showed weak or non-significant correlations with the total test score. Item 1 was not correlated, while item 7 was negatively correlated with the total score. All remaining items demonstrated significant positive correlations with the total score of the Symbol Search subtest, with significance levels ranging between $p < 0.05$ and $p < 0.01$.

Word reasoning

The results of the item intercorrelations for the Word Reasoning subtest, which consists of 28 items, indicated that items 1 through 6 were not significantly correlated with the total test score. Items 7, 8, 9, 10, 13, 14, 16, and 24 showed weak correlations, while item 11 was negatively correlated with the total score. All remaining items demonstrated significant positive correlations with the total score of the Word Reasoning subtest, with significance levels ranging between $p < 0.05$ and $p < 0.01$.

Coding

The results of the item intercorrelations for the Coding subtest, which consists of 59 items, revealed that items 2, 3, 4, 5, 6, 10, 25, 55, 58, and 59 showed weak or non-significant correlations with the total test score. Items 1 and 57 were negatively correlated, while item 56 showed no correlation. All remaining items demonstrated significant positive correlations with the total score of the Coding subtest, with significance levels ranging between $p < 0.05$ and $p < 0.01$.

Comprehension

The results of the item intercorrelations for the Comprehension subtest, which consists of 22 items, showed that items 1, 2, 3, and 22 were not significantly correlated with the total test score. Items 5, 6, and 8 demonstrated weak correlations. All remaining items were significantly positively correlated with the total score of the Comprehension subtest, with significance levels ranging between $p < 0.05$ and $p < 0.01$.

Similarities

The results of the item intercorrelations for the Similarities subtest, which consists of 24 items, showed that items 18 through 24 were not significantly correlated with the total test score. All other items demonstrated significant positive correlations with the total score of the Similarities subtest, with significance levels ranging between $p < 0.05$ and $p < 0.01$.

Sensory vocabulary

The results of the item intercorrelations for the Sensory Vocabulary subtest, which consists of 38 items, showed that items 1 through 9, 11, 12, 14, 15, and 16 were not significantly correlated with the total test score. Items 10, 13, 18, 19, 20, 21, 36, and 38 showed weak correlations. All remaining items demonstrated significant positive correlations with the total score of the Sensory Vocabulary subtest, with significance levels ranging between $p < 0.05$ and $p < 0.01$.

Assembling things

The results of the item intercorrelations for the Assembling Things subtest, which consists of 14 items, showed that item 1 was not correlated with the total test score, while item 3 showed a weak correlation. All other items demonstrated significant positive correlations with the total score of the Assembling Things subtest, with significance levels ranging between $p < 0.05$ and $p < 0.01$.

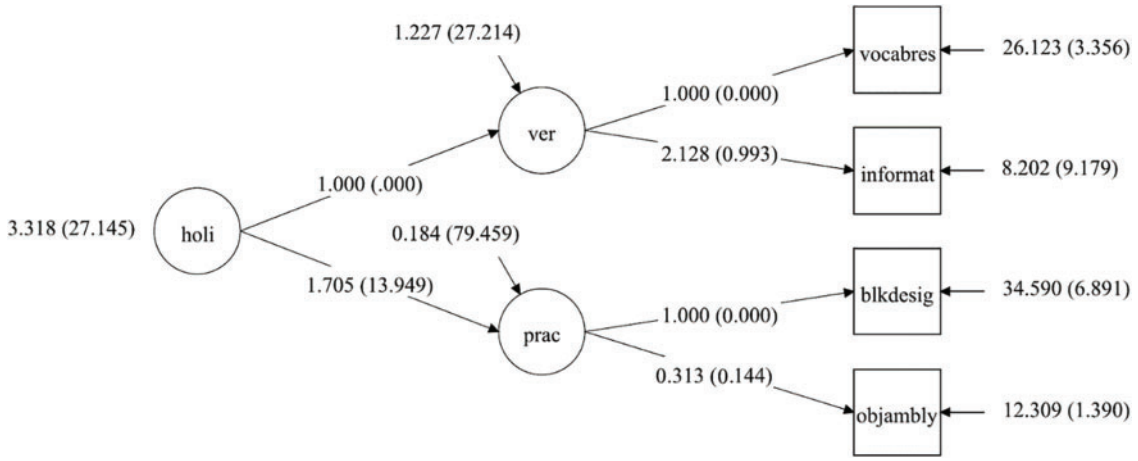

Figure 1. Confirmatory factor analysis of total intelligence and its verbal and performance factors for the first age group

Naming pictures

The results of the item intercorrelations for the Picture Naming subtest, which consists of 30 items, showed that item 1 was not correlated with the total test score, while items 2, 4, 10, and 12 demonstrated weak correlations. All other items showed significant positive correlations with the total score of the Picture Naming subtest, with significance levels ranging between $p < 0.05$ and $p < 0.01$.

Confirmatory factorial validity indices of the WPPSI-III

Confirmatory factor analysis was conducted on the intelligence scale for the first age group. The scale consisted of two factors: verbal intelligence, including the sensory vocabulary and information test, and practical intelligence, including the designing cubes and assembling objects. The model demonstrated acceptable fit indices as follows: CFI = 0.88, TLI = 1.00, Root Mean Square Error of Approximation (RMSEA) = 0.0001, Standard Root Mean Square Residual (SRMR) = 0.042.

They are all acceptable values. Figure 1 presents a diagram of the model used in the analysis. Secondly, Confirmatory factor analysis was conducted for the intelligence scale of the second age group. In this model, the scale consisted of two factors: verbal intelligence, including the information, verbal vocabulary, and word reasoning, and practical intelligence, including the designing cubes, matrices, and image concepts. The matching indicators are as follows: CFI = 0.982, TLI = 0.962, Root Mean Square Error of Approximation (RMSEA) = 0.087, Standard Root Mean Square Residual (SRMR) = 0.031.

They are all acceptable values. Figure 2 presents a diagram of the model used in the analysis. Thirdly: Confirmatory factor analysis was conducted for the linguistic complex of the second category, comprising two factors: sensory vocabulary and naming of images. The matching indicators are as follows: CFI = 0.954, TLI = 1.000, Root Mean Square Error of Approximation (RMSEA) = 0.000, Standard Root Mean Square Residual (SRMR) = 0.000. They are all acceptable values.

Figure 3 shows a diagram of the model used in the analysis. Confirmatory factor analysis was conducted for the linguistic complex of the second category, consisting of two factors: sensory vocabulary and image naming. The matching indicators are as follows: CFI = 0.956. TLI = 1.000, Root Mean Square Error of Approximation (RMSEA) = 0.000, Standard Root Mean Square Residual (SRMR) = 0.000. They are all acceptable values.

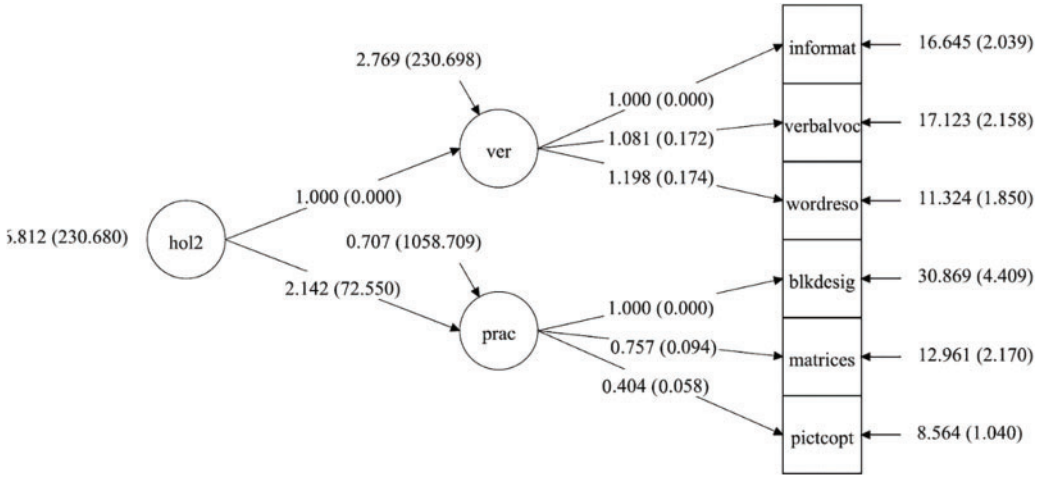

Figure 2. Confirmatory factor analysis of total intelligence and its verbal and performance factors for the second category

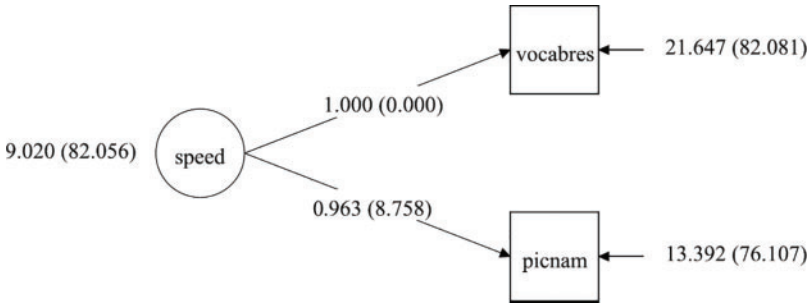

Figure 3. Confirmatory factor analysis of cognitive speed for the second age group

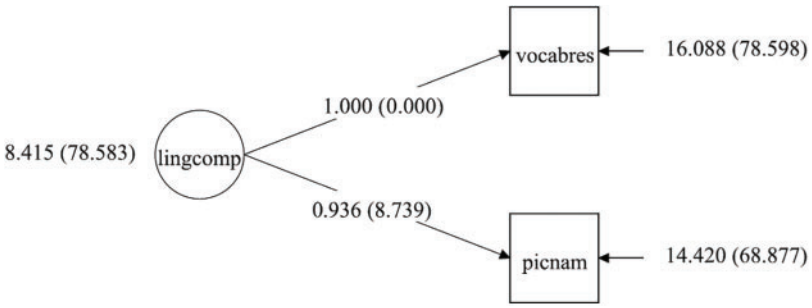

Figure 4. Confirmatory factor analysis of the linguistic complex of the first category

Figure 4 presents a diagram of the model used in the analysis.

Reliability of WPPSI-III subtest scores

Table 3 presents the reliability estimates of the WPPSI-III subtest scores, including the Spearman-Brown and Guttman coefficients and Cronbach's alpha values.

Table 3 shows the stability of the WPPSI-III subtest items based on the split-half method. The reliability coefficients calculated using the Spearman-Brown formula ranged from 0.44 to 0.80, while the Guttman split-half coefficients ranged from 0.25 to 0.75. Additionally, Cronbach's alpha coefficients ranged from 0.54 to0.93. These values indicate moderate to high internal consistency for the subsets.

Age effects on WPPSI-III raw scores

Table 4 presents the results of a one-way analysis of variance (ANOVA) examining differences in the raw scores of the WPPSI-III, across different age groups.

As shown in Table 3, all F values are all statistically significant at $p < 0.01$: verbal intelligence (9.914), practical intelligence (9.854), processing speed (18.904), linguistic complex (5.350), and total intelligence (4.752) The results showed that children in the youngest age group (2:06–2:08) obtained the lowest average scores in verbal intelligence, practical intelligence, linguistic composite, and processing speed. In contrast, children in the oldest age group (4:00–4:02) obtained the highest average scores across all these intelligence measures.

Discussion

The WPPSI-II scores demonstrated acceptable levels of reliability for research use with Sudanese kindergarten and primary school children. The findings of the present study are generally consistent with previous research, including the original American study by Wechsler (2002), which reported strong psychometric properties for the WPPSI-III. Similar patterns were also observed in European contexts (Corral et al., 2015; Karakas & Yalin, 2003), Asian studies (Chen & Zhu, 2008; Rasheed et al., 2018), research from Arab countries (Ejbara, 2024; Al-Mustafa, 2021; Al-Mustafa & Al-Rusan, 2013; Hadi & Murad, 2014), a South American study (Figueiredo & Nascimento, 2007), and the sole African study from South Africa (Holding et al., 2019).

This study further examined the internal consistency of the WPPSI-III within the Sudanese context, contributing preliminary data to the growing international literature; however, some aspects of validity may require further investigation across diverse populations. The results indicated an overall satisfactory level of internal consistency across most subtest items of the Wechsler Scale, suggesting reasonable validity in this sample. Nevertheless, certain subtests—particularly "Cube Design," "Information," and "Matrix"—showed weaker correlations with total scores, highlighting the need for potential review or adaptation of these specific items to better suit the local context. Moreover, future studies involving larger and more diverse samples are recommended to confirm and extend these findings.

Confirmatory factor analysis (CFA) confirmed the validity of the hypothesized factor structure of the WPPSI-III in the Sudanese context. This result is consistent with the results of previous studies conducted in Western and Asian societies (Wechsler, 2002; Chen and Zhu, 2008). It is also similar to the study done by Ejbara (2024) in Saudi Arabia, which is the only Arab study that has investigated its factor structure by confirmatory factor analysis. However, some subtests showed relatively weak correlations with their hypothesized factors, suggesting

Table 3. Reliability of the subtests of (WPPSI-III)

Test number	Cronbach's alpha	Guttman	Spearman-Brown	Subtest
1	0.87	0.61	0.70	Cube design
2	0.85	0.52	0.63	Information
3	0.87	0.36	0.43	Matrices
4	0.54	0.40	0.50	Verbal vocabulary
5	0.81	0.67	0.70	Image Concepts
6	0.93	0.53	0.54	Search for symbols
7	0.81	0.29	0.50	Word inference
8	0.93	0.75	0.75	Coding
9	0.83	0.66	0.68	Understanding
10	0.84	0.42	0.44	Complete pictures
11	0.91	0.67	0.88	Similarities
12	0.80	0.25	0.49	Sensory vocabulary
13	0.86	0.70	0.80	Collect Things
14	0.89	0.66	0.78	Image naming

Table 4. One-way analysis of variance (ANOVA) for differences in intelligence scores according to age

Probability value	Value (f)	Mean squares	Degree of freedom	Sum of squares		
0.00	9.91	435.80	18	8280.23	Between groups	Verbal intelligence
		43.96	365	16,001.27	Within groups	
			383	24,281.50	Total	
0.00	9.85	350.25	18	6304.45	Between groups	Practical intelligence
		35.54	365	12,972.79	Within groups	
			383	19,277.24	Total	
0.00	18.90	2833.06	18	50,995.06	Between groups	Linguistic complex
		149.86	365	54,700.29	Within groups	
			383	105,695.35	Total	
0.00	5.35	218.60	13	1311.611	Between groups	Processing speed
		40.86	263	7640.637	Within groups	
			191	8952.247	Total	
0.00	4.75	148.55	18	2673.826	Between groups	Total
		31.26	365	11,410.734	Within groups	Intelligence
			384	14,084.560	Total	

possible cultural and linguistic influences on children's cognitive performance. Therefore, future studies should explore the possibility of modifying some of the test items to enhance their suitability for non-Western contexts.

The WPPSI-III scores had inconsistent convergent and categorical validity with the Draw-a-Man test and the Colored Progressive Matrices (CPM). This result may indicate that the two tests measure different cognitive domains: the Wechsler focuses on verbal and performance abilities, with many of its subscales being verbal, whereas the matrices test focuses on fluid intelligence and visual reasoning skills. The study revealed significant differences in WPPSI-III scores across age groups, with older children recording higher cognitive performance, as would be expected. This indicates the sensitivity of the scale in discriminating between different age levels, which is essential when using it as an assessment tool. This trend has been reported in previous normative studies (Wechsler, 2002; Canivez, 2009), highlighting the importance of appropriate age criteria when testing intelligence. These results emphasize the importance of taking chronological age into account when applying intelligence scales. However, the sharp increase in scores between younger and older age groups also suggests the need for further research into the cognitive development trajectories of Sudanese children, particularly in relation to environmental and educational factors.

Finally, the WPPSI-III demonstrated high reliability and validity, with Cronbach's alpha values exceeding the generally accepted threshold of 0.70 (Nunnally & Bernstein, 1994). The high reliability coefficients for both the verbal and performance subtests indicate that the scale provides stable and consistent measures of intelligence in preschool and lower elementary school children up to 7 years and 3 months of age. However, some subtests, such as cube design and information, demonstrated relatively lower reliability estimates, which may be attributed to linguistic and cultural factors that influence item comprehension. Future research should conduct item-level analyses to discover and improve culturally sensitive test items.

Limitations and future directions

Despite its contributions, the current study faces some limitations. First, the sample size was relatively small compared to that of previous studies. Therefore, future studies should use larger sample size to determine the stability of the findings we report here. Second, the study relied primarily on quantitative methods; the use of qualitative methods, such as interviews with teachers and parents could provide additional insights into the ecological validity of the WPPSI-III. Finally, as this was a cross-sectional study, future longitudinal research is needed to explore the predictive validity of WPPSI-III scores concerning academic achievement and socio-emotional development over time.

Conclusion

The results of the study indicate that the Wechsler Preschool and Primary Scale of Intelligence (WPPSI-III)—Third Edition demonstrates comparable validity and reliability of scores to those reported in other countries, suggesting its potential as a suitable tool for assessing cognitive abilities in early childhood within the Sudanese context, including lower primary school grades up to the age of 7 years and 3 months. Future research should focus on improving the wording of some of the test items, to strengthen the scale and improve its contextual relevance. In addition, future studies should expand or add psychometric indicators that were not included in the current study, such as differential performance of items, and gender differences. Overall, these current efforts contribute meaningfully to the ongoing discourse on developing fair and culturally responsive intelligence assessments worldwide.

Acknowledgement: The authors extend their appreciation to the Ongoing Research Funding Program, number (ORF2025R705), King Saud University, Riyadh, Saudi Arabia, for funding this work.

Funding Statement: The authors extend their appreciation to the Ongoing Research Funding Program, number (ORF2025R705), King Saud University, Riyadh, Saudi Arabia, for funding this work.

Author Contributions: The authors confirm contribution to the paper as follows: study conception and design: Rwaa Omer Ali Ahmed, FadlAlMawla AbdulRadi; data collection: Ayman Mohamed Taha Abdelaziz Ahmed, Rwaa Omer Ali Ahmed; analysis and interpretation of results: Salaheldin Farah Attallah Bakhiet, Ismael Salamah Albursan, FadlAlMawla AbdulRadi; draft manuscript preparation: Salaheldin Farah Attallah Bakhiet, Ismael Salamah Albursan. All authors reviewed the results and approved the final version of the manuscript.

Availability of Data and Materials: Data available on request from the authors.

Ethics Approval: Ethical approval for the study was obtained from the Faculty of Education Research Board subcommittee on Research Ethics, University of Khartoum, ensuring that all procedures complied with established ethical standards for research involving human participants. Prior to data collection, informed consent was obtained from the parents or legal guardians of all participating children, who were provided with detailed information about the purpose, procedures, potential risks, and benefits of the study.

Conflicts of Interest: The authors declare no conflicts of interest to report regarding the present study.

References

Al-Hussein, A. (2005). *Adaptation and Standardization of the Wechsler Scale of Children's Intelligence, third edition, in Khartoum State* [Unpublished master's thesis]. Sudan: Al-Neelain University.

Al-Hussein, A. (2008). *Adaptation and Standardization of the Wechsler Children's Intelligence Scale, third edition, in the northern states* [Unpublished doctoral thesis]. Sudan: Al-Neelain University.

Al-Khatib, M., Al-Mutawakkil, M., Bakhiet, S., & Abd Al-Rahim, N. (2021). Standardization of the colored progressive matrices test in the sudanese environment for children aged (4–6) years.

Al-Khatib, M., Al-Mutawakkil, M., & Hussein, A. (2006). *Codification of the colored sequential matrices test for students of the first cycle in the basic stage in Khartoum State*. Khartoum: Sudan Currency Printing Company Ltd.

Al-Mustafa, G. (2021). Derivation of the final image criteria for the Wechsler Scale: The Jordanian image of children's intelligence for preschool and primary school for the age group 4–6.5 years. *International Journal of Quality Assurance*, *4*(1), 13–23.

Al-Mustafa, G., & Al-Rusan, F. (2013). *The effectiveness of a Jordanian image of the Wechsler Intelligence Scale for preschool children on the Jordanian environment for the age group (4–6) years* [Ph.D. thesis]. Jordan: College of Graduate Studies, World Islamic Sciences and Education University.

Badri, H. (1997a). *A trial of standardization of Columbia Mental Maturity Scale (CMMS) in Khartoum state* [Unpublished M.A. thesis]. Sudan: University of Khartoum.

Badri, M. (1997b). *Psychology of children's drawings, human drawing tests and their applications on children in Arab countries*. 2nd ed. Jordan: Dar Al-Furqan.

Bornman, J., Sevcik, R. A., Romski, M., & Pae, H. K. (2021). Measuring early cognitive development in South African children: Adaptation and validation of the WPPSI-IV. *South African Journal of Childhood Education*, *11*(1). https://scielo.org.za/pdf/sajce/v11n1/15.pdf.

Canivez, G. L. (2009). Factor structure of the WPPSI-III and comparison with other measures of intelligence. *Psychological Assessment*, *21*(3), 479–487. https://doi.org/10.1037/a0016617

Chen, H., & Zhu, J. (2008). Adaptation and validation of the Wechsler Preschool and Primary Scale of Intelligence-Third Edition (WPPSI-III) in China. *Chinese Journal of Clinical Psychology*, *16*(5), 505–508.

Corral, S., Arribas, D., Santamaría, P., Sueiro, M. J., & Pereña, J. (2015). *WPPSI-III: Escala de Inteligencia de Wechsler para Preescolar y Primaria–III*. Madrid: TEA Ediciones.

Ejbara, M. (2024). Psychometric characteristics of a Saudi version of wechsler intelligence scale for preschool and young children. *Dirasat: Educational Sciences*, *51*(2), 35–52. https://doi.org/10.35516/edu.v51i2.6661

Figueiredo, V. L. M., & Nascimento, E. (2007). Adaptação e validação da Escala de Inteligência Wechsler para Crianças–Terceira Edição (WISC-III) e WPPSI-III no Brasil. *Psicologia: Reflexão e Crítica*, *20*(2), 226–235.

Hadi, F. A., & Murad, S. A. (2014). Standardization of the Wechsler Intelligence Scale for preschool and primary school children. *Kuwait: Journal of Gulfand Arabian Peninsula Studies*, *40*(155), 205–270.

Holding, P., Abubakar, A., Kitsao-Wekulo, P., Kvalsvig, J., Mupere, E., et al. (2019). Assessing early cognitive development in African children: The adaptation of WPPSI-III in the MAL-ED study. *Journal of Psychoeducational Assessment*, *37*(3), 254–268. https://doi.org/10.1177/0734282919850040

Hussein, M. (1988). *Standardization and sudanese adaptation of the wechsler scale of children's intelligence—revised* [Unpublished master's thesis]. Sudan: University of Khartoum.

Jascenoka, J., Korsch, F., Petermann, F., & Petermann, U. (2015). Kognitive Leistungsprofile von Kindern mit motorischen Entwicklungsstörungen und ADHS im Vorschulalter [Cognitive profiles of preschool children with developmental coordination disorders and ADHD]. *PraxKinderpsychol Kinderpsychiatr*, *64*(2), 117–134. https://doi.org/10.13109/prkk.2015.64.2.117

Karakas, S., & Yalin, A. (2003). The adaptation and standardization of the Wechsler Preschool and Primary Scale of Intelligence-Third Edition (WPPSI-III) in Turkey. *Turkish Journal of Psychology*, *18*(52), 1–15.

Karino, C., Laros, J., & de Jesus, G. (2011). Evidências de validadeconvergente do SON-R 2½-7[a] com o WPPSI-III e WISC-III Evidences of convergent validity of SON-R 2½-7[a] with WPPSI-III and WISC-III. *Psicologia: Reflexão E Crítica*, *24*(4), 621–629. https://doi.org/10.1590/s0102-79722011000400001

Khaleefa, O. H. (1987). *Standardization of the wechsler adult intelligence scale – revised for use in sudan* [Unpublished master's thesis]. Sudan: Faculty of Arts, University of Khartoum.

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory*. 3rd ed. New York: McGraw-Hill.

Rasheed, M., Pham, S., Memon, U., Siyal, S., Obradović, J., &, et al. (2018). Adaptation of the Wechsler Preschool and Primary scale of intelligence-III and lessons learned for evaluating intelligence in low-income settings. *International Journal of School & Educational Psychology*, *6*(3), 197–207. https://doi.org/10.1080/21683603.2017.1302851

Renaud, F., Béliveau, M., Akzam-Ouellette, M., Jauvin, K., & Labelle, F. (2022). Comparison of the Wechsler preschool and Primary scale of intelligence-third edition and the Leiter-R intellectual assessments for clinic-referred children. *Journal of Psychoeducational Assessment*, *40*, 825–838. https://doi.org/10.1177/07342829221105388

Ruan-Iu, L., Pendergast, L., Rasheed, M., Tofail, F., Svensen, E., et al. (2020). Assessing early childhood fluid reasoning in low- and middle-income nations: Validity of the wechsler preschool and primary scale of intelligence across seven MAL-ED sites. *Journal of Psychoeducational Assessment*, *38*(2), 256–262. https://doi.org/10.1177/0734282919850040

Wechsler, D. (2002). *The wechsler preschool and primary scale of intelligence*. Third Edition (WPPSI-III). San Antonio, TX, USA: The Psychological Corporation.