

Article

The Adaptation of the Wechsler Intelligence Scale for Children—5th Edition (WISC-V) for Indonesia: A Pilot Study

Whisnu Yudiana ^{1,2,3,4} , Marc P. H. Hendriks ^{1,5,6} , Christiany Suwartono ⁵ , Shally Novita ^{2,3,4}, Fitri Ariyanti Abidin ^{2,7} and Roy P. C. Kessels ^{1,8,9,*} 

- ¹ Donders Institute for Brain, Cognition and Behaviour, Radboud University, 6525 GD Nijmegen, The Netherlands; whisnu.yudiana@donders.ru.nl (W.Y.); marc.hendriks2@donders.ru.nl (M.P.H.H.)
- ² Department of Psychology, Faculty of Psychology, Universitas Padjadjaran, Jatinangor 45363, West Java, Indonesia; s.novita@unpad.ac.id (S.N.); fitri.ariyanti.abidin@unpad.ac.id (F.A.A.)
- ³ Center for Psychological Innovation and Research, Faculty of Psychology, Universitas Padjadjaran, Jatinangor 45363, West Java, Indonesia
- ⁴ Centre for Psychometrics Study, Faculty of Psychology, Universitas Padjadjaran, Jatinangor 45363, West Java, Indonesia
- ⁵ Faculty of Psychology, Atma Jaya Catholic University of Indonesia, Jakarta 12930, Indonesia; christiany.suwartono@atmajaya.ac.id
- ⁶ Academic Centre for Epileptology, Kempenhaeghe, 5591 VE Heeze, The Netherlands
- ⁷ Center for Relationship, Family Life and Parenting Studies, Faculty of Psychology, Universitas Padjadjaran, Jatinangor 45363, West Java, Indonesia
- ⁸ Vincent van Gogh Institute for Psychiatry, 5803 DN Venray, The Netherlands
- ⁹ Radboudumc Alzheimer Center, Radboud University Medical Center, 6500 HB Nijmegen, The Netherlands
- * Correspondence: roy.kessels@donders.ru.nl

Abstract

The Wechsler Intelligence Scale for Children (WISC) is a widely used instrument for assessing cognitive abilities in children. While the latest fifth edition (WISC-V) has been adapted in various countries, Indonesia still relies on the outdated first edition, a practice that raises substantial concerns about the validity of diagnoses, outdated norms, and cultural bias. This study aimed to (1) adapt the WISC-V to the Indonesian linguistic and cultural context (WISC-V-ID), (2) evaluate its psychometric properties in a pilot study with an Indonesian sample, (3) reorder the item sequence of the subtests according to the empirical item difficulty observed in Indonesian children's responses, and (4) evaluate the factor structure of the WISC-V-ID using confirmatory factor analysis. The adaptation study involved a systematic translation procedure, followed by psychometric evaluation with respect to gender, age groups, and ethnicity, using a sample of 221 Indonesian children aged 6 to 16 years. The WISC-V-ID demonstrated good internal consistency. Analysis of item difficulty revealed discrepancies in item ordering compared to the original WISC-V, suggesting a need for item reordering in future studies. In addition, the second-order five-factor model, based on confirmatory factor analysis, indicated that the data did not adequately fit the model, stressing the need for further investigation. Overall, the WISC-V-ID appears to be a reliable measure of intelligence for Indonesian children, though a comprehensive norming study is necessary for full validation.

Keywords: psychological assessment; intelligence testing; WISC-V; adaptation; reliability



Received: 1 May 2025
Revised: 17 June 2025
Accepted: 20 June 2025
Published: 24 June 2025

Citation: Yudiana, Whisnu, Marc P. H. Hendriks, Christiany Suwartono, Shally Novita, Fitri Ariyanti Abidin, and Roy P. C. Kessels. 2025. The Adaptation of the Wechsler Intelligence Scale for Children—5th Edition (WISC-V) for Indonesia: A Pilot Study. *Journal of Intelligence* 13: 76. <https://doi.org/10.3390/jintelligence13070076>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Intelligence is a core cognitive concept in neuropsychology, typically measured by standardized tests. Some researchers view intelligence as a hierarchically structured con-

struct composed of several cognitive processes and domains, such as verbal comprehension, fluid reasoning, visual-spatial ability, working memory, and information processing speed (Daniel 1997; Deary et al. 2010; Schneider and McGrew 2018; Wechsler 2014a). Others, however, emphasize the general intelligence factor (g factor) as a central component in the structure of cognitive abilities (Jensen 1998; Reeve and Charles 2008). Intelligence plays a crucial role in various domains of human life, including planning and problem-solving in daily activities, education, professional careers, and clinical practice (Priyamvada et al. 2024). Specifically in children, intelligence is essential for characterizing and classifying individuals based on their cognitive strengths and weaknesses (Daniel 1997), identifying learning problems and disabilities (Daniel 1997; Maki and Adams 2019), and determining the most effective teaching and instructional methods in the classroom (Miller et al. 2021). Clinicians such as neuropsychologists, developmental psychologists, or clinical psychologists utilize intelligence assessments to evaluate the cognitive profiles of children, including those with intellectual disabilities (Billard et al. 2021; Chen et al. 2023), developmental disorders such as Attention Deficit Hyperactivity Disorder (Pritchard et al. 2012), or brain disorders such as epilepsy (Leonard et al. 2023).

Intelligence tests differ in content, format, and underlying theoretical frameworks. Their primary aim is to assess different intellectual domains, which may vary among children, even in those with similar general cognitive abilities (Caemmerer et al. 2020). The Wechsler Intelligence Scales for Children (WISC) are among the most widely used tests for measuring children's intellectual abilities worldwide (Benson et al. 2019; McGill et al. 2020; Van de Vijver et al. 2019). The first version of the WISC was published in the USA in 1949 (Wechsler 1949). Since then, the test has undergone numerous refinements and modifications, both in terms of items, subtests, and norms, in response to updated conditions and theoretical developments, particularly driven by the rapid growth of research in cognitive neuroscience and functional brain imaging (Kaufman et al. 2016; McGill et al. 2020). These modifications were also motivated by the Flynn effect, which refers to substantial IQ gains from one generation to another within the twentieth century (Flynn 2007).

The most recent version of the Wechsler Intelligence Scales for Children, the fifth edition (WISC-V, Wechsler 2014a), was initially published in the USA in 2014. The test comprises 21 subtests, including 10 primary subtests, 6 secondary subtests, and 5 supplemental subtests. Since its release, the WISC-V has been translated and adapted worldwide. In most countries, such as Canada (Wechsler 2014b), the United Kingdom (Wechsler 2016c), Australia and New Zealand (Wechsler 2016b), and Taiwan (Wechsler 2018), the WISC-V includes 16 subtests, primary and secondary subtests combined. In other countries, such as Spain (Wechsler 2015), France (Wechsler 2016a), Germany (Wechsler 2017), and Chile (Rodríguez-Cancino and Concha-Salgado 2024), the 15-subtest version was published, excluding the secondary subtest Picture Concepts. In the Netherlands (Hendriks et al. 2017), a 14-subtest version was published, excluding the secondary subtests, Information and Comprehension. These adaptations and studies reflect the generalizability of the intelligence structure proposed in the WISC (Van de Vijver et al. 2019; Wilson et al. 2023). Despite its potential applications in assessing various aspects of children's cognitive abilities, reports on the adaptation of the WISC-V in Southeast Asian countries, including Indonesia, remain relatively scarce. To date, the only Wechsler test that has been adapted and standardized for use in the Indonesian context is the Indonesian version of the Wechsler Adult Intelligence Scale-IV (WAIS-IV-ID) (Suwartono 2016), which also includes the development of a short-form version (Suwartono et al. 2023).

The adaptation of the WAIS-IV-ID followed the guidelines for test adaptation set by the International Test Commission (Hernández et al. 2020; International Test Commission 2017). Currently, the process is in its final stages before the test can be used in Indone-

sia. In contrast, intelligence measures for children in Indonesia are, to date, limited to unauthorized translations of the first edition of the WISC (Wechsler 1949) and the WISC-R (Wechsler 1974), conducted at the time without formal permission from the test publisher. Furthermore, the widespread use of the WISC in Indonesia—such as for clinical and educational practice, including in the curricula of psychology training programs at most universities—has not been accompanied by a well-documented written manual and, more importantly, has never been standardized and normed in the Indonesian population.

The use of the first edition of the WISC without standardization or normative data for the Indonesian population raises concerns about its validity in assessing intelligence in the Indonesian context. One study suggests that relying on non-representative norms may lead to misclassification and an invalid understanding of children's cognitive abilities (Lozano-Ruiz et al. 2021). For example, a child may be classified as below average due to the use of inappropriate, culturally biased norms. Other research has emphasized the importance of adopting local norms to ensure accurate interpretation of cognitive test results (Casaletto et al. 2015). In addition, it is difficult to argue that the current version of the WISC, which was developed in 1949 and 1974 without an authorized translation, can be used without any concern in the 21st century. Therefore, the current versions of the WISC used in Indonesia carry the risk of measurement errors, misdiagnosis of children's intellectual abilities, and wrong decisions in indicating interventions (i.e., treatment programs for learning difficulties). Therefore, a formal, authorized adaptation of the Wechsler Intelligence Scale for Children (i.e., the WISC-V) for Indonesia is necessary and urgent. The purposes of this article are to (1) describe the adaptation processes of the WISC-V for use in the Indonesian context, (2) evaluate the psychometric properties of the WISC-V-ID using an Indonesian sample, (3) reorder the item sequence of the subtests according to the empirical item difficulty observed in Indonesian children's responses, and (4) evaluate the factor structure of the WISC-V-ID using confirmatory factor analysis.

2. Materials and Methods

2.1. Participants

This pilot study was conducted in the province of West Java, Indonesia, the most populous province in Indonesia, comprising 18% of Indonesia's population in 2020 (BPS 2022). The study included 221 children aged 6 to 16 ($M = 11.40$, $SD = 3.16$), of whom 51% (113) were girls. Of these, 4.98% were in kindergarten, 49.32% were in primary school (years 1 to 6), 27.15% attended middle schools (years 7 to 9), and 18.55% were in high school (years 10 to 11). Regarding ethnicity, 42.08% were Sundanese, 24.43% Batak, 18.10% Javanese, 9.05% were Chinese, and the remainder belonged to other ethnicities. All participants were native Indonesian speakers without apparent physical or intellectual disabilities that could interfere with test administration.

2.2. Instrument

In the present study, the fifth version of the Wechsler Intelligence Scale for Children—5th edition was used to adapt for Indonesian speakers (referred to as the WISC-V-ID). Similar to the original version, the test comprises 10 primary subtests, namely Block Design (BD), Similarities (SI), Matrix Reasoning (MR), Digit Span (DS), Coding (CD), Vocabulary (VC), Figure Weights (FW), Visual Puzzles (VP), Picture Span (PS), and Symbol Search (SS). Originally, these primary subtests are used to estimate five Primary Index scores: Fluid Reasoning, Verbal Comprehension, Visual Spatial, Working Memory, and Processing Speed. The test also includes six secondary subtests: Information (IN), Picture Concepts (PC), Letter-Number Sequencing (LN), Cancellation (CA), Comprehension (CO), and Arithmetic (AR).

2.3. Adaptation Process of the WISC-V-ID

The translation and adaptation of WISC-V-ID followed the guidelines of the International Test Commission ([Hernández et al. 2020](#); [International Test Commission 2017](#)) and the International Neuropsychological Society ([Nguyen et al. 2024](#)) and involved the following phases:

2.3.1. Permission for Translation and Adaptation

Permission was obtained from the intellectual property rights holder, Pearson Education South Asia Pte Ltd, Sydney, Australia, through a Statement of Work (SOW) agreement with Universitas Padjadjaran, Indonesia, dated 27 March 2023.

2.3.2. Expert Review

Expert consultations identified the specific components requiring translation and adaptation, which include two aspects. First, the Indonesian translation of instructions, administration procedures, and scoring guidelines. Second, item translations and adaptation for selected subtests of Verbal Comprehension (SI, VO, IN, and CO) and Fluid Reasoning (AR).

2.3.3. Forward and Backward Translation

Two bilingual translation teams conducted forward and backward translations, ensuring cultural and linguistic equivalence. Each team consisted of three members fluent in English and Bahasa Indonesia, two of whom had a PhD in psychology and one of whom was a clinical child psychologist. All the translators also had experience in psychometrics and had previously conducted test adaptations. The following steps were taken to ensure the quality of both forward and backward translations:

The first step involved the forward translation of the administration procedures and scoring guidelines for 16 subtests, as well as all items from five subtests (SI, VO, IF, CO, and AR). Initially, all translators worked independently to review, translate, and modify the items to ensure alignment with Indonesian culture and local context while maintaining equivalence with the original items. Where specific nomenclature or terminology was unfamiliar in the Indonesian context (e.g., the word “Winter”), it was replaced with a more familiar nomenclature (e.g., “Rain”). The translation coordinator then compiled all forward translations into a single document. Any discrepancies in terminology or items were discussed to reach a consensus, with the involvement of an Indonesian language expert to ensure an accurate and equivalent adaptation process.

In step 2, the backward translation was conducted by three independent translators. All items with discrepancies in meaning between the Indonesian and English versions were discussed to arrive at a single translation result. Next, an independent committee reviewed and discussed the forward and backward translation results and the original items to achieve the most accurate adaptation.

Across the five subtests of the original Verbal Comprehension Index (VCI), the percentage of items undergoing adaptation for the Indonesian context varied from 3% to 31%. The VO subtest underwent the most significant changes. Some items in the original version assessed C1–C2 English proficiency, assuming advanced language skills. However, due to the difficulty of finding Indonesian equivalents of similar levels of difficulty, these items were replaced with more commonly used words with simpler meanings. Additionally, several example answers were added to key questions related to “remedy”, as Indonesian children were likely to interpret the term in the context of improving a low exam grade rather than its original meaning related to “cure” or “medicine”. Adjustments were also made to other VCI subtests. In the SI subtest, items referencing unfamiliar concepts were modified. For example, besides an item about US seasons being adapted to reflect Indonesia’s two-

season climate, the item concerning sibling relationships (e.g., “brother and sister”) was also adapted because the Indonesian language does not specify gender. In addition, an item referring to a specific time-measuring instrument unfamiliar to Indonesian children was replaced with a more commonly recognized Indonesian timekeeping tool. The IN and CO subtests were similarly modified. Geographic locations unfamiliar to Indonesian children were replaced with recognizable cities in the IN subtest. In the CO subtest, an item related to the children’s reasons for traveling was replaced with an activity more familiar to Indonesian school children. Finally, the pronouns in the AR subtest were adjusted to sound more natural in Indonesian, as the originals were based on US child names.

Step 3 involved a small-scale field test of the first version of the Indonesian WISC-V (WISC-V-ID) with five Indonesian children as a target population. The administration was administered by members of the adaptation team and observed by two junior psychologists acting as data collectors and assessors. The purpose of the field test was to assess how feasible the adapted administration procedures and items were when applied in the Indonesian context. During testing, administrators provided additional information, simplified instructions, and made modifications as needed when children experienced difficulties in understanding. After testing the children, the adaptation team and assessors reviewed both the testing experience and the children’s feedback, then adjustments were made as necessary to improve clarity and usability. All adjustments were documented to finalize the administration procedures, which were then documented as the final translated version of the WISC-V-ID.

2.4. Procedure

The data collection procedures were registered and approved by the Research Ethics Committee of Universitas Padjadjaran. Permission was obtained from the school principals and teachers, written consent was obtained from one or both parents, and assent was also obtained from the participating children. Data collection was conducted individually by Master’s students or graduates with a Bachelor’s degree in Psychology who had undergone a two-day training on the administration and scoring of the adapted WISC-V-ID. The training included role-playing exercises before examiners commenced data collection with children. To ensure the standardization and quality of the test administration, each examiner was required to submit recordings of their first two data collection sessions and await feedback before proceeding with further testing. Any identified errors were reviewed and discussed with the examiners for improvement, as outlined in the administration manual ([Wechsler 2014a](#)).

The test was administered following the guidelines outlined in the administration manual ([Wechsler 2014a](#)). Each child was tested individually, beginning with 10 primary subtests. After a brief break, the test continued with six secondary subtests. In this pilot study, the item order mirrored the original US version, and discontinue rules were not applied; all items were administered to each child. On average, the test administration ranged from 2 to 3 h, depending on the child’s pace. After completion, the children received a voucher or lunch box and snacks valued at Rp50,000 (approximately \$3.00 or €2.80).

2.5. Analyses

The analyses were conducted to evaluate the psychometric properties of the WISC-V-ID, focusing on item discrimination, difficulty, and test reliability. Thirteen of sixteen subtests of the WISC-V-ID were analyzed, while CD, SS, and CA were excluded from the analysis since they rely on time constraints rather than individual correct answers.

The majority of subtests in the WISC-V-ID consist of dichotomous items; however, five out of the sixteen subtests (i.e., BD, SI, PS, VO, and CO) contain polytomous items.

Item discrimination measures the ability of the items to differentiate between high- and low-achieving children (Cohen and Swerdlik 2009). In this study, item discrimination was estimated using the item-total correlation for each subtest. Correlations were primarily based on Pearson product-moment correlation for polytomous items, whereas biserial point correlations were used for dichotomous items. Coefficients were expected to be positive, and items with negative coefficients were subject to revision or removal (Cohen and Swerdlik 2009).

Item difficulty (p) was typically estimated using the average item score. For dichotomous items, this index reflects the proportion of children who answered the item correctly, with values ranging from 0 to 1 (where 1 indicates an easy item) (Cohen and Swerdlik 2009). This analysis is crucial for reordering the WISC-V-ID items based on their difficulty. Reordering was necessary because, in the WISC-V and previous versions, participants did not complete all subtest items but instead progressed based on ability, with discontinue rules in place to stop testing after consecutive incorrect answers (Wechsler 2014a).

Cronbach's α coefficients were calculated to assess internal consistency, reflecting the extent to which items within each subtest measure the same underlying construct. However, given the limitations of α in accurately estimating the reliability of multidimensional scales, which are common in neuropsychological assessments (Watkins 2017), McDonald's ω was also computed in this study. McDonald's ω was selected because it relies on fewer assumptions and is increasingly recommended for reporting internal consistency, particularly in scales that may not meet the strict assumptions required by α (Dunn et al. 2014). In the US version, reliability coefficients were reported separately by age (Wechsler 2014a). However, due to constraints in the sample size and to ensure more homogeneous characteristics within each group, this pilot study only reported reliability estimates across three broader age groups: 6–9 years, 10–12 years, and 13–16 years. Each group comprised approximately 66 to 88 participants. Since CD, SS, and CA assess processing speed based on all items taken together, reliability estimates could not be computed for these subtests. Several recommendations have been proposed for interpreting the internal consistency of psychological tests. A commonly cited minimum threshold is 0.70, while values between 0.80 and 0.90 are often recommended when tests are used for individual-level decision-making (Cohen and Swerdlik 2009; Kaplan and Saccuzzo 2012). In clinical settings, high reliability (with a value of at least 0.95) is critical due to the potential impact of test results on an individual's diagnosis (Cohen and Swerdlik 2009; Kaplan and Saccuzzo 2012; Nunnally and Bernstein 1994). To aid interpretation, the standard error of measurement was also reported.

As mentioned in the Section 2.4, the test was administered using the original US version without applying the discontinue rules. However, to further examine the quality of the test, this study also calculated internal consistency estimates after reordering the items based on the observed item difficulty in the sample and applying the discontinue rules as recommended in the test manual (Wechsler 2014a). This approach was used to conduct a comparison of reliability estimates under different scoring conditions, providing additional evidence regarding the internal consistency of the WISC-V-ID. To evaluate the effects of item ordering and the application of discontinue rules, a 2×2 factorial analysis was conducted. In addition to the two primary conditions, reliability estimates were also calculated for two supplementary conditions: the original item order with discontinue rules applied, and the reordered items without rules being applied (See Appendix A, Tables A2 and A3).

This pilot study involved conducting confirmatory factor analysis (CFA) to provide evidence of the WISC-ID's validity evidence based on its internal structure. Two models were evaluated. The first model, based on the WISC-V test manual (Wechsler 2014a), specifies a second-order five-factor structure representing the following broad cognitive

abilities: (1) Verbal Comprehension (Gc), measured by VO, SI, IN, and CO; (2) Visual Spatial (Gv), measured by BD and VP; (3) Fluid Reasoning (Gf), measured by MR, FW, and PC; (4) Working Memory (Gwm), measured by DS, PS, LN, and AR; (5) Processing Speed (Gs), measured by CD, SS, and CA. These five first-order factors were modeled under a higher-order general intelligence factor (g). The second model was a modified version informed by prior findings in the WISC-V manual, which indicated that the AR subtest may cross-load onto both Gf and Gc (Wechsler 2014a; Weiss et al. 2015). CFA was conducted for each age group, as well as for the full sample of 6–16-year-olds evaluated on the original version.

Multiple indices were employed to assess model fit (Gignac 2007; Kline 2016). The chi-square (χ^2), root mean square error of approximation (RMSEA), and standardized root mean square residual (SRMR) were used to evaluate the absolute fit of the model. RMSEA and SRMR values of 0.06 or less are indicative of a good model fit (Hu and Bentler 1999), whereas an RMSEA value of 0.10 or higher is typically considered evidence of a poor fit (MacCallum et al. 1996). However, the chi-square statistic is known to be highly sensitive to large sample sizes and may overestimate model misfits (Kline 2016). Relative model fit was assessed using the comparative fit index (CFI) and the Tucker–Lewis index (TLI). Values of 0.95 or higher are generally considered to reflect a good fit to the data (Hu and Bentler 1999). Based on the model, internal consistency estimates of reliability were calculated. In Addition, correlations were calculated between raw subtest scores and the children’s ages as a measure of the validity, with positive correlations expected.

All analyses were conducted using R statistical software (R Core Team 2021). Descriptive statistics and reliability analyses were performed using the *psych* package (Revelle 2023). The 2×2 factorial analysis was performed using the *factorial2x2* package (Leifer and Troendle 2019), while confirmatory factor analyses were conducted using the *lavaan* (Rosseel 2012) and the *semTools* packages (Jorgensen et al. 2025).

3. Results

Table 1 summarizes the descriptive statistics of each of the WISC-V-ID subtests for the total sample ($N = 221$). When comparing the maximum test scores achievable by children and the mean correct scores, it was evident that, on average, nearly half of the items were answered correctly by the children. The FW ($M = 21.76$, $SD = 4.80$), MR ($M = 18.77$, $SD = 4.20$), and PS ($M = 28.43$, $SD = 8.04$) subtests were the three highest-scoring subtests, with an average of around 56–64% of the items answered correctly by children. Conversely, the CD ($M = 47.22$, $SD = 18.29$), CO ($M = 15.63$, $SD = 5.90$), and IN ($M = 16.63$, $SD = 4.74$) subtests had the three lowest scores, with an average of around 40–44% of the items answered correctly by children.

Table 2 summarizes the item discrimination and difficulty indexes for each subtest of the WISC-V-ID. The correlation between items and total scores varied in strength across the subscales, with the average item discrimination index ranging from 0.34 to 0.47. For dichotomous subtests, the average item difficulty index ranged from 0.49 to 0.64. For polytomous subtests scored on a 0–2 scale (SI, VO, PS, and CO), the average item difficulty ranged from 0.87 to 1.09. In addition, the average item difficulty for the BD subtest, which is scored on a 0–7 scale, was 2.17. However, a total of eight items exhibited negative item discrimination indexes, that is, four items in the MR subtest, two in the FW subtest, one in the PC subtest, and one in the AR subtest. Seven out of these eight items had a difficulty level categorized as hard ($p < 0.30$), while only one item from the AR subtest was categorized as easy ($p = 0.80$) (this criterion was developed according to Cohen and Swerdlik (2009)). The item difficulty analysis was performed to reorder the items in the WISC-V-ID according to their difficulty levels.

Table 1. Descriptive data of the WISC-V-ID subtest scores.

Subtest	Maximum Score of Test	Correct Answer		
		Range	Mean	SD
Block Design (BD)	58	5–56	28.38	10.63
Similarities (SI)	46	2–41	21.82	6.90
Matrix Reasoning (MR)	32	3–28	18.77	4.20
Digit Span (DS)	54	4–38	24.48	6.45
Coding (CD)	117	13–97	47.22	18.29
Vocabulary (VC)	54	4–47	25.48	10.48
Figure Weights (FW)	34	3–31	21.76	4.80
Visual Puzzles (VP)	29	2–26	15.69	4.76
Picture Span (PS)	49	1–46	28.43	8.04
Symbol Search (SS)	60	3–60	28.28	9.32
Information (IN)	31	1–23	16.63	4.74
Picture Concepts (PC)	27	1–21	13.22	3.98
Letter-Number Sequencing (LN)	30	3–24	15.71	4.12
Cancellation (CA)	128	15–118	61.12	21.13
Comprehension (CO)	38	2–34	15.63	5.90
Arithmetic (AR)	34	4–29	17.77	5.71

Note: The results based on the total sample of 221 children aged 6–16.

Table 2. Item discrimination and difficulty for the WISC-V-ID subtest scores.

Subtest	Range of Score	Item Discrimination			Item Difficulty		
		Range	Mean	SD	Range	Mean	SD
Block Design (BD)	0–7	0.11–0.75	0.45	0.22	0.13–3.69	2.17	1.12
Similarities (SI)	0–2	0.25–0.68	0.43	0.13	0.05–1.93	0.94	0.70
Matrix Reasoning (MR)	0–1	−0.12–0.63	0.34	0.20	0.06–1.00	0.59	0.32
Digit Span (DS)	0–1	0.07–0.56	0.37	0.12	0.00–0.99	0.49	0.38
Coding (CD)		Not calculated					
Vocabulary (VC)	0–2	0.08–0.73	0.47	0.17	0.22–1.91	0.87	0.40
Figure Weights (FW)	0–1	−0.08–0.65	0.38	0.21	0.14–0.99	0.64	0.32
Visual Puzzles (VP)	0–1	0.02–0.62	0.39	0.18	0.06–0.99	0.54	0.33
Picture Span (PS)	0–2	0.11–0.6	0.44	0.12	0.01–1.94	1.09	0.63
Symbol Search (SS)		Not calculated					
Information (IN)	0–1	0.03–0.67	0.44	0.17	0.00–0.99	0.50	0.35
Picture Concepts (PC)	0–1	0.07–0.60	0.34	0.14	0.01–0.97	0.49	0.33
Letter-Number Sequencing (LN)	0–1	0.06–0.71	0.40	0.17	0.01–0.99	0.54	0.39
Cancellation (CA)		Not calculated					
Comprehension (CO)	0–2	0.14–0.63	0.45	0.14	0.01–1.95	0.82	0.59
Arithmetic (AR)	0–1	−0.01–0.72	0.40	0.18	0.01–1.00	0.52	0.33

Note: The results based on the total sample of 221 children aged 6–16.

Internal consistency was estimated using Cronbach's α and McDonald's ω , based on the original item order, with all items administered and no discontinue rules applied. Estimates were calculated for the overall sample and separately for each age group, as detailed in Appendix A, Table A1. Across the full sample, Cronbach's α values ranged from 0.76 to 0.89, and McDonald's ω values ranged from 0.73 to 0.90. The MR subtest consistently demonstrated the lowest internal consistency across both indices and age groups. For children aged 6–9, reliability estimates were generally high, with coefficients ranging from 0.74 to 0.90 across most subtests. In the 10–12 age group, the MR and CO subtests yielded values below the commonly accepted threshold of 0.70, indicating a need for further analysis. Additionally, the FW subtest yielded a notably low value of $\omega = 0.42$. Among children aged 13–16, three subtests (MR, PC, and LN) produced α and

ω coefficients ranging from 0.61 to 0.63 for both, which fall below the generally accepted standard for internal consistency.

Table 3 shows the internal consistency of the subtests after the items were reordered and the discontinue rules applied. Internal consistency estimates based on Cronbach's α and McDonald's ω were high for the overall age group, ranging from 0.81 to 0.94. Across all age groups, Cronbach's α coefficients met acceptable standards, ranging from 0.72 to 0.91. The lowest value was observed for the CO subtest in the 10–12 age group. Similar results were found using McDonald's ω , which ranged from 0.70 to 0.93. However, the LN subtest in the 13–16 age group warrants attention, as the ω value was 0.47, falling below the commonly accepted reliability threshold.

Table 3. Cronbach's α , McDonald's ω , and standard error of measurement (SEM) by age group and overall sample following reordering and application of discontinue rules.

Subtest	N of Items	Cronbach's α by Age Group				McDonald's ω by Age Group			
		6–9 (n = 67)	10–12 (n = 66)	13–16 (n = 88)	Overall (n = 211)	6–9 (n = 67)	10–12 (n = 66)	13–16 (n = 88)	Overall (n = 211)
BD	13	0.81 (4.01)	0.78 (4.85)	0.74 (4.74)	0.81 (4.79)	0.85 (3.56)	0.84 (4.14)	0.84 (3.72)	0.87 (3.97)
SI	23	0.88 (2.51)	0.77 (2.44)	0.80 (2.56)	0.86 (2.63)	0.90 (2.29)	0.76 (2.49)	0.81 (2.50)	0.88 (2.51)
MR	32	0.89 (1.60)	0.85 (1.66)	0.88 (1.22)	0.88 (1.66)	0.90 (1.52)	0.86 (1.60)	0.80 (1.57)	0.89 (1.59)
DS	54	0.90 (2.01)	0.82 (1.94)	0.80 (2.13)	0.89 (2.11)	0.91 (1.91)	0.84 (1.83)	0.80 (2.13)	0.90 (2.06)
CD		Not Calculated							
VC	29	0.84 (2.35)	0.89 (7.63)	0.86 (3.36)	0.93 (3.16)	0.85 (2.28)	0.90 (7.28)	0.88 (3.11)	0.94 (2.89)
FW	34	0.90 (1.57)	0.91 (4.38)	0.88 (1.53)	0.91 (1.67)	0.92 (1.40)	0.93 (3.86)	0.89 (1.47)	0.93 (1.5)
VP	29	0.85 (1.62)	0.91 (3.84)	0.88 (1.68)	0.89 (1.71)	0.86 (1.56)	0.91 (3.84)	0.90 (1.53)	0.91 (1.59)
PS	26	0.88 (2.84)	0.84 (2.85)	0.83 (2.90)	0.88 (2.98)	0.89 (2.72)	0.86 (2.67)	0.84 (2.81)	0.90 (2.71)
SS		Not Calculated							
IN	31	0.85 (1.40)	0.85 (1.37)	0.80 (1.54)	0.90 (1.57)	0.87 (1.30)	0.86 (1.32)	0.84 (1.37)	0.91 (1.37)
PC	27	0.85 (1.54)	0.84 (1.62)	0.79 (1.61)	0.86 (1.62)	0.86 (1.49)	0.85 (1.57)	0.80 (1.58)	0.87 (1.55)
LN	30	0.90 (1.55)	0.83 (1.40)	0.72 (1.46)	0.87 (1.55)	0.91 (1.47)	0.82 (1.44)	0.47 (2.02)	0.88 (1.53)
CA		Not Calculated							
CO	19	0.77 (2.00)	0.72 (2.35)	0.80 (2.46)	0.85 (2.35)	0.79 (1.91)	0.70 (2.43)	0.82 (2.33)	0.85 (2.35)
AR	34	0.90 (1.68)	0.83 (1.84)	0.82 (1.89)	0.90 (1.88)	0.89 (1.77)	0.85 (1.73)	0.84 (1.78)	0.91 (1.79)

Note: Bold = internal consistency; regular = standard error of measurement. BD = Block Design; SI = Similarities; MR = Matrix Reasoning; DS = Digit Span; CD = Coding; VC = Vocabulary; FW = Figure Weights; VP = Visual Puzzles; PS = Picture Span; SS = Symbol Search; IN = Information; PC = Picture Concepts; LN = Letter–Number Sequencing; CA = Cancellation; CO = Comprehension; AR = Arithmetic.

A 2×2 factorial analysis examined the effects of item ordering (Original vs. Reordered) and discontinue rule application (Without vs. With) on Cronbach's α and mean scores. The analysis showed no effect of item ordering on Cronbach's α ($F(1, 12) = 0.005$, $p = 0.94$), but applying discontinue rules significantly increased Cronbach's α ($F(1, 12) = 20.53$, $p < 0.001$; $M = 0.88$ [$SD = 0.03$] vs. 0.84 [$SD = 0.04$]). Meanwhile, a follow-up analysis of mean subtest scores found that the application of the discontinue rules led to lower scores ($F(1, 12) = 35.38$, $p < 0.001$; $M = 19.38$ vs. 20.29). Interestingly, the mean score for the group with reordered items and applied to discontinue rules was slightly higher than those for the group with the original item ordering and applied to discontinue rules ($M = 19.38$ vs. 19.20), though this difference was not statistically significant ($t(12) = -1.00$, $d = -0.03$).

Table 4 shows the goodness-of-fit indices for the confirmatory factor analyses of Model 1 and Model 2. Model 1 is a second-order five-factor model, while Model 2 retains the same structure but allows the AR subtest to cross-load onto both Gc and Gf. The results indicate that Model 1 did not meet the recommended fit criteria (Hu and Bentler 1999): RMSEA values exceeded the suggested cutoff point of 0.06 in all age groups and in the overall sample. In contrast, Model 2 showed a slight improvement in the fit, particularly

for the overall age sample. This model demonstrated acceptable fit indices (CFI = 0.96, TLI = 0.95, SRMR = 0.04), although the RMSEA remained marginally above the recommended cutoff at 0.07. These findings suggest that the model fit improves when the AR subtest is permitted to load on both Gf and Gc, indicating that AR may tap into multiple cognitive domains. Notably, among the age groups, the 6–9 age group showed a marginal fit to the model (CFI = 0.95, TLI = 0.94, RMSEA = 0.07, SRMR = 0.06), while the other age groups exhibited lower fit indices.

Table 4. Goodness of fit statistic for confirmatory factor analysis.

Age Group	Goodness of Fit Index						
	χ^2	df	CFI	TLI	RMSEA	SMSR	AIC
Model 1							
6–9	141.80	99	0.94	0.92	0.08	0.06	6239
10–12	142.31	99	0.92	0.90	0.08	0.08	6168
13–16	179.23	99	0.87	0.84	0.10	0.09	8216
All Ages	240.10	99	0.95	0.94	0.08	0.04	20,944
Model 2							
6–9	128.38	97	0.95	0.94	0.07	0.06	6230
10–12	130.72	97	0.93	0.92	0.07	0.08	6160
13–16	157.98	97	0.90	0.88	0.08	0.09	8198
All Ages	210.38	97	0.96	0.95	0.07	0.04	20,919

Additionally, Figure 1 presents the standardized factor loadings for the five-factor model of the WISC-V in the overall sample. The loadings from the first-order factors to the observed variables were relatively high, ranging from 0.41 to 0.92. Notably, the AR subtest has a very low loading on Gf (0.10), compared to higher loadings on Gc (0.41) and Gwm (0.42). This pattern contrasts with findings from the US standardization sample, where AR had its lowest loading on Gc (0.16; see Wechsler 2014a). At the second-order level, the model produced an overestimated loading of Gf on the general factor (g), of 1.03, exceeding the theoretical maximum of 1.00, which may indicate an estimation issue or model misfit. This finding is consistent with similar results reported in prior research (Wechsler 2014a).

The reliability coefficients for the composite scores are presented in Table 5. For the overall sample, all composite scores showed reliability above 0.80, indicating good consistency to support individual-level decisions. The general intelligence factor (g) demonstrated excellent reliability at 0.96. However, examining the age groups more closely reveals that a few composite scores are cause for concern. For instance, the Gf composite had reliability values of 0.65 for children aged 10–12 and 0.69 for those aged 13–16. These reliability coefficients were below the commonly accepted threshold. The Gs composite score for the 13–16 age group was particularly low at 0.59.

Additionally, Table 6 demonstrates that age was significantly and positively correlated with all subtests of the WISC-V-ID. Overall, the correlations ranged from 0.44 to 0.76. Specifically, age showed a high correlation with all Verbal Comprehension subtests (SI, VC, IN, and CO) and some Processing Speed subtests (CD and CA).

Table 5. Reliability estimates for WISC-V-ID composite scores.

Composite	Age Group			
	6–9	10–12	13–16	Overall
Verbal Comprehension (Gc)	0.90	0.86	0.88	0.93
Visual Spatial (Gv)	0.79	0.83	0.74	0.84

Table 5. Cont.

Composite	Age Group			
	6–9	10–12	13–16	Overall
Fluid Reasoning (Gf)	0.73	0.65	0.69	0.80
Working Memory (Gwm)	0.90	0.82	0.78	0.91
Processing Speed (Gs)	0.79	0.81	0.59	0.89
Full Scale IQ	0.93	0.92	0.92	0.96

Note: Composite score based on model 1.

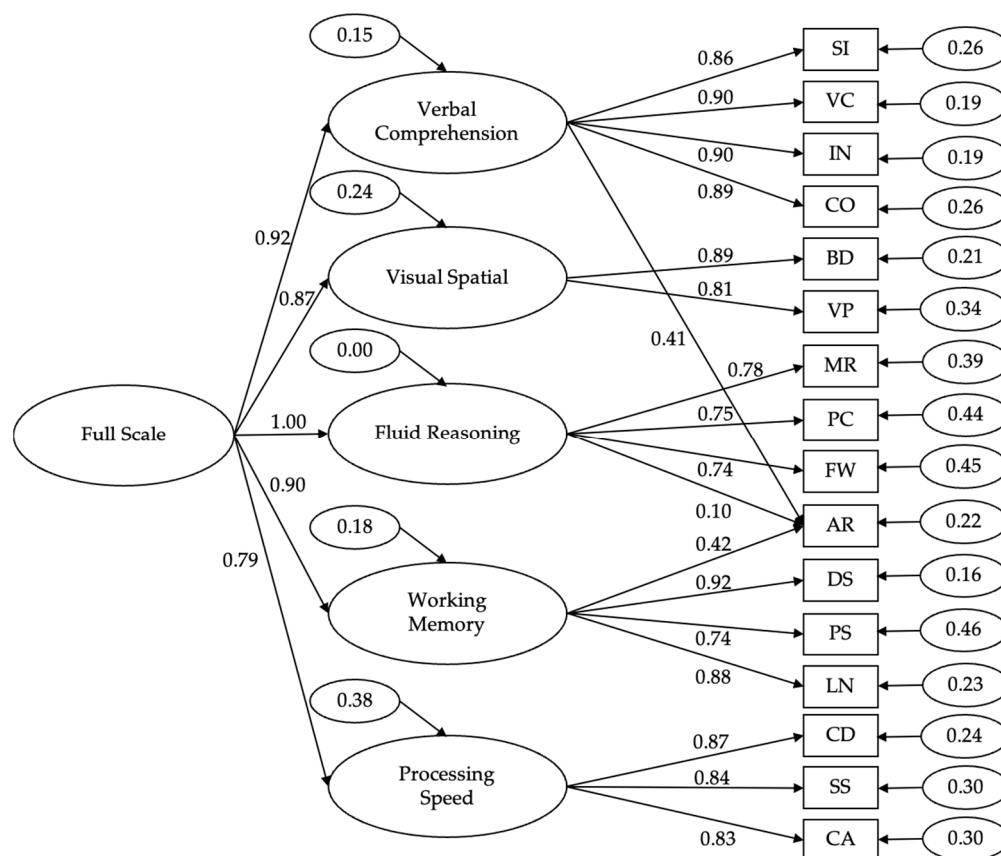


Figure 1. Five-factor second-order model for the WISC-V-ID. Note: Bold = internal consistency; regular = standard error of measurement. BD = Block Design; SI = Similarities; MR = Matrix Reasoning; DS = Digit Span; CD = Coding; VC = Vocabulary; FW = Figure Weights; VP = Visual Puzzles; PS = Picture Span; SS = Symbol Search; IN = Information; PC = Picture Concepts; LN = Letter–Number Sequencing; CA = Cancellation; CO = Comprehension; AR = Arithmetic.

Table 6. Correlations between the WISC-V-ID subtest scores and age.

Subtest	BD	SI	MR	DS	CD	VC	FW	VP	PS	SS	IN	PC	LN	CA	CO	AR
Age *	0.53	0.61	0.55	0.65	0.76	0.71	0.46	0.44	0.53	0.58	0.72	0.53	0.64	0.72	0.68	0.67

Note: * All correlations were statistically significant ($p < 0.001$). Note: Bold = internal consistency; regular = standard error of measurement. BD = Block Design; SI = Similarities; MR = Matrix Reasoning; DS = Digit Span; CD = Coding; VC = Vocabulary; FW = Figure Weights; VP = Visual Puzzles; PS = Picture Span; SS = Symbol Search; IN = Information; PC = Picture Concepts; LN = Letter–Number Sequencing; CA = Cancellation; CO = Comprehension; AR = Arithmetic.

4. Discussion

This study aimed to describe the cultural adaptation and translation process of the WISC-V and evaluate the psychometric properties of the WISC-V-ID in an Indonesian

sample of 221 children, focusing on item difficulty, item discrimination, and reliability for each subtest. Overall, the results provided evidence of a successful adaptation process.

Specifically, this study aimed to reorder item difficulty based on the observed performance of Indonesian children aged 6 to 16. Findings indicated that item difficulty order in some of the WISC-V-ID subtests was similar to the US-English version ([Wechsler 2014a](#)), while others differed. This difference excludes the Processing Speed subtests (CD, SS, and CA) due to their time-based format, which did not require item-level adaptations. Following a previous study, the item reordering in the WISC-V-ID for this study was based on two factors: the conceptual framework of item sequencing in the US-English version and the observed item difficulty in this pilot study ([Dang et al. 2011](#)). For instance, the item order remained unchanged in the DS, PS, and LN subtests. These subtests assess working memory, where items are sequenced based on the increasing number of stimuli children need to remain active within working memory. Therefore, no order changes were made. Similarly, the BD subtest also maintained its original order. Data showed that BD's item difficulty increased as the number of blocks used increased and the complexity of the stimuli went up. Additionally, minor adjustments were made to the item order in the MR, FW, VP, PC, and AR subtests, where on average, items were reordered only one or two positions up or down.

In contrast, subtests measuring the Verbal Comprehension Index underwent the most significant changes in item order, content, and difficulty. This was especially the case for the VC subtest. As previously mentioned during the adaptation process, approximately nine words originally classified as C1–C2 level in English were adapted to simpler meanings in Indonesian. This adaptation was necessary, in part, due to the limitations of Indonesian vocabulary in expressing complex concepts often found in global communication ([Paauw 2009](#)). For example, the item “frugal”, which relates to the concept of “prudently saving or sparing”, was translated into “*hemat*” in Indonesian. The word is a commonly used word that closely aligns with the meaning of “economical”, but there is no more complex term in Indonesian that fully captures the connotation of “frugal”. The item moved from position #26 to #13. The data revealed that most of these words originally classified as difficult became moderately difficult after adaptation. This suggests the chosen replacements were more commonly used by some of the children. In the SI subtest, items generally moved one to five positions up or down. The item asking about the similarity between “sour” and “salty” (likely easier for Indonesian children) moved from position #10 to #5. This suggests that the concept of taste might be easier for children to grasp because it relates to food, a concept that is more culturally important. Meanwhile, the item asking about the relation between a “shirt” and “shoes” moved from position #4 to #10, thus suggesting a higher difficulty level. This shift might be related to cultural clothing norms in Indonesia, where pants might be more common than shoes as paired clothing items for children.

The CO subtest required small changes in item order. However, an interesting finding was observed within this subtest. Two proverb items demonstrated differing patterns of difficulty. One item (#11, which addresses personal responsibility in learning) was more difficult for Indonesian children than for US children and was moved to Item #16. In contrast, Item #15 (which involves how to handle difficult situations) was found to be easier for Indonesian children and was moved to an earlier position, to Item #12. The accuracy in interpreting proverbs may be influenced by contextual understanding, abstract reasoning, metaphorical thinking ([Nippold et al. 1998](#)), and familiarity with the proverb's topic or phrasing in the Indonesian context ([Maulana 2020](#)). In this case, the question in Item #15 may appear to have been more culturally familiar to Indonesian children than the one in Item #11. In addition, the IN subtest also had some interesting changes in item order. One notable example is the item about an ancient limestone statue located in Egypt. This item

moved from position #20 to #15, suggesting it became easier for Indonesian children. The possible reason for this is that familiarity with statues through stories or Islamic cultural references might have played a role.

Reporting of both Cronbach's α and McDonald's ω reliability coefficients has become increasingly important, particularly for newly developed tests including items with a wide range of factor loadings, both high and low (Dunn et al. 2014; Hayes and Coutts 2020). The analysis revealed that internal consistency estimates obtained using Cronbach's α and McDonald's ω were generally comparable across most subtests and age groups. Differences ranged from 0.01 to 0.04 points for the overall sample. Consistent with several previous studies, α tended to produce slightly lower values than ω . This reflects the conceptual distinction whereby α represents a lower bound of reliability, while generally ω provides a more accurate and typically higher estimate (Hayes and Coutts 2020; Malkewitz et al. 2023; Viladrich et al. 2017). It is important to understand that these estimates are critical because reliability directly influences the calculation of the standard error of measurement, which in turn affects the interpretation of results and decision-making processes (Viladrich et al. 2017).

Based on the overall sample, this study showed that the reliability of all WISC-V-ID subtests was satisfactory, with 11 out of 13 subtests exhibiting coefficients exceeding 0.80 (based on Cronbach's α or McDonald's ω). The DS and VC subtests showed the highest reliability estimates ($\omega = 0.90$). However, the MR and PC subtests, which assess fluid reasoning using non-verbal stimuli (identifying relationships among objects), had slightly lower reliability coefficients, ranging from Cronbach's α or McDonald's $\omega = 0.73$ to 0.79. Similar results were previously found when adapting the Wechsler Preschool and Primary Scale of Intelligence III (WPPSI-III) for a low-income setting, where BD, MR, and PC were the three subtests with the lowest reliability (Rasheed et al. 2018). The slightly lower reliability of these subtests may be explained by their relative difficulty compared to other subtests (see Table 2). In particular, several difficult items displayed negative item discrimination, further impacting reliability. In the current study, all items were administered to the children without discontinue rules. Unlike verbal tests, where participants can say "I don't know," the multiple-choice format (for MR and PC) lacked a "don't know" option. This likely prompted children to guess on items they were unsure about, particularly on difficult items, negatively impacting the test's reliability (Paek 2015; Zimmerman and Williams 2003). Removing four items with negative discrimination indices from the MR subtest could increase reliability estimates from Cronbach's α from 0.76 to 0.80 and McDonald's ω from 0.73 to 0.77. In contrast, removing one item with a negative discrimination index from the PC subtest resulted in a slight increase in Cronbach's α from 0.78 to 0.79, with no change observed in McDonald's ω . However, an independent cross-validation sample is required to test these assumptions empirically, as conclusions cannot be drawn from the current analyses alone.

Internal consistency was also examined across three age groups: 6–9, 10–12, and 13–16 years. The results showed that the reliability values of nine out of thirteen subtests—including BD, SI, DS, FW, VP, PS, IN, and AR—had comparable reliability values to the overall Cronbach's α coefficients, indicating consistent test quality across age groups. However, concerns were raised regarding the Cronbach's α coefficients. An interesting finding was the substantial discrepancy whereby Cronbach's α was notably higher than McDonald's ω for the MR and FW subtests in the 10–12 age group. Several factors may explain this difference. Firstly, correlated measurement errors across items can inflate Cronbach's α estimates (Dunn et al. 2014; Hayes and Coutts 2020). Another possible explanation is the presence of skewed item distributions and multifactorial measurement structures, which can also affect reliability estimates (Malkewitz et al. 2023). However, this

phenomenon was observed only in this specific age group and subtests, indicating the need for further examination.

Further analysis revealed a clear improvement in reliability after reordering the items according to difficulty and applying discontinue rules. These rules aim to improve test efficiency while maintaining subtest reliability. Research suggests that stopping the test after two or three incorrect responses is appropriate for fixed-format tests (Basaraba et al. 2020). Our study found similar results: applying the discontinue rules outlined in the manual after two or three incorrect responses (Wechsler 2014a), along with reordering the items from easiest to most difficult, resulted in a good to excellent reliability range for the test. Based on these reliability coefficients, items with negative discrimination indices in the MR and PC subtests were retained, since applying discontinue rules improved the overall quality of the subtests (see Table 3). Retaining these items also ensured that the test contained a sufficient number of items that met the ceilings and floors that were established in the US version (Miller and McGill 2016). A factorial 2×2 analysis revealed that the application of discontinue rules significantly increased the reliability of the WISC-V-ID. Additionally, item reordering slightly increased the mean reliability compared to the original item order. These findings suggest that item reordering and the application of discontinue rules could be beneficial in future research. Applying discontinue rules is important for reducing testing time and participant fatigue. Any potential reduction in scores resulting from not administering all items can be addressed by developing norms based on the application of discontinue rules.

The manual reported that, based on confirmatory factor analysis (CFA), a secondary five-factor model incorporating 16 primary and secondary subtests was supported as the structural framework of the WISC-V. Furthermore, model fit indices improved significantly when the AR subtest loaded on both the Gc and Gf factors. This factor structure was also found to be consistent across different age groups (Wechsler 2014a). However, this pilot study could not fully replicate the secondary five-factor model in the data sample, for either the full sample or within age groups. This suggests that the WISC-V-ID may not fit well with the second-order five-factor model. The inadequacy of the second-order five-factor model has also been observed in various standardized samples, including the US version, and has been subject to debate by several researchers (Canivez et al. 2016; Lecerf and Canivez 2018; Pauls and Daseking 2021; Watkins et al. 2018). Most studies have found that a four-factor model provides a better fit than the five-factor model (Canivez et al. 2016) or a bifactor model with four group factors and one general factor (Lecerf and Canivez 2018; Pauls and Daseking 2021; Watkins et al. 2018). In these studies, Gf and Gv abilities were combined into a single factor.

Slightly different results were found when the AR subtest was loaded onto Gc and Gf: the model fit improved for the overall sample, but not for the individual age groups. Besides the questionable fit of the second-order five-factor structure, the poor model fit across age groups may be due to the small sample size in this study, which, per age group, was below the recommended minimum of 100 participants as suggested (Kline 2016). This finding suggests that the model structure based on age groups still requires further examination. In this model, AR was more strongly associated with Gc and Gwm than with Gf, which differs from the US version (Wechsler 2014a). The composite reliabilities demonstrated good consistency, supporting their use for individual-level decisions.

To further support the validity of the WISC-V-ID, the correlation between subtest scores and age was examined. The results showed that the correlations ranged from moderate to high. The positive correlations between subtest scores and age likely reflect the rapid development of cognitive functions observed from early childhood to early adolescence (Mous et al. 2017; Śmigasiewicz et al. 2021). The subtests showing the strongest correlations

with age were CD and CA, which measure processing speed. This positive correlation likely reflects the well-documented trend of processing speed abilities improving with age in children and young adults (Ferrer et al. 2013; Hale 1990; Kail 1991; Śmigajewicz et al. 2021). As children age, their processing speed abilities continue to mature, leading to faster information processing, reaction times, improved matching skills, and better mental rotation abilities. Other subtests that correlated highly with age are SI, VC, IN, and CO, which measure Verbal Comprehension. The high correlation might be driven by a strong education effect on crystallized intelligence tests (Schroeders et al. 2015) and the strong language development from the age of 6 onwards that helps to define the word correctly (Corthals 2010; Korkman et al. 1999). Moreover, the FW subtest, which measures fluid reasoning, was found to have a lower correlation with age compared to other subtests. Previous research suggests a non-linear pattern for fluid reasoning development, as opposed to crystallized intelligence, with rapid increases in childhood followed by a gradual rise in adolescence due to the uncertain development of how the ability to reason about relationships increases (Fry and Hale 2000). This non-linear pattern might explain the FW subtest's lower correlations with age in this study.

Overall, the results of this pilot study suggest that the WISC-V-ID could be feasible and reliable for use in Indonesia. However, several limitations in the current study suggest areas for future research. Firstly, this study primarily employed a classical test theory approach to examine test quality. Future research could benefit from applying item response theory (IRT) to provide more precise information on item characteristics, differential item functioning, and the effects of item reordering and the application of discontinue rules on test information—an approach that has been utilized in previous studies. Secondly, the pilot study involved a relatively small sample size. A larger sample is recommended to better identify the factor structure of the WISC-V-ID and to advance efforts to standardize and norm the test for a broader Indonesian population. Such studies should employ confirmatory factor analysis (CFA) based on the Cattell–Horn–Carroll (CHC) model (see Schneider and McGrew 2018), exploratory structural equation modeling (ESEM), bifactor models, and multi-group invariance testing to address the cross-cultural generalizability of the WISC-V factor structure.

Author Contributions: Conceptualization, W.Y., M.P.H.H., C.S., and R.P.C.K.; methodology, W.Y., M.P.H.H., C.S., F.A.A., and S.N.; formal analysis, W.Y.; investigation, W.Y.; resources, M.P.H.H. and C.S.; data curation, W.Y.; writing—original draft preparation, W.Y.; writing—review and editing, W.Y., M.P.H.H., C.S., F.A.A., S.N., and R.P.C.K.; supervision, M.P.H.H., C.S., and R.P.C.K.; project administration, W.Y.; funding acquisition, M.P.H.H. All authors have read and agreed to the published version of the manuscript.

Funding: Data collection was supported by the Center for Higher Education Funding and Assessment (PPAPT), the Ministry of Higher Education, Science, and Technology of the Republic of Indonesia, and the Indonesia Endowment Fund for Education (LPDP) under the 'Beasiswa Pendidikan Indonesia (BPI)' scholarship program.

Institutional Review Board Statement: The study was conducted according to the guidelines of the Declaration of Helsinki and approved by the Research Ethics Committee of Universitas Padjadjaran No. 505/UN6.KEP/EC/2023, approved on 17 April 2023.

Informed Consent Statement: Informed consent for the administration of the WISC-V-ID was obtained from parents of all children and all participants.

Data Availability Statement: The informed consent forms did not specifically ask for permission to store and share the data in a public repository. However, the fully anonymized data are available from the corresponding author upon reasonable request.

Acknowledgments: We thank Aliya Ulfa and Made Nisa Adriana Widiastini for coordinating data collection in Indonesia, Witriani for her consultation during the translation process, and Adiyo Roebianto for valuable discussions on data analysis. We also express our gratitude to Pearson Education South Asia Pte Ltd. for their approval.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A

Table A1. Cronbach's α and McDonald's ω coefficients of reliability and standard error of measurement (SEM) by age group and overall age in the original order and without application of discontinue rules.

Subtest	N of Items	Cronbach's α by Age Group				McDonald's ω by Age Group			
		6–9 (<i>n</i> = 67)	10–12 (<i>n</i> = 66)	13–16 (<i>n</i> = 88)	Overall (<i>n</i> = 211)	6–9 (<i>n</i> = 67)	10–12 (<i>n</i> = 66)	13–16 (<i>n</i> = 88)	Overall (<i>n</i> = 211)
BD	13	0.77 (4.12)	0.76 (4.89)	0.72 (4.88)	0.79 (4.87)	0.80 (3.84)	0.82 (4.23)	0.82 (3.91)	0.85 (4.12)
SI	23	0.84 (2.70)	0.72 (2.58)	0.76 (2.67)	0.84 (2.74)	0.86 (2.52)	0.73 (2.53)	0.77 (2.61)	0.86 (2.58)
MR	32	0.78 (2.11)	0.61 (2.03)	0.63 (1.92)	0.76 (2.04)	0.78 (2.11)	0.43 (2.46)	0.63 (1.92)	0.73 (2.18)
DS	54	0.89 (2.01)	0.81 (1.99)	0.79 (2.18)	0.89 (2.15)	0.90 (1.92)	0.81 (1.99)	0.79 (2.18)	0.90 (2.04)
CD		Not Calculated							
VC	29	0.79 (2.92)	0.83 (3.40)	0.82 (3.46)	0.89 (3.43)	0.81 (2.78)	0.84 (3.30)	0.83 (3.37)	0.90 (3.32)
FW	34	0.80 (2.03)	0.79 (1.91)	0.79 (1.79)	0.83 (1.96)	0.72 (2.40)	0.42 (3.18)	0.80 (1.74)	0.84 (1.91)
VP	29	0.77 (1.89)	0.86 (1.83)	0.80 (1.87)	0.85 (1.87)	0.79 (1.81)	0.87 (1.76)	0.82 (1.77)	0.86 (1.78)
PS	26	0.85 (2.99)	0.81 (2.93)	0.79 (2.97)	0.85 (3.07)	0.86 (2.89)	0.83 (2.77)	0.80 (2.90)	0.87 (2.90)
SS		Not Calculated							
IN	31	0.81 (1.54)	0.79 (1.52)	0.76 (1.62)	0.88 (1.64)	0.83 (1.46)	0.81 (1.45)	0.79 (1.51)	0.89 (1.57)
PC	27	0.79 (1.83)	0.74 (1.85)	0.61 (1.81)	0.78 (1.85)	0.80 (1.78)	0.75 (1.81)	0.61 (1.81)	0.79 (1.82)
LN	30	0.87 (1.63)	0.73 (1.43)	0.63 (1.48)	0.85 (1.58)	0.89 (1.50)	0.70 (1.51)	0.62 (1.50)	0.85 (1.60)
CA		Not Calculated							
CO	19	0.74 (2.12)	0.69 (2.39)	0.76 (2.58)	0.83 (2.43)	0.76 (2.04)	0.69 (2.39)	0.78 (2.47)	0.84 (2.36)
AR	34	0.88 (1.80)	0.79 (1.95)	0.75 (1.98)	0.88 (1.99)	0.89 (1.72)	0.80 (1.90)	0.77 (1.90)	0.89 (1.90)

Note: BD = Block Design; SI = Similarities; MR = Matrix Reasoning; DS = Digit Span; CD = Coding; VC = Vocabulary; FW = Figure Weights; VP = Visual Puzzles; PS = Picture Span; SS = Symbol Search; IN = Information; PC = Picture Concepts; LN = Letter–Number Sequencing; CA = Cancellation; CO = Comprehension; AR = Arithmetic.

Table A2. Cronbach's α for the WISC-V-ID: Effects of item order and discontinue rules application for sample age 6–16 years.

Subtest	N of Items	Original— Without Discontinue Rules	Original— With Discontinue Rules	Reordered— Without Discontinue Rules	Reordered— With Discontinue Rules
BD	13	0.79 (4.87)	0.81 (4.79)	0.79 (4.87)	0.81 (4.79)
SI	23	0.84 (2.74)	0.87 (2.65)	0.84 (2.74)	0.86 (2.63)
MR	32	0.76 (2.04)	0.88 (1.68)	0.76 (2.04)	0.88 (1.66)
DS	54	0.89 (2.15)	0.89 (2.15)	0.89 (2.15)	0.89 (2.11)
VC	29	0.89 (3.43)	0.93 (3.14)	0.89 (3.43)	0.93 (3.16)
FW	34	0.83 (1.96)	0.92 (1.66)	0.82 (1.96)	0.91 (1.67)
VP	29	0.85 (1.87)	0.89 (1.71)	0.85 (1.87)	0.89 (1.71)
PS	26	0.85 (3.07)	0.86 (3.02)	0.85 (3.07)	0.88 (2.98)
IN	31	0.88 (1.64)	0.90 (1.57)	0.88 (1.66)	0.90 (1.57)
PC	27	0.78 (1.85)	0.86 (1.64)	0.78 (1.85)	0.86 (1.62)
LN	30	0.85 (1.58)	0.88 (1.55)	0.85 (1.58)	0.87 (1.55)
CO	19	0.83 (2.43)	0.86 (2.31)	0.83 (2.43)	0.85 (2.35)
AR	34	0.88 (1.99)	0.90 (1.87)	0.88 (1.98)	0.90 (1.88)

Note: BD = Block Design; SI = Similarities; MR = Matrix Reasoning; DS = Digit Span; CD = Coding; VC = Vocabulary; FW = Figure Weights; VP = Visual Puzzles; PS = Picture Span; SS = Symbol Search; IN = Information; PC = Picture Concepts; LN = Letter–Number Sequencing; CA = Cancellation; CO = Comprehension; AR = Arithmetic.

Table A3. Mean and standard deviation for the WISC-V-ID: Effects of item order and discontinue rules application for sample age 6–16 years.

Subtest	Original— Without Discontinue Rules	Original— With Discontinue Rules	Reordered— Without Discontinue Rules	Reordered— With Discontinue Rules
BD	28.38 (10.63)	27.83(11.05)	28.38 (10.63)	27.83 (11.05)
SI	21.82 (6.90)	21.05 (7.33)	21.82 (6.90)	21.08 (7.24)
MR	18.77 (4.20)	16.87 (4.93)	18.77 (4.20)	16.78 (4.80)
DS	24.48 (6.45)	24.00 (6.53)	24.48 (6.45)	24.00 (6.53)
VC	25.48 (10.48)	22.04 (12.14)	25.48 (10.48)	24.36 (11.79)
FW	21.76 (4.80)	19.95 (5.83)	21.76 (4.80)	20.15 (5.68)
VP	15.69 (4.76)	14.76 (5.28)	15.69 (4.76)	14.77 (5.29)
PS	28.43 (8.04)	28.03 (8.17)	28.43 (8.04)	27.55 (8.62)
IN	16.63 (4.74)	15.87 (4.86)	16.63 (4.74)	16.07 (4.93)
PC	13.22 (3.98)	12.08 (4.32)	13.22 (3.98)	12.01 (4.31)
LN	15.71 (4.12)	15.35 (4.5)	15.71 (4.12)	15.35 (4.50)
CO	15.63 (5.90)	14.80 (6.09)	15.63 (5.90)	15.05 (6.06)
AR	17.77 (5.71)	16.96 (5.95)	17.77 (5.71)	17.00 (5.97)

Note: BD = Block Design; SI = Similarities; MR = Matrix Reasoning; DS = Digit Span; CD = Coding; VC = Vocabulary; FW = Figure Weights; VP = Visual Puzzles; PS = Picture Span; SS = Symbol Search; IN = Information; PC = Picture Concepts; LN = Letter–Number Sequencing; CA = Cancellation; CO = Comprehension; AR = Arithmetic.

References

- Basaraba, Deni L., Paul Yovanoff, Pooja Shivraj, and Leanne R. Ketterlin-Geller. 2020. Evaluating Stopping Rules for Fixed-Form Formative Assessments: Balancing Efficiency and Reliability. *Practical Assessment, Research & Evaluation* 25: 8. Available online: <https://scholarworks.umass.edu/pare/vol25/iss1/8> (accessed on 9 March 2024).
- Benson, Nicholas F., Randy G. Floyd, John H. Kranzler, Tanya L. Eckert, Sarah A. Fefer, and Grant B. Morgan. 2019. Test Use and Assessment Practices of School Psychologists in the United States: Findings from the 2017 National Survey. *Journal of School Psychology* 72: 29–48. [\[CrossRef\]](#)
- Billard, Catherine, Camille Jung, Arnold Munnich, Sahawanatou Gassama, Monique Touzin, Anne Mirassou, and Thiébaud-Noël Willig. 2021. External Validation of BMT-i Computerized Test Battery for Diagnosis of Learning Disabilities. *Frontiers in Pediatrics* 9: 733713. [\[CrossRef\]](#)
- BPS. 2022. *Analisis Profil Penduduk Indonesia Mendeskripsikan Peran Penduduk Dalam Pembangunan*. Badan Pusat Statistik. Available online: <https://www.bps.go.id/id/publication/2022/06/24/ea52f6a38d3913a5bc557c5f/analisis-profil-penduduk-indonesia.html> (accessed on 24 April 2024).
- Caemmerer, Jacqueline M., Timothy Z. Keith, and Matthew R. Reynolds. 2020. Beyond Individual Intelligence Tests: Application of Cattell-Horn-Carroll Theory. *Intelligence* 79: 101433. [\[CrossRef\]](#)
- Canivez, Gary L., Marley W. Watkins, and Stefan C. Dombrowski. 2016. Factor Structure of the Wechsler Intelligence Scale for Children–Fifth Edition: Exploratory Factor Analyses with the 16 Primary and Secondary Subtests. *Psychological Assessment* 28: 975–86. [\[CrossRef\]](#) [\[PubMed\]](#)
- Casaletto, Kaitlin B., Anya Umlauf, Jennifer Beaumont, Richard Gershon, Jerry Slotkin, Natacha Akshoomoff, and Robert K. Heaton. 2015. Demographically Corrected Normative Standards for the English Version of the NIH Toolbox Cognition Battery. *Journal of the International Neuropsychological Society* 21: 378–91. [\[CrossRef\]](#)
- Chen, Yuhe, Simeng Ma, Xiaoyu Yang, Dujuan Liu, and Jun Yang. 2023. Screening Children’s Intellectual Disabilities with Phonetic Features, Facial Phenotype and Craniofacial Variability Index. *Brain Sciences* 13: 155. [\[CrossRef\]](#)
- Cohen, Ronald Jay, and Mark Swerdlik. 2009. *Psychological Testing and Assessment: An Introduction to Tests and Measurement*, 7th ed. New York: McGraw Hill.
- Corthals, Paul. 2010. Nine- to Twelve-year Olds’ Metalinguistic Awareness of Homonymy. *International Journal of Language & Communication Disorders* 45: 121–28. [\[CrossRef\]](#)
- Dang, Hoang-Minh, Bahr Weiss, Amie Pollack, and Minh Cao Nguyen. 2011. Adaptation of the Wechsler Intelligence Scale for Children-IV (WISC-IV) for Vietnam. *Psychological Studies* 56: 387–92. [\[CrossRef\]](#)
- Daniel, Mark H. 1997. Intelligence Testing: Status and Trends. *American Psychologist* 52: 1038–45. [\[CrossRef\]](#)

- Deary, Ian J., Lars Penke, and Wendy Johnson. 2010. The Neuroscience of Human Intelligence Differences. *Nature Reviews Neuroscience* 11: 201–11. [CrossRef] [PubMed]
- Dunn, Thomas J., Thom Baguley, and Vivienne Brunsden. 2014. From Alpha to Omega: A Practical Solution to the Pervasive Problem of Internal Consistency Estimation. *British Journal of Psychology* 105: 399–412. [CrossRef]
- Ferrer, Emilio, Kirstie J. Whitaker, Joel S. Steele, Chloe T. Green, Carter Wendelken, and Silvia A. Bunge. 2013. White Matter Maturation Supports the Development of Reasoning Ability through Its Influence on Processing Speed. *Developmental Science* 16: 941–51. [CrossRef] [PubMed]
- Flynn, James R. 2007. *What Is Intelligence?: Beyond the Flynn Effect*, 1st ed. New York: Cambridge University Press.
- Fry, Astrid F., and Sandra Hale. 2000. Relationships among Processing Speed, Working Memory, and Fluid Intelligence in Children. *Biological Psychology* 54: 1–34. [CrossRef]
- Gignac, Gilles E. 2007. Multi-Factor Modeling in Individual Differences Research: Some Recommendations and Suggestions. *Personality and Individual Differences* 42: 37–48. [CrossRef]
- Hale, Sandra. 1990. A Global Developmental Trend in Cognitive Processing Speed. *Child Development* 61: 653. [CrossRef] [PubMed]
- Hayes, Andrew F., and Jacob J. Coutts. 2020. Use Omega Rather than Cronbach's Alpha for Estimating Reliability. But... *Communication Methods and Measures* 14: 1–24. [CrossRef]
- Hendriks, Marc P. H., Ruiter Selma, Mark Schittekatte, and Anne-Marie Bos. 2017. *WISC-V-NL Technische Handleiding*. Nederlandse Bewerking. Amsterdam: Pearson.
- Hernández, Ana, María D. Hidalgo, Ronald K. Hambleton, and Juana Gómez-Benito. 2020. International Test Commission Guidelines for Test Adaptation: A Criterion Checklist. *Psicothema* 32: 390–98. [CrossRef] [PubMed]
- Hu, Li-tze, and Peter M. Bentler. 1999. Cutoff Criteria for Fit Indexes in Covariance Structure Analysis: Conventional Criteria versus New Alternatives. *Structural Equation Modeling: A Multidisciplinary Journal* 6: 1–55. [CrossRef]
- International Test Commission. 2017. *The ITC Guidelines for Translating and Adapting Tests*, 2nd ed. Available online: https://www.intestcom.org/files/guideline_test_adaptation_2ed.pdf (accessed on 24 April 2024).
- Jensen, Arthur Robert. 1998. *The g Factor: The Science of Mental Ability*. Human Evolution, Behavior, and Intelligence. Westport: Praeger.
- Jorgensen, Terrence D., Sunthud Pornprasertmanit, Alexander M. Schoemann, and Yves Rosseel. 2025. semTools: Useful Tools for Structural Equation Modeling (version 0.5-7). Available online: <https://cran.r-project.org/web/packages/semTools/index.html> (accessed on 30 May 2025).
- Kail, Robert. 1991. Developmental Change in Speed of Processing during Childhood and Adolescence. *Psychological Bulletin* 109: 490–501. [CrossRef]
- Kaplan, Robert M., and Dennis P. Saccuzzo. 2012. *Psychological Testing: Principles, Applications, and Issues*. Toronto: Wadsworth Cengage Learning.
- Kaufman, Alan S., Susan Engi Raiford, and Diane L. Coalson. 2016. *Intelligent Testing with the WISC-V*. Hoboken: John Wiley & Sons.
- Kline, Rex B. 2016. *Principles and Practice of Structural Equation Modeling*, 4th ed. New York: Guilford publications.
- Korkman, Marit, Sarianna Barron-Linnankoski, and Pekka Lahti-Nuuttila. 1999. Effects of Age and Duration of Reading Instruction on the Development of Phonological Awareness, Rapid Naming, and Verbal Memory Span. *Developmental Neuropsychology* 16: 415–31. [CrossRef]
- Lecerf, Thierry, and Gary L. Canivez. 2018. Complementary Exploratory and Confirmatory Factor Analyses of the French WISC-V: Analyses Based on the Standardization Sample. *Psychological Assessment* 30: 793–808. [CrossRef]
- Leifer, Eric, and James Troendle. 2019. Factorial2x2: Design and Analysis of a 2×2 Factorial Trial. Available online: <https://cran.r-project.org/web/packages/factorial2x2/factorial2x2.pdf> (accessed on 30 May 2025).
- Leonard, Skyler, Elizabeth C. Loi, and Emily K Olsen. 2023. Pediatric Epilepsy Patients Demonstrate Stability and Variability in Verbal Learning and Memory Functions Over Time. *Journal of Pediatric Neuropsychology* 9: 127–40. [CrossRef]
- Lozano-Ruiz, Alvaro, Ahmed F Fasfous, Inmaculada Ibanez-Casas, Francisco Cruz-Quintana, Miguel Perez-Garcia, and María Nieves Pérez-Marfil. 2021. Cultural Bias in Intelligence Assessment Using a Culture-Free Test in Moroccan Children. *Archives of Clinical Neuropsychology* 36: 1502–10. [CrossRef] [PubMed]
- MacCallum, Robert C., Michael W. Browne, and Hazuki M. Sugawara. 1996. Power Analysis and Determination of Sample Size for Covariance Structure Modeling. *Psychological Methods* 1: 130–49. [CrossRef]
- Maki, Kathrin E., and Sarah R. Adams. 2019. A Current Landscape of Specific Learning Disability Identification: Training, Practices, and Implications. *Psychology in the Schools* 56: 18–31. [CrossRef]
- Malkewitz, Camila Paola, Philipp Schwall, Christian Meesters, and Jochen Hardt. 2023. Estimating Reliability: A Comparison of Cronbach's α , McDonald's Ω t and the Greatest Lower Bound. *Social Sciences & Humanities Open* 7: 100368. [CrossRef]
- Maulana, Andri. 2020. Cross Culture Understanding in EFL Teaching: An Analysis for Indonesia Context. *Linguists: Journal of Linguistics and Language Teaching* 6: 98. [CrossRef]
- McGill, Ryan J., Thomas J. Ward, and Gary L. Canivez. 2020. Use of Translated and Adapted Versions of the WISC-V: Caveat Emptor. *School Psychology International* 41: 276–94. [CrossRef]

- Miller, Daniel C., and Ryan J. McGill. 2016. Review of the WISC–V. In *Intelligent Testing with the WISC-V*. Edited by Alan S. Kaufman, Susan Engi Raiford and Diane L. Coalson. Hoboken: Wiley Online Library, pp. 645–62.
- Miller, Laura T., Emily C. Bumpus, and Scott Lee Graves. 2021. The State of Cognitive Assessment Training in School Psychology: An Analysis of Syllabi. *Contemporary School Psychology* 25: 149–56. [CrossRef]
- Mous, Sabine E., Nikita K. Schoemaker, Laura M. E. Blanken, Sandra Thijssen, Jan Van Der Ende, Tinca J. C. Polderman, Vincent W. V. Jaddoe, Albert Hofman, Frank C. Verhulst, Henning Tiemeier, and et al. 2017. The Association of Gender, Age, and Intelligence with Neuropsychological Functioning in Young Typically Developing Children: The Generation R Study. *Applied Neuropsychology: Child* 6: 22–40. [CrossRef] [PubMed]
- Nguyen, Christopher Minh, Shathani Rampa, Mathew Staios, T. Rune Nielsen, Busisiwe Zapparoli, Xinyi Emily Zhou, Lingani Mbakile-Mahlanza, Juliet Colon, Alexandra Hammond, Marc Hendriks, and et al. 2024. Neuropsychological Application of the International Test Commission Guidelines for Translation and Adapting of Tests. *Journal of the International Neuropsychological Society* 30: 621–34. [CrossRef]
- Nippold, Marilyn A., Susan L. Hegel, Linda D. Uhden, and Silvia Bustamante. 1998. Development of Proverb Comprehension in Adolescents: Implications for Instruction. *Journal of Children's Communication Development* 19: 49–55. [CrossRef]
- Nunnally, Jum, and Ira H. Bernstein. 1994. *Psychometric Theory*, 3rd ed. New York: McGraw-Hill.
- Paauw, Scott. 2009. One Land, One Nation, One Language: An Analysis of Indonesia's National Language Policy. *University of Rochester Working Papers in the Language Sciences* 5: 2–16.
- Paek, Insu. 2015. An Investigation of the Impact of Guessing on Coefficient α and Reliability. *Applied Psychological Measurement* 39: 264–77. [CrossRef]
- Pauls, Franz, and Monika Daseking. 2021. Revisiting the Factor Structure of the German WISC-V for Clinical Interpretability: An Exploratory and Confirmatory Approach on the 10 Primary Subtests. *Frontiers in Psychology* 12: 710929. [CrossRef]
- Pritchard, Alison E., Carly A. Nigro, Lisa A. Jacobson, and E. Mark Mahone. 2012. The Role of Neuropsychological Assessment in the Functional Outcomes of Children with ADHD. *Neuropsychology Review* 22: 54–68. [CrossRef]
- Priyamvada, Richa, Ranjan Rupesh, Soumya Sharma, and Suprakash Chaudhury. 2024. Efficacy of Cognitive Rehabilitation in Various Psychiatric Disorders. In *A Guide to Clinical Psychology: Therapies*. New York: Nova Medicine and Health, pp. 167–76.
- Rasheed, Muneera A., Sofia Pham, Uzma Memon, Saima Siyal, Jelena Obradović, and Aisha K. Yousafzai. 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: 197–207. [CrossRef]
- R Core Team. 2021. *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing. Available online: <https://www.R-project.org/> (accessed on 1 March 2024).
- Reeve, Charlie L., and Jennifer E. Charles. 2008. Survey of Opinions on the Primacy of g and Social Consequences of Ability Testing: A Comparison of Expert and Non-Expert Views. *Intelligence* 36: 681–88. [CrossRef]
- Revelle, William. 2023. Psych: Procedures for Psychological, Psychometric, and Personality Research. Available online: <https://cran.r-project.org/web/packages/psych/index.html> (accessed on 1 March 2024).
- Rodríguez-Cancino, Marcela, and Andrés Concha-Salgado. 2024. The Internal Structure of the WISC-V in Chile: Exploratory and Confirmatory Factor Analyses of the 15 Subtests. *Journal of Intelligence* 12: 105. [CrossRef]
- Rosseel, Yves. 2012. Lavaan: An R Package for Structural Equation Modeling. *Journal of Statistical Software* 48: 1–36. [CrossRef]
- Schneider, W. Joel, and Kevin S. McGrew. 2018. The Cattell–Horn–Carroll Theory of Cognitive Abilities. In *Contemporary Intellectual Assessment: Theories, Tests, and Issues*, 4th ed. New York: The Guilford Press, pp. 73–163.
- Schroeders, Ulrich, Stefan Schipolowski, and Oliver Wilhelm. 2015. Age-Related Changes in the Mean and Covariance Structure of Fluid and Crystallized Intelligence in Childhood and Adolescence. *Intelligence* 48: 15–29. [CrossRef]
- Suwartono, Christiany. 2016. Alat Tes Psikologi Konteks Indonesia: Tantangan Psikologi Di Era MEA. *Jurnal Psikologi Ulayat* 3: 1–6. [CrossRef]
- Suwartono, Christiany, Marc P. H. Hendriks, Lidia L. Hidajat, Magdalena S. Halim, and Roy P. C. Kessels. 2023. The Development of a Short Form of the Indonesian Version of the Wechsler Adult Intelligence Scale—Fourth Edition. *Journal of Intelligence* 11: 154. [CrossRef] [PubMed]
- Śmigasiewicz, Kamila, Mathieu Servant, Solène Ambrosi, Agnès Blaye, and Boris Burle. 2021. Speeding-up While Growing-up: Synchronous Functional Development of Motor and Non-Motor Processes across Childhood and Adolescence. *PLoS ONE* 16: e0255892. [CrossRef]
- Van de Vijver, Fons J. R., Lawrence G. Weiss, Donald H. Saklofske, Abigail Batty, and Aurelio Prifitera. 2019. A Cross-Cultural Analysis of the WISC-V. In *WISC-V. Clinical Use and Interpretation*. London: Academic Press, pp. 223–44. [CrossRef]
- Viladrich, Carme, Ariadna Angulo-Brunet, and Eduardo Doval. 2017. Un Viaje Alrededor de Alfa y Omega Para Estimar La Fiabilidad de Consistencia Interna. *Anales de Psicología* 33: 755. [CrossRef]
- Watkins, Marley W. 2017. The Reliability of Multidimensional Neuropsychological Measures: From Alpha to Omega. *The Clinical Neuropsychologist* 31: 1113–26. [CrossRef]

- Watkins, Marley W., Stefan C. Dombrowski, and Gary L. Canivez. 2018. Reliability and Factorial Validity of the Canadian Wechsler Intelligence Scale for Children—Fifth Edition. *International Journal of School & Educational Psychology* 6: 252–65. [\[CrossRef\]](#)
- Wechsler, David. 1949. *Wechsler Intelligence Scale for Children*. Wechsler Intelligence Scale for Children. San Antonio: Psychological Corporation.
- Wechsler, David. 1974. *Manual for the Wechsler Intelligence Scale for Children—Revised*. San Antonio: The Psychological Corporation.
- Wechsler, David. 2014a. *Wechsler Intelligence Scale for Children—Fifth Edition*. Bloomington: Pearson Clinical Assessment.
- Wechsler, David. 2014b. *Wechsler Intelligence Scale for Children—Fifth Edition: Canadian*. Toronto: Pearson Canada Assessment.
- Wechsler, David. 2015. *Escala de Inteligencia de Wechsler Para Niños-V*. Madrid: Pearson Educación.
- Wechsler, David. 2016a. *Wechsler Intelligence Scale for Children and Adolescents—Fifth Edition: Adaptation Française*. Paris: ECPA.
- Wechsler, David. 2016b. *Wechsler Intelligence Scale for Children, Fifth Edition: Australian and New Zealand Standardized Edition*. Sydney: Pearson.
- Wechsler, David. 2016c. *Wechsler Intelligence Scale for Children—Fifth UK Edition*. London: Harcourt Assessment.
- Wechsler, David. 2017. *Wechsler Intelligence Scale for Children—Fifth Edition (WISC-V). Technisches Manual. Deutsche Fassung von F. Petermann*. Frankfurt a. M.: Pearson.
- Wechsler, David. 2018. *Wechsler Intelligence Scale for Children—Fifth Edition. Technical and Interpretive Manual (Taiwan Version)*. Taipei: Chinese Behavioral Science Corporation.
- Weiss, Lawrence G., Donald H. Saklofske, James A. Holdnack, and Aurelio Prifitera. 2015. *WISC-V Assessment and Interpretation: Scientist-Practitioner Perspectives*. San Diego: Academic Press.
- Wilson, Christopher J., Stephen C. Bowden, Linda K. Byrne, Louis-Charles Vannier, Ana Hernandez, and Lawrence G. Weiss. 2023. Cross-National Generalizability of WISC-V and CHC Broad Ability Constructs across France, Spain, and the US. *Journal of Intelligence* 11: 159. [\[CrossRef\]](#)
- Zimmerman, Donald W., and Richard H. Williams. 2003. A New Look at the Influence of Guessing on the Reliability of Multiple-Choice Tests. *Applied Psychological Measurement* 27: 357–71. [\[CrossRef\]](#)

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.