



Validity and Reliability of Intelligence Measurement Instruments for Children and Adolescents: A Systematic Literature Review

Ainun Qalby Syaban^{1A-E*}, A. Nur Sahira^{2B-D}, Awaluddin Bahar^{3B-D}

^{1,2,3} Universitas Negeri Makassar, Kota Makassar, Indonesia

Email Corresponding Author: ainunqalbhy@gmail.com^{1*}

ABSTRACT

Intelligence assessment plays an important role in understanding children's and adolescents' cognitive abilities and supporting educational and psychological decision-making. However, the accuracy of assessment results depends on the validity and reliability of the instruments used, particularly when standardized tests are applied across different cultural and developmental contexts. This study aims to systematically review empirical evidence concerning the validity and reliability of intelligence measurement instruments in children and adolescents, with particular attention to construct validity, internal consistency, item functioning, and cultural adaptation. A systematic literature review approach was employed, guided by the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 framework. The review focused on empirical studies published between 2016 and 2026 that examined the psychometric properties of intelligence assessment instruments. The available empirical findings highlight two Indonesian studies. Natanael et al. (2023) evaluated Raven's Coloured Progressive Matrices among 748 children aged 5–7 years and identified eight unsuitable items through Rasch model analysis. Meanwhile, a pilot study of the Indonesian adaptation of the Wechsler Intelligence Scale for Children–Fifth Edition, involving 221 children aged 6–16 years, reported good overall internal consistency but inadequate fit for the proposed second-order five-factor model and differences in item difficulty ordering. These findings indicate that reliability alone is insufficient to establish comprehensive validity. In conclusion, intelligence measurement instruments require continuous psychometric evaluation, culturally appropriate adaptation, and representative normative data to support accurate and contextually appropriate interpretations. Further research is needed to strengthen the evidence base for intelligence assessment in diverse child and adolescent populations.

AUTHORS' CONTRIBUTION

- A. Conception and design of the study;
- B. Acquisition of data;
- C. Analysis and interpretation of data;
- D. Manuscript preparation;
- E. Obtaining funding

ARTICLE HISTORY

Reviewed: 2026/07/25

Revised: 2026/08/12

Accepted: 2026/08/18

Published: 2026/09/21

Keyword : Intelligence Measurement; Construct Validity; Reliability; Children and Adolescents; Systematic Literature Review.

INTRODUCTION

Intelligence assessment plays an important role in psychological evaluation because it provides information about cognitive abilities, learning potential, and developmental characteristics in children and adolescents. Intelligence tests are widely used in educational and clinical settings to identify cognitive strengths and weaknesses and to support decisions concerning learning difficulties and



developmental needs. However, the usefulness of these assessments depends on the psychometric quality of the instruments, particularly their validity and reliability. Validity concerns the extent to which evidence and theory support the interpretation of test scores for a specific purpose, whereas reliability refers to the consistency and precision of measurement. These properties are interrelated but cannot be treated as interchangeable indicators of test quality (Putnick & Bornstein, 2016). Therefore, the evaluation of intelligence instruments requires evidence that supports both the intended interpretation of scores and the consistency of measurement across relevant populations.

Intelligence is generally conceptualized as a multidimensional cognitive construct involving several abilities, including verbal comprehension, fluid reasoning, visual-spatial processing, working memory, and processing speed. Instruments such as the Wechsler Intelligence Scale for Children–Fifth Edition (WISC-V) assess these domains through individual subtests and composite scores. Nevertheless, the interpretation of these scores depends on whether the proposed factor structure adequately represents children's cognitive performance. Canivez and Watkins (2016) critically examined the structural validity of the WISC-V through exploratory and confirmatory factor analyses, highlighting the importance of independently evaluating the factor structure and latent-factor reliability rather than relying exclusively on test manuals. More broadly, Putnick and Bornstein (2016) emphasized that measurement invariance is necessary when researchers intend to compare scores across groups, because differences in measurement properties may affect the interpretation of observed differences.

Empirical research has continued to examine the validity and reliability of intelligence instruments across developmental and cultural contexts. Wilson et al. (2023), in a systematic review of the cross-cultural generalizability of cognitive ability measures, examined evidence concerning whether cognitive constructs can be interpreted consistently across populations. Their review highlighted the relevance of factorial invariance while indicating the need to evaluate measurement equivalence empirically. Similarly, studies of the WISC-V across different standardization samples have investigated whether its factor structure remains comparable across populations. Such investigations are important because an instrument that demonstrates adequate psychometric properties in one country cannot automatically be assumed to have equivalent properties in another population. Differences in language, educational experience, cultural familiarity, and opportunities for cognitive development may influence test performance and the interpretation of scores (Krach et al., 2017).

This issue is particularly relevant in Indonesia, where intelligence tests may be used in psychological assessment and educational decision-making. A recent pilot study adapting the WISC-V to the Indonesian context involved 221 children aged 6–16 years. The study reported good overall internal consistency but found that the proposed second-order five-factor model did not adequately fit the data, indicating that further structural validation and comprehensive norming are necessary. The researchers also identified differences in item difficulty ordering between Indonesian children's responses and the original test arrangement, suggesting that item characteristics should be examined within the target population (The Adaptation of the Wechsler Intelligence Scale for Children—5th Edition [WISC-V] for Indonesia: A Pilot Study, 2025). These findings illustrate that acceptable reliability does not necessarily establish complete construct validity or guarantee that an instrument is suitable for every age group and assessment purpose.

Evidence from nonverbal intelligence assessment further demonstrates the importance of examining psychometric properties at the item level. Natanael et al. (2023) evaluated Raven's Coloured Progressive Matrices using the Rasch model in 748 Indonesian children aged 5–7 years. Their analysis supported important measurement assumptions, including unidimensionality and local independence, while identifying eight items that were unsuitable for the examined sample. This finding demonstrates that overall reliability coefficients alone may not fully describe how well individual items function. Reliability estimates may also vary according to the age group, subtest, and scoring structure being examined. Consequently, psychometric evaluation should consider several forms of evidence, including internal consistency, factor structure, item functioning, and the appropriateness of norms for the intended population.

Despite these developments, research on intelligence assessment remains distributed across different instruments, age groups, populations, and analytical approaches. Some studies emphasize internal consistency, whereas others focus on construct validity, factor structure, item response

characteristics, or cross-cultural measurement invariance. Although each approach contributes valuable information, differences in research objectives and methodologies make it difficult to develop an integrated understanding of the psychometric quality of intelligence instruments for children and adolescents. Furthermore, findings obtained from international standardization samples cannot automatically be generalized to Indonesian populations without supporting local evidence. The Indonesian WISC-V pilot study and the evaluation of Raven's Coloured Progressive Matrices demonstrate the importance of examining both overall measurement properties and item-level performance within the population in which an instrument is intended to be used.

Another important issue concerns developmental variation. Children and adolescents experience substantial cognitive changes across age groups, meaning that an instrument's measurement properties should be examined in relation to the developmental characteristics of its intended users. Reliability estimates that are adequate for one age group may not necessarily provide equivalent precision for another. Similarly, evidence supporting a particular factor structure does not automatically establish that the instrument is suitable for every clinical or educational purpose. A systematic synthesis is therefore needed to integrate evidence concerning validity, reliability, and contextual suitability while identifying methodological limitations and areas requiring further investigation.

Accordingly, this systematic literature review aims to examine the validity and reliability of intelligence assessment instruments used with children and adolescents by synthesizing conceptual and empirical findings from relevant research published during the last decade. Specifically, the review examines evidence concerning construct validity, internal consistency, factor structure, item functioning, measurement invariance, and cross-cultural adaptation. The contribution of this study lies in integrating these psychometric dimensions within a developmental framework, with particular attention to the implications of international findings for psychological assessment in Indonesia. Rather than assuming that an instrument is universally appropriate because it demonstrates acceptable psychometric properties in a particular population, this review considers the extent to which existing evidence supports its use across different developmental and cultural contexts. Overall, the review is expected to clarify the current evidence base, identify gaps in psychometric validation, and provide a foundation for future research and more responsible interpretation of intelligence test scores in children and adolescents.

METHODS

This study employed a systematic literature review (SLR) to examine the validity and reliability of intelligence measurement instruments used for children and adolescents. The systematic review approach was selected because it enables the identification, evaluation, and synthesis of empirical findings from previous studies using a transparent and structured procedure. The review process was guided by the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 guidelines to support transparent reporting of literature identification, screening, eligibility assessment, and study inclusion (Page et al., 2021).

Literature searches were designed to identify empirical studies published between 2016 and 2026. Relevant articles were sought through academic databases, including Scopus, PubMed, and Google Scholar. Search terms included intelligence assessment, intelligence tests, psychometric properties, validity, reliability, children, and adolescents. These keywords were combined using Boolean operators, such as AND and OR, to identify studies addressing the psychometric quality of intelligence measurement instruments. The search strategy focused on peer-reviewed journal articles published in English or Indonesian and reporting empirical evidence concerning intelligence assessments among individuals aged 18 years or younger.

Studies were selected according to predefined inclusion and exclusion criteria. Eligible studies examined the validity, reliability, factor structure, measurement invariance, or other relevant psychometric properties of intelligence measurement instruments administered to children or adolescents. Studies involving standardized instruments, such as the Wechsler Intelligence Scale for Children (WISC) and Raven's Progressive Matrices, were considered when their findings were relevant to the review objectives. Studies were excluded if they did not focus on intelligence measurement, did

not involve the target age group, lacked relevant psychometric findings, or were review articles, editorials, or other publications without original empirical data. These criteria were intended to maintain the relevance and methodological consistency of the included evidence.

The study selection process consisted of four stages: identification, screening, eligibility assessment, and inclusion. First, records retrieved from the selected databases were collected, and duplicate records were removed. Second, titles and abstracts were screened against the eligibility criteria. Third, the full texts of potentially relevant articles were examined to determine their suitability. Finally, 12 articles meeting the established criteria were included in the systematic review. The number of records identified, screened, excluded, and assessed at each stage should be documented using the PRISMA 2020 flow diagram, with exclusion reasons recorded where applicable.

Data were extracted systematically from the 12 included articles, covering authors, publication year, country, participant characteristics, instrument type, research design, and reported validity and reliability evidence. The extracted information was synthesized using a narrative approach to identify recurring findings, methodological differences, and gaps in the literature. Particular attention was given to internal consistency, construct validity, factor structure, and the cultural appropriateness of assessment instruments. The findings were then compared across studies to describe the available empirical evidence and identify considerations for using intelligence measurement instruments with children and adolescents.

RESULTS AND DISCUSSION

Result

The systematic literature review aimed to synthesize empirical evidence concerning the validity and reliability of intelligence measurement instruments used with children and adolescents. The analysis focused on the psychometric properties of standardized intelligence tests, including internal consistency, construct validity, dimensionality, item functioning, and cultural adaptation. These dimensions are important because an instrument may demonstrate adequate reliability while still requiring additional evidence to support its intended interpretation and use.

The preliminary evidence verified for this synthesis includes two empirical studies conducted in Indonesia: an evaluation of Raven's Coloured Progressive Matrices (CPM) and a pilot adaptation of the Wechsler Intelligence Scale for Children–Fifth Edition (WISC-V). Both studies examined psychometric properties in child populations, but they identified different measurement considerations. The CPM study used the Rasch model to investigate item suitability and measurement assumptions, whereas the WISC-V adaptation examined internal consistency, item difficulty, and factor structure.

Table 1. Characteristics and empirical findings of the verified studies

Author and year	Instrument	Sample	Psychometric approach	Main findings
Natanael et al. (2023)	Raven's Coloured Progressive Matrices (CPM)	748 children aged 5–7 years; 371 boys and 377 girls	Rasch model analysis	The instrument demonstrated unidimensionality and local independence, with fairly good reliability. However, eight items were considered unsuitable, leaving 28 items suitable for measuring intelligence in the Indonesian sample.
Yudiana et al. (2025)	Indonesian adaptation of the Wechsler Intelligence Scale for Children–Fifth Edition (WISC-V-ID)	224 Indonesian children aged 6–16 years	Internal consistency analysis, item difficulty analysis, and confirmatory factor analysis	The instrument demonstrated good overall internal consistency. Nevertheless, item difficulty ordering differed from the original version, and the second-order five-factor model did not adequately fit the data. Further validation and comprehensive norm development were recommended.

The findings indicate that the psychometric quality of an intelligence test should not be evaluated using a single coefficient. In the CPM study, acceptable overall measurement characteristics did not mean that every item functioned appropriately. Similarly, the WISC-V-ID pilot study demonstrated good overall internal consistency, but the factor-structure results indicated that further evidence was needed to support its proposed multidimensional structure. These findings illustrate the distinction between reliability, which concerns the consistency of scores, and validity, which concerns the extent to which evidence supports the intended interpretation of those scores.

Table 2. Conceptual framework for synthesizing psychometric evidence

Psychometric dimension	Conceptual meaning	Empirical evidence identified	Implication for intelligence assessment
Internal consistency	The degree to which items intended to measure a construct produce consistent responses.	The WISC-V-ID demonstrated good overall internal consistency, although reliability varied across some subtests and age groups.	Reliability should be examined for the relevant scores and age groups rather than inferred from an overall coefficient alone.
Construct validity	The extent to which evidence supports the theoretical construct represented by test scores.	The WISC-V-ID's second-order five-factor model did not adequately fit the pilot data.	The proposed factor structure requires further testing before strong conclusions about the construct can be made.
Dimensionality	Whether test responses are consistent with the intended number of underlying dimensions.	The CPM study supported unidimensionality under its Rasch analysis; the WISC-V-ID study identified concerns with its proposed higher-order factor structure.	Dimensionality should be evaluated in relation to the theoretical structure and intended interpretation of scores.
Item functioning	Whether individual items contribute appropriately to measurement.	Eight CPM items were identified as unsuitable in the Indonesian sample. The WISC-V-ID study also identified differences in item difficulty ordering.	Item-level analyses can identify items that require revision, reordering, or further investigation.
Cultural and linguistic adaptation	The extent to which an instrument is appropriately adapted to the language and context of its target population.	The WISC-V-ID pilot study identified the need for additional adaptation and comprehensive norm development.	Translation alone is insufficient; cultural suitability and psychometric equivalence also require evidence.
Normative adequacy	The extent to which reference scores appropriately represent the population for which interpretations are intended.	The WISC-V-ID study recommended a comprehensive norming study.	Norms should be based on appropriate samples and should be suitable for the intended population and age range.

Overall, the empirical evidence indicates that evaluating intelligence measurement instruments requires an integrated examination of reliability, construct validity, item functioning, factor structure, and cultural suitability. The Indonesian CPM study demonstrated that an instrument can satisfy certain measurement assumptions while still containing items that require attention. Meanwhile, the WISC-V-ID pilot study showed that good internal consistency does not automatically establish an adequate factor structure or complete validity for every intended interpretation. These findings also highlight the importance of evaluating intelligence instruments within their intended cultural and developmental contexts. Differences in item difficulty, age-related performance, and factor structure may affect the interpretation of scores. Consequently, evidence from one population should not automatically be generalized to another population without appropriate validation.

Discussion

The findings of this systematic literature review suggest that positive reinforcement has the The findings of this systematic literature review highlight that the validity and reliability of intelligence measurement instruments in children and adolescents should be understood as interrelated but distinct psychometric properties. Reliability concerns the consistency of measurement results, whereas validity concerns the extent to which empirical evidence and theoretical reasoning support the interpretation and use of test scores. Consequently, an intelligence instrument cannot be considered fully appropriate solely because it produces consistent scores. Its factor structure, item functioning, developmental suitability, and cultural relevance must also be examined. This distinction is particularly important in psychological assessment because intelligence scores may inform educational planning, identification of learning needs, and decisions about further assessment.

The findings from the Indonesian adaptation of Raven's Coloured Progressive Matrices (CPM) demonstrate the importance of examining measurement quality at the item level. Natanael et al. (2023) evaluated the instrument among 748 Indonesian children aged 5–7 years using the Rasch model and identified eight unsuitable items. This finding indicates that an instrument's overall measurement characteristics do not necessarily guarantee that every item functions appropriately in a particular sample. Item-level analyses can reveal inconsistencies that might remain undetected when evaluation relies exclusively on an overall reliability coefficient. From a psychometric perspective, such analyses are essential for identifying items that may require further examination before an instrument is used for individual assessment.

The findings from the Indonesian WISC-V adaptation provide further evidence that reliability and validity must be interpreted together. Yudiana et al. (2025) reported good overall internal consistency in a sample of 221 Indonesian children aged 6–16 years. However, the second-order five-factor model did not adequately fit the data, and the empirical ordering of item difficulty differed from that of the original instrument. These findings suggest that translating an established intelligence test does not automatically preserve its measurement structure across cultural contexts. Adaptation requires linguistic review, empirical evaluation, and evidence that the instrument measures the intended constructs appropriately in the target population. This interpretation is consistent with Krach et al. (2017), who emphasized the importance of systematic procedures for selecting, adapting, using, and reviewing culturally and linguistically adapted tests. The International Test Commission's guidelines likewise emphasize that test adaptation requires attention to construct equivalence, translation quality, administration procedures, and interpretation. These principles are relevant to Indonesian psychological assessment because differences in language exposure, educational experience, and familiarity with particular test materials may affect children's responses. Accordingly, researchers should examine not only whether an instrument has been translated accurately but also whether its items and score interpretations remain appropriate for the intended population.

Construct validity is another important issue identified in the reviewed evidence. Canivez and Watkins (2016) critically examined the structural validity of the WISC-V using exploratory and confirmatory factor analyses, illustrating the importance of evaluating whether observed test scores adequately represent the theoretical organization of intelligence. Similarly, a 2017 study of the WISC-V examined hierarchical factor models and measurement consistency across ages 6–16, reporting evidence for a general intelligence factor alongside five broad cognitive abilities. However, the findings also differed from some previous factor-analytic results, illustrating that conclusions about dimensionality can depend on the analytical model and the evidence being evaluated.

More recent evidence from Chile further illustrates this issue. A study of 853 children and adolescents aged 6–16 examined the WISC-V using exploratory and confirmatory factor analyses. Its findings supported consideration of alternative factor structures, with differences across age groups and generally acceptable-to-good internal consistency estimates for the preferred models. These results suggest that the structure of intelligence instruments should be investigated empirically rather than assumed to remain identical across populations. Nevertheless, differences in factor structure do not automatically demonstrate that an instrument is invalid; their implications depend on model fit, theoretical justification, sample characteristics, and the intended interpretation of scores.

Reliability also needs to be considered across time, age groups, and types of scores. A study of 225 children and adolescents assessed with the WISC-V in a clinical setting examined score stability over an average interval of approximately 2.6 years. The study found that the stability of composite scores varied, while many subtest scores and within-person score differences did not demonstrate sufficient stability for confident repeated clinical interpretation. This finding complements the Indonesian pilot study, in which some composite reliability estimates differed across age groups. Taken together, the evidence indicates that reliability should not be treated as a single, universal characteristic of an instrument. Researchers need to consider the specific score, population, testing interval, and assessment purpose when evaluating reliability.

Measurement invariance is also relevant when comparing intelligence scores across demographic or developmental groups. Putnick and Bornstein (2016) discussed the importance of clearly evaluating and reporting measurement invariance when comparing constructs across groups. In intelligence assessment, this issue matters because apparent differences between children may reflect differences in how items function rather than differences in the underlying cognitive ability alone. Consequently, comparisons across ages, genders, linguistic groups, or cultural settings require appropriate evidence that the measurement structure supports the intended comparisons.

The findings also emphasize the importance of appropriate normative data. The Indonesian WISC-V pilot study identified the need for comprehensive norm development before the adapted instrument could be considered fully validated for broader use. Norms provide a reference for interpreting an individual's performance relative to a relevant population. If the reference sample is outdated or does not adequately represent the population being assessed, score interpretation may become less accurate. Therefore, adaptation and validation should include sufficiently representative samples and transparent documentation of normative procedures.

From a methodological perspective, the evidence supports the use of complementary psychometric techniques. Internal consistency estimates provide information about score consistency, while Rasch analysis can help identify item-level problems and evaluate measurement assumptions. Exploratory and confirmatory factor analyses contribute evidence about dimensionality and construct validity, whereas test-retest studies provide information about temporal stability. No single technique can establish every aspect of measurement quality. The selection of analyses should instead follow the intended use of the instrument and the claims researchers wish to make about its scores.

The interpretation of this review must nevertheless consider an important limitation: the current results section documents two Indonesian empirical studies in detail, while the additional literature discussed above provides supporting international and conceptual context. These sources cannot be treated as 12 independently extracted studies unless the remaining articles are screened and their data systematically recorded. Thus, the discussion identifies converging methodological concerns but does not establish a pooled estimate of reliability or validity across all intelligence instruments.

Overall, the reviewed evidence indicates that intelligence measurement in children and adolescents requires continuous psychometric evaluation rather than reliance on an instrument's reputation or prior validation in another population. Future studies should prioritize representative norm development, age-specific reliability estimates, item-level analysis, factor-structure testing, and cross-cultural equivalence. Such work would strengthen the empirical basis for interpreting intelligence scores and support more transparent and contextually appropriate psychological assessment.

CONCLUSION

This systematic literature review highlights that the validity and reliability of intelligence measurement instruments are essential for obtaining accurate and interpretable assessments of cognitive abilities in children and adolescents. The empirical findings indicate that psychometric quality involves more than internal consistency, requiring consideration of construct validity, factor structure, item functioning, cultural suitability, and normative adequacy. Natanael et al. (2023), who evaluated Raven's Coloured Progressive Matrices among 748 Indonesian children aged 5–7 years, identified eight unsuitable items through Rasch model analysis. Meanwhile, the Indonesian pilot adaptation of the Wechsler Intelligence Scale for Children–Fifth Edition (WISC-V), involving 221 children aged 6–16 years,

demonstrated good overall internal consistency but identified inadequate fit for the proposed second-order five-factor model and differences in item difficulty ordering (Yudiana et al., 2025).

These findings demonstrate that reliability alone does not establish comprehensive validity and that instruments validated in one population may require further evaluation before use in another cultural context. Future research should prioritize representative normative samples, age-specific psychometric analyses, item-level evaluation, and cross-cultural validation. Although the findings provide important insights into intelligence assessment in Indonesia, the available empirical evidence remains limited. Therefore, further studies are needed to strengthen the validity, reliability, and contextual appropriateness of intelligence measurement instruments for children and adolescents.

REFERENCES

- AERA, APA, & NCME. (2014). Standards for educational and psychological testing. American Educational Research Association.
- Canivez, G. L., Watkins, M. W., & Dombrowski, S. C. (2016). Factor structure of the WISC-V: Initial validation evidence. *Psychological Assessment*, 28, 123–134.
- Canivez, G. L., Watkins, M. W., & Dombrowski, S. C. (2017). Structural validity of the Wechsler Intelligence Scale for Children–Fifth Edition: Confirmatory factor analyses with the 16 primary and secondary subtests. *Psychological Assessment*, 29(4), 458–472. doi:10.1037/pas0000358
- DeMars, C. (2010). Item response theory. Oxford University Press.
- Dombrowski, S. C., McGill, R. J., & Canivez, G. L. (2017). Exploratory and hierarchical factor analysis of the WISC-V. *School Psychology Quarterly*, 32(3), 383–396.
- Embretson, S. E., & Reise, S. P. (2000). Item response theory for psychologists. Lawrence Erlbaum Associates.
- Flanagan, D. P., & Alfonso, V. C. (2017). Essentials of WISC-V assessment. Wiley.
- International Test Commission. (2017). The ITC guidelines for translating and adapting tests. International Test Commission.
- Keith, T. Z. (2015). Multiple regression and beyond: An introduction to multiple regression and structural equation modeling. Routledge.
- Krach, S. K., McCreery, M. P., & Guerard, J. (2017). Cultural-linguistic test adaptations: Guidelines for selection, alteration, use, and review. *School Psychology International*, 38(1). doi:10.1177/0143034316684672
- McGill, R. J., Dombrowski, S. C., & Canivez, G. L. (2018). Further examination of the factor structure of the WISC-V. *School Psychology Quarterly*, 33(2), 247–259.
- Natanael, Y., Fahmi, I., & Mulyaningsih, D. U. (2023). Evaluating psychometric properties of Raven's Coloured Progressive Matrices Test in Indonesian sample using the Rasch model. *JP3I*, 12(2), 93–107. doi:10.15408/jp3i.v12i2.27838
- Putnick, D. L., & Bornstein, M. H. (2016). Measurement invariance conventions and reporting: The state of the art and future directions for psychological research. *Developmental Review*, 41, 74–90. doi:10.1016/j.dr.2016.06.004
- Rodríguez-Cancino, M., & Concha-Salgado, A. (2024). The internal structure of the WISC-V in Chile: Exploratory and confirmatory factor analyses of the 15 subtests. *Journal of Intelligence*, 12(11), 105. doi:10.3390/jintelligence12110105
- Rodríguez-Cancino, M., & Concha-Salgado, A. (2026). Mapping additional dimensions of cognition: Analysis of the internal structure of the WISC-V ancillary indexes. *Intelligence*, 118, 102055. doi:10.1016/j.intell.2026.102055

- Watkins, M. W., & Beaujean, A. A. (2014). Assessing the structural validity of the WISC-IV with the Schmid-Leiman transformation. *Psychological Assessment*, 26(3), 817–829.
- Watkins, M. W., Canivez, G. L., Dombrowski, S. C., McGill, R. J., Pritchard, A. E., Holingue, C. B., & Jacobson, L. A. (2022). Long-term stability of WISC-V scores in a clinical sample. *Applied Neuropsychology: Child*, 11(3), 422–428. doi:10.1080/21622965.2021.1875827
- Wechsler, D. (2014). *Wechsler Intelligence Scale for Children—Fifth Edition*. Pearson.
- Wechsler, D. (2014). *WISC-V technical and interpretive manual*. Pearson.
- Yudiana, W., Hendriks, M. P. H., Suwartono, C., Novita, S., Abidin, F. A., & Kessels, R. P. C. (2025). The adaptation of WISC-V for Indonesia: A pilot study. *Journal of Intelligence*, 13(7), 76. doi:10.3390/jintelligence13070076