RESEARCH ARTICLE OPEN ACCESS

# Beyond Primary Assessment on the WISC-V: An Investigation of the Structural Validity of the Ancillary Scores

Ryan J. McGill[1] | Stefan C. Dombrowski[2] | Gary L. Canivez[3]

[1]Department of School Psychology and Counselor Education, William & Mary School of Education, Virginia, USA | [2]Department of Graduate Education, Leadership, and Counseling, Rider University, New Jersey, USA | [3]Department of Psychology, Eastern Illinois University, Illinois, USA

**Correspondence:** Ryan J. McGill (rmcgill@wm.edu)

**Received:** 9 September 2024 | **Revised:** 17 January 2025 | **Accepted:** 15 February 2025

**Keywords:** ancillary indexes | CFA | structural validity | WISC-V

## ABSTRACT

The present study examined the posited structure of the Wechsler Intelligence Scale for Children-Fifth Edition (WISC-V) ancillary index scores with normative sample participants aged 6-16 years ($N = 2200$) using a series of confirmatory factor analyzes (CFA) with maximum likelihood estimation. CFA results supported the retention of auditory working memory (AWM) but not quantitative reasoning (QR) as narrow dimensions in an extended WISC-V measurement model. Additional results from models explicating the structures for each of the posited ancillary composite-level indexes (nonverbal [NVI], general ability [GAI], cognitive processing [CPI]) provided support, in part, that these indexes represent global dimensions with differing degrees of generality. Though some of these scores may be used in the manner intended by the test publisher (e.g., comparing and contrasting performance on different composites, specific learning disability identification), provisional limitations for using the ancillary indexes as a focal point of clinical decision-making are discussed.

Despite long-standing calls for paradigm shift, the administration and interpretation of intelligence tests remains ubiquitous in school psychology training and practice (Goforth et al. 2021). In particular, the Weschler Intelligence Scale for Children, now in its fifth edition (WISC-V; Wechsler 2014a), is ubiquitously used for this purpose. In a recent survey of test use among practicing school psychologists, it was reported that the WISC-V was administered an average of 3.49 times per month, far outpacing usage rates for other competing measures (Benson et al. 2019). Despite the immediate embrace of the instrument by practitioners soon after its publication, independent reviews soon emerged raising important questions as to what the test measures (see Kaufman et al. 2016). Given the modifications made to the WISC-V interpretive structure, it bears considering the events leading to its validation and eventual publication to provide important context for the current study.

## 1 | An Instrument in Search of a Theory?

As noted by Wasserman and Kaufman (2016), Wechsler formulated and organized initial versions of the Wechsler Scales largely for pragmatic purposes rather than adherence to any theoretical orientation. This has led many in the assessment literature to regard the scales as being atheoretical (e.g., Freeman and Chen 2019), a perceived limitation in an era of instrumentation where formal theories of intelligence now serve as blueprints for the development of commercial ability measures (Kamphaus et al. 2018). This lack of a coherent theoretical alignment has plagued interpretation of prior editions (e.g., Reinecke et al. 1999; Saklofske et al. 2006). Nevertheless, the organization of tests and the underlying structure of the WISC has come to be regarded as a veritable theory unto itself, leading some scholars to refer to various WISC structures as representing so-called "Wechsler Theory" (e.g., McGill et al. 2020).

**Summary**

- Results indicate that the QRI should be interpreted with caution, if at all, in clinical practice as it was not able to be located in the present study.
- While the GAI and CPI appear to function reasonably as composite scores, they should not be interpreted as reflecting any psychological construct.
- The NVI score is not unidimensional.

Consonant with other IQ tests, it is noted in the Technical and Interpretive Manual (Wechsler 2014b) that one of the major goals for the development of the WISC-V was to update the theoretical foundations of the instrument based on structural intelligence and working memory models as well as information furnished from neurodevelopmental and neurocognitive research. As consistent with the documentation of its predecessor (i.e., WISC-IV [Wechsler 2003]), the WISC-V features several structural validity studies as influencing its development; however, the technical documentation regarding the WISC-V theoretical foundations do not reference any one theory as being persuasive (Williams et al. 2003)[1]. Based on the information reviewed, it was concluded that "These studies provide overwhelming evidence for a hierarchical model with general intelligence at the top and various related yet distinguishable broad abilities at the level beneath" (Wechsler 2014b, p. 23). In particular, it was suggested that there was consensus for separate Visual-Spatial and Fluid Reasoning dimensions among the broad abilities which would necessitate splitting the former Perceptual Reasoning Index from the WISC-IV to better represent that distinction as part of the WISC-V revision. As will be discussed below, this modification to the organization of the instrument did not occur in isolation.

## 1.1 | 2013 JPA Special Issue: What's Past Is Prologue?

Anticipated changes to the WISC-V structure were presaged in a 2013 special issue of the *Journal of Psychoeducational Assessment* (JPA) on Wechsler Theory and practice. Specifically, Weiss et al. (2013) re-examined the WISC-IV normative data using confirmatory factor analysis (CFA) and provided results supporting the retention of a new five-factor model for the instrument. The final validation model split Perceptual Reasoning into separate Visual-Spatial and Fluid Reasoning dimensions. However, this required the specification of an intermediary Inductive Reasoning factor mediating the influence of Picture Concepts and Matrix Reasoning on Fluid Reasoning, leaving Arithmetic as the sole measured variable directly influenced by that dimension. This alternative model was similar to one previously produced by Keith et al. (2006) for the WISC-IV that appeared to have inspired the specification of a series of ad hoc index-level scores aligning with that structure in subsequent WISC-IV interpretive guidebooks (e.g., Flanagan and Kaufman 2009). Though that model did not contain an Inductive Reasoning factor, it seemed to provide a slightly better global fit to the WISC-IV data than the model furnished by Weiss et al. (2013). Nevertheless, both studies indicated that some form of a five-factor model should be preferred over the publisher suggested four-factor model and that while "a possible fifth Wechsler factor has sometimes been cast as a contest between the Wechsler and CHC models of intelligence…adding a fluid reasoning factor to the Wechsler model has been a systematic research goal since 1997" (Weiss et al. 2013, pp. 128–129).

Whereas most of the commentaries contained in the special issue were largely supportive of this development, a critique was provided by Canivez and Kush (2013). In their points of contention, they noted that the target model contained numerous departures from desired simple structure (Thurstone 1954) via the specification of an intermediary factor as well as a host of cross-loadings of subtests on multiple latent factors. As but one example, portions of Matrix Reasoning were simultaneously explained by Inductive Reasoning, Verbal Comprehension, and Fluid Reasoning, albeit indirectly. If this model is accepted, it remains unclear how a clinician could disentangle the varying degrees of influence on Matrix Reasoning when attempting to interpret that measure in clinical practice. Further, these post hoc modifications to initial model estimation did not yield appreciable improvement in global model fit calling into question their empirical justification. Accordingly, Canivez and Kush (2013) suggested that these model re-specifications appeared to be "an attempt to improve the fit of the model to the data (and not the theory)" (p. 163).

More concerning, Canivez and Kush (2013) noted that the final validation model retained by Weiss et al. (2013) also contained a second-order standardized loading of 1.0 between *g* and Fluid Reasoning indicating that these dimensions are statistically isomorphic. Whereas path coefficients that equal unity are technically permissible in CFA[2], psychometric best practice, which favors parsimony, suggests that the redundant variable should be omitted as the model is likely overfit (Brown 2015). That same problematic loading was also observed in the five-factor model furnished by Keith et al. (2006), suggesting that there may be some underlying identification issues that preclude the identification of Fluid Reasoning on the WISC-IV. If true, this would cast doubt on attempts to advance a five-factor interpretive model for that instrument or future versions without substantial modifications to the composition of the test. Discussion featured in the JPA special issue was prescient as similar lines of criticism have been raised about the validation of the internal structure of the WISC-V since its publication.

## 1.2 | Organization and Structure of the WISC-V

The WISC-V is composed of 10 "primary" subtests (Vocabulary, Similarities [SI], Block Design [BD], Visual Puzzles [VP], Matrix Reasoning [MR], Figure Weights [FW], Digit Span [DS], Picture Span [PS], Coding [CD], and Symbol Search [SS]) that combine to form five primary index-level scores (Verbal Comprehension [VC], Fluid Reasoning [FR], Visual-Spatial [VS], Working Memory [WM], and Processing Speed [PS]) as well as an omnibus FSIQ composite[3]. Additionally, six "secondary" subtests (Information [IN], Comprehension [CO], Picture Concepts [PC], Arithmetic [AR], Letter-Number Sequencing [LN], and Cancellation [CA]) can be combined with selected primary subtests to form five ancillary index and composite scores. Whereas one of the primary or secondary subtests can be used

**TABLE 1** | Visual representation of the organization and scoring structure of the WISC-V.

| | Primary index and composites | | | | | | Ancillary index and composites | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | **FSIQ** | **VC** | **VS** | **FR** | **WM** | **PS** | **GAI** | **NVI** | **CPI** | **AWM** | **QR** |
| **Primary Subtests** | | | | | | | | | | | |
| Similarities | * | * | | | | | * | | | | |
| Vocabulary | * | * | | | | | * | | | | |
| Block design | * | | * | | | | * | * | | | |
| Visual puzzles | | | * | | | | | | | | |
| Matrix reasoning | * | | | * | | | * | * | | | |
| Figure weights | * | | | * | | | * | * | | | * |
| Digit span | * | | | | | | | | * | * | |
| Picture span | | | | | * | | | * | * | | |
| Coding | * | | | | | * | | * | * | | |
| Symbol search | | | | | | * | | | * | | |
| **Ancillary subtests** | | | | | | | | | | | |
| Information | | | | | | | | | | | |
| Comprehension | | | | | | | | | | | |
| Picture concepts | | | | | | | | | | | |
| Arithmetic | | | | | | | | | | | * |
| Letter-number Sequencing | | | | | | | | | | * | |
| Cancellation | | | | | | | | | | | |

Abbreviations: AWM = auditory working memory, CPI = cognitive processing index, FR = fluid reasoning, FSIQ = full scale IQ, GAI = general ability index, NVI = nonverbal index, PS = processing speed, QR = quantitative reasoning, VC = verbal comprehension, VS = visual-spatial, WM = working memory.

as a substitute for an invalid or missing test for the calculation of the FSIQ, substitution is not permitted for any of the primary or ancillary index scores. For a visual aid depicting the organization of the WISC-V please see Table 1.

Preliminary structural validity of the WISC-V was established using a series of CFAs reported in the Technical Manual (Wechsler 2014b). Several rival hierarchical models were explored with a general intelligence factor influencing subtests indirectly through various combinations of two to five first-order factors. The best fitting model (Model 5e, Wechsler 2014b, p. 83) included a hierarchical general intelligence dimension with five first-order factors (VC, FR, VS, WM, and PS) and the 16 primary and secondary subtests uniquely associated with one of those dimensions except for AR, which was permitted to cross-load on VC, FR, and WM. While all the five-factor models that were explored fit the data better than the rival four-factor model, analyzes of Model 5e indicate that its global fit could be improved by allowing FW to load on VS and FR. However, retention of this parameter was rejected "because it was not thought to have a clear theoretical rationale" (p. 83).

### 1.3 | Subsequent Research on the WISC-V Primary Measurement Model

Soon after the publication of the WISC-V, independent scholars began to question the veracity of the structural/interpretive model proposed by the test publisher. For example, Canivez and Watkins (2016) noted that the publisher relied exclusively on the use of CFA to evaluate internal structure with no consideration to results furnished from complimentary exploratory factor analytic (EFA) techniques. Further, these procedures produced a final validation model (Model 5e) that contained many of the same problematic parameters associated with prior attempts to explicate a five-factor structure for the WISC-IV (e.g., Weiss et al. 2013). Of additional concern, a series of independent CFA/EFA studies indicated that a rival four-factor model, cohering with previous Wechsler Theory (VC, Perceptual Reasoning [PR], WM, and PS), best explains the WISC-V normative data (Canivez et al. 2016, 2017; Dombrowski et al. 2018). Results that have been replicated in other WISC-V factor analytic studies featuring large clinical samples (e.g., Canivez et al. 2020; Dombrowski et al. 2022).

However, research from other scholars has provided support for a five-factor model although with varying degrees of allegiance to the model reported in the Technical Manual (e.g., Chen et al. 2015; Reynolds and Keith 2017). Given the number of conflicting interpretive structures that have been suggested for the WISC-V since its publication, Dombrowski et al. (2021) sought to determine the replication rate of these rival structures via Monte Carlo simulation. Results indicated that while the four-factor model originally produced by Canivez and Watkins (2016) fit the data the best, multiple versions of a five-factor model were more replicable in the 1,000 resampling runs suggesting that the four-factor model may lack stability. Although our understanding of the WISC-V primary and secondary subtest measurement model continues to evolve, the influence of the posited ancillary indexes and composites were not featured in any of the CFA models explored in the Technical Manual or in the independent EFA/CFA studies produced to this point for the WISC-V.[4] As these scores reflect dimensions thought to be nested within an extended Wechsler measurement model (Weiss et al. 2016), a brief review of their development is instructive for understanding the rationale for their specification on the WISC-V.

### 1.4 | Genesis of the Ancillary Scores

Consistent with psychometric best practice (e.g., Keith and Kranzler 1999), all the scores that can be derived from the primary subtests are structurally derived using factor analysis. By comparison, the WISC ancillary indexes are logically constructed by reconfiguring various primary subtests with secondary measures. Those configurations are as follows: QR (FW and AR), AWM (DS and LN), NVI (BD, VP, MR, FW, PS, and CD, GAI (SI, VC, BD, MR, and FW), and the CPI (DS, PS, CD, and SS). While the QR, AWM, and NVI scores are new to the WISC-V, versions of the GAI and CPI were previously developed for the WISC-III (Wechsler 1991) and WISC-IV.

The GAI was originally conceptualized as an alternative measure of global ability, featuring only the VC and PR subtests. According to Prifitera et al. (1998) it was primarily intended to be used as a reference anchor for the discrepancy model for specific learning disability identification (LD/SLD) in circumstances where it was believed that the FSIQ was disproportionally influenced by underlying processing deficits in WM and PS. It was believed that this so-called "Mark penalty" (Willis and Dumont 2002) could artificially mask IQ-achievement discrepancies for some examinees raising the risk of Type 2 decision error. Within a discrepancy model context, a Type II error would represent a false negative decision insofar as a student would not manifest with a significant discrepancy even though they may likely have a SLD. It is also important to note that the GAI was added as a post hoc interpretive feature to the Wechsler Scales thus to calculate the GAIs for the WISC-III and WISC-IV users had to rely on norms provided in external documentation for the instrument after their publication (e.g., Prifitera et al. 1998; Saklofske et al. 2005). Later, Dumont and Willis (2001) recommended using the remaining WM and PS subtests to form a complimentary index that could be compared to a score resembling the GAI. While Dumont and Willis reported provisional norms for their indexes derived from the Tellegen and Briggs (1967) formula in that same webpost[5], Weiss et al. (2006) later provided norms for that score calibrated from the WISC-IV normative data and referred to that score as the CPI for the first time.

Not surprisingly, after the formal conceptualization of the CPI, speculation about the clinical implications of significant discrepancies between the GAI and CPI (i.e., invalidating the FSIQ) began to emerge in the literature (e.g., Bremner et al. 2011), resurrecting a popular interpretive heuristic for the WISC-III that was seemingly lost when the Verbal-Performance IQ scores were dismantled from the WISC-IV. Although the GAI and CPI were largely described as formative constructions, their potential representation as latent dimensions in the Wechsler Model were outlined as part of a symposium on the GAI presented at a meeting of the American Psychological Association in 2009[6]. In a slide titled "WISC-IV Structure with GAI & CPI," Zhu (2009) conceptualized the GAI and CPI as intermediary dimensions (like the VIQ and PIQ) between the FSIQ and the four WISC-IV index scores. However, the internal CFA results presented for that model indicated that both the GAI (0.93) and CPI (1.00) were virtually redundant with the FSIQ (i.e., psychometric *g*) suggesting the model was likely overfit and thus not dispositive for adjudicating whether the GAI or CPI represent viable psychological dimensions on the WISC-IV or beyond. Particularly, given the construction of the CPI mimics, to some extent, prior attempts to group selected Wechsler subtests into unique profile configurations that promise to offer diagnostic insight for SLD identification (Smith and Watkins 2004) or Autism identification (Styck et al. 2019) which subsequent research studies have failed to support.

### 1.5 | Purpose of the Current Study

At a minimum, the inability to consistently replicate a FR dimension in previous EFA/CFA studies, coupled with the lack of representation of the ancillary indexes in any of the structural validity studies of the WISC-V that have been reported to this point, leave important questions about the construct validity of those indices unanswered. To wit, it is suggested that QR reflects "a type of fluid reasoning" (Weiss et al. 2016, p. 17) and thus should be represented as a "minor" factor that is nested within FR in the WISC-V measurement model. Although the structure for the primary measurement model has been explored many times, users are unable to extrapolate from these analyzes the information necessary to determine whether the ancillary indexes reflect viable psychological dimensions on the WISC-V. Further, previous attempts to model previous ancillary indexes (i.e., GAI and CPI) on previous editions of the WISC have not yielded compelling evidence that they represent reflective constructs (Edwards 2011). Despite these unknowns, speculation about the use of these scores for clinical decision making has not been tempered (Dale et al. 2023; Giofrè et al. 2017). Consequently, the goal of this study is to examine the structural validity of the WISC-V ancillary indexes using best practice CFA techniques to evaluate the following research questions more specifically:

1. Do the ancillary index-level scores (AWM, QR) represent viable psychological constructs in an extended WISC-V measurement model?
2. Is the configuration of NVI subtests best represented by a unidimensional model?
3. Are the CPI and GAI best represented as intermediary factors within the WISC-V measurement model as suggested by Zhu (2009)?

As this is the first investigation of its kind investigating the ancillary structure of the WISC-V, it is believed that the results produced from this study will be instructive for advancing evidence-based interpretive procedures for the instrument to better inform data-based decision making for school psychologists who elect to interpret these scores in clinical practice.

## 2 | Methods

### 2.1 | Participants

Participants included 2200 children and adolescents ages 6:0 to 16:11 from the WISC-V normative sample. This sample was obtained using a stratified sampling plan designed to accord with 2012 US census estimates. Inspection of the demographic data reported in the Technical Manual (Wechsler 2014b) reveal that the data for the normative sample were consistent with the US population parameters for age, gender, race/ethnicity, parent education level (as a proxy for socioeconomic status), and geographic region. All WISC-V measures were administered to participants by licensed clinical examiners or those with verified training in child psychological assessment procedures.

### 2.2 | Measurement Instrument

The WISC-V is an individually administered, norm-referenced, test of intellectual function for children and adolescents. It is comprised of 16 primary and secondary subtests which combine to yield five primary index-level scores (VC, PR, VS, WM, and PS), five ancillary-level index scores and composites (QR, AWM, GAI, CPI, and NVI), as well as an omnibus FSIQ that is based on a differentially weighted combination of seven of the primary subtests. While all primary and secondary subtests are expressed as scaled scores ($M = 10$, $SD = 3$), primary and ancillary index and composite scores are expressed as standard scores ($M = 100$, $SD = 15$). The WISC-V also features five new "complimentary" subtests (Naming Speed Literacy, Naming Speed Quantity, Immediate Symbol Translation, Delayed Symbol Translation, and Recognition Symbol Translation) developed for "special clinical uses" (Wechsler 2014b, p. 6) and do not contribute to the measurement of intelligence and thus were not included in this study. Extensive reliability and validity evidence for the measures are reported in the Technical Manual.

### 2.3 | Procedure and Data Analyzes

Mplus 8.0 (Muthén and Muthén 1998–2017) was used to conduct CFA using maximum likelihood estimation. Given that access to the WISC-V normative data is restricted by the test publisher, the present study utilized summary data (i.e., *N*, Means, *SDs*, and correlations) reported in the WISC-V Technical and Interpretive Manual (Wechsler 2014b). Specifically, the correlation matrix for the 16 primary and secondary subtests for the total normative sample (Table 5.1, p. 74) was extracted and used to produce the necessary covariance matrices for CFA. A series of CFAs was then conducted, each designed to evaluate the unique substructures implied by the inclusion of ancillary index and composite scores in the Wechsler measurement model.

### 2.4 | WISC-V CFA Ancillary Model Specifications

First, a series of rival structures were evaluated to establish a baseline model for the WISC-V normative data. As the Arithmetic (AR) subtest has been regarded as an enigma throughout the evolution of Wechsler Theory (Karzmark 2009), several hierarchical models were explored cohering with the different ways in which that indicator has been hypothesized to align in the WISC-IV and WISC-V interpretative literature. These models include (a) a five-factor publisher validation model featuring AR cross-loading on VC, FR, and WM; (b) five-factor model consistent with previous Wechsler Theory with AR loading on WM (i.e., WISC-III [Freedom From Distractibility] and WISC-IV), and a five-factor model based on the publisher suggested organization framework (see Wechsler 2014b, p. 21) for the test in which AR can be used to substitute for one of the primary FR measures. Additionally, alternative four- and five-factor models that have been found to best fit the WISC-V normative data in the independent factor analytic literature (i.e., Canivez et al. 2017; Reynolds and Keith 2017) were also explored. Once an adequate baseline model was established, the proposed AWM and QR dimensions were iteratively added as minor factors consistent with their proposed theoretical alignment (e.g., Weiss et al. 2016).

As the GAI/CPI and NVI composites are produced from different configurations of primary and secondary subtests, each of these substructures was evaluated independently. A series of unidimensional, higher order (GAI/CPI only), and hierarchical models were explicated to determine the best approximation of the underlying structure for these hypothesized score configurations. It should be noted that in the higher-order and hierarchical models that were specified and explored, consideration of additional WISC-V factors was limited to plausible primary (e.g., FR, VS, VC) constructs to avoid identification issues (Bee et al. 2023).

### 2.5 | Assessing Model Fit

Consistent with best practice, multiple fit indices were examined to evaluate the adequacy of global fit (Lai and Green 2016). These included, the (a) chi-square ($\chi^2$), (b) comparative fit index (CFI), (c) root mean square error of approximation (RMSEA), (d) standardized root mean square residual (SRMR), and (e) Akaike's information criterion (AIC). To comport with common recommendations, the following guidelines were used for good-model fit criteria: (a) CFI $\geq 0.95$; (b) SRMR and RMSEA $\leq 0.06$ (Beribisky and Hancock 2024). While there are no comparable criteria for information-based indices like the AIC, smaller values may indicate better approximations of the true measurement model after accounting for model complexity (Marsh and Alamer 2024). As utilizing fixed cut offs for evaluating global model fit is not recommended (McNeish and Wolf 2023), meaningful differences between well-fitting models were evaluated based upon the following criteria: (a) exhibit good fit according to CFI, RMSEA, and SRMR indices; (b) demonstrate a $\Delta$CFI value $\leq 0.01$ for nested models; and/or (c) display the smallest AIC value. In addition to global fit, each model was also inspected for the presence of local strain (i.e., implausible paramaters).

## 3 | Results

### 3.1 | Extended WISC-V Ancillary Indexes

Results from global fit statistics for all the models evaluated are presented in Table 2. All attempts to establish a five-factor

**TABLE 2** | Confirmatory factor analysis fit statistics for WISC-V 16 subtest configuration with ancillary indexes for normative sample participants ages 6–16 (N = 2,200).

| Model | $\chi^2$ | *df* | *p*[a] | CFI | SRMR | RMSEA | 90% CI RMSEA | AIC |
|---|---|---|---|---|---|---|---|---|
| Ancillary Indexes | | | | | | | | |
| 1. Publisher Validation Model | | | Impermissible Model (Negative Residual Variance FR) | | | | | |
| 2. Wechsler Theory | | | Impermissible Model (Negative Residual Variance FR) | | | | | |
| 3. Publisher Organizational Framework | | | Impermissible Model (Negative Residual Variance FR) | | | | | |
| 4. Alt. Four-Factor (Canivez et al. 2017) | 577.85 | 100 | < 0.001 | 0.969 | 0.030 | 0.047 | [0.043, 050] | 162231 |
| 5. Alt. Five-factor (Reynolds and Keith 2017) | 364.64* | 97 | < 0.001 | 0.984 | 0.023 | 0.034 | [0.030, 038] | 161998 |
| 5a. Add Minor AWM | 333.24* | 96 | < 0.001 | 0.985 | 0.023 | 0.034 | [0.030, 037] | 161994 |
| 5b. Add Minor AWM and QR | | | No Convergence, Number of Iterations Exceeded | | | | | |
| Nonverbal Ability Index | | | | | | | | |
| 6. One-factor (*g*) | 122.88 | 9 | < 0.001 | 0.967 | 0.028 | 0.076 | [0.064, 088] | 63187 |
| 7. Hierarchical (FR, VS and NV) | 59.46* | 6 | < 0.001 | 0.984 | 0.019 | 0.064 | [0.050, 079] | 41109 |
| GAI and CPI | | | | | | | | |
| 8. First-Order Orthogonal | 1832.65 | 27 | < 0.001 | 0.738 | 0.215 | 0.174 | [0.168, 181] | 94649 |
| 9. First-Order Oblique | 878.00* | 26 | < 0.001 | 0.876 | 0.060 | 0.122 | [0.115, 129] | 93696 |
| 10. Intermediary Wechsler Oblique | 77.26* | 22 | < 0.001 | 0.992 | 0.019 | 0.034 | [0.026, 042] | 92903 |
| 11. Hierarchical *g* (Zhu 2009) | | | Impermissible Model (Negative Residual Variance CPI) | | | | | |

Abbreviations: AIC = Akaike information criterion, AWM = auditory working memory, CFI = comparative fit index, CPI = cognitive proficient index, FR = fluid reasoning, GAI = general ability index, NV = nonverbal, QR = quantitative reasoning, RMSEA = root mean square error of approximation, SRMR = standardized root mean square residual, VS = Visual-Spatial, WISC-V = Wechsler Intelligence Scale for Children-Fifth Edition.
[a]Values rounded to nearest hundredth.
*Statistically different ($p < 0.05$) from previous model(s).

model at baseline from the posited assignment of AR (e.g., FR and WM) in the Technical Manual or the prior WISC-IV structure produced negative residual variance for the FR dimension. For example, in the publisher validation model (Model 1), the standardized loading between FR and *g* was 1.02 which exceeds unity indicating that the FR dimension in that model does not account for any unique variance apart from general intelligence. As noted by Brown (2015), "A measurement model should not be deemed acceptable if the solution contains one or more parameter estimates that have out-of-range values" (p. 162). As a result, fit statistics for models in which Heywood cases were encountered are not reported in Table 2. It should be noted that these results cohere with the CFA results for that model reported in the Technical and Interpretive Manual in which a standardized loading of 1.0 was observed indicating that the preferred model is likely overfit.[7]

Consequently, a series of alternative four- and five-factor models were specified and examined to identify an acceptable baseline model free of issues pertaining to local fit. Fit statistics associated with the four-factor model, consistent with previous Wechsler Theory (i.e., VC, PR WM, PS), produced by Canivez et al. (2017) indicated that model fit the normative data well (CFI = 0.969). However, the five-factor model postulated by Reynolds and Keith (2017), containing a correlated residual path between FR and VS (Model 5) and AR loading directly on *g* as well as WM, provided the best fit to the data among the competing baseline models that were evaluated and was statistically distinguishable from Model 4 ($\Delta$CFI = 0.015). Based on these results, Model 5 was selected as the baseline model from which to incrementally add the hypothesized constructs associated with the ancillary indexes (i.e., QR and AWM) as part of an extended WISC-V measurement model.

Addition of AWM (Model 5a) yielded a statistical improvement in fit from the baseline model ($\Delta\chi^2 = 31.40$, $p < 0.05$) and while several fit indices were identical to those obtained from baseline, its AIC value was lower indicating that it should be judged to fit the data better in comparison to other nested models (Vrieze 2012). Inspection of local fit indicated that all parameter estimates were statistically significant and did not reveal any evidence of model misspecification (see Figure 1). Based on these results, Auditory Memory appears to be a plausible construct on the WISC-V as a minor factor nested within WM. However, Model 5b failed to converge indicating that an optimal ML solution was unable to be obtained after 10,000 iterations once QR was added as a minor factor nested within FR.

### 3.2 | NVI Composite

Model fit statistics presented in Table 2 illustrate increasingly better fit as the hypothesized structure of the NVI progressed from a unidimensional model to a more dimensionally complex hierarchical model. Whereas the one-factor model (Model 6) provided an excellent fit to the normative data, the hierarchical model (Model 7), featuring the addition of FR and VS as first-order factors, yielded a statistically significant improvement in incremental fit from the unidimensional model ($\Delta$CFI = 0.02) and was judged to be superior. Figure 2 provides a graphical depiction of the hierarchical structure retained for the NVI wherein direct paths were specified between hierarchical *g* and the remaining variables that did not contribute to the measurement of first-order constructs. Although the standardized *g* loadings for CD (0.221) and PS (0.467) were low in comparison to other parameter estimates in the model, both were statistically significant. It should be noted that identification issues preclude the specification of a correlated residual term between FR and VS (as per Reynolds and Keith 2017) in Model 7.

### 3.3 | GAI and CPI Composites

In comparison to the NVI, the GAI and CPI dimensions could theoretically be modeled as standalone composites or intermediary constructs within a multi-tier WISC-V extended measurement model. Model fit statistics reported in Table 2 for competing models depicting the GAI and CPI as orthogonal (Model 8) or oblique (Model 9) as standalone higher-order composites without a hierarchical general intelligence dimension (i.e., FSIQ) were univocally inadequate. Of the models evaluated, only the three-stratum Model 10, featuring four primary dimensions as intermediary factors (VC, FR, WM, PS) with GAI and CPI correlated at the apex of the model yielded an acceptable fit to the data (CFI = 0.992, RMSEA = 0.034). Like the NVI analyzes, BD was allowed to load directly on GAI as it did not combine with any other measure to form a viable first-order primary Wechsler dimension. Fit statistics for Model 11, a rival hierarchical model featuring *g* (i.e., FSIQ) consistent with the hypothetical WISC-IV structure postulated by Zhu (2009), are not reported due to encountering a Heywood case for CPI. Based on these results, Model 10 (see Figure 3) was judged to be the only acceptable explanation for the normative data among the models that were evaluated.

## 4 | Discussion

Consistent with a long-standing trend in intelligence test development, the number of composite- and index-level scores on the WISC-V has doubled from its predecessor without much discernable change in the composition of the test (Frazier and Youngstrom 2007). While the primary measurement model for the test has been evaluated many times, posited ancillary dimensions were not featured in any of the structural validity studies reported in the Technical Manual nor in subsequent WISC-V structures featured in the literature. In fact, beyond basic descriptions regarding their organizational features, the ancillary indexes are barely discussed at all in prominent interpretive resources for the instrument (e.g., Kaufman et al. 2016; Weiss et al. 2016). Absent additional information about these measures, users are left without a compelling empirical basis for their development or use (Beaujean and Benson 2019). Accordingly, this study utilized CFA procedures to better evaluate the construct validity of the ancillary measures on the WISC-V to address this critical gap in the literature.

Results produced from a series of CFAs provided preliminary support for the interpretation of certain ancillary scores but not others. At the index-level, support was found for the inclusion of AWM as an intermediary factor within WM in an extended

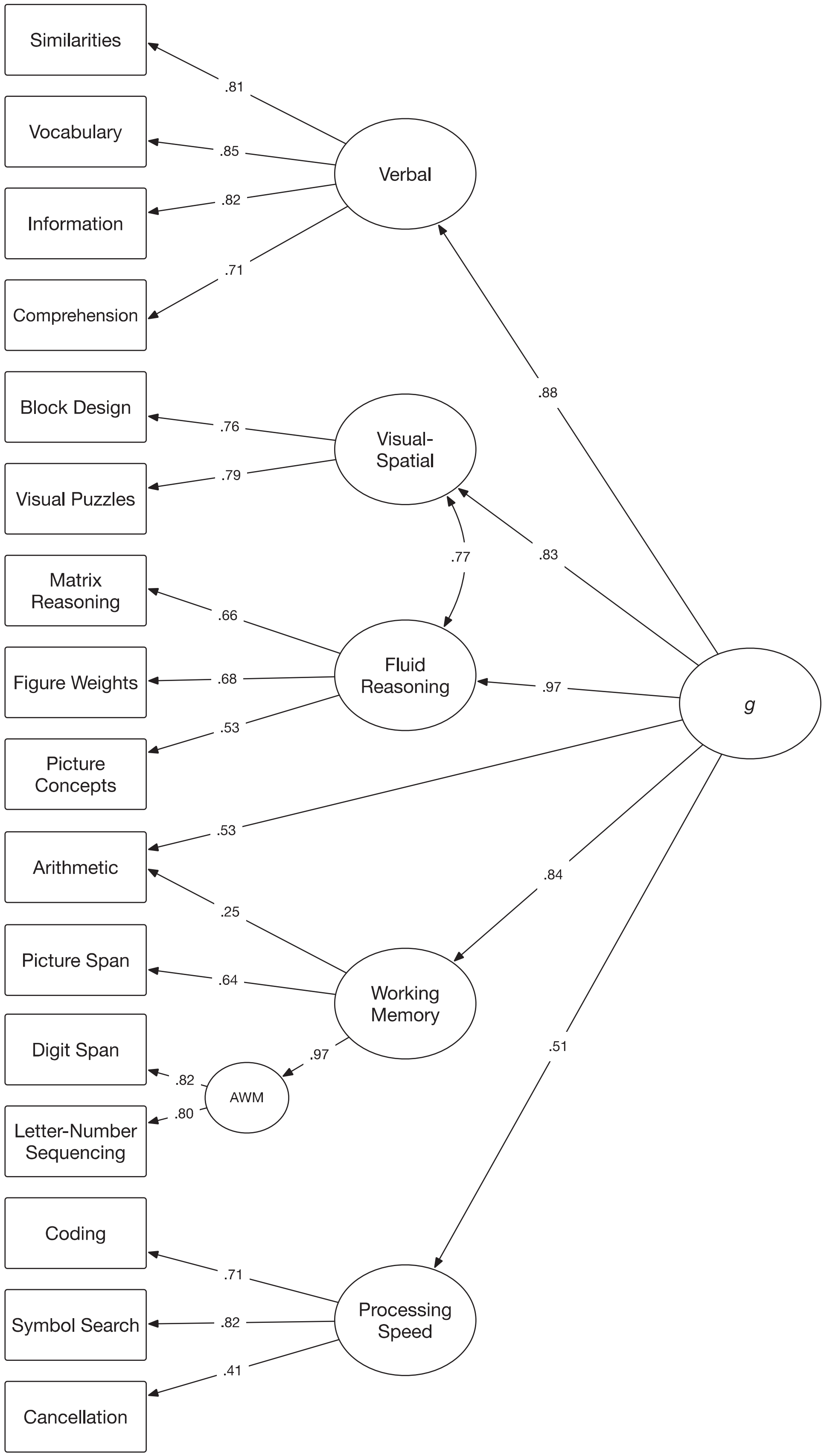


**FIGURE 1** | Extended WISC-V CFA model featuring primary and ancillary dimensions (Model 5a). AWM = Auditory Working Memory, g = General Intelligence. Residual terms omitted for clarity. Baseline WISC-V primary measurement model as per Reynolds and Keith (2017).

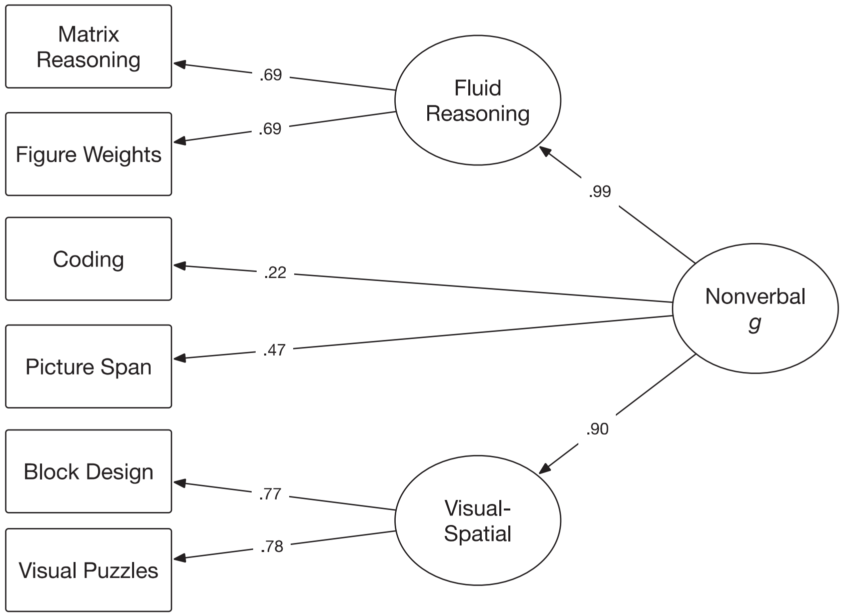


**FIGURE 2** | Final Hierarchical CFA Model for the NVI Composite (Model 7). *g* = General Intelligence. Residual terms omitted for clarity.

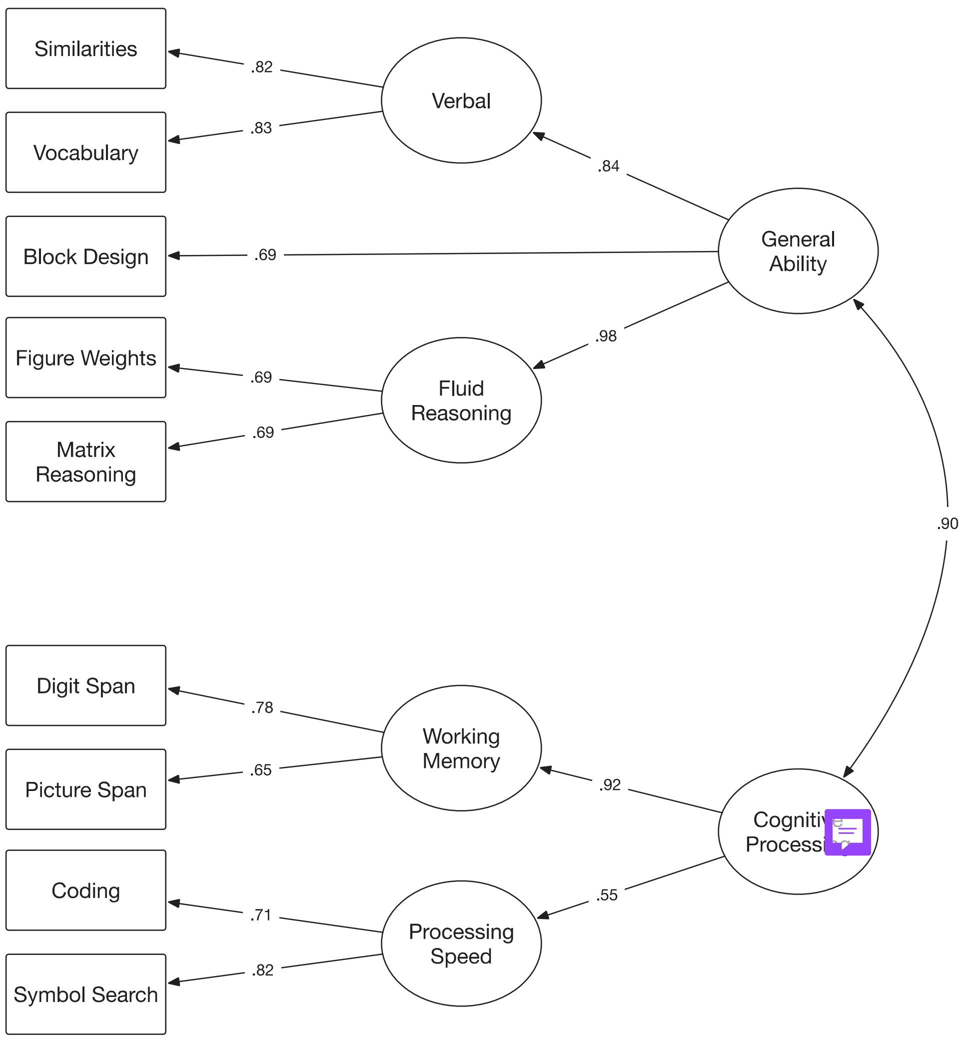


**FIGURE 3** | Final Oblique Higher-Order CFA Model for the GAI and CPI Composites (Model 10). Residual terms omitted for clarity. Model assumes that required subtests are extrapolated from administration of the primary WISC-V subtest configuration.

WISC-V measurement model. However, a rival solution featuring the inclusion of a minor QR construct within FR did not converge indicating that structural validity of that dimension is not supported. At the composite-level, the NVI, GAI, and CPI were each supported, in part, by the explication of logical substructures based on their unique subtest configurations. Taken altogether, these results raise several concerns pertaining to the interpretation of specific primary and ancillary scores on the WISC-V. These cautions are systematically reviewed in more detail below.

### 4.1 | Do the Ancillary Index-Level Scores (AWM, QR) Represent Viable Psychological Constructs in an Extended WISC-V Measurement Model?

Extended WISC-V results indicate that even though AWM can be located, it remains unclear what the index score associated with that construct represents given the fact that it represents a configuration of subtests that was previously used to construct the WM score on prior versions of the Wechsler Scales. Taken at face value, it would assume that prior interpretive nomenclature associated with the four-factor Wechsler Theory model was mis-labeled conflating a second-order construct with what is now regarded as a narrower cognitive dimension. Complicating the matter is the fact that the subtests that combine to form the AWM index were previously used as core indicators to define WM on the WISC-IV. As a result, several competing interpretations for that score can be levied based on its current and former representation within Wechsler Theory as well as theoretical depictions of associated constructs within the assessment literature. These include but are not limited to (a) a narrow CHC dimension underlying WM as postulated in CHC Theory (i.e., Auditory Short-Term Storage [*Wa*]), (b) WM version 1.0 as measured on the WISC-IV, or (c) a contrast score for identifying discrepancies between verbal and visual short-term memory capabilities. Regardless of the theoretical conceptualization for AWM that is preferred, the near isomorphic AWM/WM loading coefficient (0.97) in Model 5a suggest that clinical inferences from differences between those index scores are likely specious. It should be noted that previous research regarding the WISC–V has shown that LN is a superior indicator of WM than PS in both EFA and CFA studies (Canivez et al. 2016, 2017) yet relegated to secondary subtest status in favor of PS.

### 4.2 | Test Equivalence and the Working Memory Index

Regardless of whether users elect to interpret the ancillary AWM score, the differing conceptualizations of WM across the WISC-IV and WISC-V indicate those scores do not meet the basic assumptions for measurement invariance or test equating (Pendergast et al. 2017). This finding, in concert with other measurement issues, complicate any kind of direct comparisons between those scores even though their nominal labels indicate they are measuring the same theoretical construct (Beaujean et al. 2018). Not surprisingly, the correlation coefficient (0.59) reported in the Technical and Interpretive Manual for WM in the WISC-IV/WISC-V linking sample is substantially lower than any other comparable index or composite score in that study. This is likely the result of the paltry correlation between PS and LN (0.25) in the WISC-V normative sample.

It is important for practitioners to consider the implications of this modification for the stability of long-term decisions involving ability profiles produced from the WISC as well as other instruments in which the architecture of scores change as part of the revision process. For example, their use in various profiles of strengths and weaknesses approaches (PSW; see Alfonso and Flanagan 2018) for SLD identification. To illustrate, if an examinee is tested at Time 1 via the WISC-IV and evinces a confirmatory PSW pattern (i.e., focal cognitive/achievement weaknesses are present in an otherwise spared psychoeducational profile thus thought to be the marker for LD) is determined by a weakness in WM, equivalent test performance on the WISC-V version of WM cannot be assumed at Time 2 because that score is produced from a different combination of indicators that require fundamentally different response processes from examinees. If the weakness in WM is no longer present in the new version of the test and thus the PSW pattern is no longer confirmatory, practitioners will have a difficult time determining whether the resulting change in LD status is the product of individual differences or an artifact of transient method variance (Schneider 2013). To be clear, the evidence furnished from this investigation should not be misconstrued as a direct or indirect test of the diagnostic validity of the PSW model but the nascent empirical evidence to date that *any* unique pattern configuration among subscores on intelligence tests such as the WISC-V are stable long-term must also be considered (Styck et al. 2019). Nevertheless, school psychologists operating in jurisdictions that have recently adopted a version of that model for LD identification will soon have to confront these issues in the next wave of revisions to the instrumentation upon which much of provisional PSW eligibility decisions rest.

### 4.3 | Illusions of Meaning and the QR Index

Given that QR is presumed to represent a narrow dimension nested within FR, it is worth noting that its ability to be located in the present investigation presumes that FR can be adequately located as well on the WISC-V which has not been supported in prior structural investigations of the WISC-V primary measurement model (e.g., Canivez et al. 2016, 2017). Although adjudication of this debate continues, multiple EFA/CFA studies suggest that FR is at best a fragile broad ability incapable of explaining additional sub-constructs on the WISC-V and thus mathematically incapable of supporting additional measurement of unspecified narrow constructs. Retention of FR on the WISC-IV/WISC-V has often required the use of all available subtests with the specification of complex parameters that do not cohere with publisher theory. For example, the five-factor hierarchical model retained in the CFA investigation by Reynolds and Keith (2017), required the specification of a correlated residual term between FR and versus Rather than collapsing those terms into a complexly determined PR dimension consistent with previous Wechsler Theory, the authors speculate that this parameter likely reflects "an intermediate nonverbal general reasoning factor between the broad abilities and g" (p. 42). As a result of these complexities, the version of FR produced from this alternative model likely contains insufficient variance to be partialed even further to support the explication of QR leading to the collapse of Model 5a. In sum, these results suggest that the ancillary index associated with FR likely represents in illusory construct.

Nevertheless, the QR score may have broad appeal to practitioners given its nominal label and the desire to assess dimensions related to quantitative cognition in the schools. For example, using a low score on that index as a basis for providing access to calculation accommodations on high stakes tests. While intuitively appealing, such uses are not supported by the results in the present study as that construct does not appear to be a viable psychological dimension on the WISC-V.

### 4.4 | Is the Configuration of NVI Subtests Best Represented by a Unidimensional Model?

A case can be made that the ancillary composite scores appear to provide users of the WISC-V with a host of viable alternatives to the omnibus FSIQ as an estimate of general ability. For example, the fit statistics associated with Model 7 illustrate well that the NVI is best represented as a hierarchical global dimension that best explains the six-subtest configuration from which it is produced, in combination with the second-order contributions of FR and versus These results, coupled with the large correlation between FSIQ and the NVI (0.93) reported in the Technical and Interpretive Manual, suggest that users of the WISC-V may elect to interpret that score with a reasonable degree of confidence in the clinical situations in which use of a nonverbal estimate of *g* is justified (e.g., Ortiz 2023).

### 4.5 | Are the CPI and GAI Best Represented as Intermediary Factors Within the WISC-V Measurement Model as Suggested by Zhu (2009)?

Whereas the GAI and CPI have long been features of the Wechsler Scales, these composites were originally constructed for pragmatic reasons and do not appear to have any coherent rationale supporting their use beyond potential applications for LD identification within a discrepancy model context. While significant GAI/CPI differences are often used as justification for invalidating the FSIQ, empirical justification for this and other-related interpretive heuristics inferred from the presence of test scatter lack compelling empirical support (e.g., McGill 2016). Accordingly, Weiss et al. (2016) warn users of the WISC-V against this "very problematic practice" (p. 15). While results from the present investigation support the GAI and CPI as plausible composites for their respective subtest configurations, these scores do not appear to have a discernable structural relationship with *g* (as per Zhu 2009) thus any inferences gleaned from discrepancies with the FSIQ should be regarded as speculative.

Given the correlation between FSIQ and the GAI (0.96) clinicians have been encouraged to prefer interpretation of that composite score as a matter of course given that it likely provides users with a more parsimonious representation of *g*. Results from a recent simulation by Farmer et al. (2020) suggest that as a measure of general ability the omnibus FSIQ likely does not provide users with any discernable advantages over the GAI from a technical standpoint. However, in one of the few investigations of its kind, Grieder et al. (2022) found that the individual-level comparability of different global composites derived from the same test battery was unsatisfactory. Thus, careful deliberation is required to select the most optimal global composite to interpret on the WISC-V and clinicians are warned against simply picking the highest index score among the buffet of options that are available. In concert with the intended uses of the scores explicated by the test publisher (e.g., Weiss et al. 2016), clinicians would do well to identify a compelling justification for deviating from the FSIQ for the GAI or NVI before engaging in assessment based on their understanding of the examinee's strengths and limitations.

### 4.6 | Limitations

This study is not without limitations that should be considered when interpreting the results. Although the present study featured a large nationally representative normative sample, future investigations will be instructive for determining the degree to which these results generalize. That is, narrow dimensions such as QR that failed to emerge in this sample but may emerge in others (crDombrowski and McGill 2024). According to Fried (2020), there is an upper-bound limit as to what can be uncovered in any single CFA investigation, while factor analysis provides the statistical rationale undergirding the development of psychological test scores, additional evidence is needed to determine if the newly developed ancillary indices are useful for diagnostic or treatment decisions.

## 5 | Conclusion

Haynes et al. (2019) argue that "to make valid and useful clinical judgments, clinicians must understand the dimensionality of the constructs they assess and of the measures used to assess them" (p. 151). That is not, assuming the name associated with a score presumes that score only captures only *that* psychological dimension or that dimension can be located by a test (i.e., unidimensional [McGill et al. 2018]). The results furnished by this investigation have direct implications for the clinical interpretation of the WISC-V and provide important information about the structural validity of the newly featured ancillary indexes (QR, AWM) as well as the alternative global composites that have been featured in previous editions of the Wechsler Scales (GAI and CPI). Consistent with established test standards (e.g., American Educational Research Association et al. 2014), practitioners are encouraged to interpret these measures with caution until better evidence emerges to support their discriminate validity apart from other WISC-V dimensions.

## 6 | Impact Statement

Modern ability measures such as the Wechsler Intelligence Scale for Children-Fifth Edition (WISC-V) provide practitioners with an increasing array of scores to interpret. However, confidant use and interpretation of these scores depends on the ability to ascertain important information about their underlying psychological structure and relationships with other variables. Unfortunately, this information for the new ancillary index scores is noticeably absent from the WISC-V literature. This article reports results from the first structural validity investigation of these indices. Results provide inconsistent support for their interpretation which has implications for contexts in which these scores are featured in clinical decision-making.

---


### Data Availability Statement

The authors have nothing to report.


### Endnotes

[1]Some literature that has been influential in the development of the Cattell-Horn-Carroll (CHC; Schneider and McGrew 2018) architecture is cited in various sections of the Technical Manual (e.g., Carroll 1993) but the CHC model is not referenced directly in the discussion of structural theories informing the development of the WISC-V.

[2]A loading coefficient $> 1.0$, often referred to as a "Heywood Case," represents an "out-of-bounds" estimate and should be interpreted as evidence of model misspecification (Greene et al. 2023).

[3]The FSIQ is composed of seven primary "FSIQ" subtests (Vocabulary, SI, BD, MR, FW, DS, and CD) and is thought to be a proxy for psychometric g.

[4]Although, previous EFA studies have featured the full WISC-V battery in which the ancillary scores could, in theory, be explicated, those analyzes did not attempt to extract enough factors to accord with these dimensions or conduct the targeted analyzes necessary (i.e., Schmid and Leiman 1957 with $k > 1$ dimensions) for exploring presumed ancillary global constructs (e.g., GAI, CPI).

[5]Both composites were named "Dumont-Willis" indexes (DWIs) or DWI-1 (GAI) and DWI-2 (CPI) respectively. While the GAI was developed before the DWI-1, the emergence of the DWI-2 appears to have inspired the formal conceptualization of the CPI.

[6]The slides retain a copyright from 2007, no previous presentations can be located in the literature.

[7]While it is possible to add a mathematical constraint to artificially prevent an offending parameter from going out of bounds in CFA, it only masks the underlying issue (Dombrowski et al. 2021).

### References


Alfonso, V. C., and D. P. Flanagan. 2018. *Essentials of Specific Learning Disability Identification*. 2nd ed. Wiley.

American Educational Research Association, American Educational Research Association, American Psychological Association, and National Council on Measurement on Education. 2014. *Standards for educational and psychological testing*. American Educational Research Association.

Beaujean, A. A., and N. F. Benson. 2019. "Theoretically-Consistent Cognitive Ability Test Development and Score Interpretation." *Contemporary School Psychology* 23, no. 2: 126–137. https://doi.org/10.1007/s40688-018-0182-1.

Beaujean, A. A., N. F. Benson, R. J. McGill, and S. C. Dombrowski. 2018. "A Misuse of IQ Scores: Using the Dual Discrepancy/Consistency Model for Identifying Specific Learning Disabilities." *Journal of Intelligence* 6, no. 3: 36. https://doi.org/10.3390/jintelligence6030036.

Bee, R. M., T. Koch, and M. Eid. 2023. "A General Theorem and Proof for the Identification of Composed CfA Models." *Psychometrika* 88, no. 4: 1334–1353. https://doi.org/10.1007/s11336-023-09933-6.

Benson, N. F., R. G. Floyd, J. H. Kranzler, T. L. Eckert, S. A. Fefer, and G. 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. https://doi.org/10.1016/j.jsp.2018.12.004.

Beribisky, N., and G. R. Hancock. 2024. "Comparing RMSEA-Based Indices for Assessing Measurement Invariance in Confirmatory Factor Models." *Educational and Psychological Measurement* 84, no. 4: 716–735. https://doi.org/10.1177/00131644231202949.

Bremner, D., B. McTaggart, D. H. Saklofske, and T. Janzen. 2011. "WISC-IV GAI and CPI in Psychoeducational Assessment." *Canadian Journal of School Psychology* 26, no. 3: 209–219. https://doi.org/10.1177/0829573511419090.

Brown, T. A. 2015. *Confirmatory Factor Analysis for Applied Research*. 2nd ed. Guilford Press.

Canivez, G. L., and J. C. Kush. 2013. "WAIS-IV and WISC-IV Structural Validity: Alternate Methods, Alternate Results. Commentary on Weiss Et Al. (2013a) and Weiss Et Al. (2013B)." *Journal of Psychoeducational Assessment* 31, no. 2: 157–169. https://doi.org/10.1177/0734282913478036.

Canivez, G. L., and M. W. Watkins. 2016. "Review of the Wechsler Intelligence Scale for Children-Fifth Edition: Critique, Commentary, and Independent Analyses." In *Intelligent Testing With the WISC-V*, edited by A. S. Kaufman, S. E. Raiford, and D. L. Coalson, 683–702. Wiley.

Canivez, G. L., R. J. McGill, S. C. Dombrowski, M. W. Watkins, A. E. Pritchard, and L. A. Jacobson. 2020. "Construct Validity of the WISC-V in Clinical Cases: Exploratory and Confirmatory Factor Analyses of the 10 Primary Subtests." *Assessment* 27, no. 2: 274–296. https://doi.org/10.1177/1073191118811609.

Canivez, G. L., M. W. Watkins, and S. 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, no. 8: 975–986. https://doi.org/10.1037/pas0000238.

Canivez, G. L., M. W. Watkins, and S. C. Dombrowski. 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, no. 4: 458–472. https://doi.org/10.1037/pas0000358.

Carroll, J. B. 1993. *Human Cognitive Abilities: A Survey of Factor-analytic Studies*. Cambridge University Press.

Chen, H., O. Zhang, S. E. Raiford, J. Zhu, and L. G. Weiss. 2015. "Factor Invariance Between Genders on the Wechsler Intelligence Scale for Children-Fifth Edition." *Personality and Individual Differences* 86: 1–5. https://doi.org/10.1016/j.paid.2015.05.020.

Dombrowski, S. C., and R. J. McGill. 2024. "Clinical Assessment in School Psychology: Impervious to Scientific Reform?" *Canadian Journal of School Psychology* 39: 297–306. https://doi.org/10.1177/08295735231224052.

Dale, B. A., W. H. Finch, and K. A. R. Shellabarger. 2023. "Performance of Children With ASD on the WISC-V Ancillary Index Scales." *Psychology in the Schools* 60, no. 2: 431–440. https://doi.org/10.1002/pits.22688.

Dombrowski, S. C., G. L. Canivez, and M. W. Watkins. 2018. "Factor Structure of the 10 WISC-V Primary Subtests Across Four Standardization Age Groups." *Contemporary School Psychology* 22: 90–104. https://doi.org/10.1007/s40688-017-0125-2.

Dombrowski, S. C., R. J. McGill, G. L. Canivez, M. W. Watkins, and A. A. Beaujean. 2021. "Factor Analysis and Variance Partitioning in Intelligence Test Research: Clarifying Misconceptions." *Journal of Psychoeducational Assessment* 39, no. 1: 28–38. https://doi.org/10.1177/0734282920961952.

Dombrowski, S. C., R. J. McGill, and G. B. Morgan. 2021. "Monte Carlo Modeling of Contemporary Intelligence Test (IQ) Factor Structure: Implications for Iq Assessment, Interpretation, and Theory." *Assessment* 28, no. 3: 977–993. https://doi.org/10.1177/1073191119869828.

Dombrowski, S. C., R. J. McGill, M. W. Watkins, G. L. Canivez, A. E. Pritchard, and L. A. Jacobson. 2022. "Will the Real Theoretical Structure of the WISC-V Please Stand Up? Implications for Clinical Interpretation." *Contemporary School Psychology* 26: 492–503. https://doi.org/10.1007/s40688-021-00365-6.

Dumont, R., and J. Willis. 2001. Use of the Tellegen & Briggs Formula to Determine the Dumont-Willis Indexes (DWI-1 & DWI-2) for the WISC-IV [Archived webpage]. https://www.myschoolpsychology.com/.

Edwards, J. R. 2011. "The Fallacy of Formative Measurement." *Organizational Research Methods* 14, no. 2: 370–388. https://doi.org/10.1177/1094428110378369.

Farmer, R. L., R. G. Floyd, M. R. Reynolds, and K. S. Berlin. 2020. "How Can General Intelligence Composites Most Accurately Index Psychometric *g* and What Might Be Good Enough?" *Contemporary School Psychology* 24, no. 1: 52–67. https://doi.org/10.1007/s40688-019-00244-1.

Flanagan, D. P., and A. S. Kaufman. 2009. *Essentials of WISC-IV Assessment*. 2nd ed. Wiley.

Frazier, T. W., and E. A. Youngstrom. 2007. "Historical Increase in the Number of Factors Measured by Commercial Tests of Cognitive Ability: Are We Overfactoring?" *Intelligence* 35, no. 2: 169–182. https://doi.org/10.1016/j.intell.2006.07.002.

Freeman, A. J., and Y.-L. Chen. 2019. "Interpreting pediatric intelligence tests: A framework From Evidence-Based Medicine." In *Handbook of Psychological Assessment*, edited by G. Goldstein, D. N. Allen, and J. DeLuca, 4th ed., 65–101. Academic Press.

Fried, E. I. 2020. "Lack of Theory Building and Testing Impedes Progress in the Factor and Network Literature." *Psychological Inquiry* 31, no. 4: 271–288. https://doi.org/10.1080/1047840X.2020.1853461.

Giofrè, D., E. Toffalini, G. Altoè, and C. Cornoldi. 2017. "Intelligence Measures as Diagnostic Tools for Children With Specific Learning Disabilities." *Intelligence* 61: 140–145. https://doi.org/10.1016/j.intell.2017.01.014.

Goforth, A. N., R. L. Farmer, S. Y. Kim, S. C. Naser, A. B. Lockwood, and N. Affrunti. 2021. "Status of School Psychology in 2020: Demographics of the NASP Membership Survey." *NASP Research Reports* 5, no. 2: 301-657-0270. https://www.nasponline.org.

Greene, A. L., A. L. Watts, M. K. Forbes, R. Kotov, R. F. Krueger, and N. R. Eaton. 2023. "Misbegotten Methodologies and Forgotten Lessons From Tom Swift's Electric Factor Analysis Machine: A Demonstration With Competing Structural Models of Psychopathology." *Psychological Methods* 28, no. 6: 1374–1403. https://doi.org/10.1037/met0000465.

Grieder, S., A. Bünger, S. D. Odermatt, F. Schweizer, and A. Grob. 2022. "Limited Internal Comparability of General Intelligence Composites: Impact on External Validity, Possible Predictors, and Practical Remedies." *Assessment* 29, no. 6: 1172–1189. https://doi.org/10.1177/10731911211005171.

Haynes, S. N., G. T. Smith, and J. D. Hunsley. 2019. *Scientific Foundations of Clinical Assessment*. 2nd ed. Routledge.

Kamphaus, R. W., A. P. Winsor, E. W. Rowe, and S. Kim. 2018. "A History of Intelligence Test Interpretation." In *Contemporary Intellectual Assessment; Theories, Tests, and Issues*, edited by D. P. Flanagan and P. L. Harrison, 4th ed., 56–72. Guilford Press.

Karzmark, P. 2009. "The Effect of Cognitive, Personality, and Background Factors on the Wais-III Arithmetic Subtest." *Applied Neuropsychology* 16, no. 1: 49–53. https://doi.org/10.1080/09084280802644144.

Kaufman, A. S., S. E. Raiford, and D. L. Coalson. 2016. *Intelligence Testing With the WISC-V*. Wiley.

Keith, T. Z., J. G. Fine, G. E. Taub, M. R. Reynolds, and J. H. Kranzler. 2006. "Higher Order, Multisample, Confirmatory Factor Analysis of the Wechsler Intelligence Scale for Children-Fourth Edition: What Does It Measure?" *School Psychology Review* 35, no. 1: 108–127. https://doi.org/10.1080/02796015.2006.12088005.

Lai, K., and S. B. Green. 2016. "The Problem With Having Two Watches: Assessment of Fit When RMSEA and CFI Disagree." *Multivariate Behavioral Research* 51: 220–239. https://doi.org/10.1080/00273171.2015.1134306.

Marsh, H., and A. Alamer. 2024. "When and How to Use Set-Exploratory Structural Equation Modelling to Test Structural Models: A Tutorial Using the R Package Lavaan." *British Journal of Mathematical and Statistical Psychology* 77, no. 3: 459–476. https://doi.org/10.1111/bmsp.12336.

McGill, R. J. 2016. "Invalidating the Full Scale IQ Score in the Presence of Significant Factor Score Variability: Clinical Acumen or Clinical Illusion?" *Archives of Assessment Psychology* 6, no. 1: 49–79.

McGill, R. J., S. C. Dombrowski, and G. L. Canivez. 2018. "Cognitive Profile Analysis in School Psychology: History, Issues, and Continued Concerns." *Journal of School Psychology* 71: 108–121. https://doi.org/10.1016/j.jsp.2018.10.007.

McGill, R. J., T. J. Ward, and G. L. Canivez. 2020. "Use of Translated and Adapted Versions of the WISC-V: Caveat Emptor." *School Psychology International* 41, no. 3: 276–294. https://doi.org/10.1177/014303432090379.

McNeish, D., and M. G. Wolf. 2023. "Dynamic Fit Index Cutoffs for Confirmatory Factor Analysis Models." *Psychological Methods* 28, no. 1: 61–88. https://doi.org/10.1037/met0000425.

Muthén, L. K., and B. O. Muthén. 1998–2017. *Mplus user's guide*. 8th ed. Muthén & Muthén.

Ortiz, S. O. 2023. "Best Practices in Nondiscriminatory Assessment." In *Best Practices in School psychology*, edited by P. L. Harrison, S. L. Proctor, and A. Thomas, 7th ed., 149–161. National Association of School Psychologists.

Pendergast, L. L., N. von der Embse, S. P. Kilgus, and K. R. Eklund. 2017. "Measurement Equivalence: A Non-Technical Primer on Categorical Multi-Group Confirmatory Factor Analysis in School Psychology." *Journal of School Psychology* 60: 65–82. https://doi.org/10.1016/j.jsp.2016.11.002.

Prifitera, A., L. G. Weiss, and D. H. Saklofske. 1998. "The WISC-III in context." In *WISC-III Clinical use and Interpretation: Scientist-practitioner perspectives*, edited by A. Prifitera and D. H. Saklofske, 1–38. Academic Press.

Reinecke, M. A., D. W. Beebe, and M. A. Stein. 1999. "The Third Factor of the Wisc-III: It's (Probably) Not Freedom From Distractibility." *Journal of the American Academy of Child & Adolescent Psychiatry* 38, no. 3: 322–328. https://doi.org/10.1097/00004583-199903000-00020.

Reynolds, M. R., and T. Z. Keith. 2017. "Multi-Group and Hierarchical Confirmatory Factor Analysis of the Wechsler Intelligence Scale for Children-Fifth Edition: What Does It Measure?" *Intelligence* 62: 31–47. https://doi.org/10.1016/j.intell.2017.02.005.

Saklofske, D. H., A. Prifitera, L. G. Weiss, E. Rolfhus, and J. Zhu. 2005. "Clinical interpretation of the WISC-IV FSIQ and GAI." In *WISC-IV Clinical use and Interpretation: Scientist-Practitioner Perspectives*, edited by A. Prifitera, D. H. Saklofske, and L. G. Weiss, 33–65. Academic Press.

Saklofske, D. H., L. G. Weiss, S. E. Raiford, and A. Prifitera. 2006. "Advanced Interpretive Issues With the WISC-IV Full-Scale IQ and General Ability Index Scores." In *WISC-IV Advanced Clinical Interpretation*, edited by L. G. Weiss, D. H. Sakflofske, A. Prifitera, and J. A. Holdnack, 99–138. Academic Press.

Schmid, J., and J. M. Leiman. 1957. "The Development of Hierarchical Factor Solutions." *Psychometrika* 22: 53–61. https://doi.org/10.1007/BF02289209.

Schneider, W. J. 2013. "What If We Took Our Models Seriously? Estimating Latent Scores in Individuals." *Journal of Psychoeducational Assessment* 31, no. 2: 186–201. https://doi.org/10.1177/0734282913478046.

Schneider, W. J., and K. S. McGrew. 2018. "The Cattell-Horn-Carroll Theory of Cognitive Abilities." In *Contemporary Intellectual Assessment: Theories, Tests, and Issues*, edited by D. P. Flanagan and E. M. McDonough, 4th ed., 73–163. Guilford Press.

Smith, C. B., and M. W. Watkins. 2004. "Diagnostic Utility of the Bannatyne Wisc-III Pattern." *Learning Disabilities Research & Practice* 19, no. 1: 49–56. https://doi.org/10.1111/j.1540-5826.2004.00089.x.

Styck, K. M., M. S. Aman, and M. W. Watkins. 2019. "Diagnostic Utility of the Wisc-Iv Gai >Cpi Cognitive Score Profile for a Referred Sample of Children and Adolescents With Autism." *Contemporary School Psychology* 23: 115–125. https://doi.org/10.1007/s40688-018-0172-3.

Styck, K. M., A. A. Beaujean, and M. W. Watkins. 2019. "Profile Reliability of Cognitive Ability Subscores in a Referred Sample." *Archives of Scientific Psychology* 7, no. 1: 119–128. https://doi.org/10.1037/arc0000064.

Tellegen, A., and P. F. Briggs. 1967. "Old Wine in New Skins: Grouping Wechsler Subtests Into New Scales." *Journal of Consulting Psychology* 31, no. 5: 499–506. https://doi.org/10.1037/h0024963.

Thurstone, L. L. 1954. "An Analytical Method for Simple Structure." *Psychometrika* 19: 173–182. https://doi.org/10.1007/BF02289182.

Vrieze, S. I. 2012. "Model Selection and Psychological Theory: A Discussion of the Differences Between the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (Bic)." *Psychological Methods* 17, no. 2: 228–243. https://doi.org/10.1037/a0027127.

Wasserman, J. D., and A. S. Kaufman. 2016. "A History of Ability Tests and Theories." In *Oxford Handbook of Clinical Neuropsychology*, edited by W. B. Barr and L. A. Bieliauskas, 275–329. Oxford University Press.

Wechsler, D. 1991. *Wechsler Intelligence Scale for Children*. 3rd ed. Psychological Corporation.

Wechsler, D. 2003. *Wechsler Intelligence Scale for Children*. 4th ed. Psychological Corporation.

Wechsler, D. 2014a. *Wechsler Intelligence Scale for Children*. 5th ed. Pearson.

Wechsler, D. 2014b. *WISC-V Technical and Interpretive Manual*. Pearson.

Weiss, L. G., T. Z. Keith, J. Zhu, and H. Chen. 2013. "Wisc-IV and Clinical Validation of the Four- and Five-Factor Interpretative Approaches." *Journal of Psychoeducational Assessment* 31, no. 2: 114–131. https://doi.org/10.1177/0734282913478032.

Weiss, L. G., D. H. Saklofske, J. A. Holdnack, and A. Prifitera. 2016. "WISC-V: Advances in the Assessment of Intelligence." In *WISC-V Assessment and Interpretation: Scientist-Practitioner perspectives*, edited by L. G. Weiss, D. H. Saklofske, J. A. Holdnack, and A. Prifitera, 3–23. Academic Press.

Weiss, L. G., D. H. Saklofske, D. M. Schwartz, A. Prifitera, and T. Courville. 2006. "Advanced Clinical Interpretation of WISC-IV Index Scores." In *WISC-IV: Advanced Clinical Interpretation*, edited by L. G. Weiss, D. H. Saklofske, A. Prifitera, and J. A. Holdnack, 139–180. Academic Press.

Williams, P. E., L. G. Weiss, and E. Rolfhus. 2003. *Theoretical Model and Test Blueprint[WISC-IV Technical Report #1]*. Psychological Corporation.

Willis, J. O., and R. P. Dumont. 2002. *Guide to Identification of Learning Disabilities*. 3rd ed. Authors.

Zhu, J. 2009. "What is the GAI?" In *The What, When, and How, of the Wechsler General Ability Index [Symposium presentation]*, edited by D. H. Saklofske. Toronto, ON, CA: Meeting of the American Psychological Association.