

## Measurement Invariance of the Woodcock-Johnson® V Assessments

Tong Wu, Riverside Insights®  
Hyeonjoo Oh, Riverside Insights  
Kevin McGrew, Institute for Applied Psychometrics  
Erica LaForte, Northwestern University

*In standardized testing, it is essential that a test or test battery measures the same ability construct(s) across diverse groups of examinees. This study evaluates the measurement invariance of the Cattell-Horn-Carroll (CHC) constructs using the Horn no-g broad CHC model (10 broad CHC obliquely correlated latent factors) reported in the Woodcock-Johnson® V (WJ V™) Technical Manual (LaForte et al., 2025). The 10 broad CHC factor model (Gf, Gc, Gv, Ga, Gl, Gr, Gwm, Gs, Grw, Gq) was evaluated for structural invariance across groups differentiated by three demographic characteristics (race/ethnicity, gender, and age). Using multigroup confirmatory factor analyses (MGCFAs), configural and metric invariance across the groups was demonstrated. Our findings showed that the WJ V consistently maintains its 10 broad CHC structure and stable test-to-factor loadings across race/ethnicity, gender, and age groups, providing strong support for the test's validity across diverse populations. This study underscores the importance of establishing measurement invariance in test development and test interpretation for school psychologists.*

Copyright © 2025 by Riverside Assessments, LLC. All rights reserved. No part of this work may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying and recording, or by any information storage or retrieval system without the prior written permission of Riverside Assessments, LLC. Requests for permission to reproduce any part of the work should be sent via email to [permissions@riversideinsights.com](mailto:permissions@riversideinsights.com) or addressed via mail to Riverside Insights, Attention: Permissions, One Pierce Place, Suite 101C, Itasca, Illinois 60143.

Riverside Insights, the Riverside Insights logo, WJ IV, and Woodcock-Johnson are registered trademarks of Riverside Assessments, LLC. WJ V is a trademark of Riverside Assessments, LLC.

All other trademarks are the property of their respective owners.

# Measurement Invariance of the Woodcock-Johnson V Assessments

## Introduction

Ensuring that test scores are comparable across diverse demographic groups is a fundamental concern in large-scale standardized assessment. Measurement invariance testing provides the empirical foundation to evaluate whether a test functions equivalently across groups such as race/ethnicity, gender, and age. Without establishing invariance, observed score differences may reflect measurement artifacts rather than true differences in ability or achievement. As such, the current study adopts a rigorous multigroup confirmatory factor analysis (MG-CFA) framework to assess invariance across key populations for the *Woodcock-Johnson*<sup>®</sup> V (WJ V; McGrew, Mather, LaForte, & Wendling, 2025) test batteries.

## Test Validity

When developing educational or psychological assessments, the evaluation of test construct validity is critical to ensure accurate and fair interpretation of the scores. The *Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 2014) define validity as “the degree to which evidence and theory support the interpretations of test scores for proposed uses of tests” (p. 10). If a test lacks proper validation the results may be misleading, potentially introducing bias and reducing the overall effectiveness and reliability of the test. In creating the *Woodcock-Johnson V* (LaForte et al., 2025; McGrew, Mather, LaForte, & Wendling, 2025), a network of validity evidence (i.e., content, structural, concurrent or criterion-related) was reported in the *Woodcock-Johnson V Technical Manual* (LaForte, et al., 2025).

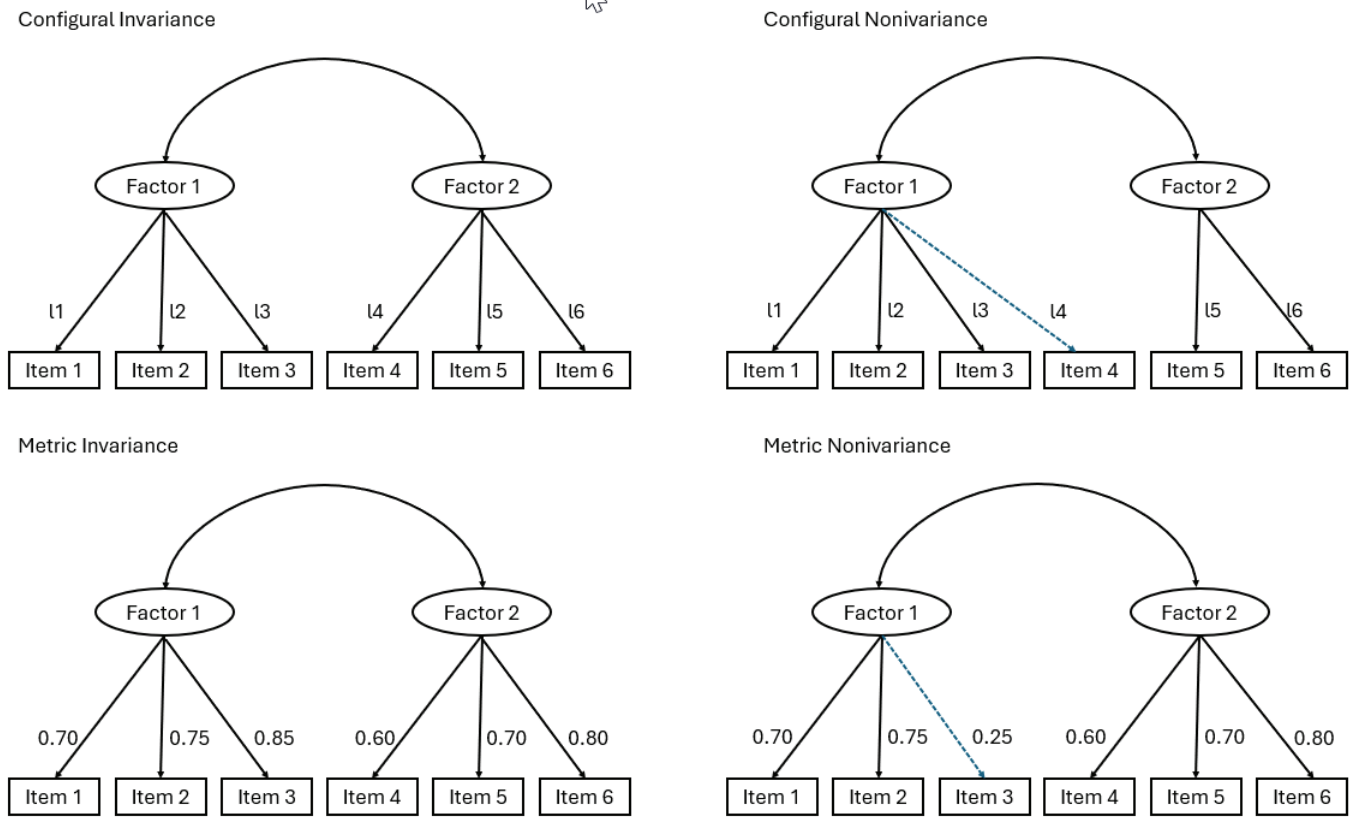
Internal structural evidence examines “the extent to which the relationships among test items and test components conform to the construct on which the proposed test score interpretations are based” (AERA et al., 2014, p. 16). Confirming that a test instrument corresponds to the hypothesized structure of the test battery is a key source of internal structural validity evidence. During the development and validation of the *Woodcock-Johnson V* (McGrew, Mather, LaForte, & Wendling, 2025), the Cattell-Horn-Carroll (CHC) theoretical model was examined, assessing the test’s normative data using a three-stage set of structural analysis procedures (LaForte et al., 2025). Four types of exploratory structural analyses (i.e., principal axis factor analysis, cluster analysis, multidimensional scaling, hierarchical exploratory psychometric network analysis) were completed with the WJ V tests in six developmentally age-differentiated model development (MD) norm samples. The results of the four exploratory structural analysis procedures were synthesized and the commonalities across the methods were used to specify and evaluate CFA MD structural models through an iterative process. The final CFA MD structural models were then cross-validated “as is” with CFA in independent cross-validation samples for the same six age groups.

## Measurement Invariance

The WJ V Technical Manual (LaForte et al., 2025) did not report the examination of measurement invariance of the 10 broad CHC Horn no-g broad CHC model. The Horn no-g broad CHC model shares similarities with the Carroll hierarchical g broad CHC model in that the measurement models are identical, but the structural models differ. Specifically, the Horn model retains the intercorrelations among the 10 broad CHC factors without incorporating a general factor. In this respect, it aligns in certain ways with contemporary no-g psychometric network models (see LaForte et al., 2025 for details). Measurement invariance is a statistical method used in educational and psychological assessments to evaluate the equivalence of relationships between observed variables (e.g., individual tests) and latent constructs (e.g., CHC constructs) across multiple groups. This process is used to determine whether the same constructs are measured consistently across different populations, such as those defined by race/ethnicity, gender, age, and other demographic factors. If measurement invariance does not hold, this suggests that a construct may have different meanings or structures for different groups or under varying conditions, making it difficult to interpret or compare results across groups or over time (Putnick & Bornstein, 2016).

A widely used approach to investigate the consistency of construct validity across groups (e.g., race/ethnicity, gender, and different developmental age groups) is multigroup confirmatory factor analyses (MGCFA; Jöreskog, 1971). Invariance testing using MGCFA can be conducted by iteratively specifying increasingly more stringent restrictions or constraints on the structural model parameters (Schmitt & Kuljanin, 2008; Vandenberg & Lance, 2000). *Configural* invariance checks whether the overall model structure is consistent across groups. *Metric* invariance assesses whether the factor loadings are the same across groups. *Scalar* invariance examines whether test intercepts remain constant across groups. In studies of measurement invariance, researchers begin with configural invariance as the baseline model, followed successively by tests for metric invariance, scalar invariance, and in some investigations, residual invariance. Given the complexity of the WJ V batteries, which include more than 60 individual tests, this study limited its focus to examining configural and metric invariance. The more stringent levels of invariance testing are beyond the scope of this paper. Figure 1 provides an illustration of measurement invariance and noninvariance at the two levels.

**Figure 1.**  
*Illustration of Configural Invariance  
 and Metric Invariance*



## Objective

This paper introduces the method of evaluating measurement invariance to school psychologists who are interested in increasing their understanding of this form of test validity evidence and its relevance for test interpretation. The measurement invariance of the complete WJ V test batteries, across different race/ethnicity, gender, and age groups, is reported.

## Methodology

To evaluate measurement invariance across key demographic groups, this study employed an MG-CFA framework. The analyses were conducted in a stepwise manner, testing increasingly restrictive models to assess the consistency of factor structure and loadings across groups. Model comparisons were based on established fit indices and recommended change thresholds.

## The *Woodcock-Johnson V* Tests

The WJ V is designed to measure general intelligence, along with broad and narrow cognitive abilities as defined by the CHC theory of cognitive abilities (Schneider & McGrew, 2018). The full WJ V system consists of 60<sup>1</sup> tests that contribute to various interpretive composites, or clusters. The WJ V has three conormed batteries—the WJ V Tests of Cognitive Abilities (WJ V COG), the WJ V Tests of Achievement (WJ V ACH), and the WJ V Virtual Test Library (WJ V VTL). The three-battery organization gives examiners the flexibility to use the two primary (COG and ACH) batteries and the VTL battery independently, or in any combination, to meet diverse assessment needs (Laforte et al., 2025). The WJ V COG battery includes 20 tests, the WJ V ACH battery includes 25 tests, and the WJ V VTL includes 15 tests.

Several changes were made in the WJ V revision.

- In the WJ V COG battery, the previous long-term retrieval (Glr) factor was split into two separate factors: long-term retrieval (Gl) and retrieval fluency (Gr).
- In the WJ V COG battery, 2 *Woodcock-Johnson IV* (WJ IV; Schrank et al., 2014) tests were dropped and 6 new WJ V tests were added.
- The WJ V ACH battery was designed to be one of the most comprehensive and flexible measures of achievement within a single battery, with the addition of 8 new WJ V ACH tests and 2 new oral language tests.
- The WJ V VTL battery was added to assess oral language, linguistic, and related cognitive and achievement abilities, further enhancing the comprehensiveness of the COG and ACH batteries.
- The WJ IV Oral Language battery was eliminated, and its tests were moved to either the WJ V COG, ACH, or VTL.

In alignment with the prior structural validity analysis of the WJ IV (McGrew, et al., 2014), the WJ V used a similar three-stage structural validity analysis framework (Laforte et al., 2025). In the first stage, the WJ V norming sample was divided into six developmentally age-differentiated groups (ages 4 to 5, 6 to 9, 10 to 14, 15 to 19, 20 to 49, and 50 to 80+), with the data at each age group randomly split into separate model development (MD) and model cross-validation (CV) samples. In the second stage, multiple exploratory structural analysis methods were conducted in the MD sample for each age group, leading to the construction of a preliminary CFA model. This MD model was then iteratively refined and evaluated with MD CFA methods that also incorporated information from structural analysis of prior editions of the WJ (i.e., WJ, WJ-R, WJ III, WJ IV), CHC theory, and the extant CHC structural research literature. In the third stage, the final MD structural models were fit “as is” in the CV samples for each age group.

Three classes of CHC models were evaluated and compared in the WJ V norm data—a Horn no-g broad CHC model, a Carroll hierarchical g broad CHC model, and a Carroll hierarchical g broad+narrow CHC model. All three models specified the same 10 broad CHC factors—Auditory Processing (Ga), Comprehension-Knowledge (Gc), Fluid Reasoning (Gf), Long-Term Storage (Gl), Retrieval Fluency (Gr), Reading and Writing (Grw), Cognitive Processing Speed (Gs), Quantitative Knowledge (Gq), Visual Processing (Gv), and Auditory Working Memory Capacity (Gwm). The models differed in how they handled the correlations (covariances) among the 10 broad ability factors (i.e., 10 obliquely correlated factors or models where a hierarchical psychometric g factor accounted for the latent correlations between the 10 first-order latent factors).

---

<sup>1</sup> Oral Vocabulary and General Information each consist of two subtests: Synonyms and Antonyms for Oral Vocabulary and Where and What for General Information.

## Norming Data

The study aimed to assess whether the CHC factorial structure of the WJ V is consistent (invariant) across various race/ethnicity, gender, and age groups. Data for these analyses were drawn from the WJ V norming sample that included examinees from 3 to more than 90 years of age. The norm data, collected from late 2021 through 2023, was based on the 2020 U.S. Census, with a sampling plan that ensured representativeness of the U.S. population, within practical limits, through stratification by census region, gender, race/ethnicity, and education level (parent education for children and examinee education for adults).

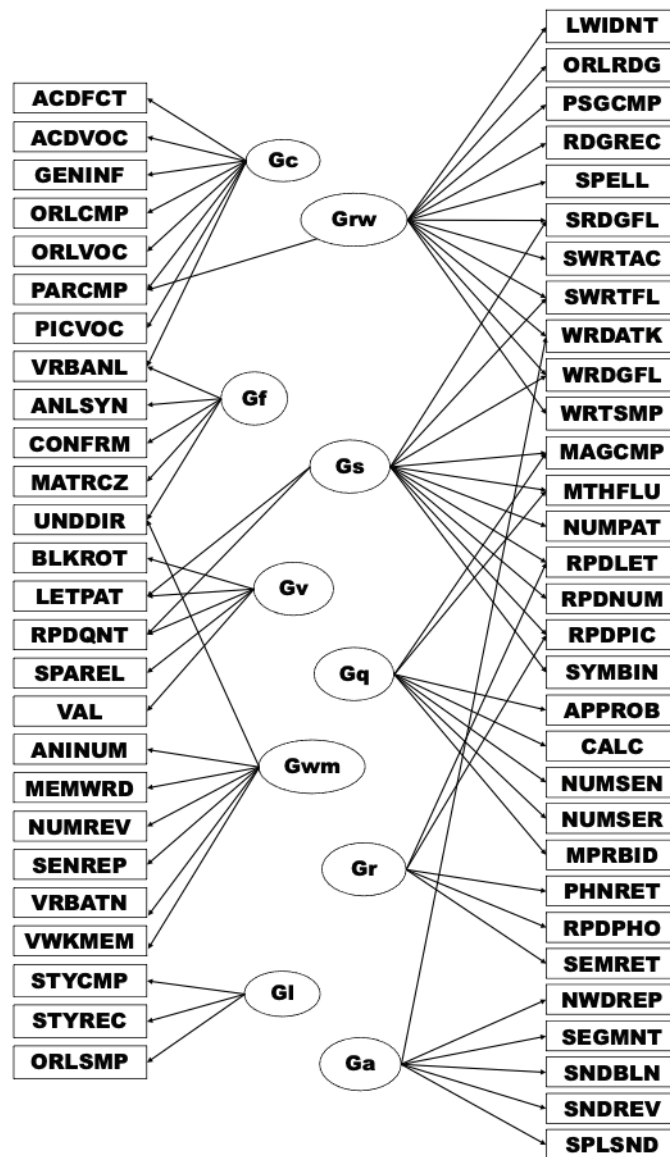
The current investigation examined the invariance across four race/ethnicity groups, two gender groups, and five age categories. A total of 5,174 participants were classified into four race/ethnicity groups: White ( $N = 2,953$ ), African American ( $N = 706$ ), Hispanic ( $N = 1,069$ ), and Other ( $N = 412$ ). Thirty-four participants who did not specify their race/ethnicity category were excluded from this study on race/ethnicity invariance. Participants were also divided by gender into female ( $N = 2,757$ ) and male ( $N = 2,381$ ); 36 individuals who did not disclose their gender were excluded due to the small sample size. The age groups analyzed included ages 6 to 9 ( $N = 949$ ), ages 10 to 14 ( $N = 1,256$ ), 15 to 19 ( $N = 1,198$ ), 20 to 49 ( $N = 905$ ), and 50 to 80+ ( $N = 866$ ). The ages 4 to 5 group was not included in this study due to the smaller number of individuals tested compared to the other groups. Additionally, the Letter Writing Fluency, Sound Deletion, and Sound Substitution tests were excluded because they are only available for limited age ranges. In total, 57 WJ V tests were included in the analysis.

## Measurement Invariance Model

This study focused on the structural invariance of the Horn no-g broad CHC model because a comparison of model fit statistics across the three CHC models showed that they were highly similar within each age group. However, based on the principle of parsimony, the Horn no-g broad CHC model was judged to be the preferred model (LaForte et al., 2025). In the original structural analysis presented in the WJ V Technical Manual, a Horn no-g broad CHC model was constructed for each age group, resulting in a total of five Horn no-g models, each with slight variations. To test measurement invariance, we first developed a CFA model based on the commonalities of the five age-group-specific models. To ensure model parsimony, we removed a few cross-loadings with lower factor loadings.

The refined Horn no-g broad CHC model (illustrated in Figure 2) was then used to assess the structural invariance of the WJ V measures. Model fit statistics were examined separately for each subgroup. This structural model included 10 correlated broad CHC factors, allowing for a detailed evaluation of the relationships between test scores and latent CHC constructs. The CFA analysis ensured that the proposed model accurately reflected the underlying theoretical framework before proceeding to the imposition of additional parameter constraints for measurement invariance testing.

**Figure 2.**  
Horn No-g Broad CHC  
Model



Notes. Auditory Processing (Ga), Comprehension-Knowledge (Gc), Fluid Reasoning (Gf), Long-Term Storage (GI), Retrieval Fluency (Gr), Reading and Writing (Grw), Cognitive Processing Speed (Gs), Quantitative Knowledge (Gq), Visual Processing (Gv), and Auditory Working Memory Capacity (Gwm). Individual test residual variances and latent factors correlations between all pairs of broad CHC factors are not included in the figure for readability purposes.

Measurement invariance was assessed hierarchically, with each level imposing stricter constraints than the previous model. First, we tested for configural invariance, which evaluates whether the basic model structure functions similarly across different subgroups. We used an unconstrained model to examine the four race/ethnicity groups, two gender groups, and five age groups. Next, we assessed metric invariance by constraining factor loadings to be equivalent across groups. If minimal differences in fit statistics are observed between the configural and metric models, this suggests that metric invariance holds. The measurement invariance analyses were conducted using the AMOS statistical software (Arbuckle, 2022).

## Estimation Method

Multicollinearity, non-normality, and the complexity of models with large numbers of variables can introduce bias in maximum-likelihood (ML) estimated standard errors and can affect chi-square tests and overall model fit indices (Brown, 2006). Given that these conditions existed in the WJ V norm data due to the large number of significantly correlated tests within certain CHC domains (e.g., Gc, Gs, Grw, Gq) (see LaForte et al., 2025), the Scale-Free Least Squares (SFLS) estimation method was used to calculate estimates and fit indices. This approach is conceptually like using the Unweighted Least Squares (ULS) algorithm on input from a correlation matrix. SFLS is advantageous when dealing with non-normal distributions, as it does not require strong multivariate normality assumptions. The SFLS method aims to eliminate the influence of scale or magnitude differences among variables.

## Results

To investigate the two gender groups, first we fitted the refined CFA model to each group separately. For females, the model yielded  $\chi^2(1462) = 12090.778, p < .001$ . For males, the model yielded  $\chi^2(1462) = 9012.746, p < .001$ . The chi-square test indicated statistical significance, but because this test is sensitive to sample size (Cheung & Rensvold, 2002), we focused on additional model fit indices such as the standardized root mean square residual (SRMR), the adjusted goodness-of-fit index (AGFI), and the parsimony adjustment to the normed fit index (PNFI). For the female group, the fit indices were as follows: SRMR = 0.0515, AGFI = 0.982, and PNFI = 0.900. For the male group, the fit indices were: SRMR = 0.0479, AGFI = 0.987, and PNFI = 0.905. These values meet the accepted criteria for an adequate model fit, as SRMR values below 0.08<sup>2</sup> (Hu & Bentler, 1999; Xia & Yang, 2019) and AGFI and PNFI values closer to 1.0 are generally considered indicators of good fit. Similar results were observed for the four race/ethnicity subgroups and five age groups, all of which met the fit criteria.

Second, we examined the fit indices for both the configural and metric invariance models across the demographic subgroups (see Table 1). For the gender groups, the configural invariance model demonstrated a good fit, allowing us to proceed with testing whether factor loadings were consistent across the two groups. When comparing the configural and metric invariance models, we found that although the chi-square difference ( $\Delta\chi^2$ ) tests indicated significant changes in fit, other fit indices remained within acceptable thresholds (Chen, 2007; Cheung & Rensvold, 2002). For metric invariance to hold, the change in SRMR ( $\Delta$ SRMR) should be less than 0.03; in our analysis,  $\Delta$ SRMR was approximately 0.001 for the gender groups, which met this criterion. While a critical value for  $\Delta$ AGFI is not explicitly established for metric invariance, it is generally accepted that changes in CFI ( $\Delta$ CFI) should be less than 0.01. However, since AMOS did not report CFI values when using the SFLS method, we relied on the assumption that  $\Delta$ AGFI would behave similarly to  $\Delta$ CFI. In this study,  $\Delta$ AGFI also met the expected criteria, indicating that factor loadings were equivalent across the gender groups, thus confirming metric invariance.

---

<sup>2</sup> The SRMR threshold is primarily established for model fit evaluation using the ML method; however, no specific threshold has been identified for SRMR under the SFLS method. Scholars, such as Xia and Yang (2019), have argued that the traditional cutoff point may fail to detect model misfit when ULS and DWLS are used. The use of the cut-off point should take into account the context of the study.

**Table 1.**

Summary of Fit Statistics for  
Testing Measurement Invariance

Group	Model	df	$\Delta df$	$\chi^2$	$\Delta \chi^2$	AGFI	$\Delta AGFI$	SRMR	$\Delta SRMR$
Race/ Ethnicity	Configural Invariance	5852		24112.242		0.981		0.052	
	Metric Invariance	6029	177	25660.150	1547.908	0.980	0.001	0.053	0.001
Gender	Configural Invariance	2926		21248.94		0.985		0.052	
	Metric Invariance	2985	59	22260.836	1011.896	0.984	0.001	0.053	0.001
Age	Configural Invariance	7315		22567.785		0.984		0.051	
	Metric Invariance	7551	236	30348.862	7781.077	0.979	0.005	0.068	0.017

Notes. *df* = degree of freedom,  $\chi^2$  = Chi-square statistic, AGFI = adjusted goodness-of-fit index, SRMR = standardized root mean square residual

Table 2 presents the factor loadings for both gender groups, indicating that these loadings were comparable across various groups when examining configural and metric invariance. Similar findings were observed for race/ethnicity and age groups in Tables 3 and 4, confirming that metric invariance was achieved across these demographic categories.

**Table 2.**

Factor Loadings Between  
Gender Groups at the  
Configural and Metric  
Levels

Test	CHC Broad Ability	Configural		Metric	
		Female	Male	Female	Male
Sound Blending	Ga	0.66	0.66	0.65	0.67
Sound Reversal	Ga	0.66	0.69	0.67	0.68
Segmentation	Ga	0.64	0.66	0.63	0.67
Spelling of Sounds	Ga	0.71	0.74	0.72	0.73
Nonsense Word Repetition	Ga	0.58	0.59	0.58	0.60
Word Attack	Ga	0.21	0.14	0.17	0.17
Oral Vocabulary	Gc	0.87	0.88	0.87	0.88
Academic Facts	Gc	0.89	0.91	0.90	0.89
Academic Vocabulary	Gc	0.89	0.90	0.89	0.90
General Information	Gc	0.76	0.74	0.74	0.76
Oral Comprehension	Gc	0.75	0.76	0.75	0.76
Picture Vocabulary	Gc	0.70	0.70	0.70	0.71
Verbal Analogies	Gc	0.55	0.55	0.55	0.56
Paragraph Reading Comprehension	Gc	0.26	0.33	0.29	0.30
Concept Formation	Gf	0.75	0.78	0.76	0.77
Analysis-Synthesis	Gf	0.73	0.75	0.72	0.76
Matrices	Gf	0.67	0.71	0.67	0.71
Understanding Directions	Gf	0.31	0.27	0.31	0.32
Verbal Analogies	Gf	0.27	0.29	0.27	0.28

**Table 2.** (cont.)  
*Factor Loadings Between  
 Gender Groups at the  
 Configural and Metric  
 Levels*

Test	CHC Broad Ability	Configural		Metric	
		Female	Male	Female	Male
Story Recall	GI	0.72	0.74	0.72	0.75
Story Comprehension	GI	0.71	0.72	0.71	0.72
Oral Language Samples	GI	0.69	0.72	0.69	0.71
Applied Problems	Gq	0.85	0.89	0.86	0.88
Number Sense	Gq	0.86	0.89	0.86	0.89
Calculation	Gq	0.81	0.84	0.81	0.84
Number Series	Gq	0.79	0.83	0.80	0.83
Math Problem Identification	Gq	0.90	0.91	0.88	0.93
Math Facts Fluency	Gq	0.32	0.31	0.30	0.32
Magnitude Comparison	Gq	0.21	0.23	0.21	0.23
Phonemic Word Retrieval	Gr	0.76	0.80	0.77	0.79
Semantic Word Retrieval	Gr	0.74	0.76	0.74	0.76
Rapid Phoneme Naming	Gr	0.56	0.58	0.55	0.59
Rapid Letter Naming	Gr	0.17	0.14	0.16	0.17
Rapid Picture Naming	Gr	0.08	0.18	0.12	0.12
Letter-Word Identification	Grw	0.79	0.83	0.81	0.81
Spelling	Grw	0.81	0.84	0.82	0.84
Oral Reading	Grw	0.82	0.83	0.81	0.84
Sentence Writing Accuracy	Grw	0.82	0.86	0.84	0.84
Written Language Samples	Grw	0.81	0.84	0.81	0.84
Word Attack	Grw	0.55	0.65	0.59	0.62
Reading Recall	Grw	0.64	0.67	0.63	0.68
Paragraph Comprehension	Grw	0.53	0.49	0.49	0.52
Passage Comprehension	Grw	0.86	0.87	0.85	0.88
Sentence Writing Fluency	Grw	0.29	0.27	0.25	0.29
Sentence Reading Fluency	Grw	0.23	0.25	0.22	0.25
Word Reading Fluency	Grw	0.13	0.18	0.14	0.16
Number-Pattern Matching	Gs	0.79	0.80	0.79	0.80
Letter-Pattern Matching	Gs	0.65	0.67	0.65	0.66
Word Reading Fluency	Gs	0.69	0.65	0.66	0.69
Magnitude Comparison	Gs	0.59	0.61	0.62	0.58
Sentence Reading Fluency	Gs	0.64	0.62	0.63	0.65
Symbol Inhibition	Gs	0.76	0.74	0.75	0.75
Rapid Number Naming	Gs	0.63	0.61	0.62	0.62
Rapid Quantity Naming	Gs	0.59	0.67	0.63	0.59
Rapid Letter Naming	Gs	0.51	0.58	0.54	0.54
Sentence Writing Fluency	Gs	0.50	0.49	0.49	0.51
Math Facts Fluency	Gs	0.48	0.49	0.49	0.46
Rapid Picture Naming	Gs	0.48	0.38	0.43	0.45
Spatial Relations	Gv	0.74	0.76	0.73	0.76
Block Rotation	Gv	0.69	0.73	0.71	0.71
Visual-Auditory Learning	Gv	0.74	0.73	0.71	0.75
Rapid Quantity Naming	Gv	0.16	0.13	0.16	0.16
Letter-Pattern Matching	Gv	0.15	0.17	0.16	0.17

**Table 2.** (cont.)  
Factor Loadings Between  
Gender Groups at the  
Configural and Metric  
Levels

Test	CHC Broad Ability	Configural		Metric	
		Female	Male	Female	Male
Verbal Attention	Gwm	0.76	0.78	0.76	0.78
Animal-Number Sequencing	Gwm	0.71	0.74	0.72	0.74
Numbers Reversed	Gwm	0.70	0.74	0.71	0.74
Memory for Words	Gwm	0.56	0.61	0.55	0.61
Sentence Repetition	Gwm	0.64	0.65	0.60	0.69
Understanding Directions	Gwm	0.45	0.51	0.44	0.47
Visual Working Memory	Gwm	0.52	0.60	0.55	0.57

**Table 3.**  
Factor Loadings Among Race/  
Ethnicity Groups at the Configural  
and Metric Levels

Test	CHC Broad Ability	Configural				Metric			
		White	African American	Hispanic	Other	White	African American	Hispanic	Other
Sound Blending	Ga	0.63	0.60	0.65	0.65	0.62	0.65	0.66	0.61
Sound Reversal	Ga	0.64	0.70	0.69	0.67	0.65	0.67	0.70	0.66
Segmentation	Ga	0.62	0.65	0.65	0.61	0.61	0.66	0.65	0.63
Spelling of Sounds	Ga	0.70	0.71	0.75	0.67	0.71	0.69	0.74	0.68
Nonsense Word Repetition	Ga	0.54	0.59	0.65	0.55	0.56	0.58	0.63	0.54
Word Attack	Ga	0.15	0.28	0.21	0.16	0.21	0.21	0.22	0.23
Oral Vocabulary	Gc	0.85	0.88	0.87	0.84	0.85	0.87	0.87	0.87
Academic Facts	Gc	0.88	0.90	0.89	0.85	0.88	0.89	0.91	0.85
Academic Vocabulary	Gc	0.87	0.90	0.89	0.86	0.86	0.89	0.92	0.86
General Information	Gc	0.71	0.75	0.74	0.68	0.71	0.74	0.75	0.68
Oral Comprehension	Gc	0.71	0.73	0.78	0.69	0.73	0.77	0.72	0.70
Picture Vocabulary	Gc	0.66	0.70	0.69	0.65	0.66	0.68	0.69	0.65
Verbal Analogies	Gc	0.54	0.57	0.57	0.54	0.54	0.57	0.56	0.53
Paragraph Reading Comprehension	Gc	0.26	0.24	0.36	0.32	0.27	0.27	0.27	0.24
Concept Formation	Gf	0.76	0.72	0.76	0.75	0.74	0.77	0.76	0.79
Analysis-Synthesis	Gf	0.72	0.75	0.72	0.74	0.72	0.71	0.76	0.70
Matrices	Gf	0.66	0.67	0.68	0.62	0.67	0.65	0.67	0.65
Understanding Directions	Gf	0.22	0.22	0.38	0.38	0.25	0.25	0.26	0.26
Verbal Analogies	Gf	0.27	0.26	0.25	0.27	0.27	0.27	0.27	0.26
Story Recall	Gl	0.71	0.76	0.72	0.65	0.70	0.77	0.72	0.67
Story Comprehension	Gl	0.68	0.73	0.69	0.64	0.68	0.70	0.69	0.69
Oral Language Samples	Gl	0.66	0.72	0.73	0.71	0.66	0.74	0.73	0.65
Applied Problems	Gq	0.85	0.88	0.87	0.83	0.86	0.87	0.85	0.86
Number Sense	Gq	0.86	0.85	0.87	0.87	0.85	0.86	0.86	0.89
Calculation	Gq	0.81	0.82	0.82	0.85	0.80	0.85	0.85	0.80
Number Series	Gq	0.79	0.82	0.82	0.82	0.81	0.78	0.81	0.82
Math Problem Identification	Gq	0.89	0.88	0.90	0.90	0.89	0.91	0.88	0.91
Math Facts Fluency	Gq	0.30	0.44	0.25	0.31	0.30	0.31	0.33	0.31
Magnitude Comparison	Gq	0.20	0.22	0.24	0.18	0.20	0.23	0.22	0.21

**Table 3.** (cont.)  
*Factor Loadings Among Race/  
 Ethnicity Groups at the Configural  
 and Metric Levels*

Test	CHC Broad Ability	Configural				Metric			
		White	African American	Hispanic	Other	White	African American	Hispanic	Other
Phonemic Word Retrieval	Gr	0.77	0.80	0.74	0.76	0.76	0.79	0.77	0.83
Semantic Word Retrieval	Gr	0.71	0.75	0.75	0.84	0.73	0.77	0.73	0.77
Rapid Phoneme Naming	Gr	0.54	0.58	0.58	0.58	0.54	0.59	0.59	0.57
Rapid Letter Naming	Gr	0.15	0.24	0.20	0.18	0.19	0.21	0.20	0.20
Rapid Picture Naming	Gr	0.10	0.17	0.13	0.07	0.12	0.14	0.14	0.14
Letter-Word Identification	Grw	0.80	0.83	0.82	0.80	0.80	0.81	0.83	0.80
Spelling	Grw	0.81	0.84	0.83	0.81	0.80	0.83	0.85	0.85
Oral Reading	Grw	0.80	0.82	0.84	0.81	0.80	0.85	0.84	0.82
Sentence Writing Accuracy	Grw	0.82	0.81	0.85	0.85	0.81	0.85	0.84	0.84
Written Language Samples	Grw	0.80	0.83	0.85	0.84	0.81	0.82	0.84	0.77
Word Attack	Grw	0.62	0.49	0.57	0.58	0.54	0.56	0.58	0.61
Reading Recall	Grw	0.63	0.66	0.67	0.63	0.64	0.65	0.66	0.61
Paragraph Comprehension	Grw	0.51	0.58	0.46	0.48	0.52	0.53	0.54	0.50
Passage Comprehension	Grw	0.83	0.89	0.88	0.85	0.84	0.86	0.86	0.87
Sentence Writing Fluency	Grw	0.28	0.28	0.22	0.22	0.24	0.29	0.27	0.24
Sentence Reading Fluency	Grw	0.21	0.23	0.21	0.31	0.21	0.24	0.24	0.22
Word Reading Fluency	Grw	0.12	0.14	0.18	0.14	0.13	0.15	0.15	0.14
Number-Pattern Matching	Gs	0.80	0.78	0.79	0.83	0.80	0.77	0.81	0.78
Letter-Pattern Matching	Gs	0.67	0.64	0.67	0.63	0.68	0.65	0.65	0.62
Word Reading Fluency	Gs	0.69	0.66	0.64	0.64	0.68	0.65	0.68	0.69
Magnitude Comparison	Gs	0.63	0.56	0.58	0.70	0.62	0.58	0.61	0.61
Sentence Reading Fluency	Gs	0.64	0.60	0.67	0.55	0.64	0.61	0.63	0.65
Symbol Inhibition	Gs	0.74	0.70	0.71	0.69	0.74	0.70	0.71	0.73
Rapid Number Naming	Gs	0.61	0.65	0.64	0.62	0.62	0.59	0.64	0.61
Rapid Quantity Naming	Gs	0.62	0.56	0.62	0.55	0.62	0.56	0.62	0.59
Rapid Letter Naming	Gs	0.56	0.48	0.52	0.46	0.53	0.48	0.52	0.50
Sentence Writing Fluency	Gs	0.48	0.45	0.54	0.56	0.51	0.49	0.48	0.47
Math Facts Fluency	Gs	0.50	0.33	0.50	0.51	0.49	0.42	0.49	0.48
Rapid Picture Naming	Gs	0.45	0.40	0.46	0.40	0.43	0.39	0.43	0.43
Spatial Relations	Gv	0.72	0.74	0.75	0.70	0.73	0.72	0.75	0.68
Block Rotation	Gv	0.70	0.70	0.69	0.70	0.69	0.71	0.71	0.63
Visual-Auditory Learning	Gv	0.72	0.70	0.76	0.67	0.72	0.71	0.73	0.74
Rapid Quantity Naming	Gv	0.15	0.16	0.14	0.16	0.15	0.15	0.16	0.15
Letter-Pattern Matching	Gv	0.17	0.11	0.16	0.17	0.16	0.17	0.16	0.15

**Table 3.** (cont.)  
*Factor Loadings Among Race/  
 Ethnicity Groups at the Configural  
 and Metric Levels*

Test	CHC Broad Ability	Configural				Metric			
		White	African American	Hispanic	Other	White	African American	Hispanic	Other
Verbal Attention	Gwm	0.74	0.73	0.78	0.74	0.74	0.76	0.78	0.71
Animal-Number Sequencing	Gwm	0.69	0.71	0.74	0.65	0.68	0.71	0.74	0.72
Numbers Reversed	Gwm	0.71	0.70	0.74	0.71	0.71	0.70	0.76	0.70
Memory for Words	Gwm	0.55	0.58	0.59	0.55	0.55	0.56	0.60	0.55
Sentence Repetition	Gwm	0.59	0.62	0.68	0.63	0.61	0.62	0.64	0.59
Understanding Directions	Gwm	0.50	0.53	0.42	0.39	0.48	0.48	0.52	0.50
Visual Working Memory	Gwm	0.55	0.57	0.56	0.56	0.55	0.57	0.57	0.53

**Table 4.**  
*Factor Loadings Among Age  
 Groups at the Configural and  
 Metric Levels*

Test	CHC Broad Ability	Configural					Metric				
		Ages 6–9	Ages 10–14	Ages 15–19	Ages 20–49	Ages 50–80+	Ages 6–9	Ages 10–14	Ages 15–19	Ages 20–49	Ages 50–80+
Sound Blending	Ga	0.61	0.61	0.69	0.71	0.67	0.68	0.66	0.64	0.64	0.67
Sound Reversal	Ga	0.64	0.66	0.69	0.66	0.72	0.65	0.64	0.66	0.71	0.74
Segmentation	Ga	0.71	0.60	0.57	0.65	0.72	0.58	0.62	0.65	0.66	0.70
Spelling of Sounds	Ga	0.72	0.74	0.69	0.76	0.75	0.78	0.70	0.71	0.73	0.74
Nonsense Word Repetition	Ga	0.61	0.58	0.61	0.57	0.59	0.60	0.57	0.57	0.61	0.62
Word Attack	Ga	0.14	0.14	0.27	0.16	0.22	0.13	0.16	0.16	0.17	0.17
Oral Vocabulary	Gc	0.88	0.86	0.86	0.89	0.89	0.81	0.87	0.89	0.91	0.89
Academic Facts	Gc	0.88	0.90	0.90	0.91	0.89	0.89	0.86	0.87	0.95	0.94
Academic Vocabulary	Gc	0.90	0.87	0.90	0.89	0.89	0.85	0.91	0.90	0.91	0.88
General Information	Gc	0.67	0.73	0.76	0.80	0.78	0.79	0.75	0.75	0.74	0.71
Oral Comprehension	Gc	0.72	0.78	0.75	0.78	0.75	0.73	0.76	0.75	0.77	0.78
Picture Vocabulary	Gc	0.67	0.66	0.71	0.73	0.75	0.71	0.70	0.72	0.71	0.68
Verbal Analogies	Gc	0.57	0.55	0.48	0.63	0.53	0.55	0.58	0.58	0.62	0.60
Paragraph Reading Comprehension	Gc	0.25	0.31	0.32	0.32	0.23	0.20	0.23	0.24	0.25	0.26
Concept Formation	Gf	0.73	0.76	0.79	0.77	0.77	0.71	0.76	0.76	0.79	0.81
Analysis-Synthesis	Gf	0.72	0.73	0.74	0.77	0.74	0.69	0.74	0.74	0.78	0.75
Matrices	Gf	0.60	0.69	0.67	0.72	0.73	0.66	0.69	0.69	0.72	0.66
Understanding Directions	Gf	0.34	0.39	0.17	0.21	0.20	0.29	0.29	0.31	0.29	0.32
Verbal Analogies	Gf	0.23	0.26	0.34	0.25	0.32	0.22	0.25	0.25	0.25	0.24
Story Recall	Gi	0.76	0.70	0.73	0.73	0.75	0.70	0.72	0.73	0.75	0.75
Story Comprehension	Gi	0.67	0.71	0.73	0.77	0.70	0.70	0.71	0.72	0.73	0.73
Oral Language Samples	Gi	0.66	0.71	0.73	0.73	0.68	0.68	0.69	0.73	0.74	0.66

**Table 4.** (cont.)  
*Factor Loadings Among Age Groups at the Configural and Metric Levels*

Test	CHC Broad Ability	Configural					Metric				
		Ages 6-9	Ages 10-14	Ages 15-19	Ages 20-49	Ages 50-80+	Ages 6-9	Ages 10-14	Ages 15-19	Ages 20-49	Ages 50-80+
Applied Problems	Gq	0.87	0.86	0.89	0.87	0.85	0.88	0.87	0.87	0.85	0.86
Number Sense	Gq	0.87	0.86	0.88	0.86	0.89	0.86	0.87	0.91	0.86	0.84
Calculation	Gq	0.79	0.84	0.86	0.82	0.83	0.83	0.82	0.83	0.81	0.85
Number Series	Gq	0.82	0.82	0.81	0.82	0.82	0.73	0.79	0.85	0.84	0.85
Math Problem Identification	Gq	0.88	0.89	0.93	0.92	0.91	0.92	0.91	0.90	0.91	0.90
Math Facts Fluency	Gq	0.37	0.28	0.28	0.24	0.35	0.38	0.36	0.36	0.33	0.35
Magnitude Comparison	Gq	0.23	0.23	0.20	0.20	0.26	0.27	0.23	0.20	0.19	0.23
Phonemic Word Retrieval	Gr	0.82	0.77	0.74	0.79	0.80	0.71	0.75	0.81	0.85	0.78
Semantic Word Retrieval	Gr	0.74	0.73	0.79	0.76	0.72	0.71	0.72	0.74	0.78	0.81
Rapid Phoneme Naming	Gr	0.46	0.52	0.57	0.64	0.63	0.61	0.55	0.55	0.55	0.56
Rapid Letter Naming	Gr	0.13	0.14	0.17	0.14	0.22	0.19	0.16	0.15	0.17	0.17
Rapid Picture Naming	Gr	0.07	0.15	0.15	0.21	0.11	0.10	0.09	0.08	0.09	0.09
Letter-Word Identification	Grw	0.89	0.81	0.77	0.77	0.82	0.78	0.80	0.80	0.85	0.83
Spelling	Grw	0.86	0.80	0.80	0.82	0.83	0.89	0.79	0.78	0.84	0.83
Oral Reading	Grw	0.85	0.81	0.80	0.81	0.82	0.92	0.82	0.77	0.80	0.79
Sentence Writing Accuracy	Grw	0.85	0.84	0.80	0.85	0.83	0.90	0.80	0.81	0.80	0.85
Written Language Samples	Grw	0.86	0.81	0.80	0.81	0.82	0.86	0.82	0.81	0.78	0.85
Word Attack	Grw	0.68	0.62	0.49	0.59	0.57	0.65	0.62	0.60	0.63	0.57
Reading Recall	Grw	0.77	0.63	0.62	0.62	0.62	0.67	0.63	0.61	0.67	0.68
Paragraph Comprehension	Grw	0.60	0.51	0.45	0.48	0.52	0.63	0.56	0.54	0.54	0.59
Passage Comprehension	Grw	0.91	0.85	0.84	0.85	0.86	0.91	0.89	0.85	0.81	0.83
Sentence Writing Fluency	Grw	0.29	0.23	0.24	0.30	0.36	0.31	0.20	0.17	0.16	0.19
Sentence Reading Fluency	Grw	0.23	0.28	0.22	0.24	0.27	0.24	0.15	0.12	0.11	0.14
Word Reading Fluency	Grw	0.24	0.16	0.10	0.13	0.16	0.22	0.14	0.11	0.11	0.14
Number-Pattern Matching	Gs	0.66	0.81	0.84	0.82	0.77	0.66	0.75	0.81	0.80	0.80
Letter-Pattern Matching	Gs	0.53	0.68	0.70	0.66	0.63	0.54	0.65	0.69	0.71	0.69
Word Reading Fluency	Gs	0.49	0.63	0.73	0.73	0.68	0.52	0.66	0.72	0.68	0.68
Magnitude Comparison	Gs	0.45	0.58	0.68	0.63	0.57	0.48	0.57	0.60	0.63	0.63
Sentence Reading Fluency	Gs	0.50	0.56	0.65	0.66	0.62	0.60	0.73	0.77	0.73	0.71
Symbol Inhibition	Gs	0.72	0.69	0.77	0.77	0.78	0.51	0.67	0.80	0.81	0.74
Rapid Number Naming	Gs	0.49	0.60	0.64	0.67	0.64	0.49	0.59	0.63	0.67	0.58
Rapid Quantity Naming	Gs	0.41	0.60	0.63	0.67	0.60	0.54	0.61	0.64	0.70	0.66
Rapid Letter Naming	Gs	0.47	0.54	0.53	0.58	0.56	0.42	0.51	0.55	0.57	0.51
Sentence Writing Fluency	Gs	0.41	0.49	0.54	0.51	0.43	0.44	0.56	0.61	0.59	0.55
Math Facts Fluency	Gs	0.40	0.50	0.51	0.55	0.47	0.29	0.39	0.46	0.46	0.42
Rapid Picture Naming	Gs	0.34	0.39	0.43	0.43	0.47	0.35	0.46	0.50	0.50	0.46

**Table 4.** (cont.)  
*Factor Loadings Among Age Groups at the Configural and Metric Levels*

Test	CHC Broad Ability	Configural					Metric				
		Ages 6–9	Ages 10–14	Ages 15–19	Ages 20–49	Ages 50–80+	Ages 6–9	Ages 10–14	Ages 15–19	Ages 20–49	Ages 50–80+
Spatial Relations	Gv	0.71	0.76	0.78	0.73	0.74	0.74	0.73	0.74	0.76	0.77
Block Rotation	Gv	0.66	0.73	0.70	0.72	0.74	0.68	0.71	0.71	0.71	0.75
Visual-Auditory Learning	Gv	0.77	0.70	0.72	0.74	0.75	0.73	0.74	0.76	0.74	0.72
Rapid Quantity Naming	Gv	0.17	0.17	0.21	0.11	0.17	0.15	0.12	0.10	0.10	0.12
Letter-Pattern Matching	Gv	0.19	0.12	0.16	0.20	0.16	0.17	0.15	0.12	0.12	0.14
Verbal Attention	Gwm	0.80	0.73	0.77	0.79	0.77	0.72	0.77	0.78	0.80	0.77
Animal-Number Sequencing	Gwm	0.69	0.72	0.74	0.73	0.75	0.72	0.73	0.73	0.74	0.70
Numbers Reversed	Gwm	0.76	0.70	0.71	0.72	0.74	0.65	0.71	0.74	0.76	0.74
Memory for Words	Gwm	0.59	0.60	0.57	0.58	0.58	0.57	0.56	0.60	0.62	0.59
Sentence Repetition	Gwm	0.63	0.65	0.64	0.65	0.67	0.67	0.66	0.65	0.63	0.63
Understanding Directions	Gwm	0.40	0.38	0.60	0.60	0.55	0.48	0.45	0.49	0.44	0.49
Visual Working Memory	Gwm	0.51	0.57	0.59	0.55	0.56	0.61	0.52	0.52	0.55	0.61

## Discussion

This study highlights the process of evaluating measurement invariance and the importance in ensuring that tests measure the same constructs consistently across different demographic groups. Our findings demonstrated that the WJ V maintains a consistent 10 broad CHC structure and test-to-CHC factor loadings across race/ethnicity, gender, and age groups, supporting the validity of the test for diverse populations.

This study has significant implications for both research and practical applications. From a research perspective, it is essential to understand how effectively the measurement model operates within different groups of examinees and whether it maintains its validity and reliability across diverse populations. Ensuring measurement consistency across groups allows researchers to make meaningful comparisons and draw accurate conclusions. In practice, this knowledge helps educators and test developers design assessments that are fair, unbiased, and applicable to a wide range of individuals, ultimately improving the overall quality and equity of measurement tools.

Practitioners must carefully assess the significance of invariance violations by determining the extent to which these violations impact the validity and fairness of the test for its intended use. Not all violations necessarily undermine the test's effectiveness; however, if they result in biased measurements, misinterpretations of scores, or unfair advantages or disadvantages for specific groups, they must be addressed. Evaluating the practical implications of these violations ensures that the test remains a reliable tool for making informed decisions, whether in educational, psychological, or other assessment settings.

The findings of this study indicate that the WJ V tests assess comparable constructs across different demographic groups. However, we did not assess scalar invariance or other stricter levels of invariance due to the WJ V's complexity and the large number of individual tests it contains. Future research should explore stricter constraints, such as examining test intercepts across demographic groups. Additionally, this study focused on examining the construct invariance for the Horn no-g broad CHC model. Further investigations of the Carroll hierarchical g broad CHC model and the Carroll hierarchical g broad+narrow CHC model would provide additional insights into whether these factor structures hold invariance across groups.

By providing a basic overview of measurement invariance methods, this paper aims to enhance school psychologists' knowledge of structural test validity. When selecting cognitive assessment tools, it is important to select test batteries that demonstrate they measure the same constructs across diverse populations. Valid test results are essential for accurately assessing and monitoring an examinee's progress throughout their educational journey. This ensures fairness and equity in educational assessments, enabling educators and psychologists to make valid and reliable inferences about students from various backgrounds.

# References

- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. AERA.
- Arbuckle, J. L. (2022). *Amos 29.0 User's Guide*. IBM SPSS.
- Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. Guilford Press.
- Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, 14(3), 464–504.
- Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*, 9(2), 233–255.
- Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural equation modeling: A Multidisciplinary Journal*, 6(1), 1–55.
- Jöreskog, K. G. (1971). Simultaneous factor analysis in several populations. *Psychometrika*, 36(4), 409–426.
- LaForte, E. M., Dailey, D., & McGrew, K. S. (2025). *Technical Manual. Woodcock-Johnson V*. Riverside Assessments, LLC.
- McGrew, K. S., LaForte, E. M., & Schrank, F. A. (2014). *Technical Manual. Woodcock-Johnson IV*. Riverside Assessments, LLC.
- McGrew, K. S., Mather, N., LaForte, E. M., & Wendling, B. J. (2025). *Woodcock-Johnson V*. Riverside Assessments, LLC.
- 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, 71–90.
- Schmitt, N., & Kuljanin, G. (2008). Measurement invariance: Review of practice and implications. *Human Resource Management Review*, 18(4), 210–222.
- Schneider, W. J., & McGrew, K. S. (2018). The Cattell-Horn-Carroll theory of cognitive abilities. In D. P. Flanagan & E. M. McDonough (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (3rd ed., pp. 73–163). Guilford Press.
- Schrank, F. A., McGrew, K. S., & Mather, N. (2014). *Woodcock-Johnson IV*. Riverside Assessments, LLC.
- Xia, Y., & Yang, Y. (2019). RMSEA, CFI, and TLI in structural equation modeling with ordered categorical data: The story they tell depends on the estimation methods. *Behavior Research Methods*, 51, 409–428.
- Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*, 3(1), 4–70.



800.323.9540

[https://www.riversideinsights.com/woodcock\\_johnson\\_v](https://www.riversideinsights.com/woodcock_johnson_v)