Psychological Assessment
2004, Vol. 16, No. 1, 85–89

1040-3590/04/$12.00 DOI: 10.1037/1040-3590.16.1.85

A Confirmatory Analysis of the Factor Structure and Cross-Age Invariance of the Wechsler Adult Intelligence Scale—Third Edition

Gordon E. Taub
University of Central Florida

Kevin S. McGrew
University of Minnesota

E. Lea Witta
University of Central Florida

In the Wechsler Adult Intelligence Scale—Third Edition (WAIS–III; D. Wechsler, 1997), the manual reports several confirmatory factor analyses in support of the instrument's latent factor structure. In practice, examiners frequently compare an examinee's score from a current administration of the WAIS–III with the results from a previous test administration. Implicit in test–retest score comparisons is evidence that scores retain similar interpretive meaning across time. Establishing an instrument's factorial invariance provides the foundation for this practice. This study investigated the factorial invariance of the WAIS–III across the instrument's 13 age groups. The overall results from this study generally support both configural and factorial invariance of the WAIS–III when the 11 primary tests are administered.

The Wechsler Adult Intelligence Scale—Third Edition (WAIS–III; Wechsler, 1997) represents the latest edition of this intelligence battery. The Wechsler Adult Intelligence Scale (WAIS; Wechsler, 1955), in its various incarnations, is perhaps the most widely used test of adult intelligence (Boake, 2002; Sattler, 2001).

Even though the basic format of the WAIS–III remains the same (68% of the items were retained from the WAIS–R; Wechsler, 1981), the WAIS–III represents a significant departure from the WAIS–R. Similar to the Wechsler Intelligence Scale for Children—Third Edition (WISC–III; Wechsler, 1991), the WAIS–III incorporates four index scores (i.e., Verbal Comprehension, Perceptual Organization, Working Memory, and Processing Speed) rather than the three scores used in the WAIS–R. The new instrument's factor structure is more in line with the design of contemporary intelligence tests and psychometrically based hierarchical models of intelligence (see Carroll, 1993, 1997; McGrew & Woodcock, 2001). The new factor structure of the WAIS–III measures specific cognitive abilities associated with each of the 11 tests, which in turn are subsumed by one of the four first-order broad factors. At the apex of the model is a general factor of intelligence or *g*.

To support the construct validity of the four-factor model of the WAIS–III, the authors offer extensive confirmatory factor analyses across all of the instrument's 13 age groups, which range from 16 to 89 years of age (The Psychological Corporation, 1997). Even with these significant revisions, however, the WAIS–III leaves a simple, yet important question unanswered. Especially pertinent to the practice of applied intellectual assessment is the lack of empirical evidence to support the interpretation of the primary WAIS–III tests as measures of the same latent first-order cognitive constructs throughout the instrument's 13 age groups. In other words, empirical evidence of the factorial invariance of the WAIS–III is not contained in the instrument's technical manual.

The purpose of this study was threefold. The first purpose was to investigate the configural invariance of the four-factor theoretical model across the WAIS–III's 13 age-differentiated groups. The second purpose was to test the invariance or stability of the four-factor *theoretical* model across the instrument's 13 age groups (e.g., is the structure of the theoretical four-factor model the same for 16–17-year-olds as it is for 85–89-year-olds?). The third purpose was to test the invariance of the *measurement* model of the WAIS–III's four-factor theoretical model across the instrument's 13 age-differentiated groups (i.e., from age 16 to 89). This was an investigation of the extent to which the 11 primary tests included in the calculation of the first-order factors of the WAIS–III are equally valid measures of the four theoretical factors across all 13 age groups.

Method

The sample for this study consisted of the standardization sample of the WAIS–III. The standardization sample of the WAIS–III mirrored characteristics of the population of late adolescents and adults in the United States, as described by U.S. Census reports in the early to mid-1990s. The WAIS–III was standardized on 2,450 individuals. Data from these 2,450 individuals were used in the present data analyses in 13 age-differentiated groups; these are grouped into ages 16 through 17, 18 through 19, 20 through 24, 25 through 29, 30 through 34, 35 through 44, 45 through 54, 55 through 64, 65 through 69, 70 through 74, 75 through 79, 80 through 84, and 85 through 89.

Gordon E. Taub and E. Lea Witta, College of Education, University of Central Florida; Kevin S. McGrew, Department of Psychology, University of Minnesota.

Correspondence concerning this article should be addressed to Gordon E. Taub, CFCS Education TR544, University of Central Florida, 4000 Central Florida Boulevard, Orlando, FL 32816. E-mail: gtaub@mail.ucf.edu

Table 1
Summary of Results of the Wechsler Adult Intelligence Scale—Third Edition's Four First-Order Factor Structure and the Stability and Invariance of the Factor Structure Across Age Groups

Model	χ^2	*df*	GFI	CFI	TLI	RMSEA	$\Delta\chi^2$	$\Delta(df)$	*p*
Configural	525.52	523	.915	.950	.932	.025			
Invariance 1	545.52	560	.912	.949	.935	.024	20.00	37	$>.05$
Invariance 2	581.32	642	.908	.949	.943	.022	55.80	119	$>.05$

Note. GFI = goodness-of-fit index; CFI = comparative fit index; TLI = Tucker–Lewis index; RMSEA = root-mean-square error of approximation; Invariance 1 = invariant on first-order factors; Invariance 2 = invariant on second- and first-order factors.

The psychometric properties of the WAIS–III subtests have been termed "excellent" (Sattler & Ryan, 1999). The highest reliabilities are found among the six standard subtests comprising the Verbal IQ Index (VIQ) score, with average reliabilities across ages ranging from .84 to .93. Average reliabilities across the five standard subtests comprising the Performance IQ Index (PIQ) are a bit lower, but still quite high, ranging from .74 to .90. For the VIQ, PIQ, and the Full-Scale IQ (FSIQ), reported average reliabilities are excellent: .97, .94, and .98, respectively. The reliabilities of the four factors of the WAIS–III range from .88 for the Processing Speed Index, to .93 for the Perceptual Organization Index, .94 for the Working Memory Index, to the highest reliability for the Verbal Comprehension Index of .96. Both the Comprehension and Picture Arrangement tests, although not included in the theoretical four-factor model, must be administered to obtain an examinee's VIQ, PIQ, and FSIQ scores.

To answer the present research questions, we used the average intercorrelations of WAIS–III subtest scale scores, as provided in the technical manual, and converted them to variance–covariance matrices for all data analyses. Using confirmatory factor analysis (CFA) via the AMOS program (Arbuckle & Wothke, 1999), we analyzed the variance–covariance matrices for the average correlation matrix, as well as the matrices from all 13 age-differentiated groups.

Three consecutively more restrictive analyses of invariance were run. In the configural model, the first set of analyses, the number of second- and first-order factors and the assignment of each of the 11 tests to their associated first-order factors were investigated. This *configural* invariance analyses specified that the same factor structure exists across all age groups (Horn, McArdle, & Mason, 1983). The paths leading from the second-order general factor to the four first-order factors and the paths leading to each of the manifest 11 subtests were specified to be the same across all of the instrument's age groups. This analysis investigated the extent to which the 11 tests measure the same latent factors from age 16 through 89.

A more restrictive test of invariance is *metric* invariance, which requires that all factor loadings are equal across all age groups (Bollen, 1989). Two tests of metric invariance were conducted. The first test, Invariance 1, specified that the paths from the four first-order latent factors to the 11 manifest tests were invariant across all 13 age groups; however, the paths from the second-order general factor to the four latent first-order factors were allowed to be free or to vary. The next set of analyses, Invariance 2, extended the constraint of paths in Invariance 1, to also include that the paths from the second-order general factor to the latent first-order factors be invariant across all 13 age groups of the WAIS–III.

The statistical analyses of the data required a comparison of the fit between each of the metric invariance models with the fit of the configural invariance model. The finding of a nonsignificant fit (as determined by the difference between the respective model chi-squares and degrees of freedom) supports the null hypothesis that there is *not* a difference between models. This finding supports interpretation of metric invariance. Several goodness-of-fit indices were used to evaluate the models, including the goodness-of-fit index (GFI), the Tucker–Lewis index (TLI, also called the nonnormed fit index), the root-mean-square error of approximation (RMSEA), and the comparative fit index (CFI; Keith, 1997; Keith & Witta, 1997). To avoid the rejection of potentially good models, such as with the large samples used in this study ($N = 2{,}450$), we used the differential fit value (DFV), a conversion of the chi-square statistic based on a sample size of 1,000, to evaluate the goodness of fit of all models (see Keith & Witta, 1997; Taub & McGrew, in press).[1]

Results

During the initial visual analysis of the configural model, a path coefficient greater than 1.0 was found between Working Memory and *g* for the 75–79 and 25–29 age groups. Another path coefficient greater than 1.0 was also found between the latent Perceptual Organization factor and *g* at the 45–54 age group. These findings suggest that Working Memory and Perceptual Organization are isomorphic with *g* at each of these respective age groups and represent a "Heywood" case (Loehlin, 1992; Long, 1983).

To provide for proper model identification in the present investigation, we specified (fixed) the Working Memory factor's loading on *g* to be unity and fixed the error variance associated with the latent Working Memory factor to zero for the 75–79 and 25–29 age groups; this was also done for the Perceptual Organization factor for the 45–54 age group. These model specifications were maintained in all subsequent analyses.

The results of the test of configural invariance are presented in Table 1 and indicate the hypothesis that the theoretical model fits the data for all WAIS–III age groups could *not* be rejected ($p > .05$). The tests of metric invariance (Invariance 1 and Invariance 2) are also presented in Table 1. The change in chi-square and degrees of freedom between the configural model, the test of configural invariance, and Invariance 1, the test of first-order metric invariance (49.04), was *not* significant ($p > .05$). Invariance 2, the most restrictive test of invariance, required the path coefficients from the second-order factor to first-order factors and from the first-order factors to the manifest tests to be invariant across all 13 age groups. This was a test of the invariance of the *theoretical* and *measurement* model of the WAIS–III. The results from this analysis also indicated that the null hypothesis could *not* be rejected. The finding of a nonsignificant chi-square difference supports the interpretation of metric invariance, with the exception of the previously noted fixed loading of 1.0+ for Working Mem-

[1] The DFV was obtained by applying the formula $((\chi^2) / (n - 1)) \times (1000 - 1)$. For example, the actual χ^2 for Model 1 was 1,288.28. The DFV was calculated by applying the formula $((1288.28) / (2450 - 1)) \times (1000 - 1) = 525.52$.

ory on *g* in the 75–79 and 25–29 age groups and for Perceptual Organization and *g* for the adult sample age 45–54. Figure 1 presents the theoretical and measurement model of the WAIS–III and the associated *standardized* path coefficients for the 18–19 age group.

Discussion

The results of this study, which are based on the correlation matrices provided in the WAIS–III's technical manual (The Psychological Corporation, 1997), indicate that the same *pattern* of loadings exists from the second-order general factor to the four first-order index factors and from the first-order latent index factors to the 11 primary tests. These results support the configural invariance of the theoretical model across the instrument's 13 age-differentiated groups (from 16 to 89 years of age).

Statistical tests of progressively more restrictive metric invariance across all age groups supported the presence of identical factor loadings from the first-order factors to the 11 manifest tests and from the latent second-order general factor to the four latent first-order factors, with the exclusion of the aforementioned 1.0+ factor loadings on *g* for the Working Memory and Perceptual Organization factors. These findings indicate that in addition to the WAIS–III tests measuring the same pattern of abilities across the instrument's wide age range, the test factor loadings do not change as a function of age.

The finding that the subsamples are invariant was surprising. This mainly stemmed from the fact that in many CFAs the authors collapsed the 13 age groups into one of three broader age groups (16–29, 30–64, and 65–89; The Psychological Corporation, 1997). Therefore, it was expected that the matrices within each of the three age groups would be invariant but that there would not be invariance across all ages (16–89 years).

It is interesting to note that the WAIS–III, like the WISC–III, was developed using a measurement model which through data

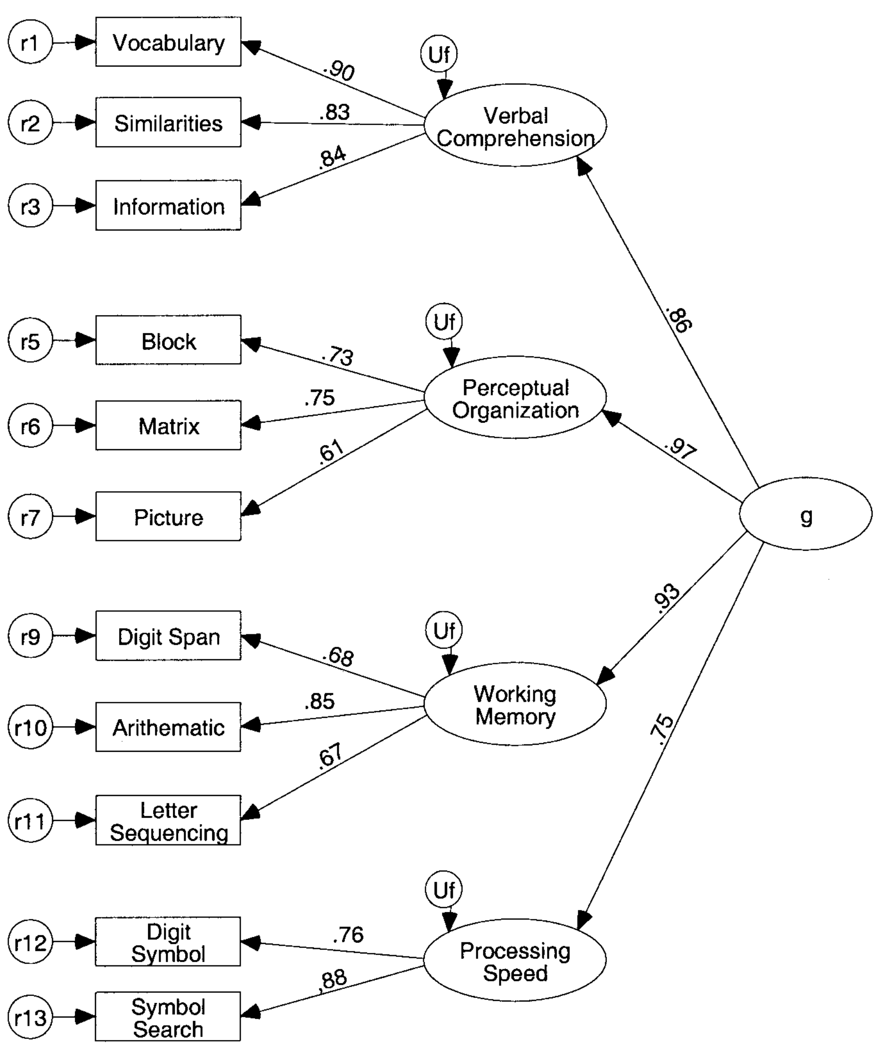

Figure 1. Theoretical and measurement model of the Wechsler Adult Intelligence Scale—Third Edition and the associated standardized path coefficients for the 18–19 age group. r = subtest unique and error variance; Uf = factor unique variance.

analysis yielded a four-factor model. In a previous study investigating the WISC–III, Keith and Witta (1997) also found the four-factor model of the WISC–III (ages 6–16) to be invariant. In contrast to using a measurement model, the Woodcock–Johnson Tests of Cognitive Abilities III, which has an age range inclusive of the WAIS–III and the WISC–III, was developed using a theoretical seven-factor model based on the Cattell–Horn–Carroll theory of intelligence. This cognitive battery also demonstrated metric invariance across all of the instrument's age groups (Taub & McGrew, in press). The use of a theoretical model assists in test development because the theoretical factor structure underlying the instrument serves as the test's blueprint. It is from this blueprint that subtests are developed to measure the various aspects of the theoretical model. Because the WAIS–III was not developed using an explicit theoretical blueprint during development, the observance of metric invariance across such a wide age span is impressive.

The results from these analyses support our contention that the WAIS–III can be used across the instrument's wide age range, and more importantly, it is possible for the practitioner to compare the results from an examinee's four-factor index scores across evaluations when the results are based on differentiated age groups. Therefore, the practitioner can be confident that each of the four first-order factors are measuring the same latent constructs from age 16 through 89. This finding meets two very important standards of the American Educational Research Association, American Psychological Association, and National Counsel on Measurement in Education, specifically, Standards 7.1 and 7.8, which recommend that test scores obtained across different subgroups be interpreted as having similar meaning only if there is empirical evidence to support the finding of invariant meaning of scores across the groups.

The observation of metric invariance is impressive; however, these positive findings need to be tempered. Although the four WAIS–III factor indices provide invariant measurement across all ages, the WAIS–III does not measure a wide breadth of human cognitive abilities. The need for an intelligence test to measure several cognitive abilities is exemplified by the contemporary Cattell–Horn–Carroll theory of intelligence (see Flanagan, McGrew, & Ortiz, 2000) and most recently by the changes to the factor structure and method of scoring the recently published Wechsler Intelligence Scale for Children—Fourth Edition (Wechsler, 2003). These cautions are consistent with Boake's (2002) conclusion that

> it is paradoxical that the Wechsler–Bellevue scale, which was a model of technical innovation in 1939, represents in its current revision one of the oldest mental tests in continuous use. The intelligence scale that is relied upon to make medical, educational, and legal decisions does not reflect advances in understanding of cognitive functioning during the past 60 years and contains tests from the 1800s. (p. 401)

There are also several limitations that suggest future study. First, specific explanations for the +1.0 loading on the Working Memory factor for the 75–79 and 25–29 age groups and on the Perceptual Organization factor for the 44–45 age group are currently undetermined. Second, the Perceptual Speed factor may be underrepresented on the WAIS–III. Although the authors state "at least two or three variables are needed to define a stable factor" (The Psychological Corporation, 1997, p. 112), standard factor-analytic rules-of-thumb recommend when identifying a factor model that at least three indicators be used for each latent factor (Floyd & Widaman, 1995; Marsh, Hau, Balla, & Grayson, 1998; Raykov & Widaman, 1995). The importance of this should not be lost on the clinician interpreting an examinee's profile scores on the WAIS–III. Because the Digit Symbol–Coding and Symbol Search subtests measure different narrow areas of the broad Perceptual Speed factor, it is possible for an individual to demonstrate a statistically significant weakness on one subtest when compared with the other. When this occurs, it is not possible for the clinician to accurately interpret an examinee's score at the broad factor level because the examinee may (a) have a weakness in the broad factor that is masked by the high score on one narrow area of Perceptual Speed or (b) have a specific weakness in the narrow area measured by the subtest but not a broad Perceptual Speed weakness. Another limitation is that the Comprehension and Picture Arrangement subtests are not included in the four-factor model. However, both tests must be administered to calculate an examinee's FSIQ. Of course, this only becomes a limitation if (a) the clinician is interpreting scores using the four-factor indices, (b) an FSIQ score is required, and (c) the calculation of a brief FSIQ score (i.e., general ability index; Tulsky, Saklofske, Wilkins, & Weiss, 2001) does not meet the examiner's needs. Psychometrically, however, an evaluation of the same model with additional indicators for the Processing Speed factor and the inclusion of both the Comprehension and Picture Arrangement subtests may be the next step for future analysis.

References

Arbuckle, J. L., & Wothke, W. (1999). *AMOS users guide Version 4.0.* Chicago: SmallWaters.

Boake, C. (2002). From the Binet–Simon to the Wechsler–Bellevue: Tracing the history of intelligence testing. *Journal of Clinical and Experimental Neuropsychology, 24,* 383–405.

Bollen, K. A. (1989). *Structural equations with latent variables.* New York: Wiley.

Carroll, J. B. (1993). *Human cognitive abilities: 25A survey of factor analytic studies.* New York: Cambridge University Press.

Carroll, J. B. (1997). The three-stratrum theory of cognitive abilities. In D. P. Flanagan, J. L. Genshaft, & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 122–130). New York: Guilford Press.

Flanagan, D. P., McGrew, K. S., & Ortiz, S. (2000). *The Wechsler Intelligence Scales and Gf-Gc theory: A contemporary approach to interpretation.* Needham Heights, MA: Allyn & Bacon.

Floyd, F. J., & Widaman, K. F. (1995). Factor analysis in the development and refinement of clinical assessment instruments. *Psychological Assessment, 7,* 286–299.

Horn, J. L., McArdle, J. J., & Mason, R. (1983). When is invariance not invariant: A practical scientist's look at the ethereal concept of factor invariance. *The Southern Psychologist, 1,* 179–188.

Keith, T. Z. (1997). Using confirmatory factor analysis to aid in understanding the constructs measured by intelligence tests. In D. P. Flanagan, J. L. Genshaft, & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 373–402). New York: Guilford Press.

Keith, T. Z., & Witta, E. L. (1997). Hierarchical and cross-age confirmatory factor analysis of the WISC–III: What does it measure? *School Psychology Quarterly, 12,* 89–107.

Loehlin, J. C. (1992). *Latent variable models: An introduction to factor, path and structural analyses.* Hillsdale, NJ: Erlbaum.

Long, J. S. (1983). *Confirmatory factor analysis.* Beverly Hills, CA: Sage.

Marsh, H. W., Hau, K. T., Balla, J. R., & Grayson, D. (1998). Is more ever too much? The number of indicators per factor in confirmatory factor analysis. *Multivariate Behavioral Research, 33,* 181–220.

McGrew, K. S., & Woodcock, R. W. (2001). *Technical manual: Woodcock–Johnson III.* Itasca, IL: Riverside.

The Psychological Corporation. (1997). *WAIS–III WMS–III technical manual.* San Antonio, TX: Author.

Raykov, T., & Widaman, K. F. (1995). Issues in applied structural equation modeling research. *Structural Equation Modeling, 2,* 289–328.

Sattler, J. M. (2001). *Assessment of Children Cognitive Applications—Fourth Edition.* La Mesa, CA: Author.

Sattler, J. M., & Ryan, J. J. (1999). *Assessment of Children, Revised: WAIS–III supplement.* Le Mesa, CA: Jerome M. Sattler.

Taub, G. E., & McGrew, K. S. (in press). A confirmatory factor analysis of CHC theory and cross-age invariance of the Woodcock–Johnson Tests of Cognitive Abilities III. *School Psychology Quarterly.*

Tulsky, D. S., Saklofske, D. H., Wilkins, C., & Weiss, L. G. (2001). Development of a general ability index for the Wechsler Adult Intelligence Scale—Third Edition. *Psychological Assessment, 13,* 655–571.

Wechsler, D. (1955). *The Wechsler Adult Intelligence Scale.* San Antonio, TX: The Psychological Corporation.

Wechsler, D. (1981). *Wechsler Adult Intelligence Scale—Revised.* San Antonio, TX: The Psychological Corporation.

Wechsler, D. (1991). *Wechsler Intelligence Scale for Children—Third Edition.* San Antonio, TX: The Psychological Corporation.

Wechsler, D. (1997). *Wechsler Adult Intelligence Scale—Third Edition.* San Antonio, TX: The Psychological Corporation.

Wechsler, D. (2003). *Wechsler Intelligence Scale for Children—Fourth Edition.* San Antonio, TX: The Psychological Corporation.

Received January 31, 2003
Revision received September 26, 2003
Accepted September 30, 2003 ■