Educational and Psychological Measurement
Volume 66 Number 6
December 2006 975-983

10.1177/0013164406288168
http://epm.sagepub.com
hosted at
http://online.sagepub.com

Factor Structure of the Wechsler Intelligence Scale for Children–Fourth Edition Among Referred Students

Marley W. Watkins
Sharise M. Wilson
Kasey M. Kotz
Maria C. Carbone
Teresa Babula
The Pennsylvania State University

Factor analysis was applied to the Wechsler Intelligence Scale for Children–Fourth Edition (WISC-IV) scores of 432 Pennsylvania students referred for evaluation for special education services to determine the factor structure of the WISC-IV with this population. A first-order, four-factor oblique solution that mirrored that found in the WISC-IV normative sample was supported. When transformed to an orthogonalized higher order model, the general factor accounted for the greatest amount of common (75.7%) and total (46.7%) variance. In contrast, the largest contribution by a first-order factor (Verbal Comprehension) was 6.5% of total variance. It was recommended that interpretation of the WISC-IV not discount the strong general factor.

Keywords: *WISC-IV; intelligence; factor analysis; special education*

Millions of students were administered the Wechsler Intelligence Scale for Children–Third Edition (WISC-III; Wechsler, 1991) to determine entitlement for special education services (Kamphaus, Petoskey, & Rowe, 2000). Prior to the WISC-III, the Wechsler Intelligence Scale for Children–Revised (WISC-R; Wechsler, 1974) was the most popular individual test of intelligence for children (Oakland & Hu, 1992). The Wechsler Intelligence Scale for Children–Fourth Edition (WISC-IV; Wechsler, 2003a) has now superseded the WISC-III and promises to, in turn, become the most widely used measure of childhood intelligence (Prifitera, Saklofske, Weiss, & Rolfhus, 2005).

Authors' Note: The generosity of the Pennsylvania school psychologists who contributed data for this study is gratefully acknowledged. Address correspondence to Marley W. Watkins, The Pennsylvania State University, Department of Educational and School Psychology and Special Education, 125 CEDAR Building, University Park, PA 16801; e-mail: mww10@psu.edu.

Given the regular application of the WISC-IV in special education settings, professional standards mandate evidence regarding that test's psychometric fitness (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 1999). Of the many forms of psychometric evidence, probably the most critical is construct validity (Messick, 2000). Although many forms of evidence relate to construct validity (i.e., test content, external relationships, etc.), internal structure evidence is especially important. That is, does empirical analysis of a test's components support the structure proposed by the test's developer across a variety of test takers? Factor analysis constitutes one major source of statistical evidence regarding the structural validity of a test's scores (Benson, 1998).

For the WISV-IV, the application of factor analysis began with its normative sample (Wechsler, 2003b). Four correlated factors were found to best describe the intercorrelations between 10 core WISC-IV subtests: (a) Verbal Comprehension (VC) was made up of Vocabulary, Similarities, and Comprehension subtests; (b) Perceptual Reasoning (PR) was formed by the Block Design, Matrix Reasoning, and Picture Concepts subtests; (c) Working Memory (WM) was composed of Digit Span and Letter-Number Sequencing subtests; and (d) Processing Speed (PS) was made up of the Coding and Symbol Search subtests. No other factor analysis of WISC-IV scores has been reported. Consequently, this study applied factor analysis to the WISC-IV scores of students referred for special education evaluation to better assess the structural validity of WISC-IV scores with this population.

Method

Participants

Four hundred thirty-two students (256 male and 176 female) served as participants. The participants' ages ranged from 6 to 16 years with a mean of 10.3 years and standard deviation of 2.7 years. Ethnic background of participants was 89.6% White, 6.3% Black, 2.5% Hispanic, and 1.6% Other. Approximately 65% of the participants were eligible for special education services: 37% with learning disabilities, 5% with mental retardation, 7% with emotional disabilities, 8% as gifted, 2% with speech disabilities, and 6% with multiple disabilities. To ensure anonymity, no other demographic data were collected.

Instrument

The WISC-IV is a revised and updated version of the WISC-III. It was standardized on a nationally representative sample of 2,200 children aged 6 to 16 years closely approximating the 2000 U.S. Census on sex, race, parent education level, and geographic region. The WISC-IV has 10 core subtests (*M* = 10, *SD* = 3) that form the four (VC, PR, WM, and PS) factor indices (*M* = 100, *SD* = 15). The Full Scale IQ (*M* = 100, *SD* = 15) is based on the sum of scores from the 10 core subtests. Extensive reliability and validity evidence was provided by Wechsler (2003b) and by Prifitera, Saklofske, and Weiss (2005).

Procedure

All 419 regular members of the Association of School Psychologists of Pennsylvania during 2003-2004 were asked via letter to participate in this study by anonymously contributing WISC-IV scores from anonymous students evaluated in their schools. Thirty responded with data on 477 students. However, 45 cases were missing data or contained invalid subtest, factor, or global IQ scores. On average, each school psychologist contributed 14 valid cases (*SD* = 10). Because responses were anonymous, there was no demographic description of the contributing school psychologists.

Statistical Analyses

Given the uncertainty surrounding the structure of the WISC-IV (Browne, 2001) and the potential for stronger structural evidence to emerge from independent replications (Goldberg & Velicer, in press), exploratory rather than confirmatory factor analysis was chosen. Common factor analysis was selected over principal components analysis because the goal of the study was to identify the latent structure of the WISC-IV (Wegener & Fabrigar, 2000). Additionally, common factor analysis may produce more accurate estimates of population parameters than does principal components analysis (Widaman, 1993). Given its relative tolerance of multivariate nonnormality and its superior recovery of weak factors, principal axis extraction was used (Briggs & MacCallum, 2003). Communalities were initially estimated by squared multiple correlations and were iterated twice to produce final communality estimates (Gorsuch, 2003). Following the advice of Velicer, Eaton, and Fava (2000) and Henson and Roberts (in press), minimum average partials (MAP; Velicer, 1976) and parallel analysis (Horn, 1965), supplemented by a visual scree test (Cattell, 1966), were used to determine the number of factors to retain for rotation. For both theoretical and empirical reasons, it was assumed that factors would be moderately correlated (Wechsler, 2003b). Thus, a Promax rotation with a k value of 4 was selected (Tataryn, Wood, & Gorsuch, 1999). Pattern coefficients ≥ .40 were predetermined to be salient to retain only those that were both statistically ($p < .01$) and practically significant (Stevens, 2002).

Wechsler (2003b) acknowledged that the three-stratum theory of Carroll (1993) was influential in revision of the WISC-IV. Carroll (2003) proposed that cognitive abilities exist at three levels or strata. Inexplicably, Wechsler (2003b) did not conduct a higher order factor analysis to verify and describe the multilevel structure of the WISC-IV. In such investigations, Carroll (1993, 1995, 1997) recommended the

Schmid-Leiman (1957) procedure, which transforms the first-order factors so that they are orthogonal to each other and to the second-order factor. Variance explained by the general intelligence factor is extracted first. The first-order factors are then residualized of all the variance present in the general factor (McClain, 1996). Following these recommendations, Schmid-Leiman (1957) orthogonalization was applied to the oblique first-order factor structure to elucidate the multilevel structure of the WISC-IV.

Results

As reported in Table 1, participants' mean WISC-IV subtest, factor, and IQ scores were slightly lower and somewhat more variable than the normative sample. Similar patterns have been found with other samples of referred students (Canivez & Watkins, 1998). Nevertheless, score distributions appeared to be relatively normal, with .51 the largest skew and .89 the largest kurtosis (Onwuegbuzie & Daniel, 2002).

Results from Bartlett's Test of Sphericity (Bartlett, 1954) indicated that the correlation matrix was not random ($\chi^2 = 2{,}419.5$; $df = 45$; $p < .001$). The Kaiser-Meyer-Olkin (KMO; Kaiser, 1974) statistic was .92, well above the minimum standard suggested by Kline (1994). Measures of sampling adequacy for each variable were also within reasonable limits. Thus, the correlation matrix was appropriate for factor analysis.

Parallel analysis and MAP criteria suggested that one factor be retained, but the scree test recommended four factors. Given that it is better to overfactor than underfactor (Wood, Tataryn, & Gorsuch, 1996), four factors were extracted. The resulting solution was examined for both substantive and statistical suitability. Fit appeared to be excellent, accounting for 62% of the total variance, with a root mean square residual (RMSR) of .017 and no residual coefficient $\geq |.05|$. As reflected in Table 2, pattern coefficients clearly identified the four factors suggested by Wechsler (2003b). Factor intercorrelations ranged from .53 between VC and PS to .81 between PR and WM, suggesting a second-order factor. All subtests demonstrated moderate to substantial loadings on the first unrotated principal factor (see Table 2), another reflection of the influence of an overarching general ability factor (Carroll, 2003). Additionally, the four-factor solution was robust across extraction (Principal Components, Maximum Likelihood) and rotation (Varimax, Oblimin) methods.

Three factors were also extracted and rotated for a statistical and substantive comparison to the four-factor solution. This caused the PR and WM factors to combine into a single factor. This three-factor solution accounted for 60% of the total variance with a RMSR of .031, and left one coefficient in the residual matrix $\geq |.05|$. Factor intercorrelations ranged from .57 to .80. Although parsimonious, this solution was not congruent with the normative WAIS-IV factor analysis (Wechsler, 2003b).

In the two-factor solution, all the VC, PR, and WM subtests coalesced into a single factor, leaving the two PS subtests to form the second factor. These two factors accounted for 57% of the total variance. The RMSR was .053 and around

Table 1
Means and Standard Deviations on Wechsler Intelligence Scale for Children–Fourth Edition (WISC-IV) Subtest, Factor, and IQ Scores of 430 Students Tested for Special Education Eligibility

Test	Mean	*SD*
SI	9.45	3.45
VO	9.18	3.22
CO	9.28	3.16
BD	8.67	3.02
PCn	9.87	3.34
MR	9.33	3.27
DS	8.75	3.09
LN	9.25	3.29
CD	8.77	2.77
SS	9.10	3.14
VC	95.8	17.2
PR	95.7	16.9
WM	93.8	16.3
PS	94.1	14.7
FSIQ	93.8	17.5

Note: SI = Similarities; VO = Vocabulary; CO = Comprehension; BD = Block Design; PCn = Picture Concepts; MR = Matrix Reasoning; DS = Digit Span; LN = Letter-Number Sequencing; CD = Coding; SS = Symbol Search; VC = Verbal Comprehension factor; PR = Perceptual Reasoning factor; WM = Working Memory factor; PS = Processing Speed factor; FSIQ = Full-Scale IQ.

15% of the residuals were sizable. When one factor was extracted, it accounted for 51% of the total variance. However, the RMSR was .091 and 28% of the residual coefficients were ≥ |.05|. Thus, the one- and two- factor solutions were marked by excessive residual variance and subtest-factor relationships dissimilar to those found by Wechsler (2003b).

Given these statistical and substantive considerations, the four-factor solution was accepted as the most adequate for this sample of referred students. To test that conclusion against the WISC-IV normative sample, the congruence coefficient (r_c), an index of factorial similarity, was calculated for each factor. Jensen (1998) suggested than an $r_c \geq +.90$ indicates “a high degree of factor similarity; a value greater than +.95 is generally interpreted as practical identity of the factors” (p. 99). Based upon these guidelines, all four first-order factors in this sample were practically identical (r_c = .96 to .99) to the WISC-IV normative sample.

The Schmid-Leiman (1957) transformation of this first-order, four-factor oblique structure is presented in Table 3. The variance of each WISC-IV subtest could be decomposed into several sizable components. The most important was general intelligence (*g*), which accounted for more variance in each of the 10 core WISC-IV subtests than any orthogonal first-order factor. For example, *g* explained 62% of the variance in

Table 2
Structure of the Wechsler Intelligence Scale for Children–Fourth Edition (WISC-IV) for Principal Axis Extraction and Promax Rotation of Four Factors Among 430 Students Tested for Special Education Eligibility

Subtest/ Factor	Unrotated First Factor	Pattern (Structure) Coefficients VC	PR	WM	PS	Communality
SI	.808	**.786** (.859)	.049 (.685)	.084 (.692)	–.054 (.452)	.743
VO	.871	**.862** (.926)	.089 (.744)	.014 (.730)	–.029 (.499)	.861
CO	.734	**.815** (.792)	–.039 (.603)	–.056 (.600)	.093 (.466)	.632
BD	.707	.001 (.583)	**.734** (.758)	–.012 (.614)	.050 (.523)	.576
PCn	.700	.115 (.610)	**.625** (.727)	.033 (.613)	–.020 (.471)	.535
MR	.760	.060 (.645)	**.722** (.800)	.037 (.668)	.003 (.531)	.643
DS	.680	.042 (.589)	.071 (.618)	**.652** (.730)	–.016 (.478)	.537
LN	.761	.106 (.667)	.059 (.686)	**.635** (.797)	.049 (.560)	.645
CD	.452	–.002 (.322)	–.027 (.399)	–.040 (.397)	**.703** (.658)	.436
SS	.637	.029 (.489)	.076 (.581)	.069 (.579)	**.653** (.763)	.597
PR		.768				
WM		.769	.807			
PS		.534	.653	.654		

Note: VC = Verbal Comprehension factor; PR = Perceptual Reasoning factor; WM = Working Memory factor; PS = Processing Speed factor; SI = Similarities; VO = Vocabulary; CO = Comprehension; BD = Block Design; PCn = Picture Concepts; MR = Matrix Reasoning; DS = Digit Span; LN = Letter-Number Sequencing; CD = Coding; SS = Symbol Search. Salient pattern coefficients (≥.40) are indicated in bold.

the Vocabulary subtest whereas the VC factor accounted for an additional 24% of that subtest's variance. Notably, the PR and WM factors explained very little variance in their constituent subtests (i.e., 7% to 9%) once general intelligence had been extracted.

The influence of general intelligence dwarfed the contributions made by the four WISC-IV first-order factors. The first-order factors accounted for 2.2% (WM) to 10.5% (VC) of common variance and 1.4% (WM) to 6.5% (VC) of total variance. In contrast, general intelligence accounted for 75.7% of common variance and 46.7% of the total variance.

Discussion

Factor analyses of the WISC-IV scores of students referred to Pennsylvania school psychologists for evaluation to determine eligibility for special education services indicated that the same oblique four-factor model proposed by Wechsler (2003b) for the general population was also appropriate for these students. Given that the three-stratum theory of Carroll (1993) influenced the development of the WISC-IV, the hypothesized multilevel structure of the WISC-IV was addressed

Table 3
Sources of Variance in the Wechsler Intelligence Scale for Children–Fourth Edition (WISC-IV) Referral Sample (*N* = 430) According to an Orthogonalized Higher Order Factor Model

	General		VC		PR		WM		PS			
Subtest	*b*	Var	*b*	Var	*b*	Var	*b*	Var	*b*	Var	h^2	u^2
SI	.732	.54	.444	.20							.73	.27
VO	.785	.62	.487	.24							.85	.15
CO	.650	.42	.460	.21							.63	.37
BD	.694	.48			.301	.09					.57	.43
PCn	.681	.46			.256	.07					.53	.47
MR	.744	.55			.296	.09					.64	.36
DS	.684	.47					.265	.07			.54	.46
LN	.756	.57					.258	.07			.64	.36
CD	.426	.18							.505	.26	.44	.56
SS	.611	.37							.469	.22	.59	.41
% total var		46.7		6.5		2.4		1.4		4.7	61.7	38.3
% common var		75.7		10.5		4.0		2.2		7.7	100	

Note: *b* = pattern/structure coefficient; Var = percentage variance explained in the subtest; h^2 = communality; u^2 = uniqueness; FSIQ = Full-Scale IQ; VC = Verbal Comprehension factor; PR = Perceptual Reasoning factor; WM = Working Memory factor; PS = Processing Speed factor; SI = Similarities; VO = Vocabulary; CO = Comprehension; BD = Block Design; PCn = Picture Concepts; MR = Matrix Reasoning; DS = Digit Span; LN = Letter-Number Sequencing; CD = Coding; SS = Symbol Search.

using the orthogonal transformation procedure recommended by Carroll (1993, 2003). These results demonstrated the primary contribution of general intelligence: It contributed 46.7% of the total variance and 75.7% of the common variance. The VC factor accounted for 6.5% of the total variance and 10.5% of the common variance. The PR and WM factors were particularly weak, accounting for only 2.4% and 1.4%, respectively, of the total variance.

Interpreting a second-level factor on the basis of first-level factors can be misleading (McClain, 1996) because performance on any subtest reflects a mixture of both general and first-order factors, which will inflate the importance of lower order factors at the expense of the higher order factor (Carretta & Ree, 2001). For example, a group of verbal subtests may be called a measure of verbal comprehension, but a proportion of its nonerror variance is contributed by the general factor. In this case, it is necessary to decompose the variance of the verbal subtests into components separately attributable to verbal comprehension and general intelligence (Gustafsson, 1994).

Following these principles, it would be a mistake to interpret the WISC-IV second-order factor on the basis of its first-order factors. Although the four oblique first-order factors provided a good fit to the data, they accounted for a small proportion of nonerror variance. This was especially true of the PR and WM factors, which combined contributed only 3.8% of the total variance. Given these explanatory contributions, recommendations favoring interpretation of the first-order factor scores over the general intelligence score (Wechsler, 2003b; Weiss, Saklofske, & Prifitera, 2005; Williams, Weiss, & Rolfhus, 2003) should be rejected.

References

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.

Bartlett, M. S. (1954). A further note on the multiplying factors for various χ^2 approximations in factor analysis. *Journal of the Royal Statistical Society, 16,* 296-298.

Benson, J. (1998). Developing a strong program of construct validation: A test anxiety example. *Educational Measurement: Issues and Practice, 17,* 10-22.

Briggs, N. E., & MacCallum, R. C. (2003). Recovery of weak common factors by maximum likelihood and ordinary least squares estimation. *Multivariate Behavioral Research, 38,* 25-56.

Browne, M. W. (2001). An overview of analytic rotation in exploratory factor analysis. *Multivariate Behavioral Research, 36,* 111-150.

Canivez, G. L., & Watkins, M. W. (1998). Long-term stability of the Wechsler Intelligence Scale for Children–Third Edition. *Psychological Assessment, 10,* 285-291.

Carretta, T. R., & Ree, J. J. (2001). Pitfalls of ability research. *International Journal of Selection and Assessment, 9,* 325-335.

Carroll, J. B. (1993). *Human cognitive abilities.* Cambridge: Cambridge University Press.

Carroll, J. B. (1995). On methodology in the study of cognitive abilities. *Multivariate Behavioral Research, 30,* 429-452.

Carroll, J. B. (1997). Theoretical and technical issues in identifying a factor of general intelligence. In B. Devlin, S. E. Fienberg, D. P. Resnick, & K. Roeder (Eds.), *Intelligence, genes, and success: Scientists respond to the Bell Curve* (pp. 125-156). New York: Springer-Verlag.

Carroll, J. B. (2003). The higher-stratum structure of cognitive abilities: Current evidence supports *g* and about ten broad factors. In H. Nyborg (Ed.), *The scientific study of general intelligence: Tribute to Arthur R. Jensen* (pp. 5-21). New York: Pergamon.

Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research, 1,* 245-276.

Goldberg, L. R., & Velicer, W. F. (in press). Principles of exploratory factor analysis. In S. Strack (Ed.), *Differentiating normal and abnormal personality* (2nd ed.). New York: Springer.

Gorsuch, R. L. (2003). Factor analysis. In J. A. Schinka & F. Velicer (Eds.), *Handbook of psychology: Vol. 2. Research methods in psychology* (pp. 143-164). Hoboken, NJ: John Wiley.

Gustafsson, J.-E. (1994). Hierarchical models of intelligence and educational achievement. In A. Demetriou and A. Efklides (Eds.), *Intelligence, mind, and reasoning: Structure and development* (pp. 45-73). New York: Elsevier.

Henson, R. K., & Roberts, J. K. (in press). Use of exploratory factor analysis in published research: Common errors and some comment on improved practice. *Educational and Psychological Measurement*.

Horn, J. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika, 30,* 179-185.

Jensen, A. R. (1998). *The g factor: The science of mental ability*. Westport, CT: Praeger.

Kaiser, H. F. (1974). An index of factorial simplicity. *Psychometrika, 39,* 31-36.

Kamphaus, R. W., Petoskey, M. D., & Rowe, E. W. (2000). Current trends in psychological testing of children. *Professional Psychology: Research and Practice, 31,* 155-164.

Kline, P. (1994). *An easy guide to factor analysis*. New York: Routledge.

McClain, A. L. (1996). Hierarchical analytic methods that yield different perspectives on dynamics: Aids to interpretation. *Advances in Social Science Methodology, 4,* 229-240.

Messick, S. (2000). Consequences of test interpretation and use: The fusion of validity and values in psychological assessment. In R. D. Goffin & E. Helmes (Eds.), *Problems and solutions in human assessment: Honoring Douglas N. Jackson at seventy* (pp. 3-20). Boston: Kluwer Academic.

Oakland, T., & Hu, S. (1992). The top 10 tests used with children and youth worldwide. *Bulletin of the International Test Commission, 19,* 99-120.

Onwuegbuzie, A. J., & Daniel, L. G. (2002). Uses and misuses of the correlation coefficient. *Research in the Schools, 9,* 73-90.

Prifitera, A., Saklofske, D. H., & Weiss, L. G. (Eds.). (2005). *WISC-IV clinical use and interpretation: Scientist-practitioner perspectives*. New York: Elsevier.

Prifitera, A., Saklofske, D. H., Weiss, L. G., & Rolfhus, E. (2005). The WISC-IV in the clinical assessment context. In A. Prifitera, D. H. Saklofske, & L. G. Weiss (Eds.), *WISC-IV clinical use and interpretation: Scientist-practitioner perspectives* (pp. 3-32). New York: Elsevier.

Schmid, J., & Leiman, J. M. (1957). The development of hierarchical factor solutions. *Psychometrika, 22,* 53-61.

Stevens, J. P. (2002). *Applied multivariate statistics for the social sciences* (4th ed.). Mahwah, NJ: Lawrence Erlbaum.

Tataryn, D. J., Wood, J. M., & Gorsuch, R. L. (1999). Setting the value of *k* in promax: A Monte Carlo study. *Educational and Psychological Measurement, 59,* 384-391.

Velicer, W. F. (1976). Determining the number of components from the matrix of partial correlations. *Psychometrika, 41,* 321-327.

Velicer, W. F., Eaton, C. A., & Fava, J. L. (2000). Construct explication through factor or component analysis: A review and evaluation of alternative procedures for determining the number of factors or components. In R. D. Goffin & E. Helmes (Eds.), *Problems and solutions in human assessment: Honoring Douglas N. Jackson at seventy* (pp. 41-71). Boston: Kluwer Academic.

Wechsler, D. (1974). *Wechsler Intelligence Scale for Children–Revised.* New York: Psychological Corporation.

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

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

Wechsler, D. (2003b). *Wechsler Intelligence Scale for Children–Fourth Edition technical and interpretive manual.* San Antonio, TX: Psychological Corporation.

Wegener, D. T., & Fabrigar, L. R. (2000). Analysis and design for nonexperimental data. In H. T. Reis & C. M. Judd (Eds.), *Handbook of research methods in social and personality psychology* (pp. 412-450). New York: Cambridge University Press.

Weiss, L. G., Saklofske, D. H., & Prifitera, A. (2005). Interpreting the WISC-IV index scores. In A. Prifitera, D. H. Saklofske, & L. G. Weiss (Eds.), *WISC-IV clinical use and interpretation: Scientist-practitioner perspectives* (pp. 71-100). New York: Elsevier Academic.

Widaman, K. F. (1993). Common factor analysis versus principal component analysis: Differential bias in representing model parameters? *Multivariate Behavioral Research, 28,* 263-311.

Williams, P. E., Weiss, L. G., & Rolfhus, E. L. (2003). *Clinical validity* (WISC-IV Technical Report No. 3). San Antonio, TX: Psychological Corporation.

Wood, J. M., Tataryn, D. J., & Gorsuch, R. L. (1996). Effects of under- and overextraction on principal axis factor analysis with varimax rotation. *Psychological Methods, 1,* 354-365.