

## CONFIRMATORY FACTOR ANALYSIS OF THE WISC-III FOR STUDENTS WITH LEARNING DISABILITIES

Marley W. Watkins  
*The Pennsylvania State University*

Joseph C. Kush  
*Duquesne University*

This study conducted confirmatory factor analyses to examine 12 competing models that attempt to explain the underlying latent constructs measured by the Wechsler Intelligence Scale for Children-Third Edition (WISC-III) in a sample of 1,201 students with learning disabilities (721 White and 480 Nonwhite). Models were selected based upon theoretical explanations of the WISC-III factor structure as well as from previously conducted empirical research conducted on the WISC-III normative sample and with independent samples of students with exceptionalities. For both White and Nonwhite groups, four models could not be statistically distinguished from one another

based upon overall fit statistics. Plausible models included the traditional first-order four-factor oblique structure of the standardization sample, the hierarchical second-order model with four first-order factors favored by Keith and Witta (1997), a bifactor second-order model with four first-order factors described by Gustafsson and Undheim (1996), and a bifactor second-order model with three first-order factors. Current results add to a growing body of evidence suggesting that WISC-III Verbal Comprehension, Perceptual Organization, and Processing Speed factors are robust across samples but the Freedom from Distractibility factor demonstrates tenuous construct validity.

Intelligence tests are administered to more than one million students each year as part of special education eligibility evaluations (Gresham & Witt, 1997). Because the Wechsler Intelligence Scale for Children-Third Edition (WISC-III; Wechsler, 1991) is the most popular individual intelligence test among school and clinical psychologists (Stinnett, Havey, & Oehler-Stinnett, 1994; Watkins, Campbell, Nieberding, & Hallmark, 1995), it is vitally important that the latent constructs measured by the WISC-III for students with disabilities be identified and compared to the constructs described for the normative sample of children. Empirical support for comparable factor structures across disabled and nondisabled populations would suggest that similar constructs or latent traits are being assessed and thereby provide preliminary support for use of the WISC-III with those groups. Alternatively, variant factor structures would indi-

---

Correspondence concerning this article should be addressed to Marley W. Watkins, The Pennsylvania State University, Department of Educational and School Psychology and Special Education, 227 CEDAR Building, University Park, PA 16802. E-mail may be directed to mww10@psu.edu.

cate that the WISC-III is measuring different attributes among these groups and, consequently, differences between groups could not be unambiguously interpreted.

Despite their widespread popularity, Wechsler scales have been criticized for their lack of a cogent theoretical foundation (Macmann & Barnett, 1994). All tests in the Wechsler family have been constructed to assess Verbal and Performance IQs in addition to the summary Full Scale IQ. Despite theoretical expectations of one or two factors, however, factor analytic studies of the Wechsler Intelligence Scale for Children-Revised (WISC-R; Wechsler, 1974) have usually found three factors and normative analyses of the current WISC-III resulted in four factors. This purported four-factor, first-order solution of the WISC-III consisted of (a) Verbal Comprehension (VC), composed of Information (IN), Similarities (SM), Vocabulary (VO), and Comprehension (CM) subtests; (b) Perceptual Organization (PO), composed of Picture Completion (PC), Picture Arrangement (PA), Block Design (BD), and Object Assembly (OA) subtests; (c) Freedom from Distractibility (FD), composed of Arithmetic (AR) and Digit Span (DS) subtests; and (d) Processing Speed (PS), composed of Coding (CD) and Symbol Search (SS) subtests (Wechsler, 1991).

The WISC-III four-factor solution was replicated in an independent nationally representative sample of 1,118 children (Roid, Prifitera, & Weiss, 1993) and among the Canadian normative sample (Roid & Worrall, 1997). However, alternative factor solutions have also been proposed for the WISC-III normative sample. Sattler (1992) analyzed the WISC-III standardization sample across 11 separate age groups and reported that a three-factor (VC, PO, and PS) model best fit the normative data. Blaha and Walbrown (1996) examined hierarchical factor solutions and found support for both two-factor and four-factor solutions, whereas Keith and Witta's (1997) hierarchical confirmatory factor analysis of the WISC-III normative sample supported the primacy of a second-order *g* factor in addition to the four first-order factors (i.e., VC, PO, FD, and PS). Thus, the reported four-factor structure of the WISC-III normative sample (Wechsler, 1991) has generally been accepted with some disputation surrounding the nature of the second-order *g* factor and the existence of the FD and PS factors (see Macmann & Barnett, 1994, for an exception to this conclusion).

Among exceptional and ethnically diverse students, the WISC-III VC and PO factors have generally been supported but the FD and PS factors have not always been found (Sullivan & Montoya, 1997). When only 10 WISC-III subtests are administered, the verbal-performance dimensions are typically produced (Kush & Watkins, 1997). However, results have been more variable when 12 subtests are included in the analyses. Kush and Watkins (1994) used 12 WISC-III subtests with 121 Mexican-American students with learning disabilities (LD). Their exploratory factor analysis supported a four-factor solution, although the FD factor was not completely congruent with the standardization solution. In contrast, Logerquist-Hansen and Barona (1994) found a three-factor solution (VC, PO, and PS) more credible among 240 children with LD. This three-factor model was also endorsed by Kush (1996) for a sample of 327 students with LD.

Poulson (1995) applied confirmatory factor analytic (CFA) techniques to the WISC-III scores of 200 students with LD and found that fit and parsimony

endorsed the two-factor model (VC and PO). More recently, Grice, Krohn, and Logerquist (1999) used CFA methods with one sample of students with LD ( $n = 280$ ) and cross-validated with an independent sample of students with LD ( $n = 240$ ). They concluded that the four-factor model was most appropriate, but not definitive because a three-factor model fit the data equally well. Grice et al. also noted that all factor models left much of the variance unexplained and suggested caution in use of the FD factor and the Symbol Search subtest. Kush, Watkins, Ward, Ward, Canivez, and Worrell (2001) also recommended caution in use of the FD and PS factors based upon their CFA of a sample of Black students referred for psychological evaluation. In this sample, three- (VC, PO, and PS) and four- (VC, PO, FD, and PS) factor solutions could not be statistically distinguished. Finally, a meta-analysis of six factor analyses of the WISC-III among 934 students with LD found a three-factor solution (verbal scale, perceptual organization, and processing speed) most acceptable (Ravert & Watkins, 2000).

In summary, empirical research examining the factor structure of the WISC-III among students with LD has produced contradictory results. These inconsistencies may occur because investigators have used a variable number of WISC-III subtests, applied differing factoring methods, and studied various exceptional populations. Additionally, sample sizes have often been small and vulnerable to sampling fluctuation effects. Because it remains unclear how many "types" of intelligence are being measured by the WISC-III, there currently exists considerable disagreement regarding its diagnostic interpretability among students with exceptionalities (Kush, 1996). This situation is especially critical for students with LD because they constitute more than 50% of students receiving special education services (U. S. Department of Education, 1999). Consequently, the present study applied CFA methods to the WISC-III among a large sample of students with LD in an attempt to elucidate the WISC-III factor structure for this population.

## METHOD

### *Participants*

Based upon Arizona Department of Education records, all 212 special education directors of Arizona school districts were contacted and asked to provide anonymous WISC-III data on students currently enrolled in their special education programs. Personnel from 40 school districts responded with anonymous data on 2,979 students in special education with current psychoeducational evaluations on file (i.e., WISC-III given within the past 3 years). Of this number, 2,274 were categorized as learning disabled. All participants were diagnosed independently by school district multidisciplinary teams (MDT) based on federal and Arizona special education rules and regulations that required the demonstration of a significant ability-achievement discrepancy. Students were determined by local MDT to exhibit LD in reading alone ( $n = 92$ ); math alone ( $n = 81$ ); written expression alone ( $n = 247$ ); reading and written expression ( $n = 339$ ); reading and math ( $n = 44$ ); math and written expression ( $n = 112$ ); and reading, math, and written expression ( $n = 286$ ).

Congruent with previous surveys of school psychologists (Canivez & Watkins, 1998), the Symbol Search and Digit Span subtests were infrequently administered. When records with missing WISC-III subtest data were excluded, 1,201 students with LD from 32 school districts remained. Males constituted 70.2% of the sample and females 29.8%. Mean age was 10.5 years ( $SD = 2.5$ ) and ranged from 6 to 16 years. Median grade placement was 4.5, with a range of kindergarten through 11. Ethnic background, as reported on school records, was 60.0% White, 21.1% Hispanic, 6.3% Black, 12.0% Native American, and 0.6% Asian/Pacific. Because data were reported anonymously from archival special education records, socioeconomic status could not be determined. However, the final sample of participants was distributed across rural, urban, and suburban school districts and was widely dispersed across all regions of the state.

### *Instruments*

The WISC-III is an individually administered test of intellectual ability for children aged 6-0 to 16-11 years. It consists of 10 mandatory subtests and 3 optional subtests ( $M = 10$ ;  $SD = 3$ ), which combine to yield Verbal (VIQ), Performance (PIQ), and Full Scale (FSIQ) IQs ( $M = 100$ ;  $SD = 15$ ). When 12 subtests are administered, it yields four (VC, PO, FD, and PS) factor index scores. Full details of the instrument are available in Wechsler (1991).

### *Analyses*

The variance-covariance matrices among the WISC-III subtest scores for the sample of 721 White students and 480 Nonwhite students with LD were submitted to EQS 5.6 for the Macintosh (Bentler & Wu, 1995) for CFA analyses. Based upon these large, multivariate normal samples, maximum likelihood estimation methods were applied.

The Goodness of Fit Index (GFI), Comparative Fit Index (CFI), and Root Mean Square Error of Approximation (RMSEA) fit statistics were selected a priori based upon coverage of diverse dimensions of model fit (Maruyama, 1998) as well as robustness across estimation method and misspecification error (Hu & Bentler, 1999). The traditional chi-square statistic was retained to allow a test of exact fit between the model and observed covariances. The GFI is analogous to a squared multiple correlation in that it represents the proportion of the observed variance and covariance explained by the model. The CFI represents the proportion of improvement in fit relative to a null model. Finally, the RMSEA reflects the covariance residuals adjusted for degrees of freedom. Thus, it is a measure of "badness of fit," whereas the GFI and CFI indexes are measures of "goodness of fit." Only RMSEA has a known sampling distribution and is well suited for confirmatory large sample analyses (Rigdon, 1996). High values of GFI and CFI (near 1.0) and low values of RMSEA (near 0) indicate good model fit (Maruyama, 1998). Hu and Bentler recommended a combination rule that requires a CFI cutoff value close to .95 and an RMSEA value near .06 to minimize Type I and Type II error rates.

As noted by MacCallum, Wegener, Uchino, and Fabrigar (1993), "without adequate consideration of alternative equivalent models, support for one model from a class of equivalent models is suspect at best and potentially

groundless and misleading” (p. 196). Thus, a wide variety of alternative models were tested. As described in Tables 1 and 2, alternatives included the normative first-order, four-factor oblique structure (see Figure 1) as well as several variations of two- and three-factor models in addition to hierarchical and bifactor higher-order models (see Figures 2 to 4).

Keith and Witte (1997) argued that hierarchical models are superior to first-order factor models and should be tested in factor analyses. As illustrated in Figure 2, a hierarchical model posits that a second-order general factor is responsible for the covariation among first-order factors that, in turn, account for the observed variation in subtests. According to Gustafsson and Undheim (1996), this bottom-up approach has been favored by American researchers and uses the first-order factors as building blocks for the second-order factor(s). This is a model implicitly recognized by the structure of the WISC-III,

Table 1  
*Labels and Descriptions of WISC-III CFA Models*

Label	Description	Source
1	One factor. All subtests loading on a general factor.	Wechsler, 1991
2	Two factors. Six verbal and six performance subtests that represent the traditional Wechsler verbal and performance dimensions.	Blaha & Walbrown, 1996; Donders & Warschausky, 1996; Keith & Witte, 1997; Konold et al., 1997; Poulson, 1995; Wechsler, 1991
3a	Three factors. Six verbal and four performance subtests with AR and DS loading on the verbal factor and CD and SS loading on the third factor.	Donders & Warschausky, 1996; Grice et al., 1999; Kush et al., 2001; Wechsler, 1991
3b	Three first-order factors. Four verbal and four performance subtests with AR, DS, CD, and SS loading on the third factor, with SS added to the third factor.	An extension of the WISC-R structure
3c	Three first-order factors. Five verbal (including AR) and five performance subtests (including CD) with DS and SS loading on the third factor.	Based upon mandatory versus supplemental subtests
3d	Three first-order factors. Four verbal and six performance subtests (including CD and SS) with AR and DS loading on the third factor.	Modified from Kamphaus et al. (1994) by retaining the SS subtest
3h1	A hierarchical model with three first-order factors (VC and PO with AR, DS, CD, and SS constituting the third first-order factor) and one second-order factor ( <i>g</i> ).	Hierarchical variant of Model 3b
3h2	A hierarchical model with three first-order factors (VC including AR and DS, PO, and PS) and one second-order factor ( <i>g</i> ).	Hierarchical variant of model 3a
3n	A bifactor model with three first-order factors (VC including AR and DS, PO, and PS) and one second-order factor ( <i>g</i> ).	Bifactor variant of models 3a and 3h2
4a	Four oblique first-order factors (VC, PO, FD, and PS).	Grice et al., 1999; Wechsler, 1991
4h	A hierarchical model with four first-order factors (VC, PO, FD, and PS) and one second-order factor ( <i>g</i> )	Keith & Witte, 1997
4n	A bifactor model with four first-order factors (VC, PO, FD, and PS) and one second-order factor ( <i>g</i> ).	Bifactor variation of Keith & Witte (1997) as per Gustafsson & Undheim (1996)

Table 2  
Subtest/Factor Configuration of Twelve WISC-III CFA Models

Model	Subtests											
	IN	SM	VO	CM	PC	PA	BD	OA	AR	DS	CD	SS
1	1	1	1	1	1	1	1	1	1	1	1	1
2	1	1	1	1	2	2	2	2	1	1	2	2
3a	1	1	1	1	2	2	2	2	1	1	3	3
3b	1	1	1	1	2	2	2	2	3	3	3	3
3c	1	1	1	1	2	2	2	2	1	3	2	3
3d	1	1	1	1	2	2	2	2	3	3	2	2
3h1	1	1	1	1	2	2	2	2	3	3	3	3
3h2	1	1	1	1	2	2	2	2	1	1	3	3
3n	1	1	1	1	2	2	2	2	1	1	3	3
4a	1	1	1	1	2	2	2	2	3	3	4	4
4h	1	1	1	1	2	2	2	2	3	3	4	4
4n	1	1	1	1	2	2	2	2	3	3	4	4

Note.—IN = Information, SM = Similarities, VO = Vocabulary, CM = Comprehension, PC = Picture Completion, PA = Picture Arrangement, BD = Block Design, OA = Object Assembly, AR = Arithmetic, DS = Digit Span, CD = Coding, SS = Symbol Search.

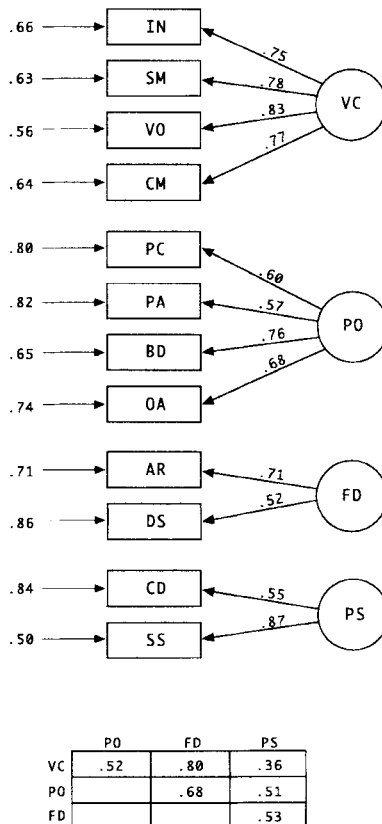
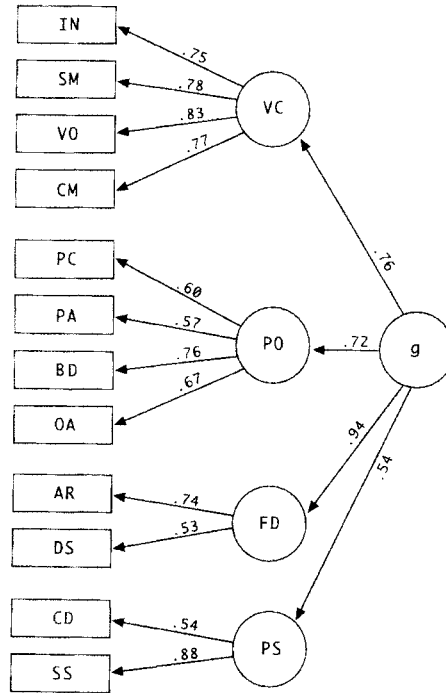
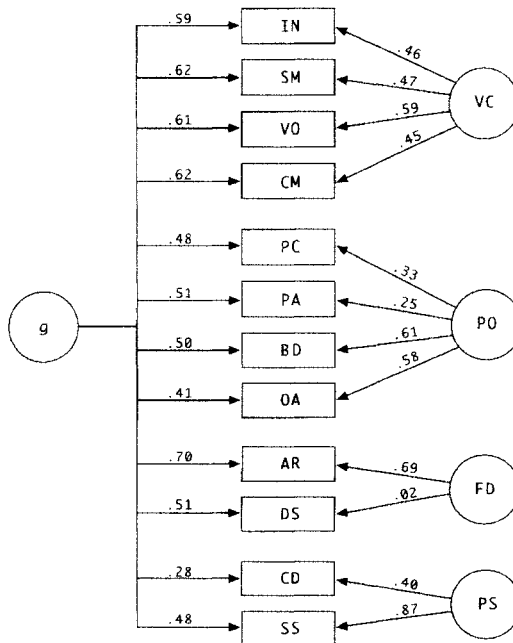


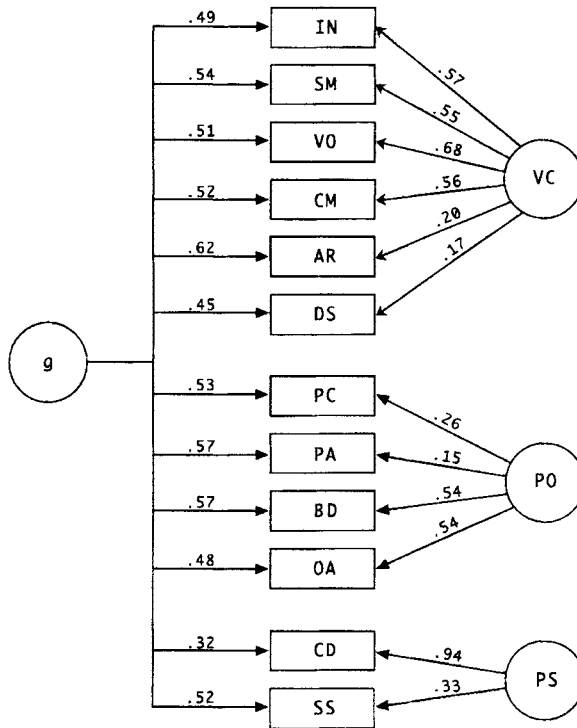
FIGURE 1. First-order, oblique four-factor model of the WISC-III for 1,210 students with learning disabilities.



**FIGURE 2. Hierarchical factor model with four orthogonal first-order factors for 1,201 students with learning disabilities.**



**FIGURE 3. Bifactor model with four orthogonal first-order factors for 1,201 students with learning disabilities.**



**FIGURE 4. Bifactor model with three orthogonal first-order factors for 1,201 students with learning disabilities.**

where the VIQ and PIQ are computed from the sums of specific subtests and the FSIQ is computed based upon the sum of all subtest scores.

An alternative hierarchical model for the WISC-III that includes a general second-order factor with direct relations to all subtests has been proposed by Gustafsson and Undheim (1996). Their bifactor model includes first-order factors that influence subtests *independent* of the general factor. Thus, the variance of each subtest is accounted for by two factors of different degrees of generality: a broad general (*g*) factor and a narrower second-order factor as represented in Figure 3. This top-down approach reflects the British factor analytic tradition whereby the general factor is fitted first and then the narrower first-order factors are extracted (Gustafsson & Undheim, 1996). More complete descriptions and explanations of CFA models are provided by Gustafsson and Undheim (1996) and Keith (1997).

## RESULTS

Descriptive statistics for the WISC-III among this sample of students with LD are presented in Table 3. As is usually found with exceptional samples (Kavale & Nye, 1985), overall scores were somewhat lower than found in the normative sample. However, variability was near normal so restriction of range did not appear to be a salient concern. Similarly, skewness and kurtosis indices suggested normal univariate score distributions.



Table 3  
*Descriptive Statistics for WISC-III Subtests and IQs*

Index	Whites <sup>a</sup>		Nonwhites <sup>b</sup>		Total <sup>c</sup>	
	Mean	SD	Mean	SD	Mean	SD
Picture Completion	9.71	2.92	8.71	2.90	9.31	2.95
Information	8.19	2.57	6.45	2.74	7.49	2.77
Coding	8.27	3.15	8.34	3.31	8.30	3.22
Similarities	8.85	2.98	6.94	3.15	8.08	3.19
Picture Arrangement	9.14	3.16	7.78	3.13	8.60	3.22
Arithmetic	7.66	2.50	6.36	2.38	7.14	2.54
Block Design	9.09	3.16	8.03	3.13	8.67	3.19
Vocabulary	8.18	2.62	6.45	3.03	7.49	2.92
Object Assembly	9.15	3.01	8.90	3.04	9.05	3.03
Comprehension	9.16	3.09	7.32	3.36	8.42	3.32
Symbol Search	9.21	3.17	8.47	3.28	8.91	3.23
Digit Span	7.37	2.46	7.00	2.49	7.22	2.48
Verbal IQ	91.2	12.1	81.5	13.2	87.3	13.4
Performance IQ	94.5	13.2	90.0	13.2	92.7	13.4
Full Scale IQ	92.0	11.8	84.2	11.8	88.9	12.4
Verbal Comprehension Index	92.7	12.4	82.9	14.1	88.8	14.0
Perceptual Organization Index	96.4	13.5	91.0	13.5	94.2	13.8
Freedom from Distractibility Index	86.9	11.8	82.0	11.5	85.0	11.9
Processing Speed Index	94.5	14.3	92.7	14.8	93.8	14.5

<sup>a</sup>*n* = 721.

<sup>b</sup>*n* = 480.

<sup>c</sup>*N* = 1,201.

### *Initial Model Reduction*

Based upon the chi-square statistic, there was no exact fit between any of the 12 models and the observed covariances. This was expected due to the well-known sensitivity of the chi-square statistic to sample size (Maruyama, 1998). CFI, GFI, and RMSEA fit statistics (see Table 4) indicated that several models were plausible fits to the WISC-III data for students with LD. Using the Hu and Bentler (1999) combination rule, which requires a CFI cutoff value of  $\geq .95$  and an RMSEA value  $\leq .06$ , four models were judged plausible fits for both White and Nonwhite samples: models 3n, 4a, 4h, and 4n.

Further discrimination among these four feasible models was accomplished via the 90% confidence interval (CI) surrounding the RMSEA (MacCallum, Browne, & Sugawara, 1996). The CI quantifies the degree of imprecision in RMSEA fit in the population and can be used to select among competing models. Those with overlapping CIs cannot be reliably distinguished, whereas RMSEA values that do not overlap can identify distinctly better fitting models. Unfortunately, none of the four plausible models was statistically distinguishable based upon RMSEA confidence intervals.

### *Final Model Refinement*

Given the identical pattern and magnitude of fit for White and Nonwhite students, these samples were combined into a total group for further analyses.

Table 4

Fit of WISC-III CFA Models for White ( $n = 721$ ), Nonwhite ( $n = 480$ ), and Total ( $N = 1,201$ ) Samples of Students with Learning Disabilities

Model	df	Whites					Nonwhites					Total				
		$\chi^2$ *	CFI	GFI	RMSEA	RMSEA 90% CI	$\chi^2$ *	CFI	GFI	RMSEA	RMSEA 90% CI	$\chi^2$ *	CFI	GFI	RMSEA	RMSEA 90% CI
1	54	650	.77	.84	.124	.115-.132	561	.73	.81	.140	.129-.150	1173	.77	.83	.131	.125-.138
2	53	336	.89	.92	.086	.077-.095	250	.89	.92	.088	.077-.099	535	.90	.93	.087	.080-.094
3a	51	214	.94	.95	.067	.057-.076	164	.94	.95	.068	.056-.080	331	.94	.95	.068	.061-.075
3b	51	283	.91	.94	.079	.070-.089	205	.92	.93	.079	.068-.091	436	.92	.94	.079	.072-.086
3c	51	345	.89	.93	.089	.081-.098	254	.89	.92	.091	.080-.102	543	.90	.93	.090	.083-.096
3d	51	268	.92	.94	.077	.068-.086	216	.91	.93	.082	.071-.093	425	.92	.94	.078	.071-.085
3h1	53	284	.91	.94	.078	.069-.087	206	.92	.93	.078	.066-.089	436	.92	.94	.078	.071-.084
3h2	52	218	.94	.95	.067	.058-.076	167	.94	.94	.068	.056-.079	342	.94	.95	.068	.061-.075
3n	43 <sup>a</sup>	114	.97	.97	.048	.037-.058	104	.97	.97	.054	.041-.068	163	.98	.98	.048	.040-.056
4a	48	143	.96	.97	.052	.043-.063	126	.96	.96	.058	.046-.071	212	.97	.97	.053	.046-.061
4h	51 <sup>b</sup>	174	.95	.96	.058	.049-.067	137	.95	.95	.059	.047-.071	253	.96	.97	.057	.050-.064
4n	44 <sup>c</sup>	135	.97	.97	.054	.043-.064	105	.97	.97	.054	.040-.067	183	.97	.98	.051	.044-.059

<sup>a</sup>Error of measurement of Coding fixed to 1 minus reliability.<sup>b</sup>Variance of Freedom from Distractibility constrained to zero.<sup>c</sup>Errors of measurement of Arithmetic and Symbol Search fixed to 1 minus reliability.\* $p < .001$ .

Model 4a reflects the first-order solution favored by the WISC-III publisher (Wechsler, 1991). However, Figure 1 indicates that the VC and FD factors correlate at .80, whereas the PO and FD factors correlate at .68. This substantial factor overlap is unexplained and, therefore, represents unsatisfactory discriminant validity (Gustafsson & Undheim, 1996; Kline, 1998). Additionally, higher-order models are preferable to first-order models on the basis of greater parsimony given equivalent fit indices (Marsh & Jackson, 1999). For these reasons, Model 4a was eliminated from further analyses and only the three higher-order models were given further consideration.

Model 4h represents the model that Keith and Witta (1997) declared to be the best fit to the WISC-III normative data. As reflected in Figure 2, results for this large sample of students with LD were very similar to those reported for the WISC-III normative sample (see Figure 2 of Keith & Witta). Given the high loading of the FD on *g*, Keith and Witta speculated that the FD factor was "a strong, perhaps the best, measure of general intelligence" (p. 104). However, a loading of .90 to .94 is extremely high and engenders doubt concerning the distinguishability of the FD factor rather than acceptance of FD as an almost perfect measure of *g*. Another perspective is provided by an inspection of Figure 3. When *g* was extracted first, the DS subtest loaded only .02 on the FD factor, whereas AR exhibited a loading of .69. Thus, the DS subtest seems to have little in common with AR, the other FD subtest, when residualized factors are examined.

With this sundering of the FD factor, a three-factor model becomes a more plausible explanation of the latent constructs being measured by the WISC-III. Figure 4 illustrates the bifactor model with three first-order factors. Its absolute fit statistics were the best of the 12 models analyzed, although not statistically superior to several competing models. Research in reading has consistently found that phonological memory for and rapid naming of digits are strongly related, perhaps causally, with word decoding (Wagner, Torgesen, & Rashotte, 1999). Thus, it is plausible that Digit Span should load onto a verbal rather than a FD factor. Likewise, the Arithmetic subtest has long been known to contain a verbal component because arithmetic problems are presented and responded to verbally (Kamphaus, 1993). However, Figure 4 indicates that the VC factor accounted for only 4% of the variance of AR and 2.9% of the variance of DS. Thus, there seems to be poor convergent validity between these two subtests (AR and DS) and the other four VC subtests among the students with LD examined in the current study.

## DISCUSSION AND CONCLUSIONS

CFA was conducted on the WISC-III subtest scores of 1,201 Arizona students enrolled in LD programs. Twelve alternative models of widely varying generality were tested for both White and Nonwhite subsamples in an attempt to determine the nature of the latent constructs being measured by the WISC-III. For both groups of students with LD, four models could not be statistically distinguished from one another based upon overall fit statistics: the traditional first-order four-factor oblique structure of the standardization sample; the hierarchical second-order model with four first-order factors favored by Keith and

Witta (1997); a bifactor second-order model with four first-order factors described by Gustafsson and Undheim (1996); and a bifactor second-order model with three first-order factors.

Additional analyses suggested that the excessively high intercorrelations between factors made the first-order four-factor solution undesirable. The second-order hierarchical model with four first-order orthogonal factors was spoiled by the indistinguishability of the FD factor from the *g* factor. The bifactor model with four first-order orthogonal factors demonstrated anomalous loadings of the DS subtest that resulted in a singleton factor. The bifactor second-order model with three orthogonal first-order factors was marred by trivial loadings of AR and DS on the VC factor and PA on the PO factor. Interestingly, similar ambiguous results have been found in other studies. For example, in an exploratory factor analysis of 505 students enrolled in gifted programs, PA, DS, and AR failed to exhibit salient loadings even when correlations were corrected for restriction of range (Watkins, Greenawalt, & Marcell, 2000). These results are also congruent with the multiple instrument factor analysis reported by McGrew and Woodcock (2001) that indicated that AR and DS loaded on separate factors if sufficient marker tests were included.

Due to their orthogonal structure, bifactor models allow easy calculation of the percent of variance accounted for by subtests and factors (Gustafsson & Undheim, 1996). This is illustrated in Table 5 for Model 4n. Similar to the results reported by Keith and Witta (1997) for the WISC-III normative sample, the highest loading subtests on each of the four first-order factors were replicated: VO on the VC factor, BD on the PO factor, AR on the FD factor, and SS on the PS factor. Likewise, the best measures of *g* were similar: AR, SM, CM, VO, and IN for the LD sample and AR, VO, IN, SM, BD, and CM for the norm sample. Based upon their analysis of the normative sample, Keith and Witta concluded that the WISC-III "is first and foremost a measure of general intelligence" (p. 105). The present results extend that conclusion to students with LD. The general factor accounted for 37.8% of the total score variance for students with LD, whereas the VC, PO, FD, and PS factors accounted for 3.7%, 3.0%, .5%, and 1.5% of the total score variance, respectively. Congruent with Grice et al. (1999), much of the variance was unexplained by the common factors. The PC, PA, DS, and CD subtests seemed to contribute disproportionately to unexplained variance.

Researchers have noted that the WISC-III subtests are too restricted in coverage to adequately represent all the factors it measures (Carroll, 1993). Multiple instrument factor analysis (McGrew & Woodcock, 2001) and theory (Flanagan, McGrew, & Ortiz, 2000) suggest that several WISC-III subtests may be single indicators of constructs that are not otherwise measured by the WISC-III. Specifically, AR appears to tap quantitative knowledge and DS seems to measure short-term memory (Keith & Witta, 1997). Consequently, the FD factor may be a statistical artifact without substantive meaning. This supposition can be tested by analyzing WISC-III subtests along with sufficient marker variables but cannot be verified by factor analyses of WISC-III subtests alone. Even though CFA has benefits over EFA, there is no clear mechanism for selecting among plausible competing models (Raykov & Marcoulides, 1999). A variety of roughly equivalent factor structures for the WISC-III has been reported (Blaha

Table 5

*Sources of Variance (Percent) in WISC-III LD Sample (N = 1,201) for Bifactor Model with Four First-Order Factors*

	<i>g</i>	VC	PO	FD	PS	Unexplained
IN	34.9	21.1				44.0
SM	38.1	21.8				40.2
VO	37.2	35.0				27.8
CM	38.3	20.4				41.2
PC	23.1		11.0			65.8
PA	25.8		6.3			68.1
BD	25.2		36.8			37.8
OA	16.9		34.1			49.0
AR	48.7			47.9		3.4
DS	25.6			.03		74.5
CD	8.0				15.6	76.4
SS	22.6				75.2	2.3
Total	37.8	3.7	3.0	.5	1.5	53.5

Note.—Rounded to tenths, may not add to 100. IN = Information, SM = Similarities, VO = Vocabulary, CM = Comprehension, PC = Picture Completion, PA = Picture Arrangement, BD = Block Design, OA = Object Assembly, AR = Arithmetic, DS = Digit Span, CD = Coding, SS = Symbol Search, VC = Verbal Comprehension, PO = Perceptual Organization, FD = Freedom from Distractibility, PS = Processing Speed.

& Walbrown, 1996; Grice et al., 1999; Keith & Witta, 1997; Kush, 1996; Kush et al., 2001; Logerquist-Hansen & Barona, 1994; Ravert & Watkins, 2000; Roid et al., 1993; Sattler, 1992; Wechsler, 1991). Thus, ambiguous factorial validity results may be the rule for the WISC-III due to its limited factor coverage.

Greater clarity may be achieved if construct validity evidence for the WISC-III includes information external to factor analysis (Comrey & Lee, 1992). Consequently, criterion-related validity evidence for the WISC-III factors should be given greater consideration. To this end, Kamphaus (1993) indicated that concurrent and predictive validity evidence for the FD and PS factors is not convincing. Similarly, attempts to demonstrate incremental validity for FD and PS factor scores, beyond the verbal and performance levels, as predictors of academic achievement have not been successful (Glutting, Youngstrom, Ward, Ward, & Hale, 1997). These two smaller factors have not demonstrated validity for prediction of behavioral dysfunction (Oakland, Broom, & Glutting, 2000; Riccio, Cohen, Hall, & Ross, 1997), nor have they displayed diagnostic accuracy with exceptional students (Watkins, Kush, & Glutting, 1997). Further, reliability coefficients of the FD and PS factors fall below the levels recommended for individual decisions (Salvia & Ysseldyke, 1998): Internal consistency, short-term stability, and long-term stability coefficients were .87, .82, and .75, respectively, for the FD factor and .85, .84, and .62, respectively, for the PS factor (Canivez & Watkins, 1998; Wechsler, 1991). When combined with the current factor analytic results, converging evidence suggests that the FD and PS indices of the WISC-III should be used with extreme caution, if at all.

## REFERENCES

- Bentler, P. M., & Wu, E. J. C. (1995). *EQS for Macintosh user's guide*. Encino, CA: Multivariate Software.
- Blaha, J., & Walbrown, F. H. (1996). Hierarchical factor structure of the Wechsler Intelligence Scale for Children-III. *Psychological Assessment, 8*, 214–218.
- Canivez, G. L., & Watkins, M. W. (1998). Long term stability of the Wechsler Intelligence Scale for Children-Third Edition. *Psychological Assessment, 10*, 285–291.
- Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. Cambridge: Cambridge University Press.
- Comrey, A. L., & Lee, H. B. (1992). *A first course in factor analysis* (2nd ed.). Hillsdale, NJ: Erlbaum.
- Donders, J., & Warschausky, S. (1996). A structural equation analysis of the WISC-III in children with traumatic head injury. *Child Neuropsychology, 2*, 185–192.
- Flanagan, D. P., McGrew, K. S., & Ortiz, S. O. (2000). *The Wechsler intelligence scales and Gf-Gc theory: A contemporary approach to interpretation*. Boston: Allyn and Bacon.
- Glutting, J. J., Youngstrom, E. A., Ward, T., Ward, S., & Hale, R. L. (1997). Incremental efficacy of WISC-III factor scores in predicting achievement: What do they tell us? *Journal of Psychological Assessment, 9*, 295–301.
- Gresham, F. M., & Witt, J. C. (1997). Utility of intelligence tests for treatment planning, classification, and placement decisions: Recent empirical findings and future directions. *School Psychology Quarterly, 12*, 249–267.
- Grice, J. W., Krohn, E. J., & Logerquist, S. (1999). Cross-validation of the WISC-III factor structure in two samples of children with learning disabilities. *Journal of Psychoeducational Assessment, 17*, 236–248.
- Gustafsson, J.-E., & Undheim, J. O. (1996). Individual differences in cognitive functions. In D. C. Berliner & R. C. Calfee (Eds.), *Handbook of educational psychology* (pp. 186–242). New York: Macmillan.
- Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*, 1–55.
- Kamphaus, R. W. (1993). *Clinical assessment of children's intelligence*. Boston: Allyn and Bacon.
- Kamphaus, R. W., Benson, J., Hutchinson, S., & Platt, L. O. (1994). Identification of factor models for the WISC-III. *Educational and Psychological Measurement, 54*, 174–186.
- Kavale, K. A., & Nye, C. (1985). Parameters of learning disabilities in achievement, linguistic, neuropsychological, and social/behavioral domains. *The Journal of Special Education, 19*, 443–458.
- 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). NY: Guilford.
- 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.
- Kline, R. B. (1998). *Principles and practice of structural equation modeling*. New York: Guilford.
- Konold, T. R., Kush, J. C., & Canivez, G. L. (1997). Factor replication of the WISC-III in three independent samples of children receiving special education. *Journal of Psychoeducational Assessment, 15*, 123–137.
- Kush, J. C. (1996). Factor structure of the WISC-III for students with learning disabilities. *Journal of Psychoeducational Assessment, 14*, 32–40.
- Kush, J. C., & Watkins, M. W. (1994, March). *Factor structure of the WISC-III for Mexican-American, learning disabled stu-*

- dents. Paper presented at the annual meeting of the National Association of School Psychologists, Seattle, WA.
- Kush, J. C., & Watkins, M. W. (1997). Construct validity of the WISC-III verbal and performance factors for Black special education students. *Assessment, 4*, 297-304.
- Kush, J. C., Watkins, M. W., Ward, T. J., Ward, S. B., Canivez, G. L., & Worrell, F. C. (2001). Construct validity of the WISC-III for White and Black students from the WISC-III standardization sample and for Black students referred for psychological evaluation. *School Psychology Review, 30*, 70-88.
- Logerquist-Hansen, S., & Barona, A. (1994, August). *Factor structure of the Wechsler Intelligence Scale for Children-III for Hispanic and non-Hispanic White children with learning disabilities*. Paper presented at the annual meeting of the American Psychological Association, Los Angeles, CA.
- MacCallum, R. C., Browne, M. W., & Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods, 1*, 130-149.
- MacCallum, R. C., Wegener, D. T., Uchino, B. N., & Fabrigar, L. R. (1993). The problem of equivalent models in applications of covariance structure analysis. *Psychological Bulletin, 114*, 185-199.
- Macmann, G. M., & Barnett, D. W. (1994). Structural analysis of correlated factors: Lessons from the verbal-performance dichotomy of the Wechsler scales. *School Psychology Quarterly, 9*, 161-197.
- Marsh, H. W., & Jackson, S. A. (1999). Flow experience in sport: Construct validation of multidimensional, hierarchical state and trait responses. *Structural Equation Modeling, 6*, 343-371.
- Maruyama, G. M. (1998). *Basics of structural equation modeling*. Thousand Oaks, CA: Sage.
- McGrew, K. S., & Woodcock, R. W. (2001). Technical manual. *Woodcock-Johnson III*. Itasca, IL: Riverside.
- Oakland, T., Broom, J., & Glutting, J. (2000). Use of freedom from distractibility and processing speed to assess children's test-taking behaviors. *Journal of School Psychology, 38*, 469-475.
- Poulson, K. M. (1995). *Factor structure of the WISC-III for unclassified, learning-disabled, and high-IQ groups*. Unpublished doctoral dissertation, Hofstra University.
- Ravert, C. M., & Watkins, M. W. (2000, March). *Meta-analysis of WISC-III factor analyses conducted with learning disabled students*. Poster session presented at the annual meeting of the National Association of School Psychologists, New Orleans, LA.
- Raykov, T., & Marcoulides, G. A. (1999). On desirability of parsimony in structural equation model selection. *Structural Equation Modeling, 6*, 292-300.
- Riccio, C. A., Cohen, M. J., Hall, J., & Ross, C. M. (1997). The third and fourth factors of the WISC-III: What they don't measure. *Journal of Psychoeducational Assessment, 15*, 27-39.
- Rigdon, E. E. (1996). CFA versus RMSEA: A comparison of two fit indexes for structural equation modeling. *Structural Equation Modeling, 3*, 369-379.
- Roid, G. H., Prifitera, A., & Weiss, L. G. (1993). Replication of the WISC-III factor structure in an independent sample [Special issue, Monograph, WISC-III series]. *Journal of Psychoeducational Assessment, 11*, 6-21.
- Roid, G. H., & Worrall, W. (1997). Replication of the Wechsler Intelligence Scale for Children-Third Edition four-factor model in the Canadian normative sample. *Psychological Assessment, 9*, 512-515.
- Salvia, J., & Ysseldyke, J. E. (1998). *Assessment* (5th ed). New York: Houghton Mifflin.
- Sattler, J. M. (1992). *Assessment of children* (3rd ed.). San Diego, CA: Author.
- Stinnett, T. A., Havey, J. M., & Oehler-Stinnett, J. (1994). Current test usage by practicing school psychologists: A national survey. *Journal of Psychoeducational Assessment, 12*, 331-350.

- Sullivan, P. M., & Montoya, L. A. (1997). Factor analysis of the WISC-III with deaf and hard-of-hearing children. *Psychological Assessment, 9*, 317–321.
- U. S. Department of Education. (1999). *Twenty-first annual report to Congress on the implementation of the Individuals with Disabilities Education Act*. Washington, DC: Author.
- Wagner, R. K., Torgesen, J. K., & Rashotte, C. A. (1999). *Examiner's manual: The Comprehensive Test of Phonological Processing*. Austin, TX: ProEd.
- Watkins, C. E., Jr., Campbell, V. L., Nieberding, R., & Hallmark, R. (1995). Contemporary practice of psychological assessment by clinical psychologists. *Professional Psychology: Research and Practice, 26*, 54–60.
- Watkins, M. W., Greenawalt, C. G., & Marcell, C. M. (2000, March). *Construct validity of the WISC-III among gifted children*. Poster presented at the annual meeting of the National Association of School Psychologists, New Orleans, LA.
- Watkins, M. W., Kush, J. C., & Glutting, J. J. (1997). Prevalence and diagnostic utility of the WISC-III SCAD profile among children with disabilities. *School Psychology Quarterly, 12*, 235–248.
- Wechsler, D. (1974). *Wechsler Intelligence Scale for Children-Revised*. San Antonio, TX: The Psychological Corporation.
- Wechsler, D. (1991). *Wechsler Intelligence Scale for Children-Third Edition*. San Antonio, TX: The Psychological Corporation.