Journal of Psychoeducational Assessment
1997, *15*, 123–137

FACTOR REPLICATION OF THE WISC-III IN THREE INDEPENDENT SAMPLES OF CHILDREN RECEIVING SPECIAL EDUCATION

Timothy R. Konold
University of Virginia

Joseph C. Kush
Duquesne University

Gary L. Canivez
Eastern Illinois University

This study examined the first-order factor structure of the Wechsler Intelligence Scale for Children-Third Edition (WISC-III; Wechsler, 1991) on three samples of children previously diagnosed with a handicapping condition. Five alternative factor models were compared through confirmatory factor analysis. Previous factor analytic studies that focused on the WISC-III's four-factor solution employed all 13 subtests in their analyses, despite the fact that only 12 subtests are combined during clinical evaluations to obtain scores on the four factors. This study investigated the WISC-III factor structure by considering the 12 subtests that actually combine to yield index scores on the dimensions of Verbal Comprehension, Perceptual Organization, Processing Speed, and Freedom from Distractibility. Results support a four-factor solution for children with disabilities.

An intelligence test is one of the most common components in the evaluation of children and adolescents (Stinnett, Havey, & Oehler-Stinnett, 1994; Wilson & Reschly, 1996). This situation remains true despite attempts to move diagnostic assessment away from traditional ability testing to methods that incorporate nontraditional, or "dynamic," methods of ability assessment (Budoff, 1987; Feuerstein, 1985; Swanson, 1994). More radical methods of assessment have also been recommended that would do away with ability testing altogether and replace it with "authentic" assessments, "curriculum-based" assessments, and other procedures designed to evaluate children's performance within a particular context and curriculum. Despite the potential advantages that nontraditional techniques might offer, traditional ability testing is likely to remain a vital part of the diagnostic armamentarium of psychologists in the twenty-first

Correspondence should be addressed to Timothy Konold, Curry School of Education, University of Virginia, 405 Emmet Street South, Charlottesville, VA 22903-2495.

century because general intelligence is among the most dominant and enduring predictors associated with scholastic and occupational success, environmental adaptation, and scientific, cultural, and political acumen (Brody, 1985; Crano, Kenny, & Campbell, 1972; Eysenck & Barrett, 1985; Jencks, 1972; Jensen, 1980; Terman & Oden, 1959).

Wechsler's series of intelligence tests are at the forefront of traditional ability testing. In fact, the Wechsler Intelligence Scale for Children-Revised (WISC-R; Wechsler, 1974) and its successor, the Wechsler Intelligence Scale for Children-Third Edition (WISC-III; Wechsler, 1991), are among the most widely administered instruments for evaluating the intellectual functioning of school-aged children (Goh, Teslow, & Fuller, 1981; Lutey & Copeland, 1982; Stinnett et al., 1994). Both the WISC-R and WISC-III are predicated on a hierarchical model of intelligence, where underlying the general factor (i.e., the Full Scale IQ) are various components through which intelligence can be expressed and interpreted.

Substantial disagreement exists regarding the proper level of interpretation of the WISC-III. A recent issue of the *School Psychology Quarterly* (Witt, 1994) was devoted to the debate surrounding the diagnostic utility of score interpretations beyond the Full Scale IQ (FSIQ). Some authors used factor analytic evidence to support positions that the WISC-III measures little more than general ability (Macmann & Barnett, 1994). Others argued the opposite position, pointing out that the majority of factor analyses completed with the WISC-III, and its predecessor the WISC-R, actually serve to support score interpretations extending beyond the FSIQ (Kaufman, 1994; Keith, 1994).

The debate surrounding the WISC-III's preferred level of interpretation emanates, in part, from an analogous debate surrounding the factor structure of the WISC-R. During initial factor analyses completed with the WISC-R standardization sample, Kaufman (1975) identified two significant factors at six age levels and three significant factors at five age levels. The largest dimension was labeled "Verbal Comprehension." The second largest dimension was called "Perceptual Organization," and the smallest, third dimension was called "Freedom from Distractibility."

Subsequent research with the WISC-R showed that the two-factor model was stable across age (Conger, Conger, Farrell, & Ward, 1979), gender (Reynolds & Gutkin, 1980), and ethnicity (Dean, 1980; Gutkin & Reynolds, 1981; Reschly, 1978). Likewise, the two-factor model held for a variety of special populations, including children diagnosed with brain damage (Kaufman, 1990), children classified as gifted (Sapp, Chissom, & Graham, 1985), and children with limited language skills (Taylor, Ziegler, & Partenio, 1984).

Alternatively, a three-factor solution is also well documented for the WISC-R. These three factors were derived for samples obtained from both regular and special education (Kaufman, 1975; Reynolds & Kaufman, 1990) and for specific subgroups receiving special education, such as children with learning disabilities (Juliano, Haddad, & Carroll, 1988; Naglieri, 1981).

The controversy surrounding the number of interpretable factors became more complex with the development of the WISC-III. A new subtest, Symbol Search (SS), was added to the WISC-III in an attempt to strengthen the previously identified third factor obtained with the WISC-R. However, results from factor analyses presented in the WISC-III manual were not as expected. Instead of loading on the third factor, SS combined with the Coding (CD) subtest to create a new factor. Thus, not only did the third factor (Freedom from Distractibility; FD) change in composition, but also an additional fourth factor (Processing Speed; PS) emerged. However, these factors (FD and PS) account for small proportions of variance (2% to 3% and 4% to 5%, respectively; Wechsler, 1991).

The WISC-III manual presents several validity studies that support the presence of a first-order, four-factor solution (Wechsler, 1991). Independent factor analyses of the WISC-III also indicate that four factors best account for subtest score variation (Keith & Witta, 1994, in Keith, 1994; Roid, Prifitera, & Weiss, 1993). However, few studies examined how the four-factor solution holds for special populations. Because of the change in composition of the FD factor and the addition of the PS factor, previous first-order factor studies on special populations from the WISC-R do not readily generalize to the WISC-III.

To date, three studies have investigated the construct validity of the WISC-III's first-order factor structure on samples of learning-disabled children. In two instances, evidence supported a four-factor solution (Roid et al., 1993; Wechsler, 1991). However, a third study found support for only the Verbal and Performance factors (Kush, 1996). Despite these efforts, Allen and Thorndike (1995) stated that "it will be for future research to determine whether the four-factor solution is replicable across studies with diverse populations" (p. 8).

Previous studies that concluded that the WISC-III's first-order factor structure was best defined by four factors utilized all 13 subtests in their analyses (Roid et al., 1993; Wechsler, 1991), despite the fact that the subtest Mazes is not included in the calculation and clinical interpretation of any of the four factors as presented in the WISC-III manual. Moreover, in an independent study, 6,424 WISC-III protocols obtained from psychodiagnostic evaluations conducted across six states were examined (Glutting, Konold, McDermott, Kush, & Watkins, 1995); all protocols had scores from the 10 mandatory subtests, while less than 1% included Mazes. Thus, the generalizability of studies investigating the number of abilities measured by the 13 WISC-III subtests to actual practice remains in question. Support for the existence of a four-factor solution should come through investigations that focus on the composition of factors as intended for use by clinicians. The inclusion of an additional subtest (Mazes) that is not used in the calculation of the WISC-III's factor scores may serve to under- or overestimate the number of abilities that operate to influence children's responses to the 12 WISC-III subtests. Thus, both clinical and pragmatic evidence suggest that Mazes be excluded from studies examining the WISC-III's first-order factor structure.

The present study examined the stability of the WISC-III's first-order factor structure on three samples of children previously diagnosed with one or more handicapping conditions. Whereas the first sample was obtained by the authors, two additional samples were obtained from other sources in order to assess the consistency of results across other samples of children comprising exceptional groups. The second and third samples were acquired from a dissertation (Bell, 1994) and from a paper presented at the annual convention of the American Psychological Asso-ciation (Logerquist-Hansen & Barona, 1994), respectively.

Sample 2 was previously used to investigate the effects of model modifications on the WISC-III's two-, three-, and four-factor solutions (Bell, 1994). Emphasis was placed on identifying the modified model that best reproduced the observed covariance matrix. Results were discussed in support of a four-factor solution. Sample 3 was formerly used to explore the convergence of the WISC-III factor structure between Hispanics and non-Hispanics (Logerquist-Hansen & Barona, 1994). Both alpha and canonical factor analyses were considered. Results of these exploratory analyses were discussed in support of a congruent three-factor solution for Hispanics and non-Hispanics.

The focus of the current study was to examine the first-order factor structure of the WISC-III on samples of children who were receiving special education. This analysis differed from those previously employed on Samples 2 and 3 in that our investigation was strictly confirmatory in nature.

Inasmuch as this investigation was confined to hypotheses surrounding the WISC-III's first-order factor structure, only 12 of the 13 subtests were examined. Five alternative factor models were compared. All models were analogous in design to those previously examined on "normal" children (Roid et al., 1993; Wechsler, 1991).

METHOD

Participants

WISC-III scores of children in the study represent three independent data collection efforts. The first sample was obtained by the authors in cooperation with school districts in the state of Arizona. Children in the second sample were also from Arizona, and children in the third sample were reported as coming from the southwestern region of the United States. Samples 2 and 3 were obtained in the form of correlation matrices.

Representations of gender, race, and child classifications for each of the three samples are presented in Table 1. Sample 1 was composed of 229 children ranging in age from 6 to 15 years ($M = 10.3$). Eighty-one percent of this sample were diagnosed as learning disabled (LD). The remaining children were diagnosed as mildly mentally retarded (MIMR), emotionally disabled (ED), speech language impaired (SLI), other health impaired (OHI), and moderately mentally retarded (MOMR).

Sample 2 was composed of 246 children ranging in age from 6 to 13 years ($M = 10.3$). All children in this sample were diagnosed as LD. Sample 3 consisted of 240 children ranging in age from 8 to 13 years ($M = 10.4$). As with the second sample, all children in this sample were diagnosed as LD. However, ethnicity was defined solely in terms of Hispanic origin.

Measure

The WISC-III is composed of 13 subtests; 10 are required to be administered, and 3 are optional (all *Ms* = 10; *SDs* = 3). These subtests combine to yield Verbal, Performance, and Full Scale IQs (all *Ms*=100, *SDs* = 15). In addition, four index scores are obtained by administering 12 of the 13 subtests: the Verbal Comprehension Index (VCI), which is composed of Information, Similarities, Vocabulary, and Comprehension; the Perceptual Organization Index (POI), which consists of Picture Completion, Picture Arrangement, Block Design, and Object Assembly; and the Freedom from Distractibility (FDI) and Processing Speed (PSI) indices, which are composed of Arithmetic and Digit Span, and Coding and Symbol Search, respectively. Mazes was excluded from our analyses because clinicians are encouraged to combine only 12 of the 13 subtests in order to obtain the four index scores (Wechsler, 1991).

Table 1
Demographics and Child Classifications

	Sample 1 (*N* = 229)	Sample 2 (*N* = 246)	Sample 3 (*N* = 240)
Gender			
Male	69%	69%	76%
Female	31%	31%	24%
Race			
Anglo	51%	67%	–
African American	10%	13%	–
Hispanic	33%	19%	50%
Non-Hispanic	–	–	50%
Native American	4%	1%	–
Asian	2%	–	–
Classification			
LD	81%	100%	100%
MIMR	8%	–	–
ED	7%	–	–
SLI	3%	–	–
OHI	<1%	–	–
MOMR	<1%	–	–

Note.–LD = learning disabled; MIMR = mildly mentally retarded; ED = emotionally disabled; SLI = speech language impaired; OHI = other health impaired; MOMR = moderately mentally retarded.

Analyses

Covariance matrices of the 12 subtests comprising the WISC-III were investigated through a series of five confirmatory factor models within each of three samples. These covariance matrices were obtained from the correlations and standard deviations of the samples. The five incremental fit models were analogous to those previously investigated on nonhandicapped children (Roid et al., 1993; Wechsler, 1991).

The one-factor model allowed all 12 subtests to load on a single dimension. This model served as a baseline against which alternative models (e.g., models comprised of two, three, four, and five factors) could be compared (Bentler & Bonett, 1980).

A two-factor model consisted of the verbal and performance dimensions. Six verbal subtests were included on the verbal factor: Information, Vocabulary, Similarities, Comprehension, Arithmetic, and Digit Span. Six performance subtests were aligned with the performance factor: Block Design, Object Assembly, Picture Completion, Picture Arrangement, Coding, and Symbol Search.

A three-factor model maintained a verbal factor similar in composition to the verbal factor investigated in the two-factor model. The second and third factors consisted of Block Design, Object Assembly, Picture Completion, and Picture Arrangement; and Coding and Symbol Search, respectively. This model was previously investigated through a comparison of nested models on normal children (Roid et al., 1993; Wechsler, 1991).

The four-factor model was analogous to the model advocated for use by clinicians in the WISC-III manual, as previously defined, and the five-factor model aligned with that proposed by Woodcock (1990). For the five-factor model, a Verbal Comprehension factor was comprised of Information, Vocabulary, Similarities, and Comprehension; a Perceptual Organization factor consisted of Block Design, Object Assembly, Picture Completion, and Picture Arrangement; a Processing Speed factor included Coding and Symbol Search; and two additional factors of Numerical Ability and Memory each consisted of one subtest, Arithmetic and Digit Span, respectively.

Each of the five models was investigated using maximum likelihood estimation through LISREL 8 (Joreskog & Sorbom, 1993). Maximum likelihood estimation is consistent with previous confirmatory factor analytic studies conducted on the WISC-III (Roid et al., 1993; Wechsler, 1991).

Several indices of model fit were investigated to compensate for the biases of each (Macmann & Barnett, 1994; Marsh, Balla, & McDonald, 1988). The chi-square (χ^2) index tests the closeness of fit between the observed covariance matrix and that of a hypothesized model. The relative fit of competing models was also gauged through the χ^2/df ratio. Large drops in the χ^2/df ratio "indicate that the changes made in the model represent a real improvement" (Joreskog & Sorbom, 1993, p. 29). In addition, both the adjusted goodness-of-fit index (AGFI) and the standardized root mean squared residual (RMSR) were used to compare competing models (Joreskog & Sorbom, 1993). The AGFI provides a measure of model fit that assesses the amount of variation/covariation in the sample covariance matrix that is predicted by the model (Bollen, 1989). It also provides an adjustment to control for the complexity of the hypothesized model. This index should range from 0 to 1.0, with larger values indicative of better model fit (Joreskog & Sorbom, 1993). The standardized RMSR measures the residual variance of the sample covariance matrix. Negative values indicate that the model is overpredicting the covariance matrix, and positive values indicate that the model is underpredicting the covariance matrix (Bollen, 1989). Good model fit is indicated by RMSRs close to 0.

The Tucker-Lewis index (TLI; Tucker & Lewis, 1973) provides a measure of model improvement by comparing a hypothesized model against a viable null model. In addition, the TLI provides an adjustment for model complexity. Following the recommendations of Bollen (1989), two null models were used to gauge model improvement as measured by the TLI. The first null model investigated was the independence model. Each of the five nested models was evaluated through contrasts to a null model that specified the absence of relationships among the WISC-III's 12 subtests. TLI values of .90 or greater generally indicate good model fit for these types of comparisons (Bentler & Bonnett, 1980). The second set of comparisons utilized the one-factor model as the null model, and competing models (i.e., two- three-, four-, and five-factor models) were evaluated by comparison (Roid et al., 1993; Wechsler, 1991). Lastly, the comparative fit index (CFI; Bentler, 1990) provides a measure of model improvement that ranges from 0 to 1.0. The CFI has been found to remain stable in small samples (e.g., $N = 50$) and overcomes the downward bias associated with other measures of fit (Bentler, 1990).

RESULTS

Means and standard deviations across the 12 subtests for each of the three samples are presented in Table 2. Although restriction of range is particularly troublesome when conducting factor analytic studies because unstable estimates may be obtained (Gorsuch, 1983), an examination of Table 2 suggests that this was not an issue. Standard deviations for each of the subtests were all close to 3, consistent with data from the standardization sample.

Table 2
Means and (Standard Deviations) of WISC-III Subtests across Samples

WISC-III Subtests	Sample 1	Sample 2	Sample 3
Picture Completion	8.3 (3.5)	9.4 (3.0)	9.3 (2.8)
Information	6.4 (2.7)	7.8 (2.6)	7.5 (3.0)
Coding	7.9 (3.2)	8.8 (3.3)	8.1 (2.9)
Similarities	6.7 (3.4)	8.5 (2.8)	8.2 (3.0)
Picture Arrangement	7.9 (3.4)	7.8 (3.2)	8.8 (3.1)
Arithmetic	6.6 (2.7)	7.0 (2.5)	7.2 (2.5)
Block Design	8.1 (3.6)	8.2 (3.0)	8.8 (3.2)
Vocabulary	6.5 (3.3)	8.1 (2.7)	7.8 (2.8)
Object Assembly	8.5 (3.4)	9.0 (3.0)	9.4 (3.1)
Comprehension	7.0 (3.4)	9.1 (3.3)	8.7 (3.0)
Symbol Search	8.5 (3.4)	9.3 (3.1)	9.1 (2.8)
Digit Span	6.7 (2.6)	7.4 (2.4)	7.1 (2.7)

Table 3 provides goodness-of-fit statistics for each sample across the five models under investigation. The largest improvement was obtained by adding a second factor to the one-factor model. However, the four-factor model demonstrated the best overall fit as indicated by virtually all measures. Completely standardized structural coefficients for the four-factor model are presented in Table 4. Given the analogous findings across the three samples, samples are compared below according to the method used to evaluate model fit and improvement.

Not surprisingly, the incremental addition of factors resulted in a successive decrease in chi-square across all three samples. A more reasonable approach to the use of chi-square in comparing nested models was provided by the χ^2/df ratio. This measure indicated that, when model complexity was taken into account, the four-factor solution provided better fit than either the one-, two-, three-, or five-factor models. The AGFI also increased from the one-factor model to the four-factor model across all three samples. Thereafter, it either remained stable when a five-factor solution was considered (Sample 1) or dropped slightly (Samples 2 and 3).

Table 3
Model Evaluation Statistics

	Goodness-of-fit statistics					Model improvement statistics				
	χ^2	*df*	χ^2/df	AGFI	RMSR	χ^2	*df*	TLI_0	TLI_1	CFI
Sample 1										
One-factor model	246.48*	54	4.56	.75	.07	—	-	.83	—	.86
Two-factor model	157.85*	53	2.98	.84	.06	88.63*	1	.90	.44	.92
Three-factor model	131.31*	51	2.57	.86	.05	26.54*	2	.92	.56	.94
Four-factor model	101.40*	48	2.11	.89	.05	29.91*	3	.95	.69	.96
Five-factor model	100.23*	46	2.18	.89	.05	1.17	2	.94	.67	.96
Sample 2										
One-factor model	266.50*	54	4.94	.75	.10	—	-	.69	—	.75
Two-factor model	146.67*	53	2.77	.87	.07	119.83*	1	.86	.55	.89
Three-factor model	99.46*	51	1.95	.91	.05	47.21*	2	.93	.76	.94
Four-factor model	83.62*	48	1.74	.92	.04	15.84*	3	.94	.81	.96
Five-factor model	83.46*	46	1.81	.91	.04	.16	2	.94	.79	.96
Sample 3										
One-factor model	267.01*	54	4.94	.75	.09	—	-	.72	—	.77
Two-factor model	170.15*	53	3.21	.85	.07	96.86*	1	.84	.44	.87
Three-factor model	122.11*	52	2.39	.88	.06	48.04*	1	.90	.65	.92
Four-factor model	90.29*	48	1.88	.91	.05	31.82*	4	.94	.78	.95
Five-factor model	86.93*	46	1.89	.90	.05	3.36	2	.94	.77	.95

Note.—All values have been rounded to the second decimal place for ease of presentation. AGFI = adjusted goodness-of-fit index; RMSR = standardized root mean square residual; TLI_0 = Tucker-Lewis index evaluated with independence model; TLI_1 = Tucker-Lewis index evaluated with one-factor model; CFI = comparative fit index.
* $p < .05$.

Table 4
Completely Standardized Structural Coefficients

	Factor Loadings											
	Verbal Comprehension Index			Freedom from Distractibility Index			Perceptual Organization Index			Processing Speed Index		
	Sample 1	Sample 2	Sample 3	Sample 1	Sample 2	Sample 3	Sample 1	Sample 2	Sample 3	Sample 1	Sample 2	Sample 3
Information	.80	.71	.75									
Vocabulary	.86	.83	.79									
Similarities	.79	.81	.81									
Comprehension	.79	.70	.68									
Arithmetic				.84	.68	.80						
Digit Span				.64	.38	.58						
Block Design							.82	.72	.65			
Object Assembly							.71	.61	.67			
Picture Completion							.74	.61	.62			
Picture Arrangement							.69	.65	.59			
Coding										.59	.43	.46
Symbol Search										.79	.99	.99
	Factor Correlations											
Freedom from Distractibility Index	.85	.69	.72									
Perceptual Organization Index	.77	.50	.60	.82	.58	.58						
Processing Speed Index	.41	.39	.22	.59	.44	.38	.67	.39	.46			

Residuals for all five models were relatively low across the three samples (standardized RMSRs $\leq .10$). However, a small drop in the standardized RMSR was observed across samples as the number of factors increased from a one-factor model to a three-factor model in Sample 1, and from a one-factor model to a four-factor model for Samples 2 and 3. Thereafter, the standardized RMSR remained stable.

Model improvement indices are also presented in Table 3. These measures gauge model improvement by comparing a hypothesized model to a null or baseline model. Successive model improvement was investigated by comparing a more restrictive model (e.g., one-factor model) to a less restrictive model (e.g., two-factor model). The largest improvement in model fit was observed from the one-factor model to the two-factor model in Samples 1, $\chi^2_d(1) = 88.63$, $p < .05$; 2, $\chi^2_d(1) = 119.83$, $p < .05$; and 3, $\chi^2_d(1) = 96.86$, $p < .05$. Thereafter, the addition of a third factor provided significant improvement over the two-factor model for each of the three samples, all *p*s < .05, and the four-factor model provided significant improvement over the three-factor model, all *p*s < .05. No significant improvement was observed with the addition of a fifth factor, all *p*s > .05.

The TLI was also used to evaluate competing models against a viable null model. This measure increased within each of the three samples as additional factors were considered, irrespective of whether comparisons were directed toward the independence model (TLI_0) or the one-factor model (TLI_1). However, both the TLI_0 and TLI_1 dropped slightly when a fifth factor was added to the four-factor solution (see Table 3). The largest increase in the TLI_0 occurred when an additional factor was added to the one-factor model. This was observed for Samples 1, 2, and 3 (TLI_0 difference = .07, .17, and .12, respectively). However, with the exception of Sample 1, the TLI_0 estimates for the one- and two-factor models failed to exceed the recommended value of .90 (Bentler & Bonett, 1980). All other models (i.e., three-, four- and five-factor models) demonstrated significant improvement over the independence model (TLIs $\geq .90$).

Contrasting the hypothesized two-, three-, four-, and five-factor models with the one-factor model indicated support for a four-factor model across all three samples. Differences between the TLI_0 and TLI_1 estimates can be attributed to differences in baseline models. More restrictive baseline models (i.e., TLI_0) generally result in better fit than less restrictive baseline models (i.e., TLI_1) (Bollen, 1989). Moreover, the relatively low estimates obtained by comparing each of the competing models with the one-factor model parallel similar findings presented in the WISC-III manual (Wechsler, 1991).

The amount of covariance among the 12 WISC-III subtests accounted for by a one-factor model ranged from 75% (Sample 2 CFI = .75) to 86% (Sample 1 CFI = .86). Whereas the four- and five-factor models accounted for the most covariance among the 12 WISC-III subtests in Samples 1, 2, and 3 (CFIs = .96, .96, and .95, respectively), the five factor-model accounted for no more covariance than that explained by the four-factor model (see Table 3).

Lastly, a caveat should be noted with respect to the estimation of the three-factor model on Sample 3. In estimating this model, a negative error variance was observed on the Symbol Search subtest. The model was subsequently adjusted by "fixing" the variance of this subtest to .05, thereby allowing the remaining parameters of the model to be estimated (Bollen, 1989).

DISCUSSION

Results of the present study indicated remarkable convergence across all three samples. Across all groups, the one-factor solution demonstrated poorer fit in comparison to all competing models. However, the structure of this investigation was not appropriately set up to substantiate the existence of a general factor. The one-factor solution was presented to provide a baseline against which the models comprised of two through five factors could be compared (Bentler & Bonett, 1980). Thus, no claim is made as to the existence or nonexistence of a one-factor solution.

The two-factor solution was found to provide estimates in support of the VIQ-PIQ dichotomy of aptitude found during the standardization of the WISC-III (Wechsler, 1991) and elsewhere (Kaufman, 1990; Roid et al., 1993). Support for this contention comes by way of contrast between the one-factor model and the two-factor model: The two-factor model was found to provide better incremental fit over the one-factor model than that observed through any other single factor increase comparison, irrespective of the measure employed.

A solution defined by three factors was questionable. Whereas several statistical indices of fit were indicative of improvement over the two-factor model, other evidence suggested that the three-factor solution was not viable. As previously noted, an indication of specification error was observed with respect to Symbol Search for Sample 3. The error variance for Symbol Search on the third factor was found to be negative. Further investigation into Samples 1 and 2 indicated a near similar situation, despite the fact that the three-factor solution converged without incident in Samples 1 and 2. It would appear that the three-factor model is simply an under- or overestimate of the number of factors underlying the 12 subtests of the WISC-III. In light of this specification error, estimates for the three-factor model should be interpreted with caution.

The existence of four factors underlying the 12 subtests of the WISC-III appears well substantiated. The four-factor model provided the best overall fit within each of the three samples as measured by the χ^2/df ratio and the TLI_1. In addition, the four-factor model provided better estimates than those obtained for the one-, two-, and three-factor models and either matched or exceeded the estimates obtained for the five-factor model as measured by the AGFI, RMSR, TLI_0, and CFI.

Little support was found for the five-factor solution. Indications of model improvement for the five-factor model were either negligible or nonexistent in comparison to the four-factor solution. Parsimony suggests that the addition of a fifth factor provides negligible insight into the 12 WISC-III subtests.

Factor analytic research on the WISC-III has resulted in disagreement on whether one (Macmann & Barnett, 1994) or four (Keith & Witta, 1994, in Keith, 1994; Roid et al., 1993; Wechsler, 1991) factors best explain its subtest variation. The results of our investigation clearly support interpretation of the WISC-III's four index scores. This finding prompted one reviewer to question why our results did not align with Macmann and Barnett's (1994) contention that the subtests located on the WISC-III are best explained by a single factor. Several plausible explanations exist for this discrepancy. First, Macmann and Barnett limited their restricted factor analyses to comparisons between a one- and two-factor model; a four-factor solution was not investigated. Moreover, their discussion in support of a one-factor model placed little emphasis on the results of their restricted factor analyses. Second, we dealt exclusively with children who were receiving special education. Third, and perhaps most important, the choice of fit statistics may lead researchers to draw different conclusions for a given set of data (Keith, 1994). This was illustrated when the same data Macmann and Barnett used to support a one-factor model was reanalyzed in support of a four-factor model (Keith & Witta, in press).

The present investigation used multiple measures of model fit and model improvement, the results of which provide consensus on two points. First, irrespective of the fit measure employed, the largest indication of model improvement was gained when an additional factor was added to the one-factor model (i.e., the VIQ-PIQ dichotomy). In other words, the two-factor model provided a larger indication of model improvement in comparison to the one-factor model than was evidenced through comparisons between the two- and three-factor models, three- and four-factor models, and four- and five-factor models. Second, the four-factor solution yielded better *overall* fit in comparison to all other competing models.

The fact that a two-factor model provides superior fit than a one-factor model suggests that two factors should be preferred over one. However, because the four-factor solution clearly demonstrated the best fit across the three independent samples of children receiving special education, the interpretation of a four-factor solution is preferred. Thus our results align with the preponderance of evidence that supports a first-order four-factor solution (Keith, 1994; Roid et al., 1993; Wechsler, 1993). Future research should begin to examine the predictive validity of these factor scores in forecasting academic achievement.

Several limitations regarding the three samples should be noted. The first concerns the heterogeneity of classifications. Whereas children in Sample 1 were primarily identified as LD (i.e., 81%), the remaining children constituted one of five alternative classifications (i.e., MIMR, ED, SLI, OHI, or MOMR). In addition, both within and between sample variability may be present among children classified as LD to the extent that the criteria used to label these children were not consistent and that they exhibit different types of LDs. Second, each of the samples was obtained from the southwestern region of the country. Limitations arise because eligibility requirements may vary across different regions of the country. Because we were unable to obtain the criteria used to classify children or the individual types of LDs present in our samples, the degree to which the children in our samples differ among themselves is unknown. Thus, while our results demonstrate replication across three independent samples, different solutions may exist in the presence of more homogeneously defined classifications. Future research should focus on exploring this possibility.

It was previously noted that in an independent study the majority of clinicians generally administer only 10 of the WISC-III's 13 subtests. Whereas clinicians are encouraged to administer all 13 subtests, common practice appears to align more with the fact that only 10 of these 13 subtests are actually being utilized for diagnostic appraisals. Thus, future investigations on the WISC-III should also focus on the implications of deriving various composite scores when only 10 of the 13 subtests are administered.

REFERENCES

Allen, S. R., & Thorndike, R. M. (1995). Stability of the WPPSI-R and WISC-III factor structure using cross-validation of covariance structure models. *Journal of Psychoeducational Assessment, 13,* 3–20.

Bell, W. M. (1994). *A confirmatory factor analysis of the Wechsler Intelligence Scale for Children-Third Edition on children with learning disabilities.* Unpublished doctoral dissertation, Mississippi State, Mississippi.

Bentler, P. M. (1990). Comparative fit indexes in structural model. *Psychological Bulletin, 107,* 238–256.

Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin, 88,* 588–606.

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

Brody, N. (1985). The validity of tests of intelligence. In B. Wolman (Ed.), *Handbook of intelligence* (pp. 353–389). New York: Wiley.

Budoff, M. (1987). The validity of learning potential assessment. In C. S. Lidz (Ed.), *Dynamic assessment: An interactional approach to evaluating learning potential* (pp. 52-81). New York: Guilford Press.

Conger, A. J., Conger, J. C., Farrell, A. D., & Ward, D. (1979). What can the WISC-R measure? *Applied Psychological Measurement, 3,* 421–436.

Crano, W. D., Kenny, J., & Campbell, D. T. (1972). Does intelligence cause achievement? A cross-lagged panel analysis. *Journal of Educational Psychology, 63,* 258–275.

Dean, R. (1980). Factor structure of the WISC-R with Mexican-American children. *Journal of School Psychology, 18,* 234–239.

Eysenck, H. J., & Barrett, P. (1985). Psychophysiology and measurement of intelligence. In C. R. Reynolds & V. L. Wilson (Eds.), *Methodological and statistical advances in the study of individual differences* (pp. 1–49). New York: Plenum Press.

Feuerstein, R. (1985). The importance of the role of assessment in successful integration programs: A dynamic alternative to traditional psychometric approaches. *Educational and Child Psychology, 2,* 138–144.

Glutting, J. J., Konold T. R., McDermott, P. A., Kush, J. C., & Watkins, M. W. (1995). *Structure and associated characteristics of a normative subtest taxonomy for the WISC-III's standardization sample.* Manuscript submitted for publication.

Goh, D. S., Teslow, C. J., & Fuller, G. B. (1981). The practice of psychological assessment among school psychologists. *Professional Psychology, 12,* 696–706.

Gorsuch, R. L. (1983). *Factor analysis* (2nd ed.). Hillsdale, NJ: Erlbaum.

Gutkin, T. B., & Reynolds, C. R. (1981). Factorial similarity of the WISC-R for White and Black children from the standardization sample. *Journal of Educational Psychology, 73,* 227–231.

Jencks, C. (1972). *Inequality: A reassessment of the effect of family and schooling in America.* New York: Basic Books.

Jensen, A. R. (1980). *Bias in mental testing.* New York: Free Press.

Joreskog, K. G., & Sorbom, D. (1993). *LISREL 8: User's reference guide.* Chicago: Scientific Software.

Juliano, J. M., Haddad, F. A., & Carroll, J. L. (1988). Three-year stability of WISC-R factor scores for Black and White, female and male children classified as learning disabled. *Journal of School Psychology, 26,* 317–325.

Kaufman, A. S. (1994). A reply to Macmann and Barnett: Lessons from the blind men and the elephant. *School Psychology Quarterly, 9,* 199–207.

Kaufman, A. S. (1990). *Assessing adolescent and adult intelligence.* Boston, MA: Allyn and Bacon.

Kaufman, A. S. (1975). Factor analysis of the WISC-R at eleven age levels between 6 and 16 years. *Journal of Clinical and Consulting Psychology, 43,* 135–147.

Keith, T. Z. (1994). Intelligence is important, intelligence is complex. *School Psychology Quarterly, 9,* 209–221.

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

Kush, J. C. (1996). Factor structure of the WISC-III for students with learning disabilities. *Journal of Psychoeducational Assessment, 14,* 32–40.

Logerquist-Hansen, S., & Barona, A. (1994). *Factor structure of the Wechsler Intelligence Scale for Children-III for Hispanic and non-Hispanic White children with learning disabilities.* Paper presented at the meeting of the American Psychological Association, Los Angeles, CA.

Lutey, C., & Copeland, E. P. (1982). Cognitive assessment of the school-aged child. In C.R. Reynolds & T.B. Gutkin (Eds.), *The handbook of school psychology* (pp. 121–155). New York: John Wiley.

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., Balla, J. R., & McDonald, R. P. (1988). Goodness-of-fit indices in confirmatory factor analysis: The effect of sample size. *Psychological Bulletin, 103,* 391–410.

Naglieri, J. A. (1981). Factor structure of the WISC-R for children identified as learning disabled. *Psychological Reports, 49,* 891–895.

Reschly, D. (1978). WISC-R factor structure among Anglos, Blacks, Chicanos, and Native American Papagos. *Journal of Consulting and Clinical Psychology, 46,* 417–422.

Reynolds, C. R., & Gutkin, T. B. (1980). Stability of the WISC-R factor structure across sex at two age levels. *Journal of Clinical Psychology, 36,* 775–777.

Reynolds, C. R., & Kaufman, A. S. (1990). Assessment of children's intelligence with the Wechsler Intelligence Scale for Children-Revised (WISC-R). In C. R. Reynolds & R. W. Kamphaus (Eds.), *Handbook of psychological and educational assessment of children: Intelligence and achievement.* New York: Guilford.

Roid, G. H., Prifitera, A., & Weiss, L. G. (1993). Replication of the WISC-III factor structure in an independent sample. *Journal of Psychoeducational Assessment WISC-III Monograph,* 6–21.

Sapp, G. L., Chissom, B., & Graham, E. (1985). Factor analysis of the WISC-R for gifted students: A replication and comparison. *Psychological Reports, 57,* 947-951.

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.

Swanson, L. H. (1994). The role of working memory and dynamic assessment in the classification of children with learning disabilities. *Learning Disabilities Research and Practice, 9,* 190–202.

Taylor, R. L., Ziegler, E. W., & Partenio, I. (1984). An investigation of WISC-R Verbal-Performance differences as a function of ethnic status. *Psychology in the Schools, 21,* 437–441.

Terman, L. M., & Oden, M. (1959). *The gifted group at mid-life.* Stanford University Press.

Tucker, L. R., & Lewis, C. (1973). A reliability coefficient for maximum likelihood factor analysis. *Psychometrika, 38,* 1–10.

Wechsler, D. (1974). *WISC-R manual.* New York: The Psychological Corporation.

Wechsler, D. (1991). *WISC-III manual.* San Antonio, TX: The Psychological Corporation.

Wilson, M. S., & Reschly, D. J. (1996). Assessment in school psychology training and practice. *School Psychology Review, 25,* 9–23.

Witt, J. C. (Ed.). (1994). *School Psychology Quarterly, 9*(3).

Woodcock, R. W. (1990). Theoretical foundations of the WJ-R measures of cognitive ability. *Journal of Psychoeducational Assessment, 8,* 231–258.