In: Intelligence Quotient
Editor: Joseph C. Kush

ISBN: 978-1-62618-728-3

Chapter 4

SUBTESTS, FACTORS, AND CONSTRUCTS: WHAT IS BEING MEASURED BY TESTS OF INTELLIGENCE?

Marley W. Watkins[1,*] and Christina M. Ravert[2,†]
[1]Department of Educational Psychology, Baylor University, Waco, Texas, US
[2]Manheim Township School District, Lancaster, Pennsylvania, US

ABSTRACT

Confirmatory factor analyses were conducted to investigate hypotheses regarding the Arithmetic and Digit Span subtests that form the third factor of the Wechsler Intelligence Scale for Children. Previous research suggested that these subtests may be measures of working memory, quantitative ability/reasoning, or some other ability. When the Wechsler Intelligence Scale for Children-Third Edition was administered to a referral sample of adolescents in conjunction with marker tests for memory and quantitative reasoning, a five factor solution was optimal. The resulting factors were verbal comprehension, perceptual organization, processing speed, quantitative reasoning, and memory. Based on these results, the WISC Arithmetic and Digit Span subtests are measures of different abilities with Arithmetic being a measure of quantitative reasoning and Digit Span a measure of memory. Given these results, interpretation of the Arithmetic subtest as a measure of memory may not be accurate.

Keywords: Intelligence, CFA, WISC-III, WISC-IV, Arithmetic, Digit Span

---

* This report is based on the dissertation of the first author that was supervised by the second author. The contributions of professors Barbara Schaefer, Hoi Suen, and Lisa Conyers are gratefully acknowledged. An abbreviated version was presented at the International Test Commission meeting in Amsterdam, Netherlands.
Correspondence concerning this chapter should be addressed to Marley W. Watkins, Department of Educational Psychology, Baylor University, One Bear Place #97301, Waco, TX 76798-7301. E-mail: marley_watkins@baylor.edu.
† christina_ravert@mtwp.net..

There appears to be a general consensus that child versions of the Wechsler Intelligence Scale measure verbal comprehension and perceptual organization/reasoning abilities (Zachary, 1990), but there has been less agreement as to the existence and nature of additional abilities measured by the Wechsler child scale. For example, the first factor analysis of the Wechsler Intelligence Scale for Children (WISC; Wechsler, 1949) found verbal comprehension and perceptual organization dimensions as well as a third factor that was loaded by the Arithmetic and Digit Span subtests (Cohen, 1959). This third factor was labeled Freedom from Distractibility (FD) and its interpretation as a memory factor was explicitly disclaimed.

Factor analyses of the Wechsler Intelligence Scale for Children-Revised (WISC-R; Wechsler, 1974) also produced three factors labeled Verbal Comprehension (VC), Perceptual Organization (PO), and Freedom from Distractibility (Kaufman, 1975). The first two factors were consistent with those found with the WISC, but the FD factor was formed by the Arithmetic, Digit Span, and Coding subtests. After endorsing the FD label, Kaufman (1975) suggested that this factor might be a measure of numerical ability.

Wielkiewicz (1990) reported that a wide range of hypotheses had been advanced to account for low scores on the FD factor, including problems with concentration/distractibility, motivation, and memory. In fact, a joint factor analysis of the WISC-R and the Differential Ability Scales (Elliott, 1990) revealed that the FD factor was not an intact entity (Stone, 1992). A summary of the results of eight studies that jointly factor analyzed the WISC-R and the Woodcock Johnson-Revised Tests of Cognitive Abilities and Achievement (Woodcock & Johnson, 1989) also concluded that the WISC-R FD subtests were actually measures of separate abilities rather than common indicators of attention or distractibility (Woodcock, 1990). Specifically, Woodcock reported that Arithmetic was a measure of quantitative ability, Coding was a measure of processing speed, and Digit Span was a measure of short-term memory.

To help strengthen the FD factor, a new Performance subtest, Symbol Search, was added to the WISC when it was next revised. It was assumed that the structure of the new Wechsler Intelligence Scale for Children-Third Edition (WISC-III; Wechsler, 1991) would consist of VC, PO, and FD factors. In actuality, the Coding and Symbol Search subtests split away to form a new fourth factor, named Processing Speed (PS), while the Arithmetic and Digit Span subtests remained as dual measures of the FD factor.

Over time, evidence has accumulated to suggest that the FD factor might not be a measure of attention/distractibility (Cohen, Becker, & Campbell, 1990; Oakland, Broom, & Glutting, 2000; Reinecke, Beebe, & Stein, 1999; Riccio, Cohen, Hall, & Ross, 1997). For example, Riccio et al. (1997) correlated WISC-III factor index scores with behavioral and neuropsychological measures and found that the FD factor did not significantly correlate with any of the behavioral scales that measured attention. Similar findings were reported by Lowman, Schwanz, and Kamphaus (1996), who found a non-significant relationship between FD factor scores and hyperactivity and attention problem scale scores from the teacher-report form of the Behavior Assessment System for Children (BASC-TRS; Reynolds & Kamphaus, 1992).

Given the weakness of the evidence supporting an attention/distractibility explanation of the WISC third factor, Keith and Witta (1997) argued that Quantitative Reasoning would be a better name because of its very high loading on *g* (.90), the obvious numerical content of Arithmetic and Digit Span subtests, and the higher loading of Arithmetic (.82) than Digit

Span (.52) on the FD factor. Carroll (1997) also opined that Quantitative Reasoning might be a better name for the FD factor. In contrast, Kranzler (1997) posited that Working Memory Efficiency provided a more appropriate label for the FD factor. Prifitera, Weiss, and Soklofske (1998) agreed that Working Memory was a better name for the FD factor given the tasks required by the FD subtests. These opinions were investigated by a joint confirmatory factor analysis of the WISC-III and Woodcock Johnson-III cognitive and achievement tests (Woodcock, McGrew, & Mather, 2001), which supported a model in which Arithmetic loaded with math achievement measures on a quantitative knowledge factor and the Digit Span subtest loaded on a short-term memory factor (Phelps, McGrew, Knopik & Ford, 2005).

In recognition of the debate surrounding the FD factor, the Wechsler Intelligence Scale for Children-Fourth Edition (WISC-IV; Wechsler, 2003) third factor was renamed Working Memory (WM). Additionally, a new subtest, Letter-Number Sequencing, was added to that factor, the Digit Span subtest was retained as a core subtest, and the Arithmetic subtest was lengthened and made an optional rather than mandatory WM subtest. Subsequent factor analyses of the WISC-IV normative sample found "problems with placement of Arithmetic on the Working Memory factor" (Keith, 2005, p. 595). Flanagan and Kaufman (2004) also expressed concern about the Arithmetic subtest, suggesting that it is a measure of quantitative knowledge rather than working memory.

After more than 50 years and four versions of the WISC, there remain four major hypotheses about what is measured by the Arithmetic and Digit Span subtests. Specifically, both subtests measure memory, both subtests measure quantitative reasoning, one subtest measures memory and the other measures quantitative reasoning, or both subtests measure some other ability. One way to test these hypotheses and identify the construct(s) that underlie the WISC third factor is to administer tests that correspond with the quantitative knowledge and short-term memory hypotheses in conjunction with the WISC. This is similar to the confirmatory cross battery technique used to develop the integrated intelligence theory (Flanagan, McGrew, & Ortiz, 2000). However, the tests used in the cross battery approach were subtests selected from other intelligence measures, and many of those subtests lacked sufficient reliability and validity evidence to verify that they are indeed psychometrically sound measures of the hypothesized constructs (Phelps et al., 2005). Therefore, it is important to use tests (*marker variables*) that have been well researched and found to be psychometrically adequate measures of the respective constructs hypothesized to be measured by the third factor. As noted by Child (2006), "marker variables are those about which we have reliable knowledge of their properties and are deliberately placed in factor analyses to form a nucleus around which less well-defined variables can be assessed" (p. 56).

The Educational Testing Service (ETS) developed a kit of brief factor-referenced cognitive tests (Ekstrom, French, Harman, & Dermen, 1976) that Carroll (1985) recommended for use as cognitive marker variables. Only factors that were deemed sufficiently established were included. To select the factors to consider for the kit, a panel of 20 prominent persons working in the field of factor analysis and human assessment set the necessary criteria (Ekstrom, 1973). A total of 72 marker variables for 23 cognitive factors were included in the 1976 edition of the Kit of Factor-Referenced Cognitive Tests (Ekstrom et al., 1976). Marker variables for two of these cognitive factors (memory span and general reasoning) were utilized in the present study. It should be noted that the general reasoning factor consists solely of marker variables that are quantitative in nature and can thus be considered a quantitative reasoning factor (Carroll, 1993).

If the WISC third factor is really a measure of either short-term memory or quantitative reasoning ability, then the data obtained from this study should fit a model where the Digit Span and Arithmetic subtests align with that construct as represented by its marker variables. It is also possible that the data will best fit a model where the two third factor subtests align with measures of separate constructs as suggested by Woodcock (1990) and Flanagan et al. (2000). In that case, the data would best fit a model where the Arithmetic subtest loads on a factor consisting of the quantitative reasoning marker variables and the Digit Span subtest loads on a factor comprised of the memory marker variables. If the third factor truly measures attention or some other ability, then the data should best fit a model where the third factor subtests remain a separate factor.

Therefore, the purpose of this study is to investigate the nature of the third factor of the WISC. The WISC-III was chosen because its administration routinely included both Digit Span and Arithmetic subtests whereas Arithmetic is an optional subtest on the WISC-IV and optional subtests are rarely administered in clinical practice (Canivez & Watkins, 1998). Marker variables reflecting the constructs of short-term memory and quantitative reasoning were also administered and the resulting data were analyzed using confirmatory factor analysis to identify the construct(s) being measured by the third factor.

METHOD

Participants

A total of 87 (51 male and 36 female) students in grades 6-12 from the mid-Atlantic region of the country participated in this study. Ethnicity of the participating students was reported to be 85% White, 9% Black, 4% Hispanic, and 2% Other. The majority of the students were in grades six through eight (92%) with a mean age of 13.4 years (*SD* = 1.3). Both special education identified (21% with learning disabilities, 13% with mental retardation, 8% with serious emotional disabilities, 8% gifted, and 5% with other heath impairments) and referred but nonidentified students (45%) were included in the sample.

Instruments

The WISC-III is appropriate for children between the ages of 6 years, 0 months through 16 years, 11 months. It contains 13 subtests, but only 10 are mandatory. If both Digit Span and Symbol Search are administered in conjunction with the ten required subtests, four factor indices can be computed: VC, PO, FD, and PS. Each of the four factor indices has a mean of 100 and a standard deviation of 15. These factor indices were derived based on the results of exploratory and confirmatory factor analyses of the standardization sample data using 12 of the WISC-III subtests. Full details of the WISC-III and its standardization are presented in Wechsler (1991). Additional reliability and validity data are provided by Sattler (2001) as well as Zimmerman and Woo-Sam (1997).

A Kit of Factor-Referenced Cognitive Tests was developed by the Educational Testing Service (Ekstrom et al., 1976) to provide researchers with a means of identifying cognitive factors in factor analytic studies. The use of common variables (marker variables) that assess identified constructs can facilitate the interpretation of other factors that contain those variables (Carroll, 1997; Ekstrom et al., 1976). The specific tests or marker variables that were included in the 1976 edition of the kit were selected after reviewing the 1963 version of the kit and the relevant literature published between editions of the kit. In addition, the variables were subjected to field experiments to verify that they were useful markers for their respective factors.

This study utilized marker variables of memory span and general reasoning factors. Visual Number Span and Auditory Letter Span, which are appropriate for students in grades 6-16, were selected from the memory span factor. With the Visual Number Span test, the student is visually presented with a series of digits of varying lengths at a rate of 1 digit per second. Immediately after the visual presentation of the series, the student is asked to write the series of numbers from memory. The Auditory Letter Span test requires the student to listen to a series of letters and then write the series from memory after it has been presented.

Two quantitative reasoning marker variables appropriate for students in grades 6-12 were also selected. For the Arithmetic Aptitude Test, the student is asked to read word problems that require only arithmetic to solve, and instructed to solve the problem and choose the correct answer from among five choices that are presented within a ten-minute time limit. In contrast, for the Necessary Arithmetic Operations test the student is asked to read a word problem and select the arithmetic operation(s) needed to solve the problem. Four answer choices are provided for each problem, and the student is given five minutes to complete the test.

Psychometric evidence on these marker variables was provided by Ekstrom et al. (1976). For example, reliability estimates for Visual Number Span and Auditory Letter Span were .63 and .65, respectively, and the memory span factor and/or its marker variables had been referenced in 21 studies prior to publication of the kit. A reliability estimate of .82 was reported for the Arithmetic Aptitude Test and .72 for the Necessary Arithmetic Operations test. The general reasoning factor and its marker variables had been referenced in 67 studies prior to publication of the kit. The Arithmetic Aptitude Test and Necessary Arithmetic Operations were timed tasks making them inappropriate for calculating an internal consistency coefficient in this sample. The remaining marker tests, Auditory Letter Span and Visual Number Span, had alpha coefficients of .74 and .80, respectively, in this sample.

Procedures

Data were collected by alumni and graduate students of a mid-Atlantic doctoral school psychology program following approval by the university’s IRB. A total of 235 non-retired alumni were asked to obtain parental consent for additional testing on students referred for a psycho-educational evaluation that contained the WISC-III. Alumni could either collect all WISC-III and marker variable data themselves or allow doctoral school psychology students to collect part or all of the data in their school districts. A total of 36 cases were collected by 22 alumni whereas the remaining 51 cases were collected by graduate students in four school districts. In all cases, the WISC-III was administered first, followed by the marker variables.

Analysis

Confirmatory factor analysis (CFA) using maximum likelihood estimation was applied to the covariance matrix using Mplus version 7 for the Macintosh. The obtained solutions were checked for convergence and the adequacy of the parameter estimates and their associated standard errors were examined prior to considering the reported fit indices. According to Hu and Bentler (1998), values ≥ .95 for CFI, ≤ .08 for SRMR, and ≤ .06 for RMSEA indicate that there is a relatively good fit between the hypothesized model and the sample data. Consequently, these cutoff values were applied in the current study. Models were also evaluated with the Bayesian Information Criterion (BIC; Schwarz, 1978), which is a parsimony corrected measure where the smallest value indicates the relatively best fit among the tested models.

Various experts have proposed inconsistent 'rules of thumb' to determine the minimum sample size needed for factor analysis, but recent research has revealed that "when communalities are high, good recovery of population factors can be achieved with relatively small samples" (MacCallum, Widaman, Preacher, & Hong, 2001). In most studies, communalities ≥ .60 were considered high. Given a median communality of .66 for the current variables, good recovery of population factors is feasible with a smaller sample size. Based on model parameters, a sample size of 78 would be needed to detect poor fit (Loehlin, 2003). Therefore, the current sample size of 87 should be sufficient.

Each of the four models selected for this study was designed to test a specific hypothesis about the third factor or the subtests loading on that factor. Given that there is evidence to suggest that the Arithmetic and Digit Span subtests are measures of memory and/or quantitative reasoning, models were developed to reflect each of these hypotheses.

Model I

The first model was designed to test the hypothesis that the WISC third factor is really a measure of memory span (MS). In this model, five factors were specified: VC, PO, PS, Short-term Memory Span (MS), and Quantitative Reasoning (QR). The VC factor was hypothesized to have significant loadings from four WISC-III subtests including Information, Vocabulary, Similarities, and Comprehension. Likewise, four WISC-III subtests were hypothesized to load on the PO factor including Picture Completion, Picture Arrangement, Block Design, and Object Assembly. The PS factor was hypothesized to consist of the WISC-III Coding and Symbol Search subtests, and the MS factor was hypothesized to include the WISC-III Arithmetic and Digit Span subtests as well as the memory span marker variables including the Visual Number Span Test and the Auditory Letter Span Test (Ekstrom et al., 1976). The quantitative reasoning (QR) factor was hypothesized to be made up of the marker variables for the general reasoning factor including the Arithmetic Aptitude Test Part 1 and the Necessary Arithmetic Operations Test (Ekstrom et al., 1976).

Model II

The second model was developed to test the hypothesis that the WISC third factor really reflects a quantitative reasoning construct. As with the first model, this model also consisted of five factors: VC, PO, PS, MS, and QR. The VC, PO, and PS factors were specified to have the same subtest composition as in the first model; however, the MS factor was hypothesized to contain only the two marker variables for memory span (Ekstrom et al., 1976). In this model, the WISC-III Arithmetic and Digit Span subtests were specified to load on the QR factor along with the two marker variables for quantitative reasoning (Ekstrom et al., 1976).

Model III

The third model was constructed to test the hypothesis that the WISC Arithmetic subtest is a measure of quantitative reasoning and the Digit Span subtest is a measure of memory. Five factors were also specified for this model including VC, PO, PS, MS, and QR. The first three factors were specified to have the same subtest composition that they had in Models I and II. However, in this third model the MS factor was hypothesized to include the two memory span marker variables as well as the WISC-III Digit Span subtest, and the QR factor was specified to have significant loadings from the WISC-III Arithmetic subtest and the two quantitative reasoning marker variables.

Table 1. Fit Statistics for Plausible Models Employing Twelve Subtests from the Wechsler Intelligence Score for Children-Third Edition and Four Marker Variables from Memory and Quantitative Reasoning Factors

Model	df	χ^2	CFI	SRMR	RMSEA	RMSEA 90% CI	BIC
I	94	148.75	.930	.077	.082	.056-.106	7947.01
II	94	138.63	.943	.062	.074	.046-.099	7936.89
III	94	111.34*	.978	.050	.046	.000-.077	7909.60
IV	89	132.51	.944	.060	.075	.046-.101	7953.10
III-Hierarchical	99	119.14*	.974	.055	.048	.000-.078	7895.07

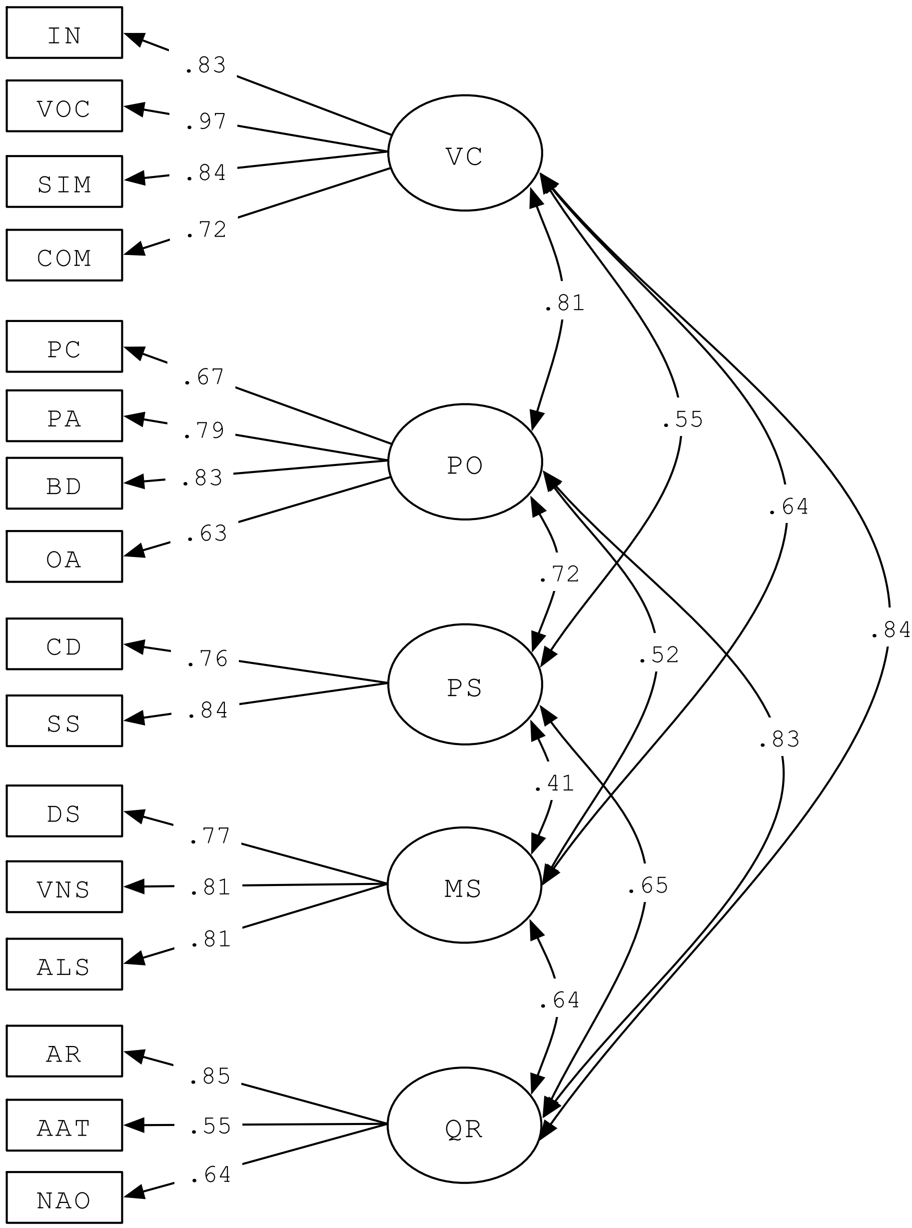
* $p \geq .05$.

Model IV

The fourth model was specified to consider the possibility that none of the evaluated hypotheses were correct, and the WISC third factor is a measure of some other construct such as attention or executive processes. Six factors were specified in this model including VC, PO, PS, MS, QR, and FD. The first three factors were specified to have the same composition that they had in Models I, II, and III. The MS and QR factors were hypothesized to be formed by their respective marker variables (Ekstrom et al. 1976), and the FD factor was specified to include the WISC-III Arithmetic and Digit Span subtests.

RESULTS

Means (with standard deviations in parentheses) for the WISC-III Full Scale IQ, Verbal IQ, and Performance IQ were 96.8 (17.9), 97.5 (17.9), and 96.7 (17.5), respectively. An examination of the univariate skewness and kurtosis values for the 16 variables under consideration indicated that all were within accepted limits for normality (Fabrigar, Wegener, MacCallum, & Strahan, 1999).



Note. IN = WISC-III Information subtest, VOC = WISC-III Vocabulary subtest, SIM = WISC-III Similarities subtest, COM = WISC-III Comprehension subtest, PC = WISC-III Picture Completion subtest, PA = WISC-III Picture Completion subtest, BD = WISC-III Block Design subtest, OA = WISC-III Object Assembly subtest, CD = WISC-III Coding subtest, SS = WISC-III Symbol Search subtest, DS = WISC-III Digit Span subtest, AR = WISC-III Arithmetic subtest, VNS = Visual Number Span marker variable, ALS = Auditory Letter Span marker variable, AAT = Arithmetic Aptitude Test marker variable, NAO = Necessary Arithmetic Operations marker variable, VC = Verbal Comprehension factor, PO = Perceptual Organization factor, PS = Processing Speed factor, MS = Memory factor, and QR = Quantitative Reasoning factor.

Figure 1. Most Plausible Model Employing Twelve Subtests from the Wechsler Intelligence Score for Children-Third Edition (WISC-III) and Four Marker Variables from Memory and Quantitative Reasoning Factors.

Two other conditions for multivariate normality are that all linear combinations of variables follow a normal distribution and all subsets of variables in the data set are normally distributed (Stevens, 2009). This was verified by examining the scatterplots of all possible variable pairs. All scatterplots had an elliptical shape. In addition, multivariate kurtosis was examined with the Normalized Estimate (Mardia, 1974) value calculated with EQS 6.1. Normalized Estimate values between -3 and +3 indicate that the data is multivariate normal (Bentler & Wu, 2002). The Normalized Estimate value for the sample data was -.51, indicating that that the data was multivariate normal.

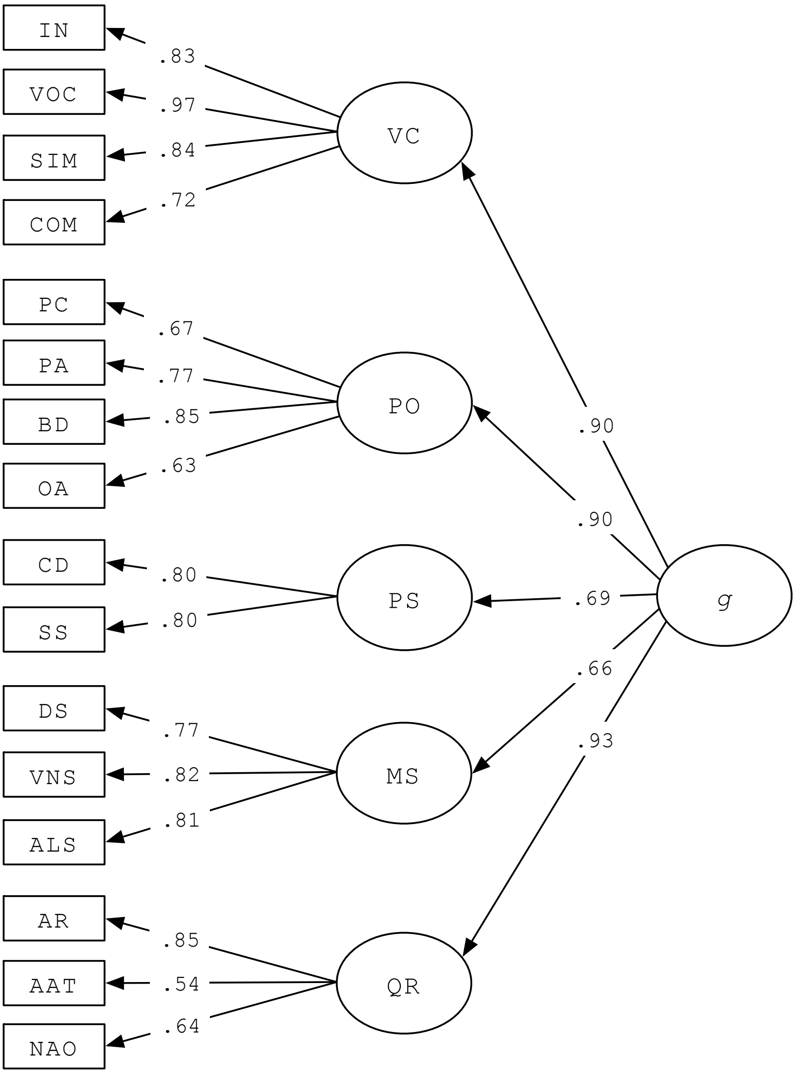
A review of model fit statistics (see Table 1) indicates that only Model III met a priori guidelines for good fit (i.e., CFI $\geq$ .95, SRMR $\leq$ .08, and RMSEA $\leq$ .06) and was the only model with a nonsignificant chi-square value. Model III was also relatively superior to the other models based on CFI (Chen, 2007) and BIC (Raftery, 1995) comparisons. Model III specified the WISC-III Arithmetic subtest as a measure of quantitative reasoning and the Digit Span subtest as a measure of memory.

Model III (with standardized coefficients) is illustrated in Figure 1. The five factors were substantially correlated, with coefficients ranging from .41 between the PS and MS factors to .84 between the VC and QR factors. The correlations among first order factors might be explained by a second order factor that would be more parsimonious than the first order structure of Model III (Thompson, 2004). In recognition of the importance of a higher order factor in investigations of cognitive ability (Jensen, 2002), a hierarchical structure was added to Model III. As anticipated, this higher order variant of Model III was an excellent fit to the data (see Table 1). The standardized coefficients displayed in Figure 2 indicate that all five first order factors were strongly related to the higher order general intelligence factor.

DISCUSSION

Confirmatory factor analyses were conducted to investigate several hypotheses regarding the Arithmetic and Digit Span subtests that formed the third factor of the WISC. Previous research suggested that the subtests may be measures of working memory, quantitative ability/reasoning, or some other ability. Alternatively, the two WISC subtests may be measures of different abilities. When the WISC-III was administered in conjunction with marker tests for memory and quantitative reasoning, a five factor solution was optimal. Results from the current study suggested that the WISC-III Arithmetic and Digit Span subtests are measures of different abilities with Arithmetic being a measure of quantitative reasoning and Digit Span a measure of memory.

These results are in line with theory and research developed from cross battery assessment of the Wechsler and Woodcock Johnson scales (Flanagan et al., 2000; McGrew & Woodcock, 2001; Phelps et al., 2005; Woodcock, 1990; Woodcock et al., 2001) and are consistent with the theory and research results reported by Carroll (1993, 1997), who indicated that the Arithmetic subtest may be a poor representation of fluid intelligence and Digit Span may represent general memory. Additional support for the current results was provided by Keith (2005) who compared results from hierarchical and nested-factors CFA models in the WISC-IV standardization sample and found that the loading of Arithmetic on the Working Memory factor varied considerably between the two models (hierarchical model = .80 and nested-factors model = .11).



Note. IN = WISC-III Information subtest, VOC = WISC-III Vocabulary subtest, SIM = WISC-III Similarities subtest, COM = WISC-III Comprehension subtest, PC = WISC-III Picture Completion subtest, PA = WISC-III Picture Completion subtest, BD = WISC-III Block Design subtest, OA = WISC-III Object Assembly subtest, CD = WISC-III Coding subtest, SS = WISC-III Symbol Search subtest, DS = WISC-III Digit Span subtest, AR = WISC-III Arithmetic subtest, VNS = Visual Number Span marker variable, ALS = Auditory Letter Span marker variable, AAT = Arithmetic Aptitude Test marker variable, NAO = Necessary Arithmetic Operations marker variable, VC = Verbal Comprehension factor, PO = Perceptual Organization factor, PS = Processing Speed factor, MS = Memory factor, QR = Quantitative Reasoning factor, and g = General Intelligence factor.

Figure 2. An Hierarchical Version of the Most Plausible Model Employing Twelve Subtests from the Wechsler Intelligence Score for Children-Third Edition (WISC-III) and Four Marker Variables from Memory and Quantitative Reasoning Factors.

The low loading of Arithmetic on the Working Memory factor in the nested-factors model suggests that it may not be a measure of working memory. In contrast, exploratory and confirmatory factor analyses in the WISC-IV manual suggested that the Arithmetic subtest does associate with the working memory subtests. However, no other quantitative tasks were included in those analyses. In addition, an inspection of the WISC-IV exploratory factor analytic loadings broken down by age group reveals a decline in Arithmetic's loading on the Working Memory factor as age increases (.73 for 6-7 year-olds, .57 for 8-10 year-olds, .39 for 11-13 year-olds, and .40 for 14-16 year-olds).

In agreement with the results from the current study, Wechsler (2003) placed the Digit Span subtest of the WISC-IV on a Working Memory factor. Contrary to these results, Wechsler (2003) specified the Arithmetic subtest of the WISC-IV as an optional measure of Working Memory. Even after modifying the Arithmetic subtest to increase demands on working memory, researchers have questioned the "placement of Arithmetic on the Working Memory factor" (Keith, 2005, p. 595). Flanagan and Kaufman (2004) also expressed concern about the WISC-IV Arithmetic subtest, suggesting that it is a measure of quantitative knowledge rather than working memory. Although the Arithmetic subtest is no longer a required subtest on the WISC-IV, it can be substituted for Digit Span or Letter-Number Sequencing. Given the current results as well as the concerns of Keith (2005) and Flanagan and Kaufman (2004), it may not be prudent to compute a WISC-IV working memory index score when the Arithmetic subtest is substituted for either WM core subtest, nor to interpret the Arithmetic subtest as a measure of working memory until additional research has specifically investigated its factorial validity. After substitution, "the underlying construct intended to be measured by the index may change" (Kaufman, Flanagan, Alfonso, & Mascolo, 2006, p. 291) and "the validity of the resulting Indexes and Full Scale IQ is unknown" (Ryan & Glass, 2006, p. 190).

This study was not without its limitations, which must be considered when examining the results and their implications. First, the size of the sample was small for CFA methods. CFA is generally regarded as a large sample methodology; however, more recent research suggests that it can be used with smaller samples under favorable conditions such as high variable communalities (MacCallum et al., 2001). Fortunately, communalities were relatively high for the current sample.

A second limitation of the current sample is that it was far less representative of the population of exceptional students in the United States than desired. Almost the entire sample was from the Mid-Atlantic section of the United States and some exceptionalities were not represented, including visual impairment, autism, and multiple disabilities. In addition, most of the students were White and of middle school age. Therefore, it is unclear how these results would generalize to the population of students referred for psychoeducational testing and to the general population.

Finally, using volunteer professionals and graduate students to collect data may have impacted the results. Some participants were tested by professional school psychologists and others by graduate students while still others may have been tested by more than one examiner across time. The impact of these disparate procedures is unknown, but examiner familiarity effects have been demonstrated (Fuchs & Fuchs, 1986). Given these limitations, additional research in this area is needed to validate findings with other Wechsler scales. Until then, clinicians should be circumspect when using the Wechsler Arithmetic subtest as a measure of working memory.

REFERENCES

Bartlett, M. S. (1954). A note on the multiplying factors for various chi square approximations. *Journal of the Royal Statistical Society, 16 (Series B),* 296-298.

Bentler, P. M., & Wu, E. J. C. (2002). *EQS 6 for Windows user's guide*. Encino, CA: Multivariate Software.

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. (1985). Exploratory factor analysis: A tutorial. In D. K. Detterman (Ed.), *Current topics in human intelligence, Vol. 1: Research methodology* (pp. 25-58). Norwood, NJ: Ablex.

Carroll, J. B. (1993). What abilities are measured by the WISC-III? In B. A. Bracken (Ed.), *Journal of Psychoeducational Assessment Monograph Series Advances in Psychoeducational Assessment Wechsler Intelligence Scale for Children: Third Edition* (pp. 134-143). Germantown, TN: Psychological Corporation.

Carroll, J. B. (1997). Commentary on Keith and Witta's hierarchical cross-age confirmatory factor analysis of the WISC-III. *School Psychology Quarterly, 12,* 108-109.

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling*, *14*, 464-504.

Child, D. (2006). *The essentials of factor analysis* (3rd ed.). New York: Continuum.

Cohen, J. (1959). The factorial structure of the WISC at ages 7-6, 10-6, and 13-6. *Journal of Consulting Psychology, 23,* 285-299.

Cohen, M., Becker, M. G., & Campbell, R. (1990). Relationships among four methods of assessment of children with Attention Deficit-Hyperactivity Disorder. *Journal of School Psychology, 28,* 189-202.

Ekstrom, R. B. (1973). *Cognitive factors: Some recent literature.* (Technical Report No. 2). Princeton, NJ: Educational Testing Service.

Ekstrom, R. B., French, J. W., Harman, H. H., & Dermen, D. (1976). *Kit of factor referenced cognitive tests.* Princeton, NJ: Educational Testing Service.

Elliott, C. D. (1990). *Introductory and technical handbook for the Differential Ability Scales*. San Antonio, TX: Psychological Corporation.

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods, 4,* 272-299.

Flanagan, D. P., & Kaufman, A. S. (2004). *Essential of WISC-IV assessment*. Hoboken, NJ: Wiley.

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

Fuchs, D., & Fuchs, L. S. (1986). Test procedure bias: A meta-analysis of examiner familiarity effects. *Review of Educational Research, 56,* 243-262.

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

Hu, L., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods, 3,* 424-453.

Jensen, A. R. (2002). Psychometric *g*: Definition and substantiation. In R. J. Sternberg & E. L. Grigorenko (Eds.), *The general factor of intelligence: How general is it?* (pp. 39-53). Mahwah, NJ: Erlbaum.

Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational & Psychological Measurement, 20,* 141-151.

Kamphaus, R. W. (1993). *Clinical Assessment of children's intelligence.* Boston, MA: Allyn and Bacon.

Kaufman, A. S. (1975). Factor analysis of the WISC-R at 11 age levels between 6½ and 16½ years. *Journal of Consulting and Clinical Psychology, 43,* 135-147.

Kaufman, A. S., Flanagan, D. P., Alfonso, V. C., & Mascolo, J. T. (2006). [Review of the *Wechsler Intelligence Scale for Children, Fourth Edition (WISC-IV)*]. *Journal of Psychoeducational Assessment, 24,* 278-295. 56.

Keith, T. Z. (2005). Using confirmatory factor analysis to aid in understanding the constructs measured by intelligence tests. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (2nd ed.; pp. 581-614). New York: Guilford.

Keith, T. Z., Fine, J. G., Taub, G. E., Reynolds, M. R., & Kranzler, J. H. (2006). Higher order, multisample, confirmatory factor analysis of the Wechsler Intelligence Scale for Children-Fourth Edition: What does it measure? *School Psychology Review, 35,* 108-127.

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.

Kranzler, J. H. (1997). What does the WISC-III measure? Comments on the relationship between intelligence, working memory capacity, and information processing speed and efficiency. *School Psychology Quarterly, 12,* 110-116.

Loehlin, J. C. (2003). *Latent variable models: An introduction to factor, path, and structural analysis* (4th ed.). Mahwah, NJ: Erlbaum.

Lowman, M. G., Schwanz, K. A., & Kamphaus, R. W. (1996). WISC-III third factor: Critical measurement issues. *Canadian Journal of School Psychology, 12,* 15-22.

MacCallum, R. C., Widaman, K. F., Preacher, K. J., & Hong, S. (2001). Sample size in factor analysis: The role of model error. *Multivariate Behavioral Research, 36,* 611-637.

Mardia, K. V. (1974). Applications of some measures of multivariate skewness and kurtosis in testing normality and robustness studies. *Sankhya, B36,* 115-128.

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

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.

Phelps, L., McGrew, K. S., Knopik, S. N., & Ford, L. (2005). The general (g), broad, and narrow CHC stratum characteristics of the WJ-III and WISC-III tests: A confirmatory cross-battery investigation. *School Psychology Quarterly, 20,* 66-88.

Prifitera, A., Weiss, L. G., & Soklofske, D. H. (1998). The WISC-III in context. In A. Prifitera & D. H. Soklofske (Eds.), *WISC-III clinical use and interpretation: Scientist-practitioner perspectives* (pp. 1-38). San Diego, CA: Academic Press.

Raftery, A. E. (1995). Bayesian model selection in social research. *Sociological Methodology, 25,* 111-163.

Reinecke, M. A., Beebe, D. W., & Stein, M. A. (1999). The third factor of the WISC-III: It's (probably) not Freedom from Distractibility. *Journal of the American Academy of Child & Adolescent Psychiatry, 38,* 322-328.

Reynolds, C. R., & Kamphaus, R. W. (1992). *Behavior Assessment System for Children – Teacher Rating Scales (BASC-TRS).* Circle Pines, MN: American Guidance Service.

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.

Ryan, J. J., & Glass, L. A. (2006). Substituting supplementary subtests for core subtests on reliability of WISC-IV indexes and full scale IQ. *Psychological Reports, 98,* 187-190.

Sattler, J. M. (2001). *Assessment of children: Cognitive applications* (4th ed.). San Diego, CA: Author.

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics, 6,* 461–464.

Stevens, J. P. (2009). *Applied multivariate statistics for the social sciences* (5th ed.). New York: Routledge.

Stone, B. J. (1992). Joint confirmatory factor analyses of the DAS and WISC-R. *Journal of School Psychology, 30,* 185-195.

Thompson, B. (2004). *Exploratory and confirmatory factor analysis: Understanding concepts and applications.* Washington, DC: American Psychological Association.

Wechsler, D. (1949). *Wechsler Intelligence Scale for Children*. New York: Psychological Corporation.

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

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

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

Wielkiewicz, R. M. (1990). Interpreting low scores on the WISC-R third factor: It's more than distractibility. *Psychological Assessment, 2,* 91-97.

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

Woodcock, R. W., & Johnson, M. B. (1989). *Woodcock Johnson Psycho-Educational Battery-Revised.* Allen, TX: DLM Teaching Resources.

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock Johnson III tests of cognitive abilities.* Itasca, IL: Riverside Publishing.

Zachary, R. A. (1990). Wechsler's intelligence scales: Theoretical and practical considerations. *Journal of Psychoeducational Assessment, 8,* 276-289.

Zimmerman, I. L., & Woo-Sam, J. M. (1997). Review of the criterion-related validity of the WISC-III: The first five years. *Perceptual and Motor Skills, 85,* 531-546.

Zwick, W. R., & Velicer, W. F. (1986). Comparison of five rules for determining the number of components to retain. *Psychological Bulletin, 99,* 432-442.

NEUROSCIENCE RESEARCH PROGRESS

INTELLIGENCE QUOTIENT

TESTING, ROLE OF GENETICS AND THE ENVIRONMENT AND SOCIAL OUTCOMES

JOSEPH C. KUSH
EDITOR

nova publishers
New York

For permission to use material from this book please contact us:
Telephone 631-231-7269; Fax 631-231-8175
Web Site: http://www.novapublishers.com

NOTICE TO THE READER

Additional color graphics may be available in the e-book version of this book.

Library of Congress Cataloging-in-Publication Data

ISBN: 978-1-62618-782-5 (eBook)

Library of Congress Control Number: 2013937394

Published by Nova Science Publishers, Inc. † New York