

The Woodcock-Johnson III and the Cognitive Abilities Test (Form 6): A Concurrent Validity Study

David F. Lohman[‡]
University of Iowa
March 2003

This study investigated the concurrent validity of the Woodcock-Johnson III (WJ-III; Woodcock, McGrew, & Mather, 2001) and Form 6 of the Cognitive Abilities Test (CogAT; Lohman & Hagen, 2001). A total of 178 students in grades 2, 5, and 9 were administered 13 tests from the WJ-III and the appropriate level of the CogAT. Interbattery confirmatory factor analyses showed that the general factors on the two batteries correlated $r = .82$. Correlations between broad-group clusters on the WJ-III and battery-level scores on the CogAT generally supported the construct interpretations of each, but also suggested important differences in the abilities measured by both batteries.

The Cognitive Abilities Test (CogAT) is one of the most widely used group ability tests, and the Woodcock-Johnson is one of the most widely used individually administered ability tests. However, there are no published reports of the concurrent validity of these two test batteries. The purpose of this study was to investigate the relationships between the latest editions of each test: Form 6 of the CogAT (Lohman & Hagen, 2001), and the Woodcock-Johnson III (Woodcock, McGrew, & Mather, 2001).

The Woodcock-Johnson III Tests of Cognitive Abilities (WJ-III) is the most recent revision of a series of tests that was first published in 1977. The WJ-III is based on the Cattell-Horn-Carroll (CHC) three-stratum theory of cognitive abilities. CHC theory combines Cattell-Horn's Gf-Gc (Cattell, 1971; Horn, 1989) and Carroll's three-stratum theories of human abilities (Carroll, 1993). This hierarchical theory posits a large array of specific or stratum I abilities (Carroll, 1993, identified 69). These narrow abilities may be grouped into eight broad or stratum II abilities. Stratum II abilities in turn define a general (g) cognitive ability factor at the third level. Although abilities at all three strata are represented in the WJ-III, the primary focus is on the measurement of the broad CHC factors at stratum II. The stratum III g score on the WJ-III is estimated from the first principal component of the scores for stratum II abilities.

Form 6 of the CogAT is the latest revision of a test that was first published in 1954 as the Lorge-Thorndike Intelligence test (Lorge & Thorndike, 1954). Like the WJ-III, the CogAT has undergone several important revisions over the years. It is also based on a

hierarchical model of abilities. Unlike the WJ-III, however, the CogAT focuses on the g factor at the third stratum and the stratum II fluid reasoning (Gf) abilities that load most highly on g. Carroll (1993) argues that the Gf factor is defined by three reasoning abilities: (1) *sequential reasoning*—verbal, logical, or deductive reasoning; (2) *quantitative reasoning*—inductive or deductive reasoning with quantitative concepts; and (3) *inductive reasoning*—typically measured with figural tasks. These correspond roughly with the three CogAT batteries: verbal reasoning, quantitative reasoning, and figural/nonverbal reasoning. Each reasoning ability is estimated from two tests in grades K-2 and from three tests in grades 3-12. Using multiple measures reduces the impact of test format and allows for reliable separation of three highly correlated reasoning abilities. A general ability score is estimated from the average of scale scores across the three batteries.

Although the CogAT and the WJ-III are both based on hierarchical models of human abilities, the CogAT focuses on general reasoning abilities, whereas the WJ-III attempts to measure a much broader collection of stratum II abilities in CHC theory. The WJ-III g score is the first principal component of a large and varied test battery. The CogAT, on the other hand, emphasizes depth. All nine of its tests measure reasoning abilities. Its g score is simply the centroid of the three battery scores. There is some question then, about the extent to which the composite and battery-level scores on CogAT will predict the general and more specific cluster scores on WJ-III.

Method

The sample for this study consisted of a total of 178 second, fifth, and ninth grade students who were administered both tests. Average ages were 7.8, 10.9, and 14.9 years for the second, fifth, and ninth graders, respectively. The sample was reasonably diverse: 71.5 percent were White, 11.7 percent Black, 7.3 percent Hispanic, and 4.0 percent other ethnic backgrounds.

[‡] I am grateful to Dr. Scott Bishop of Riverside Publishing Company for recruiting examiners and supervising the difficult and time-consuming task of collecting the data for this study. I also thank Sara Clough for assistance in preparing the data files and Patricia Martin for assistance in preparing the manuscript.

Ethnic background was not indicated for the remaining 5.6 percent of the sample. There were approximately the same number of males and females across all grades ($N = 90$ females, $N = 88$ males) and at Grade 2 ($N = 43$ females, $N = 44$ males). However, there were more males in the grade 5 sample ($N = 38$ of 66) but more females in the grade 9 sample ($N = 18$ of 25). All students were administered the nine tests in the WJ-III Standard Battery, plus four additional tests: Planning, Analysis-Synthesis, Applied Problems, and Quantitative Concepts. The Planning and Analysis-Synthesis tests were included to enhance the measurement of the Visual-Spatial Thinking (Gv) and Fluid Reasoning (Gf) factors. Applied Problems and Quantitative Concepts were included as measures of the Mathematics factor (Gq). WJ-III tests were administered by school psychologists.

Within approximately 2-3 weeks of taking the WJ-III, students were also administered the appropriate level of CogAT either by their teacher or the school psychologist. The second grade students took Level 2 of the Primary Battery, whereas the fifth and ninth grade students took levels C and F (respectively) of the Multilevel Battery. Although scale scores for the CogAT Primary Battery (grades K-2) and Multilevel Battery (grades 3-12) are vertically equated on a common scale, the two batteries differ in mode of administration (the Primary Battery requires no reading and has no time limits), as well as in the content of all but one subtest. Grades 2 and 5 were selected in order to investigate the possibility of differences in the abilities measured by the two CogAT batteries. Grade 9 was included simply to represent the post-elementary school population. Unfortunately, it proved almost impossible to recruit students at this age for the study. Therefore, scores for the 25 grade 9 students are only included in those analyses that pool scores across grades.

Results

Generalizations about relationships between test batteries depend on the extent to which the study sample faithfully represents the population. Univariate means and standard deviations provide some evidence on the representativeness of the sample. These are reported in Table 1. In general, the students in this sample were somewhat above average in ability. Mean CogAT SAS scores ranged from 103.7 to 105.3 across the three batteries. WJ-III standard scores were generally somewhat higher.

Average scores that depart from the population mean can signal restricted score variability. Although final distributions for some of the WJ-III tests showed *SDs* well below the population *SD* of 15, most tests in

both batteries showed only moderately less variability than was observed in the standardization samples. (Population *SD* is 16 for CogAT.)

The most important concern, however, is whether the patterns of relationships among subtests in each battery mirror population covariances. The best estimate of the population variance-covariance matrix for the tests in each battery is given in the standardization sample. For the WJ-III, we used the covariance matrices for 6-8 year olds and 9-13 year olds (McGrew & Woodcock, 2001). For the CogAT, we used the covariance matrices for levels 2 and C (Lohman & Hagen, 2002). We then compared these covariance matrices with the corresponding matrix for second or fifth graders in our sample.

There are a number of ways to compare covariance matrices. The most stringent test is simply to test the equality of the two covariance matrices. For even moderately large matrices, however, the probability that all k variances and $k(k - 1)/2$ covariances will be the same is exceedingly unlikely (Bollen, 1989). At the other extreme, one can test whether the factor structures of the two matrices have the same general form or show what some call configural invariance (Horn & McArdle, 1992). Two models are said to have the same form if the model for each group has the same parameter matrices with the same dimensions, and the same location of fixed, free, and constrained parameters. For example, two confirmatory factor models would have the same form if both specified three factors, and the same paths between factors and observed variables. However, factor loadings and correlations among the factors could vary. Different degrees of equivalence may be identified between these two extremes.

For the CogAT data, we assumed that covariances among tests in the standardization sample could be treated as population values. This seemed reasonable, given the size of the standardization samples ($N = 16,235$ and 15,146 at levels 2 and C, respectively). We then tested whether the parameters of confirmatory models fit to these data also fit the CogAT data in our grade 2 and grade 5 samples. This is a fairly stringent test of factorial equivalence.

The procedure used with the WJ-III data was a bit different. The assumption that the 156 covariances in each of the two standardization matrices could be treated as population values seemed unlikely, given the smaller sample size and fact that each pooled scores for several ages. Therefore, we tested whether a common model could be fit simultaneously to our data and the standardization data sets. In judging fit for these models, we sometimes adopted a somewhat more liberal criteria than is commonly employed when

Table 1
Descriptive Statistics for All Grades (N = 178)

Instrument Scores	Min	Max	<i>M</i>	<i>SD</i>	Skewness		Kurtosis	
					Initial	Final	Initial	Final
WJ-III Cluster Scale Scores								
General Intellectual Ability	62	145	107.2	13.37	0.07	-0.02	0.31	0.34
Verbal Ability	73	138	106.3	11.45	-0.27	-0.25	0.55	0.57
Thinking Ability	69	151	109.0	14.31	0.07	0.00	0.57	0.29
Cognitive Efficiency	66	140	103.1	13.37	-0.41	-0.11	1.01	0.38
Fluid Reasoning	63	143	107.8	12.99	0.15	0.01	0.90	0.62
Phonemic Awareness	73	160	111.5	15.15	0.19	-0.04	0.11	-0.14
Working Memory	65	146	106.8	14.14	0.00	0.15	0.12	-0.06
Math Reasoning	55	139	105.4	10.71	-0.50	-0.58	2.66	2.83
WJ-III Test Scale Scores								
Verbal Comprehension	73	138	106.3	11.45	-0.27	-0.25	0.55	0.57
Visual-Auditory Learning	55	145	102.1	15.10	0.36	0.17	3.94	1.27
Spatial Relations	55	145	103.2	12.52	-1.84	-0.67	11.09	3.80
Sound Bending	76	151	111.3	15.72	0.57	0.17	0.92	-0.28
Concept Formation	57	141	106.4	14.67	-0.41	-0.41	1.08	1.08
Visual Matching	54	144	103.5	15.74	-1.23	-0.25	4.67	0.40
Numbers Reversed	67	144	102.4	12.81	-0.19	0.20	1.62	0.42
Incomplete Words	67	146	107.3	13.59	0.01	0.22	0.82	0.54
Auditory Working Memory	74	144	110.0	14.23	-0.01	-0.08	0.14	-0.18
Analysis-Synthesis	65	145	107.8	12.66	0.57	0.10	2.42	1.37
Planning	85	145	108.6	9.81	3.86	0.82	24.46	2.37
Applied Problems	61	145	108.0	11.66	-0.13	-0.21	1.47	1.32
Quantitative Concepts	55	142	104.2	12.64	-0.27	-0.19	1.11	1.28
CogAT Standard Age Scores								
Verbal SAS	71	150	105.3	14.58	0.33	0.38	0.13	0.19
Quantitative SAS	71	139	103.7	14.78	-0.01	0.08	-0.43	-0.46
Nonverbal SAS	71	144	103.8	15.54	0.01	0.07	-0.60	-0.59
Composite SAS	73	143	105.0	15.16	0.18	0.28	-0.50	-0.43

Note. Minimum, maximum, mean, and standard deviation are for final sample, after replacement of outliers and missing values. Means and standard deviations were generally unaltered by these replacements. Exceptions were the three tests with large initial skewness and kurtosis: Planning, Visual Matching, and Spatial Relations.

fitting models to data. For example, Hu and Bentler (1999) recommend a cutoff close to .95 for the Tucker-Lewis index and of .06 for RMSEA. Our concern was not to decide how best to characterize either the sample or the standardization data. Rather we sought to establish the level of congruence or similarity we could assume between the sample and standardization data. Equivalence between sample and standardization covariance matrices is often merely assumed in validity studies.

Once the degree of equivalence was established for each test battery, we proceeded to estimate

relationships between latent variables on the WJ-III and CogAT6 in a series of models that combined the two test batteries. All confirmatory factor models were tested using AMOS 4.0 (Arbuckle, 1999).

Data Screening

For the WJ-III, we used standard scores in all analyses. Since these are normed to a mean of 100 and *SD* of 15 at each age, they can be used both for within-grade and across-grade analyses. For the CogAT, we used raw scores on each of the six (Primary Battery) or

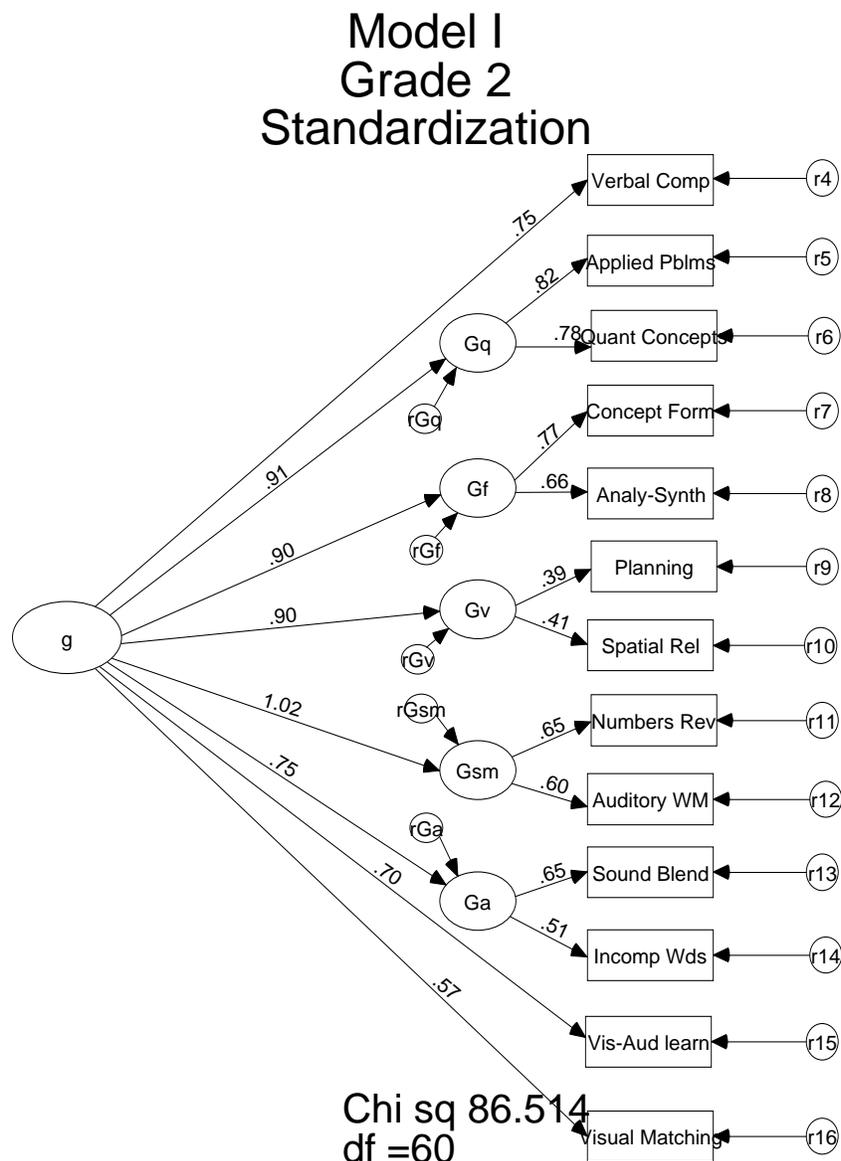


Figure 1. Model I. Hypothesized factor structure for the 13 Woodcock-Johnson III ability tests. Factor loadings and chi square are for the model fit to the covariance matrix for 6-8 year olds reported by McGrew and Woodcock (2001).

nine (Multilevel Battery) subtests for the within-grade analyses. For analyses that pooled across grades, we used the age-normed Verbal, Quantitative, and Nonverbal standard age scores (SAS).

We first screened for outliers by examining the univariate distributions for each score. Observations that differed from the mean by more than 3 *SD* were replaced by the corresponding 3 *SD* value. Overall, 1.5% of scores on the 13 WJ-III tests were replaced. Cluster scores defined by any of these altered scores were then recomputed. None of the CogAT scores were replaced. Missing scores were replaced with regression estimates. The predictor set included all other WJ-III

and CogAT tests with non-missing scores. In all 1.2% of the scores were missing, mostly on the four WJ-III supplemental tests.

Means and standard deviations reported in Table 1 are for the final data set. Skewness and kurtosis statistics, however, are reported for both the initial and final data. Distributions with absolute values of skewness greater than 3.0 are generally considered extremely skewed. Skewness for the Planning test exceeded this value. Kurtosis values greater than 10 indicate markedly peaked distributions. Kurtosis statistics for the Planning and Spatial Relations exceeded this value. Clearly, replacing outliers with

Table 2
Confirmatory Factor Models for the 13 Woodcock-Johnson tests in (a) the standardization sample, (b) our grade 2 sample, and (c) both samples simultaneously

Model ^a	χ^2	<i>df</i>	χ^2/df	TLI	AIC	RMSEA
Standardization 6-8 year olds (<i>N</i> = 325)						
I. W-J	86.5	60	1.44	.975	148.5	.037
II. No Gv	91.0	61	1.49	.972	151.0	.039
III. No Gv, Ga on Gc	74.0	60	1.23	.987	135.9	.027
Grade 2 sample (<i>N</i> = 87)						
II. No Gv	92.6	61	1.52	.860	152.6	.078
III. No Gv, Ga on Gc	78.2	60	1.30	.918	140.2	.059
Standardization and Grade 2						
III. No Gv, Ga on Gc	168.9	132	1.28	.974	268.9	.026

Note. TLI = Tucker-Lewis fit index; AIC = Akaike information criteria; RMSEA = Root Mean Square Error of Approximation; W-J = Woodcock-Johnson.

^a See text for model descriptions.

less extreme scores had a substantial impact on these and several other WJ-III subtests.

Modeling the WJ-III Data

McGrew and Woodcock (2001) report that the nine tests in the Standard battery load on seven broad group factors. The four supplementary tests we administered help determine two of these factors, as well as a Gq factor. In all, five broad group factors were measured by two tests: Mathematical Reasoning (Gq), Fluid Reasoning (Gf), Visual-Spatial Thinking (Gv), Short-Term Memory (Gsm), and Auditory Processing (Ga). These are shown in Figure 1. Three other broad group factors (Verbal Comprehension [Glr]) were measured by only one test, and so each of the three corresponding tests (Verbal Comprehension, Visual Matching, and Visual-Auditory learning) contributes only the definition of *g* in the model.

The first step in our analysis was to see if the hypothesized factor model fit the correlations among these 13 tests reported by McGrew and Woodcock (2001) for 6-8 year olds, and then for 9-13 year olds. These correspond to the second and fifth graders in this sample, respectively.

Grade 2 sample. We began with the covariance matrix for 6-8 year olds. Average within-year sample size for this matrix was 325. As would be expected, the covariance matrix for the sample of 87 Grade 2 students only poorly ($\chi^2 = 95.4$, *df* = 60, Tucker-Lewis fit index = .841, RMSEA = .083). Both the *g* to Gf and G to Gv paths were substantially greater than 1.0 in this model (1.18 and 1.12, respectively). Consequently, the

residual factors for both Gf and Gv had sizable negative variances. We then removed *g* from the model and estimated correlations among the five broad group factors shown in Figure 1. The estimated correlation between Gsm and Gv exceeded unity. We considered either collapsing Gsm and Gv into one factor, or reassigning tests to factors differently. Examination of the covariance matrix showed clearly that the Spatial Relations test could be placed on the Gsm factor, and the Planning test on the Gf factor. Indeed, in the analyses McGrew and Woodcock (2001) report, the Planning test frequently split what little common variance it shared with other tests between the Gv and Gf factors.¹

This new model (Model II in Table 2) fit the standardization matrix, but not quite as well as the original model ($\chi^2 = 91.0$, *df* = 61, Tucker-Lewis fit index = .972, RMSEA = .039). However, the residual matrix showed a large covariance between the Verbal Comprehension test and the Ga factor. This makes sense because Ga is surely related to—if not subordinate to Gc in a true hierarchical model (see Carroll, 1993, chapter 5). Further Gc was poorly represented in this battery of 13 tests.

Model III tests the hypothesis that Ga can be subsumed by Gc. (The form of this model is shown in Figure 2, but for a different analysis). The χ^2 for Model III was 74.0 (60 *df*), which represents a significant

¹ The recent addition of the Block Rotation test to the WJ-III Cognitive easel allows greatly strengthens the Gv factor. This test was not commercially available when the data for this study were gathered.

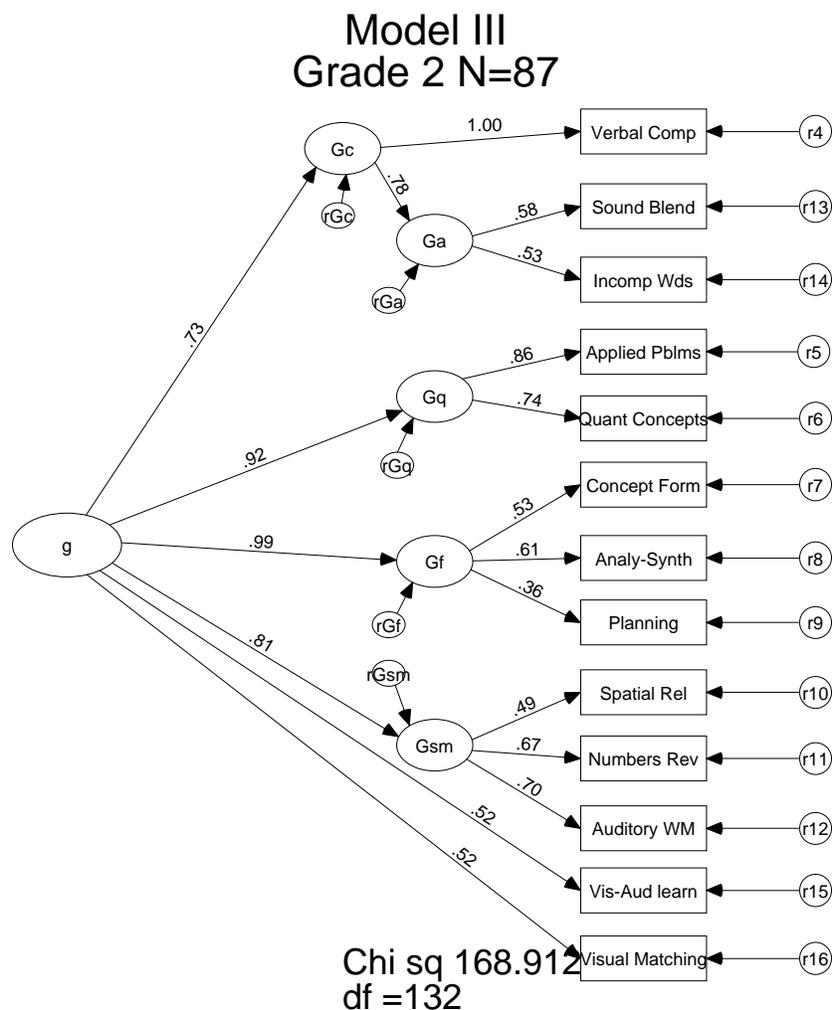


Figure 2. Model III. Final WJ-III model for the simultaneous analysis of the grade 2 test sample and the 6-8 year old standardization samples. No Gv factor. Ga subsumed under Gc. Standardized loadings are for the 87 second graders. Chi square is for the simultaneous multiple group factor analysis.

reduction in the χ^2 observed for Model II ($\Delta\chi^2 = 17.3$, $df = 1$, $p < .01$).² Fitting Model III to the covariance matrix for the second graders in our sample gave a similar χ^2 of 78.2 (60 df). Finally, we performed a simultaneous multiple group analysis on the standardization ($N = 325$) and grade 2 ($N = 87$) covariance matrices. The unstandardized paths were constrained to be the same in both matrices. The result is shown in Figure 2. Standardized path coefficients are for the sample of 87 second graders. The $\Delta\chi^2$ for this model was 16.7 (12 df) greater than the sum of the χ^2 's for separate analysis in which

² Since Model III represents more than a simple trimming of Model II, we also report AIC fit indexes in Table 2. These also show that Model III fits better than Model II in both the standardization and Grade 2 samples.

Model III was fit to each data set. This increment in χ^2 is not significant, and so we conclude that Model III fits the covariance matrices for both the standardization sample of 6-8 year olds and our sample of second graders equally well.

Although unstandardized regression coefficients were constrained to be the same in this analysis, the standardized regression coefficients differed because variances were not the same in the two samples. For the standardization sample, *g* had its highest loading ($\beta = 1.02$) on *Gsm*.³ This supports Kyllonen's (1996) argument that *g* is primarily working memory. In our

³ Although this loading is greater than 1.0, it is not significantly greater than 1.0.

Table 3
Confirmatory Factor Models for the 13 Woodcock-Johnson tests in (a) the standardization sample, (b) our grade 5 sample, and (c) both samples simultaneously

Model ^a	χ^2	<i>df</i>	χ^2/df	TLI	AIC	RMSEA
Standardization 9-13 year olds (<i>N</i> = 325)						
I. W-J	71.6	60	1.19	.989	133.6	.025
IIIa. Ga on Gc	62.0	59	1.05	.997	125.9	.013
Grade 5 sample (<i>N</i> = 66)						
IIIa. Ga on Gc	94.3	59	1.60	.848	158.3	.096
Standardization and Grade 5						
IIIa. Ga on Gc	179.5	130	1.38	.964	283.5	.032
IIIb. Ga on Gc, unconstrain g-Gsm	171.2	129	1.23	.969	277.2	.030

Note. TLI = Tucker-Lewis fit index; AIC = Akaike information criteria; RMSEA = Root Mean Square Error of Approximation; W-J = Woodcock-Johnson.

^a See text for model descriptions.

sample, however, *g* had the most influence on *Gf* ($\beta = .99$), followed closely by *Gq* ($\beta = .97$). This conforms more closely with the arguments of Gustafsson (2002) that $g = Gf$. Variation in standardized loadings across samples demonstrates that controversies about the nature of general ability may in part reflect differences in variability of test scores in the samples. It also introduces the possibility (discussed later) that *Gq* might be central to the definition of *g*.

Grade 5 sample. The same procedure was repeated for the Grade 5 data. The results are summarized in Table 3. Once again, we found that the model that placed *Ga* under *Gc* (Model IIIa) fit the data significantly better than the model that had direct paths from *g* to *Ga* and from *g* to verbal comprehension (Model I). This time, however, the hypothesized *Gv* factor could be identified in both the standardization covariance matrix for 9-13 year olds and in our Grade 5 data. Further, when the unstandardized regression coefficients value constrained to be equal in the two samples, the $\Delta\chi^2$ was 23.2 (with 12 *df*) over the sum of the chi squares for the two separate analyses. This increment in χ^2 is significant at the $p = .05$ level.

Examination of the factor loadings showed a large difference in the *g* to *Gsm* path in the two samples. Unconstraining this parameter in Model IIIb reduced the overall χ^2 to 171.2 (with 129 *df*). This is not a significant increment in χ^2 over the baseline χ^2 of 156.3 (with 118 *df*). Thus, we concluded that Model IIIb fit both data sets equally well. Further, this model closely approximates the hypothesized factor model than the model that best fit the data for the second grade samples.

To summarize, we tested whether the covariances among the 13 WJ-III tests computed for our samples of second and fifth graders differed from covariances among these tests observed in the standardization for children of roughly the same age. For second graders, we found congruence, but only after eliminating *Gv* from our model. This conforms with the hypothesis that abilities may exhibit a less differentiated structure for younger children. For both second and fifth graders, we found that *Ga* was best subsumed under a more general verbal factor. This suggests that abilities may fall in more than three strata, as Carroll (1993) has acknowledged. Nonetheless, we concluded that the covariances among WJ-III tests in our two samples did not differ markedly from those observed in the larger and more representative standardization samples.

Modeling the CogAT Data

Grade 2 sample. The CogAT Primary Battery contains six subtests, each of which help define one of the three broad abilities shown in Figure 3. This is called Model IV in Table 4. We first fit Model IV to the standardization sample ($N = 16,235$) and then to our sample of second grade students ($N = 87$). These are called Models IV-a and IV-b in Table 4. Model fits were good, which establishes what Bollen (1989) calls form equivalence. It says that the same basic model can be fit to both data sets. However, the parameters of the models are allowed to vary across the two samples. More stringent definitions of equivalence require that some or all of the model parameters be the same.

Given the enormous discrepancy in sample size between the standardization and test samples, it seemed most reasonable to treat the parameters from the analysis of the standardization data as population

Table 4

Confirmatory Factor Models for the Level 2/Grade 2 CogAT Primary Battery (Top Panel) and Level C/Grade 5 of the CogAT Multilevel Battery (Bottom Panel)

Model and Sample	χ^2	<i>df</i>	χ^2/df	TLI	RMSEA
Grade 2					
IV. CogAT Primary					
IV-a. Standardization (<i>N</i> = 16,235) unconstrained	662.7	6	110.4	.993	.082
IV-b. Grade 2 (<i>N</i> = 87) unconstrained	7.2	6	1.2	.998	.048
IV-c. Grade 2 all paths fixed	13.9	11	1.26	.997	.056
Grade 5					
V. CogAT Multilevel					
V-a. Standardization (<i>N</i> = 15,146) unconstrained	1013.0	24	42.21	.995	.052
V-b. Grade 5 (<i>N</i> = 66) unconstrained	51.5	24	2.14	.968	.133
V-c. Grade 5 all paths fixed	54.2	32	1.69	.981	.103

Note. TLI = Tucker-Lewis fit index; RMSEA = Root Mean Square Error of Approximation; CogAT = Cognitive Abilities Test.

values. We then tested whether the model that best fit the standardization data would also fit the sample data. We did this by fixing the unstandardized regression coefficients to the values obtained in the analysis of the standardization data (Model IV-a) and then fitting this model to the Grade 2 sample data. This is called Model IV-c in Table 4 and Figure 3. The increment in χ^2 over model IV-b in which factor loadings were unconstrained was not significant ($\Delta\chi^2 = 6.7$, *df* = 5). Thus, we conclude that the relationships among CogAT subtests observed in this sample of second graders are congruent with those observed in the standardization sample.

Grade 5 sample. We followed the same procedure for the Grade 5 sample. Figure 4 shows the hypothesized factor structure for the CogAT Multilevel battery. The fit statistics for this series of models are shown in the bottom half of Table 4. As expected, the hypothesized factor structure (Model Va) fit the standardization covariance matrix quite well (Tucker Lewis fit index = .995; RMSEA = .052). The same model also fit the sample of 66 fifth graders reasonably well (Tucker Lewis fit index = .968; RMSEA = .133).⁴ Figure 4 shows the standardized path coefficients for our fifth-grade sample when the unstandardized regression paths were fixed at the values obtained in the analysis of the standardization data. This is called

Model V-c in Table 4. The increment in χ^2 over Model V-b was not significant ($\Delta\chi^2 = 2.7$, *df* = 8).

In the models for both the CogAT Primary Battery (Figure 3) and Multilevel Battery (Figure 4), the general factor enters most prominently into performance on the quantitative battery, secondarily on the verbal battery, and thirdly on the nonverbal battery. The differentiation between verbal and quantitative factors is stronger on the Multilevel than on the Primary Battery. This could reflect both developmental changes in the organization of abilities or content differences between the two batteries, or both.

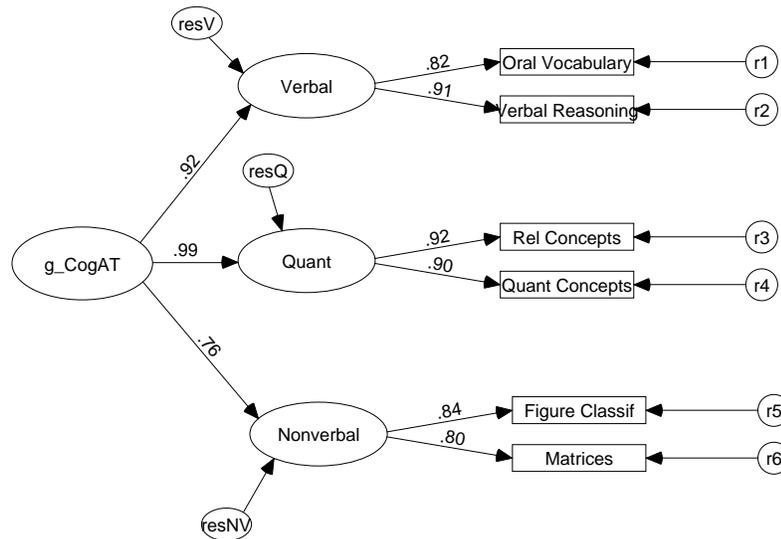
Interbattery Analyses

We have now established that the relationships among subtests on each test battery that we observed in our samples of second and fifth graders do not differ significantly from those observed in the much larger and more representative standardization samples for the two test batteries. Were this not the case, then we would have less confidence in any relationships we might report between scores or latent variables in the two test batteries.

Three different types of interbattery analyses are reported. First, we report correlations between the nine cluster scores on the WJ-III and the four CogAT SAS scores. Second, we report correlations between the broad group factors that were represented in our models of each battery. Third, we report the results of a confirmatory, interbattery factor analysis in which we estimate the correlation between the general factors on the two batteries.

⁴ It is noteworthy that the rather high value for RMSEA is actually reduced when all paths are fixed. More importantly, the goal here is merely to investigate the representativeness of the sample, not to determine the best-fitting model for the sample data.

Model IV-c All Paths Fixed Grade 2 N=87



Chi sq = 13.915
df = 11

Figure 3. Model IV-c. Final CogAT Primary Battery model for the grade 2 sample. Unstandardized loadings were fixed. Standardized factor loadings are for the sample of 87 second graders.

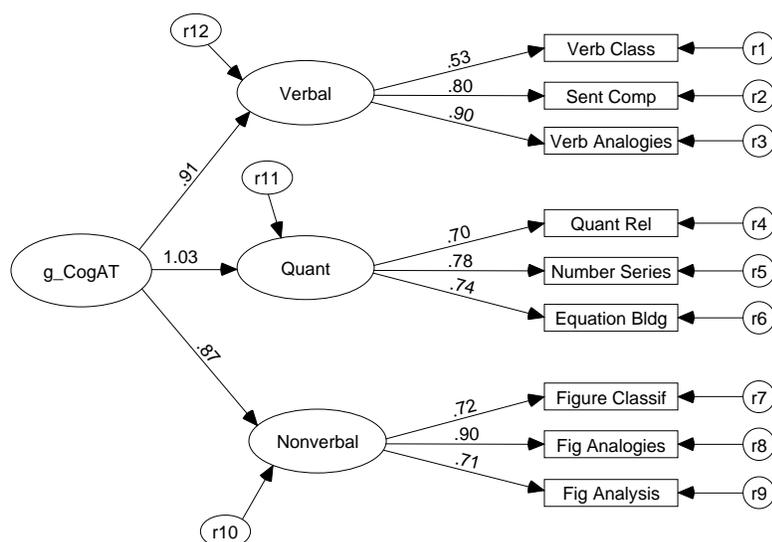
WJ-III cluster, CogAT SAS correlations. Users of either the WJ-III or the CogAT typically report cluster scores on the WJ-III or battery scores on the CogAT. Thus, the most direct comparison of the two batteries is given by the correlations between the nine WJ-III cluster scores and the four CogAT SAS scores. Using CogAT SAS scores also allows us to include all 178 cases in a single analysis. These correlations are reported in the first four columns of Table 5. Within-grade correlations are also reported for the grade 2 and the combined grade 5 and grade 9 samples.

The first question to be addressed is the extent to which the general ability scores on the two test batteries are correlated. General ability is estimated by the General Intellectual Ability (GIA) cluster on the WJ-III and by the Composite SAS score on CogAT. The correlation of $r = .68$ (see Table 5) is

about the same as the reported correlation between the WJ-III and individually administered intelligence tests. For example, Phelps (in McGrew & Woodcock, 2001) reports a correlation of $r = .71$ between the WJ-III GIA score and the Full Scale IQ on the WISC-III for a sample of 150 randomly chosen grade 3-5 students. Flanagan, Kranzler, and Keith (in McGrew & Woodcock, 2001) report a correlation of $r = .70$ between the WJ-III Brief Intellectual Ability score and the Full-Scale Score on the CAS.

On the other hand, the correlation of $r = .68$ between the GIA cluster and the CogAT Composite SAS is somewhat lower than the correlation between the CogAT and the WISC-III. Lohman (2003) found a correlation of $r = .79$ between CogAT Composite SAS scores and WISC Full Scale IQ scores for a sample of 91 sixth graders. This is not entirely unexpected, however. Both

Model V-c All paths Fixed CogAT Gr 5 N=66



Chi sq = 54.176
df = 32

Figure 4. Model V-c. Final CogAT Multilevel Battery model for the grade sample. Unstandardized loadings were fixed. Standardized factor loadings are for the sample of 66 fifth graders.

CogAT and the WISC are designed to measure primarily *g* and its major constituents, whereas the WJ-III is designed to measure seven broad-group abilities.

In general, the CogAT Verbal, Quantitative, and Nonverbal scores correlate higher with the WJ-III GIA cluster than with any of the more specific WJ-III clusters. This could in part be due to the greater reliability of the GIA score. However, the estimated reliabilities for most of the specific WJ-III clusters exceed $r_{xx'} = .90$. Therefore, it seems more likely that all three of CogAT batteries best measure what is captured by the GIA cluster and only secondarily what is measured by the five more specific clusters.

Nevertheless, there is some evidence for convergent validity. Look first at the correlations for all

subjects (columns 1-3 in Table 5). In each case, the CogAT battery-level score shows its highest correlations with the specific WJ-III clusters that measure similar abilities. CogAT Verbal correlates most highly with WJ-III Verbal Ability scores (which here is defined by one test); CogAT Quantitative correlates most highly with WJ-III Math Reasoning; and CogAT Nonverbal correlates most highly with WJ-III Fluid Reasoning.

At the Primary level (columns 5-8), the CogAT Verbal score correlates highest with the WJ-III Verbal Ability cluster ($r = .68$). The CogAT Quantitative score correlates about $r = .6$ with the WJ-III Verbal, Fluid Reasoning, and Math Reasoning clusters. And the CogAT Nonverbal score correlates highest with the Fluid Reasoning cluster.

Table 5
Correlations between Cognitive Ability Test (CogAT) Standard Age Scores (SAS) and Woodcock-Johnson III (WJ-III) Cluster Scores

WJ-III Cluster Score	CogAT SAS Score											
	Grades 2, 5, & 9 (<i>N</i> = 178)				Primary Battery Grade 2 (<i>N</i> = 87)				Multilevel Battery Grades 5 & 9 (<i>N</i> = 91)			
	V	Q	N	Composite	V	Q	N	Composite	V	Q	N	Composite
General Intellectual Ability	.62	.64	.57	.68	.63	.65	.55	.68	.63	.64	.60	.69
Verbal Ability	.62	.56	.40	.59	.68	.60	.38	.60	.58	.51	.43	.58
Thinking Ability	.52	.51	.52	.58	.55	.52	.49	.58	.51	.50	.56	.58
Cognitive Efficiency	.43	.49	.40	.48	.36	.47	.41	.47	.48	.51	.40	.50
Fluid Reasoning	.42	.56	.55	.57	.46	.60	.58	.62	.40	.53	.53	.54
Phonemic Awareness	.36	.26	.17	.30	.34	.22	.07	.22	.39	.30	.27	.37
Working Memory	.51	.56	.45	.56	.38	.44	.38	.45	.61	.68	.52	.66
Math Reasoning	.57	.58	.48	.60	.53	.62	.47	.61	.60	.55	.48	.60

Note. Grade 2 CogAT scores are from level 2 of Primary Battery; grade 5 and 9 scores are from levels C and F of the Multilevel Battery. Composite SAS scores are based on the average of verbal (V), quantitative (Q), and nonverbal (N) scale scores.

Correlations for the Multilevel Battery (columns 9-12) are less clear cut. CogAT Verbal correlates about as highly with WJ-III Verbal Ability as with WJ-III Working Memory and Math Reasoning. Similarly, CogAT Quantitative correlates highest with Working Memory ($r = .68$), followed by Math Reasoning ($r = .55$) and Fluid Reasoning ($r = .53$). Finally, CogAT Nonverbal correlates with WJ-III Thinking Ability ($r = .56$), Fluid Reasoning ($r = .53$), and Working Memory ($r = .52$).

In summary, each of the three CogAT batteries measures abilities that cut across several of the WJ-III cluster scores. All seem primarily to measure general ability (as indexed by the GIA or Working Memory clusters), and secondarily the more specific verbal, quantitative, or fluid reasoning abilities that would reasonably be linked to each battery.

Correlations between latent broad-group factors. Correlations between WJ-III cluster scores and CogAT SAS scores address the practical matter of how much discrepancy test users are likely to see in observed scores on the two tests. However, questions about relationships among the constructs measured by the two batteries are better addressed through SEM models that estimate correlations between latent variables. These are reported in Table 6. Note that Gc', Gf', and Gsm' for Grade 2 and Gc' for Grade 5 are marked with primes in this table to remind the reader that these constructs are defined somewhat differently than in McGrew and Woodcock (2001).

These correlations show several interesting patterns. First, they show that latent variables are

highly correlated on both tests. For CogAT, these correlations range from $r = .78$ between Verbal and Nonverbal in Grade 2 to $r = .92$ between Verbal and Quantitative at Grade 5. This is probably expected, since all three batteries purport to measure abstract reasoning abilities. It was not expected for the WJ-III latent variables. The median correlation among WJ-III latent variables was $r = .76$ for at Grade 2 and $r = .67$ at grade 5. Two correlations actually exceeded $r = .90$. (Interestingly, both involve the Gf factor.)

In spite of this substantial overlap among factors, there is once again reasonably good evidence that the three CogAT batteries measure predictably different abilities. For the CogAT Primary battery (Grade 2), the CogAT Verbal factor had its highest correlation ($r = .75$) with the WJ-III Gc' factor, and both the CogAT Quantitative and Nonverbal factors correlated $r = .85$ with the WJ-III Gf' factor. For the CogAT Multilevel battery (Grade 5), the CogAT Verbal factor once again had its highest correlations with Gc', followed closely by Gq. Further, both the CogAT Quantitative and Nonverbal factors showed their highest correlations with Gq. This could mean that the CogAT Primary battery measures mainly Gc and Gf, whereas the Multilevel battery measures mainly Gc and Gq. Alternatively, it could mean that the WJ-III Gq factor is actually a better measure of g than the Gf factor in this sample of 87 fifth graders. The fact that the Nonverbal Battery did not show particularly high correlations with the Gv factor at grade 5 suggests that it is not a measure of spatial visualization ability. This concurs with the observation that there are no

Table 6
Correlations among Latent Stratum II Variables for Grade 2 (Upper Diagonal) and Grade 5 (Lower Diagonal)

	V	Q	N	Gc'	Gq	Gf'	Gsm'
V	—	.90	.78	.75	.57	.66	.53
Q	.92	—	.84	.65	.69	.85	.58
N	.79	.90	—	.39	.48	.85	.60
Gc'	.77	.56	.62	—	.75	.78	.62
Gq	.74	.86	.72	.67	—	.93	.61
Gf	.38	.41	.54	.94	.45	—	.89
Gsm	.68	.72	.48	.85	.85	.57	—
Gv	.38	.51	.52	.67	.75	.63	.67

Note. Grade 2 ($N = 87$) and Grade 5 ($N = 66$) variables marked with a prime (e.g., Gc') are estimated differently in these models (see Figures 1 and 2) than in McGrew and Woodcock (2001). For definitions of Gc, Gq, Gf, Gsm, and Gv, see Figure 1 (for grade 2, upper diagonal) and Figure 2 (for grade 5, lower diagonal). V = CogAT Verbal; Q = CogAT Quantitative; N = CogAT Nonverbal.

sex differences on the CogAT Nonverbal Battery (Lohman & Hagen, 2002). However, none of the correlations in Table 6 should be taken too seriously. All depend on the way observed variables are combined to form factors. Different models and/or a wider range of observed variables to define each “broad” factor would change the estimated correlations among factors.

Interbattery factor analyses. The final set of analyses is in some ways the most important. For these models, we estimated the relationship between the latent general factors measured by each battery. Separate analyses were performed at Grade 2, Grade 5, and then for all participants (Grades 2, 5, and 9). This last analysis used the three CogAT SAS scores to define g rather than the six (Primary Battery) or nine (Multilevel battery) CogAT subtests.

The three models are shown in Figures 5, 6, and 7. Model fit statistics are reported in Table 7. Each model combines the CogAT model for a particular group (left panel) with the corresponding WJ-III model (right panel). Since we have already estimated the integrity of the two parts of each of these within-test models, the model fit statistics reported in Table 7 are of secondary interest. Instead, the purpose of fitting each model was to estimate the correlation between the two g factors. In these models, we set the variance of both general factors to 1.0 and freed all paths from g to first order factors. As is shown in Figures 5, 6, and 7, these correlations between the two g factors were $r = .76$ at grade 2, $r = .78$ at grade 5, and $r = .82$ across grades 2, 5, and 9.

Although these correlations indicate considerable overlap, they are not as high as the correlation Lohman (2003) reported between CogAT6 and the WISC-III. In

a study of 91 sixth-grade students, the CogAT Composite score correlated $r = .79$ with WISC Full Scale IQ. Latent g factors on the two tests were perfectly correlated. This concurs with an earlier study in which CogAT was found to correlate approximately $r = .77$ with IQ on the Stanford-Binet Intelligence Scale. This suggests that the general factor on the WJ-III differs somewhat from the general factors on the CogAT, the WISC-III, and the Stanford-Binet.

But are the g factors really different? We re-fit models VI, VII, and VIII, this time fixing the covariance between the two factors at 1.0. Because the variances of the g factors were also fixed at 1.0, the covariance is the same as a correlation. Each of the new models thus had one more degree of freedom than the parallel model in Table 7. A nonsignificant increase in χ^2 would indicate that one cannot reject the hypothesis that the two g factors are the same. Instead, the increases in χ^2 were all large and highly significant ($\Delta\chi^2 = 44.6, 32.0,$ and 40.8 for models VI, VII, and VIII, respectively.) Therefore, we conclude that the general factors measured by the two batteries are not the same. This could be because the WJ-III samples a broader range of stratum II abilities and thus has a better, more representative g . On the other hand, it could mean that the effort to sample broadly resulted in a general factor that is statistically broader but also psychologically more diverse.

Since the earliest days of mental testing, there have been two views about g . One view is that it represents the efficiency with which one can perform a particular set of cognitive processes—such as the “education of relations and correlates” (Spearman, 1923). Closely allied with this view is that it captures individual differences in particular functions

Model VI
CogAT g vrs WJ-III g
Grade 2 N=87

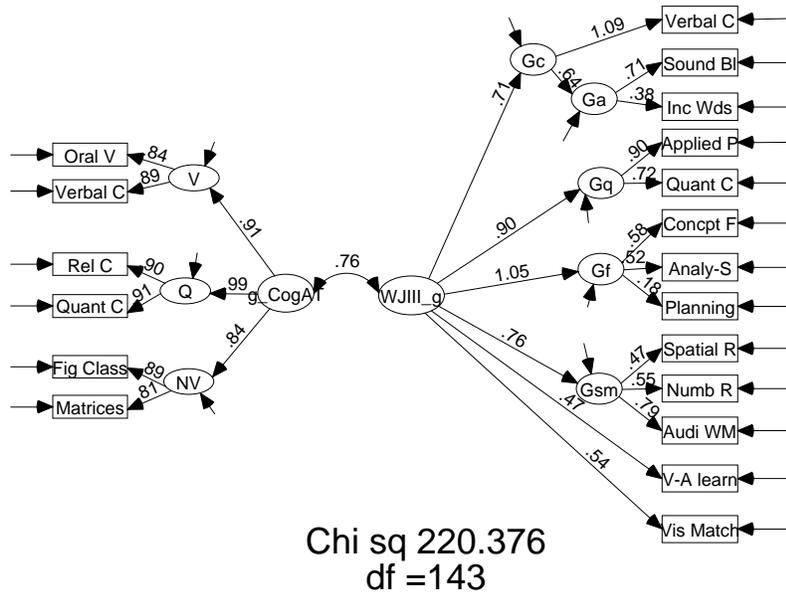


Figure 5. Model VI. Confirmatory interbattery factor analysis for the grade 2 sample. The unique and error variances of the measured variables are not shown to improve readability.

Model VII
CogAT g vrs WJ-III g
Grade 5 N=66

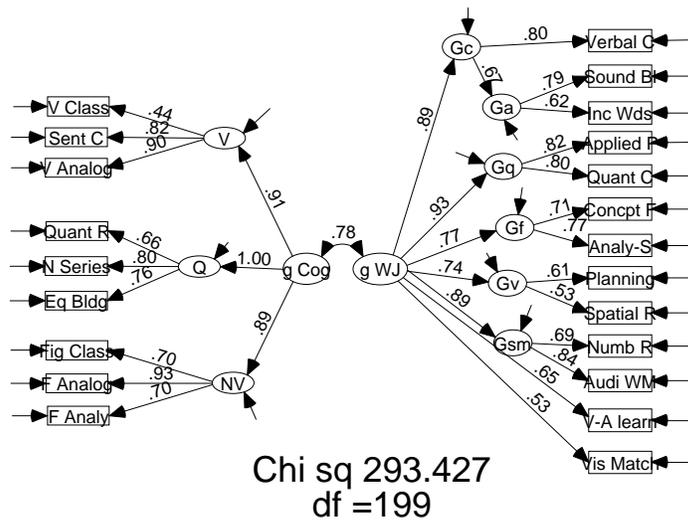


Figure 6. Model VII. Confirmatory interbattery factor analysis for the grade 5 sample. The unique and error variances of the measured variables are not shown to improve readability.

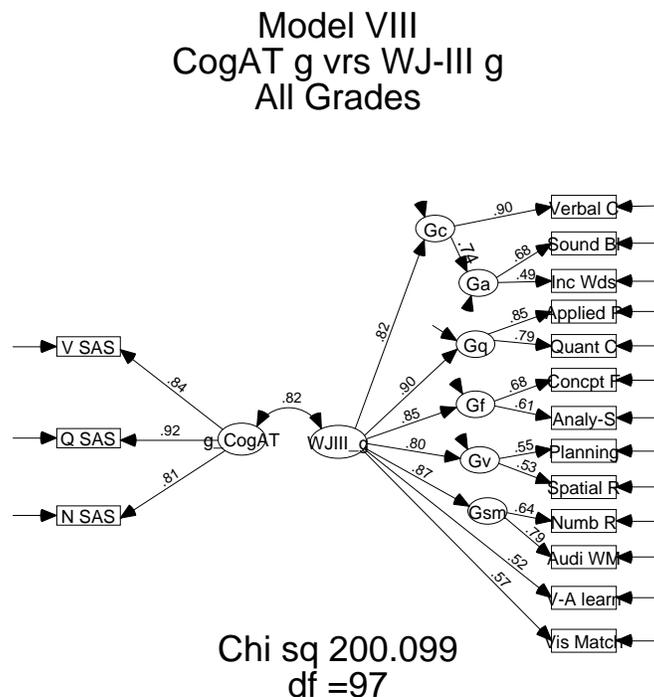


Figure 7. Model VIII. Confirmatory interbattery factor analysis for the full sample of 178 of grade 2, 5, and 9 students. The unique and error variances of the measured variables are not shown to improve readability.

or parameters of the brain. The other view is that *g* is simply a statistical abstraction, with no particular psychological meaning (Thomson, 1916). If the latter view is correct, then the best estimate of *g* would be one that averaged over a broad and representative sample of cognitive performances. If, on the other hand, the former view is correct, then the best estimate of *g* would be one that elicits the key cognitive processes in a variety of contexts.

We explored these possibilities in a model that fit a single general factor to the 13 WJ-III tests and the three CogAT SAS scores. Since the WJ-III scores outnumber the scores by a ratio of more than 4 to 1, the *g* factor should be swayed toward the *g* defined by the WJ-III

test battery. Such a result would support the hypothesis that the WJ-III defines a broader *g* than the CogAT. On the other hand, if both test batteries measure the same general factor, but the CogAT simply measures it better, then the CogAT tests should show higher loadings on the common *g* factor.

The results of this analysis, which are shown in Figure 8, clearly support the latter hypothesis. The CogAT Quantitative score had the highest *g* loading ($\beta = .83$), followed by CogAT Verbal ($\beta = .80$), WJ-III Applied problems ($\beta = .74$), CogAT Nonverbal ($\beta = .73$), WJ-III Verbal Comprehension ($\beta = .72$), and WJ-III Quantitative Concepts ($\beta = .70$). Clearly,

Table 7

Confirmatory Factor Models Estimating Correlations between the General Factors in Grade 2, Grade 5, and across Grades 2, 5, and 9

Model	Sample	χ^2	<i>df</i>	χ^2/df	TLI	RMSEA
VI. CogAT PB—WJ-III	Grade 2 (<i>N</i> = 87)	220.4	143	1.54	.880	.079
VII. CogAT ML—WJ-III	Grade 5 (<i>N</i> = 66)	293.4	199	1.47	.843	.085
VIII. CogAT—WJ-III	Grades 2, 5, 9 (<i>N</i> = 178)	200.0	97	2.06	.894	.077

Note. TLI = Tucker-Lewis fit index; RMSEA = Root Mean Square Error of Approximation; CogAT = Cognitive Abilities Test; PB = Primary Battery; ML = Multilevel Battery; WJ-III = Woodcock-Johnson III.

Model IX
Chi sq 329.672
df =104

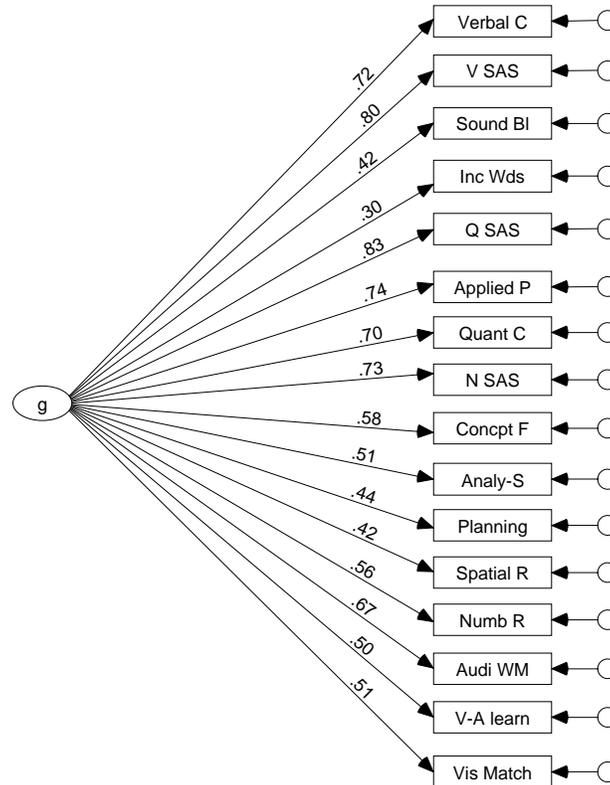


Figure 8. Model IX. Common g model for the 13 WJ-III and the 3 CogAT scores ($N = 178$).

whatever these 16 tests have in common is better measured by the CogAT than by any of the WJ-III tests. Further, it is noteworthy that, in both batteries, it is the quantitative reasoning tests rather than fluid ability, working memory, or verbal reasoning tests that have the highest g loadings.

Discussion

The purpose of this study was to investigate the concurrent validity of the Cognitive Abilities Test (Form 6) and the Woodcock-Johnson III Tests of Cognitive Abilities. We began by testing whether the covariance matrices for our samples of second and fifth grade students differed significantly from the population covariance matrices reported by the test publishers. We did this by fitting the hierarchical factor model specified by the test authors first to the standardization data, and secondly to the sample data. Separate models were fit at Grades 2 and 5 in part because the number and content of subtests on the

CogAT differs between these grades, and in part because the hypothesized model for the WJ-III fit the data for our second graders only poorly. In particular, the G_v factor could not be identified in the set of 13 WJ-III tests that were administered to second graders. Nonetheless, we were able to establish that, with the exception of one factor loading, relationships among the tests in each battery could be described by the same factor models in the standardization sample and in our data.

Given this evidence on the representativeness of our data, we investigated relationships between scores on the two test batteries in three ways. First, we examined correlations between WJ-III cluster scores and CogAT SAS scores. The WJ-III General Intellectual Ability (GIA) score correlated about as highly with the CogAT Composite score ($r = .68$) as with the general scores from individually administered ability tests. Each of the three CogAT batteries also showed convergent validity in that its highest

correlation was with the corresponding WJ-III cluster. These pairs were CogAT Verbal–WJ-III Verbal ($r = .62$), CogAT Quantitative–WJ-III Math Reasoning ($r = .58$), and CogAT Nonverbal–WJ-III Fluid Reasoning ($r = .55$). Correlations computed within the separate CogAT batteries were less straightforward. The best conclusion seemed to be that the CogAT primarily measured something shared by the various WJ-III test clusters, and only secondarily abilities unique to each.

We then examined correlations between latent variables on the two tests. These showed considerable overlap among the latent factors on both batteries, which made interpretation difficult. Nonetheless, there was clear separation among the three CogAT scores for the Multilevel Battery and for two scores (Verbal versus Quantitative/ Nonverbal) at the Primary Battery.

Finally, we investigated relationships between the general factors defined by the two test batteries in a series of confirmatory interbattery factor analyses. Relationships between the CogAT and WJ-III general factors defined by the two batteries were strong and consistent across grades. Overall, the latent g factors on the two batteries correlated $r = .82$. Although this indicates considerable overlap, it is noticeably less than the overlap reported between the CogAT and the WISC-III or between an earlier edition of the CogAT and the Stanford-Binet. This could be because the WJ-III samples a broader range of stratum II abilities and thus has a better, more representative g . On the other hand, it could mean that the effort to sample broadly resulted in a general score that is both statistically broader and psychologically more diverse. A model fit to all 16 tests supported the latter interpretation. The CogAT Quantitative score was the best measure of g , followed by the CogAT Verbal score, the WJ-III quantitative tests and the CogAT Nonverbal score.

Recent discussions of the nature of general ability have emphasized the importance of physiological processes (Jensen, 1998), the role of working memory (Kyllonen, 1996), or the congruence between a primary Inductive Reasoning factor, the stratum II Fluid Ability factor (G_f), and g (Gustafsson, 2002). However, the present study supports Keith and Witta's (1997) hypothesis that quantitative reasoning may be an even better indicator of g . Quantitative reasoning has always been represented in some form in achievement test batteries, and in aptitude tests (such as the SAT) designed to predict academic success. But a broad quantitative *knowledge* factor (G_q) was not added to G_f - G_c theory until the late 1980s (Horn, 1989). Carroll's (1993) three-stratum theory, on the other hand, considers quantitative *reasoning* to be part of a broad fluid reasoning (G_f) factor. Confirmatory factor analyses of different ability test batteries mirror this

ambivalence. Some studies find g and G_q indistinguishable [as in Keith & Bickley's (1992) factor analysis of the Stanford-Binet IV or Lohman & Hagen's (2002) factor analyses of the CogAT Primary Battery], other studies find G_q to be the best indicator of g [as in Keith & Witta's (1997) factor analyses of the WISC-III or Lohman & Hagen's (2002) factor analyses of the CogAT Multilevel Battery], and yet other studies find distinguishable g and G_q factors [as in Bickley, Keith, & Wolfle's (1995) factor analysis of the Woodcock-Johnson Psychoeducational Battery–Revised].

Paradoxically, quantitative reasoning has not been much studied because it is difficult to separate from g unless combined with tests of more specific mathematical knowledge and skill (as in the G_q factor). But it is this overlap with g that makes quantitative reasoning particularly interesting as a vehicle for understanding the nature of g . Perhaps the most salient characteristic of quantitative concepts is abstraction. Even elementary operations like counting require abstraction: two cats are in some way the same as two dogs or two of anything. The number line itself is an abstraction, especially when it includes negative numbers. Abstraction is most obvious in understanding concepts such as *variable* or, later, *imaginary number*.

Several early definitions of g emphasized abstract thinking or reasoning abilities. And the transition from concrete to abstract thinking figured prominently in Piaget's theory of intelligence. Modern definitions of g emphasize the importance of working memory resources or even of reasoning, but do not have much to say about the role of abstract thinking. These analyses suggest a closer study of quantitative reasoning might be a good place to begin in exploring this possibility.

References

- Arbuckle, J. L. (1999). *AMOS 4.0*. Chicago, IL: SmallWaters Corp.
- Bickley, P. G., Keith, T. Z., & Wolfle, L. M. (1995). The three-stratum theory of cognitive abilities: Test of the structure of intelligence across the life span. *Intelligence, 20*, 309-328.
- Bollen, K. A. (1989). *Structural equations with latent variables*. New York: Wiley.
- Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. New York: Cambridge University Press.
- Cattell, R. B. (1971). *Abilities: Their structure, growth, and action*. Boston: Houghton Mifflin.
- Gustafsson, J. -E. (2002). Measurement from a hierarchical point of view. In H. I. Braun, D. N. Jackson, & D. E. Wiley (Eds.), *The role of constructs in psychological and educational measurement* (pp. 73-96). Mahwah, NJ: Erlbaum.

- Horn, J. L. (1989). Models for intelligence. In R. Linn (Ed.), *Intelligence: Measurement, theory, and public policy* (pp. 29-73). Urbana, IL: University of Illinois Press.
- Horn, J. L. & McArdle, J. J. (1992). A practical and theoretical guide to measurement invariance in aging research. *Experimental Aging Research, 18*, 117-144.
- 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.
- Jensen, A. R. (1998). *The g factor*. Westport, CT: Praeger.
- Keith, T. Z., & Bickley, P. G. (1992). *Confirmatory factor analysis of the Stanford-Binet IV*. Unpublished manuscript.
- 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*(2), 89-107.
- Kyllonen, P. C. (1996). Is working memory capacity Spearman's *g*? In I. Dennis & P. Tapsfield (Eds.), *Human abilities: Their nature and measurement* (pp. 49-75). Mahwah, NJ: Erlbaum.
- Lohman, D. F. (2003). *The Wechsler Intelligence Scale for Children III and the Cognitive Abilities Test (Form 6): A concurrent validity study*. Unpublished manuscript.
- Lohman, D. F., & Hagen, E. (2001). *Cognitive Abilities Test (Form 6)*. Itasca, IL: Riverside.
- Lohman, D. F., & Hagen, E. (2002). *Cognitive Abilities Test (Form 6): Research handbook*. Itasca, IL: Riverside.
- Lorge, I., & Thorndike, R. M. (1954). *Lorge-Thorndike Intelligence Tests*. Boston: Houghton Mifflin.
- McGrew, K. S., & Woodcock, R. W. (2001). *Woodcock-Johnson III technical manual*. Itasca, IL: Riverside.
- Spearman, C. (1923). *The nature of "intelligence" and the principles of cognition*. London: Macmillan.
- Thomson, G. H. (1916). A hierarchy without a general factor. *British Journal of Psychology, 8*, 271-281.
- Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock-Johnson III Tests of Cognitive Abilities*. Itasca, IL: Riverside.