# **Academic Achievement *g*—Diva or a Workhorse? Further Insights from Bifactor Modeling with the WJ-IV**

Ryan J. McGill

College of William & Mary

Citation:
McGill, R. J. (2017, April). *Academic achievement—g diva or a workhorse? Further insights from bifactor modeling with the WJ-IV*. Paper presented at the annual meeting of the American Educational Research Association, San Antonio, TX.

Author Notes

Ryan J. McGill, School of Education, College of William & Mary, P.O. Box 8795, Williamsburg, VA 23187.

Correspondence concerning this paper should be addressed to Ryan J. McGill, School of Education, College of William & Mary, P.O. Box 8795, Williamsburg, VA 23187. E-Mail: rmcgill@wm.edu

**Abstract**

Using confirmatory factor analysis, this study applied different Cattell-Horn-Carroll (CHC) factor models to normative data from the Woodcock-Johnson Tests of Achievement-Fourth Edition (WJ-IV ACH) to determine if academic *g* (*ACH-g*) is best conceptualized as a superordinate or breadth factor in its influence on lower-order achievement dimensions. The bifactor model fit the WJ-IV ACH normative data better than the favored higher-order model at ages 9-13 and 14-19. Further, calculation of omega coefficients demonstrated that the domain-specific group factors exhibited poor reliability independent of *ACH-g*. These results suggest that only the general achievement dimension was sufficiently robust for clinical use and raise questions about the utility of multidimensional achievement tests such as the WJ-IV ACH for high-stakes decision making and treatment planning.

*Keywords*: WJ-IV ACH, Bifactor model, Academic-*g*, Model-based reliability

# Academic Achievement *g*—Diva or a Workhorse? Further Insights from Bifactor Modeling with the WJ-IV

The Cattell-Horn-Carroll theory of cognitive abilities (CHC; McGrew, 2009; Schneider & McGrew, 2012) is a leading method of psychometric test development and interpretation in psychology and education. CHC theory was developed as a synthesis of the *Gf*-*Gc* theory (Horn & Cattell, 1966) and Carroll's (1993) three-stratum model. It conceptualizes cognitive abilities within a hierarchical taxonomy in which elements are stratified according to breadth. The most general ability resides at the apex of the model at Stratum III and is referred to as a general factor or *g*. The next level (Stratum II) includes nine broad abilities (e.g., Fluid Reasoning [Gf], Crystallized Ability [Gc]). At the bottom of the model are over 70 narrow abilities (Stratum I) which are organized according to their mapping onto the Stratum II abilities.

Much of the research conducted in this area has used the Woodcock-Johnson Psychoeducational Battery (WJ; Woodcock & Johnson, 1989; Woodcock, McGrew, & Mather, 2001a) to help operationalize and gain insight into the structure and organization of CHC abilities on intelligence tests (Keith & Reynolds, 2010). However, it is important to note that Stratum II contains two dimensions associated with academic achievement: Reading/Writing (Grw) and Quantitative Reasoning (Gq). As such, the model has increasingly been utilized as a blueprint for developing standardized achievement tests as well as understanding how CHC-defined cognitive abilities relate to academic achievement (McGrew, 2005; Beaujean, Parkin, & Parker, 2014).

## Rise of Multifactored Achievement Tests

As a result of these advances, contemporary achievement tests have been designed using a higher-order factor model to appraise examinee performance at multiple levels (e.g., subtest scores, factor scores, global composites), providing examiners with the ability to make numerous inferences about the status of an individual's academic functioning (Schneider, 2013). The trend among publishers has been to create longer test batteries that provide users with an ever increasing number of composite indices (Glutting, Watkins, & Youngstrom, 2003). As a consequence, a considerable amount of time and resources are expended by educational professionals (e.g., school psychologists, special education teachers) to administer and interpret the wealth of information provided by these instruments (Yates & Taub, 2003). This investment is based upon the assumption that the information provided by these indices is beneficial. To wit, Lictenberger and Breaux (2010) suggest that these measures can be utilized to identify students at risk for academic failure, diagnose a learning disability, determine eligibility for special education and related services, evaluate profiles of academic strengths and weaknesses, and develop individualized academic interventions.

**Academic Achievement-*g* and its Relationship to Lower-Order Scores**

The expansive claims made by Lichtenberger and Breaux (2010) assume contemporary achievement tests measure lower-order constructs (e.g., Stratum II variables) with enough specificity to warrant confident clinical interpretation. However, the results of hierarchical exploratory factor analytic (EFA) and bifactor modeling investigations using confirmatory factor analysis (CFA) have raised significant questions about the tenability of similar claims made for multifactored intelligence tests (e.g., Canivez, 2013, 2014; Dombrowksi, Canivez, Watkins, & Beaujean, 2015; Dombrowksi, McGill, & Canivez, 2016; McGill & Spurgin, 2015; Watkins & Beaujean, 2014). In sum, these investigations have consistently found that the general factor (*COG-g*; as represented by the full scale IQ composite at Stratum III) accounts for the vast majority of reliable variance in lower-order measures and that Stratum II variables are frequently not measured well, if at all, confounding clinical interpretation. Unfortunately, related factor analytic investigations of achievement tests have been relatively limited.

In a hierarchical EFA of the WJ-III Tests of Achievement (WJ-III ACH; Woodcock, McGrew, & Mather, 2001b) using the Schmid and Leiman (1957) procedure to partition reliable variance, Dombrowski (2015) found that the general factor accounted for 44% and 70% of the total and common variance respectively in the 22 subtest total battery configuration whereas, the proportions of total (1% to 3%) and common variance (2% to 6%) attributable to the Stratum II first-order factors were consistently meager. Although the tenability of a general factor of academic achievement (*ACH-g*) has long been debated within the professional literature (e.g., Gustaffson & Balke, 1993; Rinderman & Neubauer, 2004; Spearman, 1904), results from a CFA investigation by Kaufman and colleagues (2012) found that a Stratum III *ACH-g* dimension explained tests of reading, math, and writing achievement from the Kaufman Test of Educational Achievement-Second Edition (KTEA-II; Kaufman & Kaufman, 2004) and that this general factor was related to but distinct from the Stratum III *COG-g* factor produced from cognitive measures from the WJ-III.

However, it should be noted that both investigations represented the structure of achievement using a higher-order factor model. In a higher-order model a general factor is produced from the correlations between Stratum II factors. The influence of *g* on the measured variables is mediated or filtered through Stratum II variables. As a result, the higher-order model is commonly referred to as an indirect hierarchical model in the technical literature (Canivez, 2016). An alternative to the higher-order model is a bifactor or direct hierarchical model (Holzinger & Swineford, 1937). In this model, *g* is considered to be a nested first-order variable along with the other group factors. Whereas *g* has direct influences on all tests, the remaining group factors influence only a subset of tests. Due to its advantages, the bifactor model has been recommended by researchers and has increasingly been found to provide a better fit to cognitive data when compared to the higher-order model (e.g., Beaujean, Parker, & Parkin, 2014; Chen, Hayes, Carver, Laurenceau, & Zhang, 2012; Gignac, 2007; Reise, 2012). Inexplicably, a bifactor model has yet to be applied to achievement data, suggesting that our understanding of the structure of achievement and the influence of *ACH-g* is presently incomplete.

### Objectives and Purpose of the Current Study

The purpose of the current study was to compare higher-order and bifactor representations of the CHC-based factors in the Woodcock-Johnson IV Tests of Achievement (WJ-IV ACH; Schrank, Mather, & McGrew, 2014) 20 subtest total battery configuration. Although CFA results reported in the *Technical Manual* (McGrew, LaForte, & Schrank, 2014) supported a CHC-based higher-order structure for the WJ-IV, those results were obtained by combining all measures from the Achievement, Cognitive, and Oral Language batteries together for analyses. As a separate and distinct battery, the structure of the Tests of Achievement was not explored. As previously noted, there are no equivalent validity studies that have compared direct and indirect hierarchical structural models with standardized achievement test data. Therefore the results obtained from this investigation will be important for understating the latent structure of achievement as well as the relationship and influence of *ACH-g* on lower-order achievement constructs on the WJ-IV ACH and other related measures.

## Method

### Participants

Participants were members of the WJ-IV ACH standardization sample and included a total of 3,257 individuals ranging in age from 9-19 years. The standardization sample included stratified proportional sampling across demographic variables of age, gender, race/ethnicity, parent educational level, and geographic region and close correspondence to 2010 U.S. census estimates across the stratification variables.

**Measurement Instrument**

The WJ-IV ACH is a comprehensive multidimensional individually administered achievement battery that contains 22 subtests that measure reading, writing, mathematics, and academic content knowledge (e.g., social studies). All scores on the WJ-IV ACH are normally distributed ($M$ = 100, $SD$ = 15). Extensive normative and technical information is provided in the *Technical Manual* (McGrew, LaForte, & Schrank, 2014).

**Data Analyses**

Data analyses for the present study occurred in two steps. First, WJ-IV ACH subtest correlation matrices for the two standardization sample subgroups (9-13, 14-19) were obtained from the *Technical Manual* (McGrew, LaForte, & Schrank, 2014) to conduct confirmatory factor analysis. EQS, Version 6.2 for Windows (Bentler & Wu 2012) was used to conduct CFA using maximum likelihood estimation. For each group, three first-order models (1, 3-4 factors) and two hierarchical models (direct, indirect) were specified and examined. To comport with best practice (e.g., Marsh, Hau, & Grayson, 2005; Mueller & Hancock, 2008), multiple indices were examined to evaluate the adequacy of model fit. Specifically the (a) chi-square ($\chi^2$), (b) comparative fit index (CFI), (c) root mean square error of approximation (RMSEA), (d) standardized root mean square residual (SRMR), and (e) Akaike's information criterion (AIC).

Finally, the bifactor model hypothesizes that each WJ-IV ACH subtest is influenced simultaneously by two orthogonal latent constructs: a general ability factor (*ACH-g*) and a first-order domain-specific group factor (e.g., Reading/Writing, Quantitative Reasoning, etc.). As a consequence, Omega ($\omega$) and omega-hierarchical ($\omega_h$) were estimated as model-based reliability estimates of the latent factors (Gignac & Watkins, 2013). Whereas $\omega$ estimates the variance accounted for by both of the constructs in a given domain, $\omega_h$ estimates the variance accounted for by a single target construct. Chen and colleagues (2012) stressed that "for multidimensional constructs, the alpha coefficient is complexly determined, and McDonald's omega-hierarchical ($\omega_h$; 1999) provides a better estimate for the composite score and thus should be used" (p. 228). Omega estimates were produced using the *Omega* program (Watkins, 2013). Albeit subjective, omega coefficients should at a minimum exceed .50, but .75 would be preferred (Reise, 2012; Reise, Bonifay, & Haviland, 2013).

**Results and Discussion**

Model fit statistics presented in Table 1 illustrate the increasingly better fit from one to four factors; however fit statistics for the first-order models were inadequate (CFI <.90, SRMR > .06). However, it should be noted that fit statistics obtained in the present study were substantially better than those reported for similar measurement models in the *Technical Manual* (McGrew, LaForte, & Schrank, 2014). In the oblique four-factors model, latent factor correlations ranged from .59 to .85 (ages 9-13) and .63 to .85 (ages 14-19) suggesting the presence of a higher-order dimension requiring explication (Gorsuch, 2003). For both age groups, a four-factor CHC-based direct hierarchical (bifactor) model provided statistically significant better fit ($\Delta\chi^2$, $p < .01$, lower AIC values) to the WJ-IV ACH data when compared to the oblique correlated factors and indirect hierarchical models. It should also be noted that for both age groups, the direct hierarchical model was the only model in which fit statistics were indicative of adequate fit (e.g., CFI >.90, SRMR < .06) and a substantial improvement from those reported for the final CFA validation model reported in the *Technical Manual* (McGrew, LaForte, & Schrank, 2014) for the same age groups.

Tables 2 and 3 present the decomposed subtest variance estimates of the WJ-IV ACH based on the direct hierarchical model. For both age groups, the *ACH-g* factor accounted for substantially greater portions of total and common variance relative to the four CHC-based group factors (Reading/Writing, Quantitative Reasoning, Processing Speed, and Crystallized Ability). Omega hierarchical coefficients presented in Tables 2 and 3 provide estimates of the reliability of the latent constructs with the effects of other constructs removed. Whereas, coefficients obtained for the general factor were strong, the coefficients obtained for the four group factors failed to exceed recommended thresholds for confident clinical interpretation (e.g., > .50; Reise, 2012; Reise, Bonifay, & Haviland, 2013).

**Conclusion and Major Implications**

The present results suggest that the structure of achievement on the WJ-IV ACH is best represented by a bifactor measurement model in which both the general and CHC-based group factors have direct influences on tests (see Figures 1 and 2). Consistent with Dombrowski (2015), the present results also suggest that the WJ-IV ACH provides strong measurement of general achievement however; the variance associated with more discrete CHC constructs was consistently weak, calling into question the value afforded at that level of measurement. Researchers have long questioned the clinical utility of comprehensive norm-referenced achievement measures such as the WJ-IV (Elliott & Fuchs, 1997; Good & Salvia, 1988). The present results raise additional questions about their value for high-stakes decision-making and individual treatment planning, as posited in the technical and professional literature (Lictenberger & Breaux, 2010; Mascolo, Alfonso, & Flanagan, 2014). Future studies extending these lines of research to other measurement instruments would greatly benefit clinicians as well as those researching individual differences.

## References

Beaujean, A. A., Parkin, J., & Parker, S. (2014). Comparing Cattell-Horn-Carroll factor models: Differences between bifactor and higher order factor models in predicting language achievement. *Psychological Assessment, 26,* 789-805. doi: 10.1037/a0036745

Bentler, P. M., & Wu, E. J. C. (2012). *EQS for Windows*. Encino CA: Multivariate Software.

Canivez, G. L. (2013). Psychometric versus actuarial interpretation of intelligence and related aptitude batteries. In D. H. Saklofske, C. R. Reynolds, & V. L. Schwean (Eds.). *The Oxford handbook of child psychological assessment* (pp. 84-112). New York: Oxford University Press.

Canivez, G. L. (2014). Construct validity of the WISC-IV with a referred sample: Direct versus indirect hierarchical structures. *School Psychology Quarterly, 29,* 38-51. doi: 10.1037/spq0000032

Canivez, G. L. (2016). Bifactor modeling in construct validation of multifactored tests: Implications for understanding multidimensional constructs and test interpretation. In K. Schweizer & C. DiStefano (Eds.), *Principles and methods of test construction: Standards and recent advancements* (pp. 247-271). Gottingen, Germany: Hogrefe.

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor analytic studies*. New York: Cambridge University Press.

Chen, F. F., Hayes, A., Carver, C. S., Laurenceau, J. –P., & Zhang, Z. (2012). Modeling general and specific variance in multifaceted constructs: A comparison of the bifactor model to other approaches. *Journal of Personality, 80,* 219-251. doi:10.1111/j.1467–6494.2011.00739.x

Dombrowski, S. C. (2015). Exploratory bifactor analysis of the WJ-III Achievement at school-age via the Schmid-Leiman orthogonalization procedure. *Canadian Journal of School Psychology, 30,* 34-50. doi: 10.1177/0829573514560529

Dombrowski, S. C., Canivez, G. L., Watkins, M. W., & Beaujean, A. A. (2015). Exploratory bifactor analysis of the Wechsler Intelligence Scale for Children-Fifth Edition with the 16 primary and secondary subtests. *Intelligence, 53,* 194-201. doi: 10.1016/j.intell.2015.10.009

Dombrowski, S. C., McGill, R. J., & Canivez, G. L. (2016). Exploratory and hierarchical factor analysis of the WJ-IV Cognitive at school age. *Psychological Assessment*. Advance online publication. doi: 10.1037/pas0000350

Elliott, S. N., & Fuchs, L. S. (1997). The utility of curriculum-based measurement and performance assessment as alternatives to traditional intelligence and achievement tests. *School Psychology Review, 26,* 224-233. Retrieved from http://www.nasponline.org

Gignac, G. E. (2007). Multi-factor modeling in individual differences research: Some recommendations and suggestions. *Personality and Individual Differences, 42,* 37-48. doi: 0.1016/j.paid.2006.06.019

Gignac, G. E., & Watkins, M. W. (2013). Bifactor modeling and the estimation of model-based reliability in the WAIS-IV. *Multivariate Behavioral Research, 48,* 639-662. doi: 10.1080/00273171.2013.804398

Glutting, J. J., Watkins, M. W., & Youngstrom, E. A. (2003). Multifactored and cross-battery assessments: Are they worth the effort? In C. R. Reynolds & R. W. Kamphaus (Eds.), *Handbook of psychological and educational assessment of children: Intelligence aptitude, and achievement* ($2^{nd}$ ed., pp. 343-374). New York: Guilford.

Good, R. H., & Salvia, J. (1988). Curriculum bias in published, norm-referenced reading tests:

Demonstrable effects. *School Psychology Review, 17,* 51-60. Retrieved from http:www.nasponline.org

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

Gustaffson, J. E., & Balke, G. (1993). General and specific abilities as predictors of school achievement. *Multivariate Behavioral Research, 28,* 407–434. doi: 10.1207/s15327906mbr2804_2

Holzinger, K. J., & Swineford, F. (1937). The bi–factor method. *Psychometrika, 2,* 41-54. doi:10.1007/BF02287965

Horn, J. L., & Cattell, R. B. (1966). Refinement and test of the theory of fluid and crystallized general intelligence. *Journal of Educational Psychology, 57,* 253-270. doi: 10.1037/h0023816

Kaufman, A. S., & Kaufman, N. L. (2004). *Kaufman Test of Educational Achievement-Second Edition*. Circle Pines, MN: American Guidance Service.

Kaufman, S. B., Reynolds, M. R., Liu, X., Kaufman, A. S., & McGrew, K. S. (2012). Are cognitive *g* and academic achievement *g* one and the same *g*? An exploration on the Woodcock-Johnson and Kaufman tests. *Intelligence, 40,* 123-138. doi: 10.1016/j.intell.2012.01.009

Keith, T. Z., & Reynolds, M. R. (2010). Cattell-Horn-Carroll abilities and cognitive tests: What we've learned from 20 years of research. *Psychology in the Schools, 47,* 635-650. doi: 10.1002/pits.20496

Lichtenberger, E. O., & Breaux, K. C. (2010). *Essentials of WIAT-III and KTEA-II assessment*.

Hoboken, NJ: John Wiley.

Mascolo, J. T., Alfonso, V. C., & Flanagan, D. P. (Eds.) (2014). *Essentials of planning, selecting, and tailoring interventions for unique learners*. Hoboken, NJ: John Wiley.

Marsh, H. W., Hau, K. T., & Grayson, D. (2005). Goodness of fit in structural equation models. In A. Maydeu-Olivares & J. J. McArdle (Eds.), *Contemporary psychometrics: A festschrift for Roderick P. McDonald* (pp. 275–340). Mahwah, NJ: Erlbaum

McDonald, R. (1999). *Test theory: A unified treatment*. Mahwah, NJ: Erlbaum.

McGill, R. J., & Spurgin, A. R. (2015). Exploratory higher order analysis of the Luria interpretive model on the Kaufman Assessment Battery for Children-Second Edition (KABC-II) school-age battery. *Assessment*. Advanced online publication. doi: 10.1177/1073191115614081

McGrew, K. S. (2005). The Cattell-Horn-Carroll theory of cognitive abilities: Past, present, and future. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (2nd ed., pp. 136-182). New York: Guilford.

McGrew, K. S. (2009). Editorial: CHC theory and the human cognitive abilities project: Standing on the shoulders of the giants of psychometric intelligence research. *Intelligence, 37,* 1-10. doi: 10.1016/j.intell.2008.08.004

McGrew, K. S., LaForte, E. M., & Schrank, F. A. (2014). *Woodcock-Johnson IV technical manual*. Rolling Meadows, IL: Riverside.

Mueller, R. O., & Hancock, G. R. (2008). Best practices in structural equation modeling. In J. W. Osborne (Ed.), *Best practices in quantitative methods*. Thousand Oaks, CA: Sage.

Reise, S. P. (2012). The rediscovery of bifactor measurement models. *Multivariate Behavioral Research, 47,* 667–696. doi:10.1080/00273171.2012.715555

Reise, S. P., Bonifay, W. E., & Haviland, M. G. (2013). Scoring and modeling psychological measures in the presence of multidimensionality. *Journal of Personality Assessment, 95,* 129-140. doi:10.1080/00223891.2012.725437

Rindermann, H., & Neubauer, A. C. (2004). Processing speed, intelligence, creativity, and school performance: Testing of causal hypotheses using structural equation models. *Intelligence, 32,* 573–589. doi: 10.1016/j.intell.2004.06.005

Schmid, J., & Leiman, J. M. (1957). The development of hierarchical factor solutions. *Psychometrika, 22,* 53-61. doi: 10.1007/BF02289209

Schneider, W. J. (2013). Principles of assessment aptitude and achievement. In D. H. Saklofske, C. R. Reynolds, & V. L. Schwean (Eds.). *The Oxford handbook of child psychological assessment* (pp. 286-330). New York: Oxford University Press.

Schneider, W. J., & McGrew, K. S. (2012). The Cattell-Horn-Carroll model of intelligence. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (3rd ed., pp. 99-144). New York: Guilford.

Schrank, F. A., McGrew, K. S., & Mather, N. (2014). *Woodcock-Johnson IV Tests of Achievement*. Rolling Meadows, IL: Riverside.

Spearman, C. (1904). "General intelligence," objectively determined and measured. *American Journal of Psychology, 15,* 201-293. Retrieved from http://psychclassics.yorku.ca/Spearman

Watkins, M. W. (2013). *Omega.* [Computer software]. Phoenix, AZ: Ed & Psych Associates.

Watkins, M. W., & Beaujean, A. A. (2014). Bifactor structure of the Wechsler Preschool and Primary Scale of Intelligence-Fourth Edition. *School Psychology Quarterly, 29,* 52-63. Doi: 10.1037/spq0000038

Woodcock, R. W., & Johnson, M. B. (1989). *Woodcock-Johnson Psycho-Educational Battery-Revised*. Chicago: Riverside.

Woodcock, R. W., McGrew, K. S., Mather, N. (2001a). *Woodcock-Johnson III*. Itasca, IL: Riverside.

Woodcock, R. W., McGrew, K. S., Mather, N. (2001b). *Woodcock-Johnson III Tests of Achievement*. Itasca, IL: Riverside.

Yates, B. T., & Taub, J. (2003). Assessing the costs, benefits, cost-effectiveness, and cost-benefit of psychological assessment: We should, we can, and here's how. *Psychological Assessment, 15,* 478-495. doi: 10.1037/1040-3590.15.4.478

**Table 1**

*Confirmatory Factor Analysis Fit Statistics for Woodcock-Johnson Tests of Achievement-Fourth Edition (WJ-IV ACH)*

| Model | *χ2* | *df* | *p* | CFI | SRMR | RMSEA | 90%CI RMSEA | AIC |
|---|---|---|---|---|---|---|---|---|
| Ages 9-13 (*n* = 1,572) | | | | | | | | |
| One Factor (ACH-*g*) | 5837.62 | 166 | .00 | .769 | .076 | .147 | [144, .151] | 5505 |
| Three Oblique Factors (Grw, Gq, Gc) | 4517.92 | 163 | .00 | .823 | .070 | .130 | [.127, .134] | 4191 |
| Four Oblique Factors (Grw, Gq, Gs, Gc) | 3391.98* | 160 | .00 | .868 | .064 | .113 | [.110, .117] | 3071 |
| Indirect Hierarchical | 3504.12 | 162 | .00 | .864 | .065 | .115 | [.111, .118] | 3180 |
| Direct Hierarchical | 2420.51** | 145 | .00 | .907 | .047 | .100 | [.096, .103] | 2130 |
| Ages 14-19 (*n* = 1,685) | | | | | | | | |
| One Factor (ACH-*g*) | 6572.61 | 166 | .00 | .772 | .074 | .151 | [.148, .154] | 6240 |
| Three Oblique Factors (Grw, Gq, Gc) | 4772.56 | 163 | .00 | .836 | .066 | .130 | [.126, .133] | 4446 |
| Four Oblique Factors (Grw, Gq, Gs, Gc) | 3826.95* | 160 | .00 | .869 | .061 | .117 | [.113, .120] | 3506 |
| Indirect Hierarchical | 3959.84 | 162 | .00 | .865 | .063 | .118 | [.115, .121] | 3635 |
| Direct Hierarchical | 2821.52** | 145 | .00 | .905 | .047 | .105 | [.101, 108] | 2531 |

*Note*. In the oblique four-factor models, latent factor correlations were as follows: .59 to .85 (ages 9-13), .63 to .85 (ages 14-19). CFI = comparative fit index; SRMR = standardized root mean square residual; RMSEA = root mean square error of approximation; AIC = Akaike information criterion. ACH-*g* = general achievement; Grw = Reading/Writing; Gq = Quantitative Reasoning; Gc = Academic Knowledge/Crystallized Ability; Gs = Academic Fluency/Processing Speed.

* Statistically different ($p < .01$) from previous two models.

**Statistically different ($p < .01$) from previous two models.

**Table 2**

*Standardized Loading Coefficients for Bifactor Model of the Woodcock-Johnson Tests of Achievement-Fourth Edition (Ages 9-13)*

| | Factor | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Test | ACH-*g* | Grw | Gq | Gs | Gc | $h^2$ | $u^2$ | Error | $s^2$ |
| Letter-Word Identification | .847 | .290 | | | | .802 | .198 | .060 | .138 |
| Spelling | .857 | .160 | | | | .760 | .240 | .080 | .160 |
| Passage Comprehension | .768 | .173 | | | | .620 | .380 | .110 | .270 |
| Writing Samples | .589 | .252 | | | | .410 | .590 | .100 | .490 |
| Word Attack | .619 | .684 | | | | .851 | .149 | .100 | .049 |
| Oral Reading | .748 | .164 | | | | .586 | .414 | .040 | .374 |
| Reading Recall | .627 | .094 | | | | .402 | .598 | .080 | .518 |
| Editing | .844 | .030 | | | | .713 | .287 | .090 | .197 |
| Spelling of Sounds | .573 | .399 | | | | .488 | .512 | .120 | .392 |
| Applied Problems | .696 | | .473 | | | .708 | .292 | .080 | .212 |
| Calculation | .734 | | .392 | | | .692 | .308 | .070 | .238 |
| Number Matrices | .574 | | .296 | | | .417 | .583 | .080 | .503 |
| Sentence Reading Fluency | .759 | | | .490 | | .816 | .184 | .060 | .124 |
| Math Fact Fluency | .669 | | | .392 | | .601 | .399 | .040 | .359 |
| Sentence Writing Fluency | .720 | | | .381 | | .664 | .336 | .200 | .136 |
| Word Reading Fluency | .652 | | | .533 | | .709 | .291 | .080 | .211 |
| Reading Vocabulary | .812 | | | | .237 | .716 | .284 | .120 | .164 |
| Science | .534 | | | | .618 | .667 | .333 | .160 | .173 |
| Social Studies | .589 | | | | .593 | .699 | .301 | .130 | .171 |
| Humanities | .545 | | | | .538 | .586 | .414 | .130 | .284 |
| Total Variance (%) | 48.4 | 4.3 | 2.3 | 4.1 | 5.4 | 64.5 | 35.5 | 9.7 | 25.8 |
| Common Variance (%) | 74.9 | 6.7 | 3.6 | 6.4 | 8.4 | | | | |
| $\omega$ | .966 | .933 | .819 | .901 | .883 | | | | |
| $\omega_h$ | .902 | .100 | .206 | .263 | .345 | | | | |

*Note*. ACH-*g* = general achievement, Grw = Reading/Writing, Gq = Quantitative Reasoning, Gc = Academic Knowledge/Crystallized Ability, Gs = Academic Fluency/Processing Speed. $h^2$ = communality; $u^2$ = uniqueness; error = 1-reliability estimates from McGrew, LaForte, & Schrank (2014); $s^2 = u^2$-error; $\omega$ = omega; $\omega_h$ = omega-hierarchical.

**Table 3**

*Standardized Loading Coefficients for Bifactor Model of the Woodcock-Johnson Tests of Achievement-Fourth Edition (Ages 14-19)*

| | Factor | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Test | ACH-*g* | Grw | Gq | Gs | Gc | $h^2$ | $u^2$ | Error | $s^2$ |
| Letter-Word Identification | .842 | .309 | | | | .804 | .196 | .060 | .136 |
| Spelling | .853 | .149 | | | | .750 | .250 | .080 | .170 |
| Passage Comprehension | .796 | .203 | | | | .675 | .325 | .110 | .215 |
| Writing Samples | .609 | .314 | | | | .469 | .531 | .100 | .431 |
| Word Attack | .639 | .611 | | | | .782 | .218 | .100 | .118 |
| Oral Reading | .759 | .163 | | | | .603 | .397 | .040 | .357 |
| Reading Recall | .569 | .109 | | | | .336 | .664 | .080 | .584 |
| Editing | .848 | .006 | | | | .719 | .281 | .090 | .191 |
| Spelling of Sounds | .633 | .380 | | | | .545 | .455 | .120 | .335 |
| Applied Problems | .749 | | .522 | | | .833 | .167 | .080 | .087 |
| Calculation | .768 | | .383 | | | .737 | .263 | .070 | .193 |
| Number Matrices | .622 | | .315 | | | .486 | .514 | .080 | .434 |
| Sentence Reading Fluency | .740 | | | .478 | | .776 | .224 | .060 | .164 |
| Math Fact Fluency | .625 | | | .385 | | .539 | .461 | .040 | .421 |
| Sentence Writing Fluency | .690 | | | .388 | | .627 | .373 | .200 | .173 |
| Word Reading Fluency | .604 | | | .592 | | .715 | .285 | .080 | .205 |
| Reading Vocabulary | .848 | | | | .208 | .762 | .238 | .120 | .118 |
| Science | .650 | | | | .527 | .700 | .300 | .160 | .140 |
| Social Studies | .661 | | | | .537 | .725 | .275 | .130 | .145 |
| Humanities | .630 | | | | .519 | .666 | .334 | .130 | .204 |
| Total Variance (%) | 50.8 | 4.1 | 2.6 | 4.4 | 4.4 | 66.2 | 33.8 | 9.7 | 24.1 |
| Common Variance (%) | 76.7 | 6.1 | 3.9 | 6.6 | 6.6 | | | | |
| $\omega$ | .969 | .935 | .865 | .886 | .906 | | | | |
| $\omega_h$ | .910 | .098 | .212 | .288 | .264 | | | | |

*Note*. ACH-*g* = general achievement, Grw = Reading/Writing, Gq = Quantitative Reasoning, Gc = Academic Knowledge/Crystallized Ability, Gs = Academic Fluency/Processing Speed. $h^2$ = communality; $u^2$ = uniqueness; error = 1-reliability estimates from McGrew, LaForte, & Schrank (2014); $s^2$ = $u^2$-error; $\omega$ = omega; $\omega_h$ = omega-hierarchical.

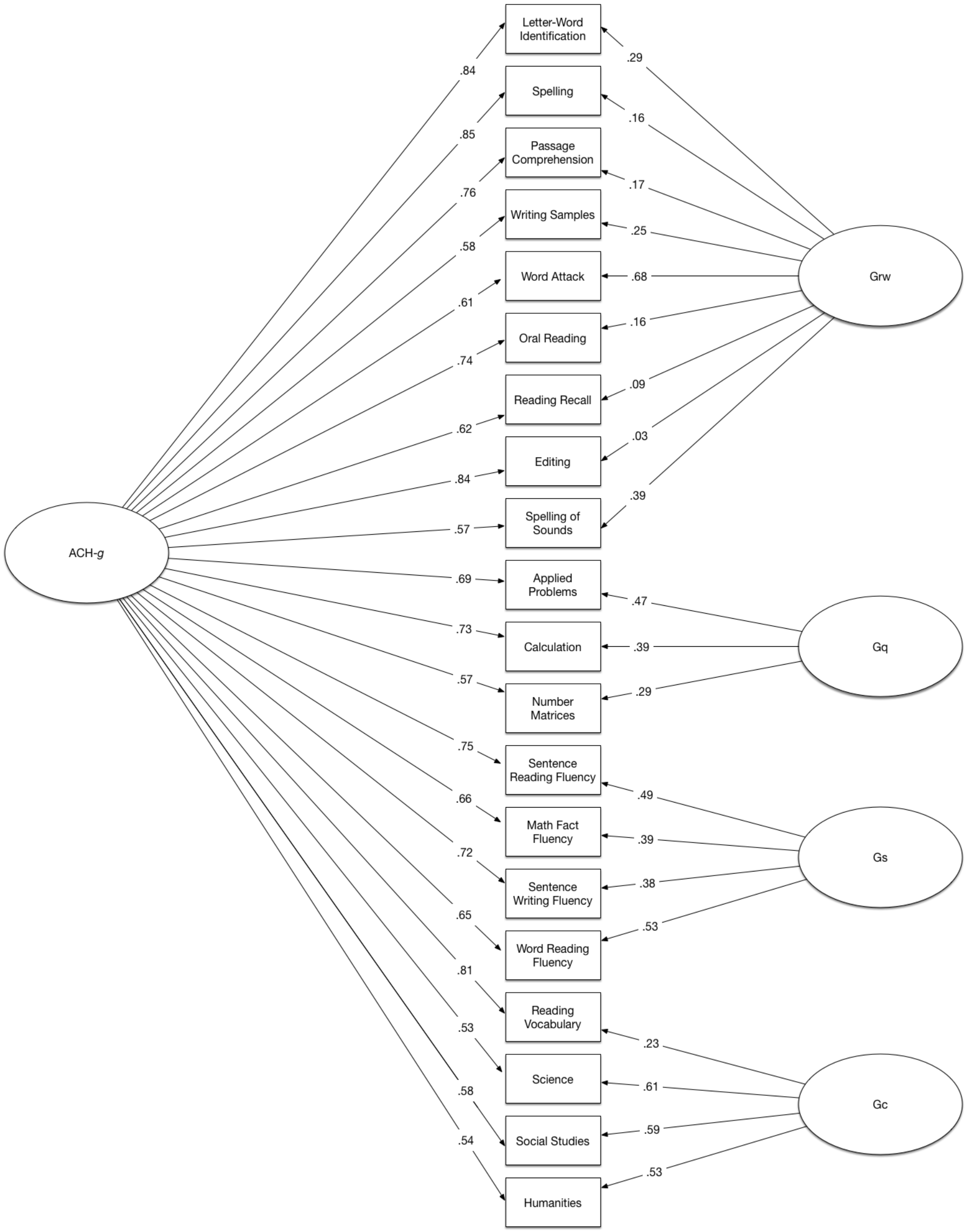


***Figure 1.*** Bifactor model for the Woodcock-Johnson Tests of Achievement-Fourth Edition (WJ-IV ACH) for ages 9-13. ACH-*g* = general achievement, Grw = Reading/Writing, Gq = Quantitative Reasoning, Gc = Academic Knowledge/Crystallized Ability, Gs = Academic Fluency/Processing Speed. Latent factor, test, and correlated test residuals were omitted for parsimony.

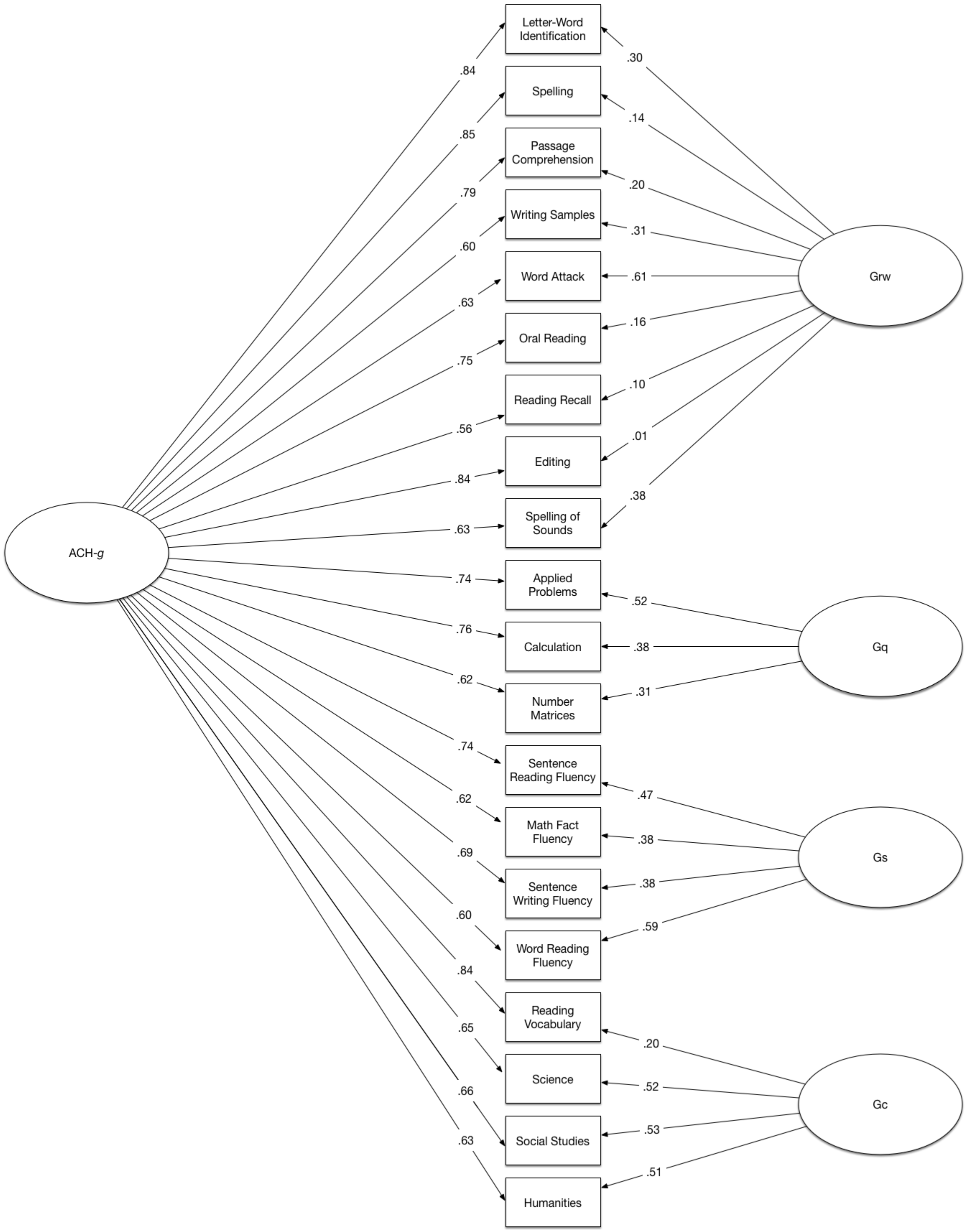


***Figure 2.*** Bifactor model for the Woodcock-Johnson Tests of Achievement-Fourth Edition (WJ-IV ACH) for ages 14-19. ACH-*g* = general achievement, Grw = Reading/Writing, Gq = Quantitative Reasoning, Gc = Academic Knowledge/Crystallized Ability, Gs = Academic Fluency/Processing Speed. Latent factor, test, and correlated test residuals were omitted for parsimony.