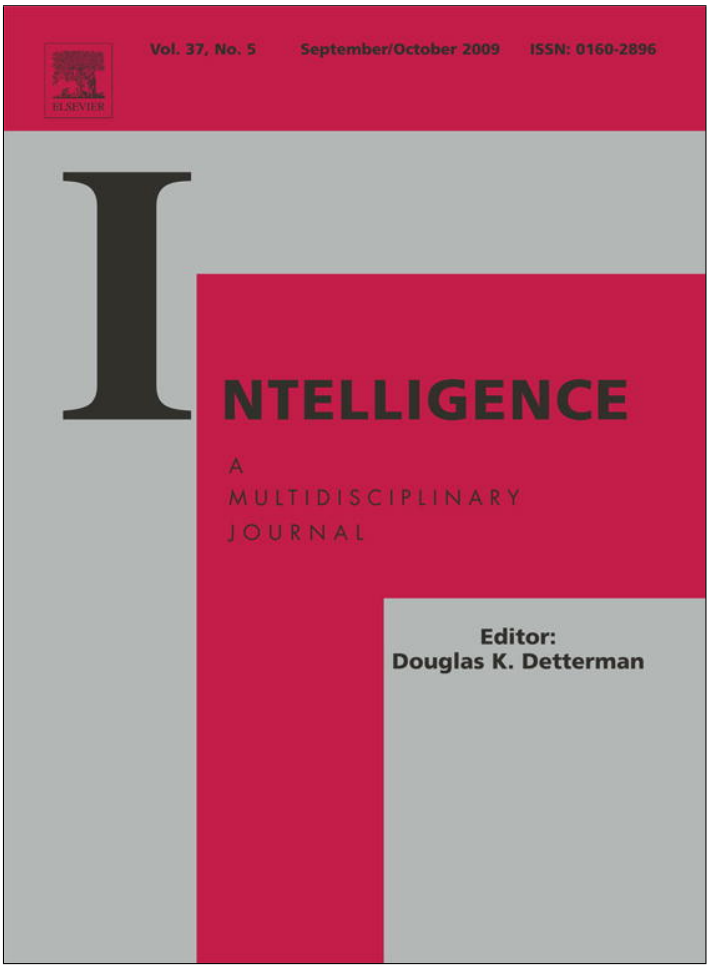
Provided for non-commercial research and education use.
Not for reproduction, distribution or commercial use.



This article appeared in a journal published by Elsevier. The attached copy is furnished to the author for internal non-commercial research and education use, including for instruction at the authors institution and sharing with colleagues.

Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier's archiving and manuscript policies are encouraged to visit:

http://www.elsevier.com/copyright

Author's personal copy

Intelligence 37 (2009) 453–465

Contents lists available at ScienceDirect

Intelligence

The dependability of general-factor loadings: The effects of factor-extraction methods, test battery composition, test battery size, and their interactions

Randy G. Floyd [a,*], Elizabeth I. Shands [a], Fawziya A. Rafael [a], Renee Bergeron [b], Kevin S. McGrew [c]

[a] The University of Memphis, United States
[b] Hawaii Department of Education, United States
[c] Woodcock-Munoz Foundation, University of Minnesota, United States

ARTICLE INFO

Article history:
Received 7 June 2007
Received in revised form 19 May 2009
Accepted 21 May 2009
Available online 11 June 2009

Keywords:
g factor
g loadings
Factor analysis
Generalizability theory
Dependability

ABSTRACT

To understand the extent to which the general-factor loadings of tests are inherent in their characteristics or due to the sampling of tests, the number of tests in the correlation matrix, and the factor-extraction methods used to obtain them, test scores from a large sample of young adults were inserted into independent and overlapping batteries of varying sizes. Principal factors analysis, maximum-likelihood estimation, and principal components analysis yielded general-factor loadings for each test. Generalizability theory analyses revealed that the characteristics of the tests consistently contributed the largest percentage of variance. Variance attributable to the factor-extraction method and its interactions was sizeable when principal components analysis was included in the analysis but negligible when it was excluded. Psychometric sampling error produced sizeable variance components in some analyses, and its effects were magnified when test batteries diminished in size. When results from principal components analysis were excluded and when the effects of psychometric sampling error were reduced, general-factor loadings were highly dependable across varying conditions.

© 2009 Elsevier Inc. All rights reserved.

1. Introduction

Many modern psychometric theories of intelligence converge in agreement that the *general factor* (a.k.a., psychometric g) meaningfully represents the majority of the positive relations among specific measures of human cognitive abilities, such as scores from the tests of intelligence test batteries (Carroll, 1993; Jensen, 1998; Spearman, 1927; Sternberg & Grigorenko, 2002). Despite this general agreement, several prominent criticisms have been levied against the construct validity of the general factor. One criticism is that the general factor is dependent on the factor-analytic methods used to extract it from a matrix of correlations, and another criticism is that the general factor is dependent on the measures used to operationalize it. These criticisms seem to have become part of virulent memes that pervade the minds of many professionals and consumers of tests results.

1.1 Factor-extraction methods

Gould (1981, 1994) asserted that the type of factor-analytic methods used to extract the general factor from a matrix of correlations affected the identification of the general factor to the extent that it undermined its meaningfulness. In fact, Gould concluded that the general factor is a statistical artifact with no representation in reality. A key target of his criticism was the factor-extraction method principal components analysis, which analyzes all variance across scores (including error variance). Snook and Gorsuch (1989), B. Thompson (2004), and others have demonstrated that principal components analysis tends to produce inflated parameter estimates. In addition, principal components analysis can produce a first component (designed to represent the general factor) with uniformly positive loadings from constituent tests when correlations among at least some of those tests are weak and not significantly different than 0 (Jensen & Weng, 1994). Despite these criticisms of principal components analysis, when an appropriate correlation matrix

* Corresponding author.
E-mail address: rgfloyd@memphis.edu (R.G. Floyd).

0160-2896/$ – see front matter © 2009 Elsevier Inc. All rights reserved.
doi:10.1016/j.intell.2009.05.003

Author's personal copy

is analyzed, values stemming from varying factor-extraction methods appear to be remarkably consistent—even when principal components analysis is used to extract the general factor. For example, Ree and Earles (1991) demonstrated both (a) correlations between 14 general component scores and factor scores from principal components analysis, principal factors analysis, and hierarchical factor analysis derived for each person and also (b) coefficients of congruence from general-factor loadings derived from varying factor-extraction methods that were very strong and near unity. Jensen and Weng (1994) also supported and extended these results through their comparisons of results from 6 to 10 different methods of factor analysis, including principal factors analysis, hierarchical exploratory factor analysis, and confirmatory factor analysis, using artificial correlation matrices and an archival correlation matrix. Their results indicated (a) very high correlations between factor scores between general factors, (b) very high congruence coefficients between general factor loadings yielded by different methods of analysis and the "true" general factor loadings used to derive the artificial correlation matrixes, and (c) consistency in the percentage of variance attributable to the general factor across methods. For instance, the Spearman correlations between general-factor loadings from 24 tests obtained using 10 different factor-extraction methods ranged from .79 to 1.0 ($M = .91$). Although general consistency across these factor-extraction methods is apparent, variation across methods is also evident.

1.2. Test battery composition and test battery size

It is sometimes argued that the general factor is dependent on the measures used to operationalize it, and it seems rational to argue, for example, that scores from a battery including a preponderance of tests of language-based abilities, when entered into factor analysis, would yield a very different general factor than when analyzing scores from a battery including a preponderance of tests of visualization abilities. Such criticism has often been levied by the most vocal opponent of the interpretation of the general factor in recent decades, John Horn (Horn, 1985, 1989; Horn & Blankson, 2005; Horn & McArdle, 2007). Consistent with this criticism, Horn has referred to the general factor and its related scores as conglomerates, mixtures measures, and hodgepodges of distinct abilities. There is some evidence of the effect of the test battery composition on the identification of the general factor. Based on Carroll's (1993) re-analysis of more than 460 data sets, he offered that "the G factor for a given data set is dependent on what lower-order factors or variables are loaded on it. One could say that a higher-order factor is 'colored' or 'flavored' by its ingredients" (p. 596). In a similar manner, Jensen and Weng (1994) conveyed, "Just as there is sampling error with respect to statistical parameters, there is *psychometric sampling error* [emphasis added] with respect to *g*, because the universe of all possible mental tests is not perfectly sampled by any limited set of tests" (p. 236).

Several studies have directly investigated this potential criticism that the general factor is dependent on the measures used to operationalize it. For example, Thorndike (1987) examined the identification of the general factor when it is formed from different samples of test scores. He employed data from 65 tests from the Army Air Forces Aviation Aircrew Classification Battery. From this battery, 48 tests were divided into six batteries of 8 tests each. The remaining 17 "probe" tests were inserted one at a time into each of the six batteries, and the general-factor loadings for each of the 17 probe tests were obtained. The median Pearson product–moment correlation coefficient between general-factor loadings across analyses using the six batteries was strong (.85). The range of correlations was .52 to .94, and two correlations were lower than .70. The average standard deviation of the general-factor loadings for the 17 tests was .07 (range = .04 to .14) across the six batteries. Based on Thorndike's results, it appears that the magnitude of the tests' loadings on the general factor is determined largely by the characteristics of the tests, rather than by characteristics of the test batteries in which they are inserted, but the influence of psychometric sampling error is apparent in the varying general-factor loadings across the test batteries.

More recently, others have examined the relations between second-order general factors extracted from varying test batteries using confirmatory factor-analytic methods and maximum-likelihood estimation, and they have demonstrated relations between these general factors that are consistently very near unity. Keith, Kranzler, and Flanagan (2001) produced a correlation of .98 between general factors derived from scores from two individually administered intelligence test batteries formed by 12 and 18 tests. Johnson, Bouchard, Krueger, McGue, and Gottesman (2004) produced correlations of .99, .99, and 1.0 between general factors formed from each of three test batteries formed by 11 to 17 tests. Most recently, Johnson, te Nijenhuis, and Bouchard (2008) produced correlations from .77 to 1.0 between general factors from each of five test batteries formed by 4 to 13 tests. Of these correlations, 7 of 10 were .95 or higher.

These results indicate similar identification of the general factor across independent test batteries. However, it is evident that disproportionate sampling of tests, biased toward specific abilities, may not allow specific variances to "average out" and for common variance, attributable to the general factor, to remain as the primary source of variance. It is logical that, as the number of test scores included in the factor analysis diminishes, the greater the effects that psychometric sampling error will have on the general factor and resulting scores. For example, in Johnson et al. (2008), the general factor formed from only 4 tests demonstrated notably lower correlations with the other general factors ($M = .85$) than did all of the other general factors with each other ($M = .95$). In addition, as indicated by the Wilks theorem (1938), as the number of test scores included in the factor analysis increases, the relations between the general factors derived from independent test scores will be strengthened (Jensen & Weng, 1994).

1.3. Purpose of the study

We sought to understand better the strength of and interactions between the effects of the factor-extraction method, the composition of the battery, and an understudied influence, the number of tests in the battery, on the identification of the general factor as well as to determine how these effects compare to differences in characteristics of

Author's personal copy

the tests under study. Like Jensen and Weng (1994) and Thorndike (1987), we chose to focus on these effects as they are manifested in the general-factor loadings from test scores for three reasons. First, these general-factor loadings are commonly reported in the literature, and second, there are established labels to guide practitioners' interpretation of these values (Floyd, McGrew, Barry, Rafael, & Rogers, 2009). In addition, and more importantly, when tests producing general-factor loadings are sufficiently diverse in terms of content, operations, stimulus input, and modes of response, as well as variable in reliability, the general-factor loadings are likely to vary notably due to the tests' characteristics, which can be contrasted with effects from the other potential influences.

To reach these goals, we implemented several innovations. First, to promote optimal measurement of the general factor and optimal psychometric sampling, we ensured that the probe tests scores stemmed from tests that were sufficiently diverse (see Jensen & Weng, 1994). We included probe tests from the individually administered cognitive ability test battery, the Woodcock–Johnson III (WJ III; Woodcock, McGrew, & Mather, 2001; Woodcock, McGrew, Mather, & Schrank, 2003), that ranged substantially across relevant characteristics. Likewise, we selected scores from our probe tests that had been analyzed using confirmatory factor analysis in several published sources (Floyd, Keith, Taub, & McGrew, 2007; McGrew & Woodcock, 2001; Phelps, McGrew, Knopik, & Ford, 2005; Taub & McGrew, 2004). The results of these prior confirmatory factor analyses collectively support the fact that the probe tests selected for the current study represent 7 of the approximately 10 broad (stratum II) abilities identified in the Cattell–Horn–Carroll theory: Comprehension–Knowledge, Long-Term Storage and Retrieval, Visual Processing, Auditory Processing, Fluid Reasoning, Processing Speed, and Short-Term Memory (see McGrew, 2009).

Second, we analyzed data from a large, nationally representative sample of young adults to prevent restricted population sampling from attenuating the relations we identify (Jensen, 1998). Third, we examined the effects of three factor-extraction methods on general-factor loadings. We selected (a) principal components analysis, (b) principal factors analysis, and (c) maximum-likelihood estimation, favored by many using confirmatory factor analysis. We believed that these methods would demonstrate some sensitivity to psychometric sampling error (Jensen & Weng, 1994). Fourth, we not only completed a partial replication of Thorndike's (1987) analysis using independent batteries to form the general factors, but we also constructed a large number of partially overlapping batteries, comprised of randomly selected tests, to form general factors. We reasoned that a large number of samples of test batteries (a) would provide results that are more reliable than those from only a few independent batteries and (b) would reduce the undermining effects of psychometric sampling error due to "biased" sampling of tests for the independent batteries. Fifth, we examined the effects of the number of tests used in identification of the general factor—beginning with batteries comprising 7 or 8 tests and decomposing these batteries to 4-test batteries and 2-test batteries.

Finally, analysis of the resulting general-factor loadings was conducted using Generalizability theory to examine the variance in general-factor loadings attributable to four systematic sources–(a) the characteristics of the probe test, (b) the factor-extraction method, (c) the composition of the test battery, and (d) the size of the test battery–as well as their interactions. Generalizability theory extends the notion of measurement error beyond that of classical test theory and offers a means to assess concurrently multiple sources of variance (Cronbach, Gleser, Nanda, & Rajaratnam, 1972; Shavelson & Webb, 1991). In addition to estimating the proportion of variance that can be accounted for by various sources (i.e., variance components), Generalizability theory yields dependability coefficients that are analogous to reliability coefficients.

2. Method

2.1. Participants

Participants were drawn from the standardization sample of the WJ III (Woodcock et al., 2001). The WJ III standardization sample was constructed using a stratified sampling plan that controlled for 10 individual variables (e.g., race, sex, educational level, occupational status) and community variables (e.g., community size, community socio-economic status) as described by the United States Census projections for the year 2000 (McGrew & Woodcock, 2001). Participants ages 20 to 39 ($n = 1409$) from the standardization sample were included. Participants in this age range were selected in order to parallel samples from the Ree and Earles (1991), Thorndike (1987), Johnson et al. (2004), and Johnson et al. (2008) studies.

2.2. Measures

All measures were drawn from the WJ III. The development, standardization, and psychometric properties of these test batteries have generally been evaluated favorably by independent reviewers (Bradley-Johnson, Morgan, & Nutkins, 2004; Cizek, 2003; Sares, 2005; Thompson, 2005). This study employed scores from 19 tests from the WJ III Tests of Cognitive Abilities, 21 tests from the WJ III Tests of Achievement, and 5 tests and 1 special composite from the WJ III Diagnostic Supplement (Woodcock, McGrew, Mather, & Schrank, 2003). The special composite is Numerical Reasoning, which represents performance on the Number Series and Number Matrices tests.[1] In contrast, 3 tests were omitted from analyses because they are derivatives of other tests included in the analyses. These tests were Visual–Auditory Learning: Delayed, Story Recall: Delayed, and Memory for Names: Delayed. Had these tests been included, shared residual variance across the pairs of tests would have produced spurious results in the factor analyses.

McGrew and Woodcock (2001) and Woodcock et al. (2003) reported estimates of reliability and evidence of validity for these 45 tests and 1 composite. Rasch analysis was used to calculate the reliability of speeded tests (i.e., Cross Out, Decision Speed, Visual Matching, Math Fluency, Pair Cancellation, Rapid Picture Naming, Reading Fluency, Retrie-

[1] At the time the data set employed in this study was constructed, these tests produced only a single composite score (Woodcock et al., 2003).

Author's personal copy

val Fluency, Writing Fluency in this study) and tests that employed multiple-point scored items (i.e., Picture Recognition, Planning, Retrieval Fluency, Spatial Relations, Story Recall, Spelling of Sounds, and Writing Samples in this study). Split-half procedures were used for the remaining tests. Composite reliabilities were calculated based on the obtained reliabilities for their component tests. As evident in Table 1, all but 12 measures demonstrated median reliability

Table 1
Test characteristics, reliability, and descriptive statistics for the Woodcock–Johnson III Tests.

WJ III test	Test type	Broad ability	Reliability	Descriptive statistics	
				M	*SD*
Verbal Comprehension	P	Gc	.94	100.31	13.86
Visual–Auditory Learning	P	Glr	.92	100.01	14.45
Spatial Relations	P	Gv	.80	99.60	14.45
Sound Blending	P	Ga	.92	100.33	14.55
Concept Formation	P	Gf	.95	100.37	13.98
Visual Matching	P	Gs	.91	100.72	12.73
Numbers Reversed	P	Gsm	.89	100.53	14.20
Analysis–Synthesis	B	Gf	.92	100.85	13.85
Academic Knowledge	B	Gc	.90	102.15	12.05
Applied Problems	B	Gq, Gf, Gc	.94	101.44	13.01
Auditory Attention	B	Ga	.86	100.78	11.30
Auditory Working Memory	B	Gsm	.82	102.40	12.46
Block Rotation	B	Gv	.84	100.85	12.90
Calculation	B	Gq	.87	101.27	13.09
Cross Out	B	Gs, Gv	.72	101.62	12.70
Decision Speed	B	Gs	.86	100.67	13.90
Editing	B	Grw	.92	99.87	14.01
General Information	B	Gc	.94	100.47	13.76
Incomplete Words	B	Ga	.85	101.26	13.48
Letter–Word Identification	B	Grw	.93	100.78	12.34
Math Fluency	B	Gq, Gs	.90	101.21	14.10
Memory for Names	B	Glr	.91	100.70	14.95
Memory for Sentences	B	Gsm, Gc	.93	100.20	13.06
Memory for Words	B	Gsm	.81	99.97	14.19
Numerical Reasoning	B	Gf	.95	101.47	12.58
Oral Comprehension	B	Gc	.90	99.76	14.66
Pair Cancellation	B	Gs	.84	102.52	14.35
Passage Comprehension	B	Grw, Gc	.82	101.18	12.82
Picture Recognition	B	Gv	.77	100.50	14.10
Picture Vocabulary	B	Gc	.87	99.39	15.00
Planning	B	Gv	.75	99.88	11.33
Quantitative Concepts	B	Gq, Gf	.93	100.93	13.20
Rapid Picture Naming	B	Glr, Gs	.97	102.14	13.80
Reading Fluency	B	Grw, Gs	.91	100.03	16.02
Reading Vocabulary	B	Grw, Gc	.92	100.35	14.96
Retrieval Fluency	B	Glr, Gs	.90	101.15	12.20
Sound Awareness	B	Ga	.74	100.30	13.67
Sound Patterns	B	Ga	.95	100.43	14.26
Spelling	B	Grw	.93	100.81	12.47
Spelling of Sounds	B	Grw, Ga	.71	102.06	12.01
Story Recall	B	Gc, Glr	.89	100.33	13.75
Understanding Directions	B	Gc, Glr, Gsm	.84	96.21	15.14
Visual Closure	B	Gv	.94	100.13	14.11
Word Attack	B	Grw, Ga	.88	100.66	12.92
Writing Fluency	B	Grw, Gs	.85	100.22	13.96
Writing Samples	B	Grw	.88	100.52	11.19

Note. P = Probe test entered one at a time into batteries. B = Test forming batteries. Gc = Comprehension–Knowledge; Glr = Long-Term Storage and Retrieval; Gv = Visual Processing; Ga = Auditory Processing; Gf = Fluid Reasoning; Gs = Processing Speed; Gsm = Short-Term Memory; Gq = Quantitative Knowledge; and Grw = Reading and Writing. Broad ability classifications stem from the confirmatory factor analysis reported in McGrew and Woodcock (2001), and labels for these factors stem from McGrew (2009).

ceptions include only Cross Out (*Mdn* = .72), Picture Recognition (*Mdn* = .77), Planning (*Mdn* = .75), Sound Awareness (*Mdn* = .74), and Spelling of Sounds (*Mdn* = .71).

2.3. Procedures

From the 45 WJ III tests and the composite, 7 probe tests were chosen to provide variables for which general-factor loadings would be obtained across varying conditions: Verbal Comprehension, Visual–Auditory Learning, Spatial Relations, Sound Blending, Concept Formation, Visual Matching, and Numbers Reversed. As evident in Table 1, all of these tests have demonstrated median reliability coefficients of .80 or higher, and each test measures a different broad cognitive ability in accordance with the Cattell–Horn–Carroll theory (McGrew & Woodcock, 2001; Taub & McGrew, 2004). We first formed independent batteries comprising 8 tests each from the remaining 39 tests from the WJ III. Four batteries were formed of 8 randomly selected tests, and a fifth battery was formed from the remaining 7 tests (see Table 2). We also selected, at random (with replacement), 8 (of the 39) tests at a time to form 30 overlapping batteries of tests (see Table 3). From these 30 batteries of 8 tests, we also formed 30 batteries of 4 tests and 30 batteries of 2 tests by first randomly selecting 4 of the 8 tests to be omitted and then randomly selecting 2 of the remaining 4 tests to be omitted (see italicized tests and underlined tests in Table 3).

Using correlation matrixes as input (presented in McGrew & Woodcock, 2001), variables from each of the probe tests were inserted one at a time into each of the 95 batteries of 8, 4, or 2 tests, and principal factors analysis, maximum-likelihood estimation, and principal components analysis were conducted. A single factor was extracted in all cases to represent the general factor. The correlations between scores from each of the probe tests and the general factor, the general-factor loadings, were the primary focus of the analysis. These general-factor loadings for the probe tests were submitted to Generalizability theory analysis to examine their consistency and dependability. Variance components were computed using SPSS 12.0, and dependability coefficients (a.k.a., phi coefficients) were calculated to provide overall indexes of dependability (Brennan, 2001; Shavelson & Webb, 1991). The variance estimate attributable to differences across the probe tests was considered universe-score variance; it was used as the numerator in the formula to calculate the dependability coefficients. The variance estimates attributable to the test battery, to the number of tests, to the factor-extraction method, to all interactions, and to residual (i.e., unexplained) variance, when each was divided by the number of variations associated with each facet, were considered error variance. The denominator of the formula consisted of the sum of the universe-score variance and error variance.

3. Results

Means and standard deviations for all WJ III test scores are shown in the right side of Table 1. Preliminary data analyses were conducted to ensure that the assumption of the factorability of the correlation matrixes was not violated, and the coefficients of congruence, which represent the relations

Author's personal copy

Table 2
Independent batteries selected randomly from the Woodcock–Johnson III.

1	2	3	4	5
Auditory Working Memory	*Pair Cancellation*	*Sound Awareness*	Memory for Names	*Analysis–Synthesis*
Incomplete Words	Memory for Words	Rapid Picture Naming	Numerical Reasoning	*Story Recall*
Decision Speed	Memory for Sentences	Visual Closure	*General Information*	Picture Recognition
Cross Out	Sound Patterns	*Block Rotation*	Picture Vocabulary	Academic Knowledge
Passage Comprehension	Auditory Attention	Spelling	*Letter–Word Identification*	Oral Comprehension
Reading Fluency	*Retrieval Fluency*	Editing	Word Attack	*Reading Vocabulary*
Writing Samples	*Planning*	*Applied Problems*	*Quantitative Concepts*	*Spelling of Sounds*
Writing Fluency	*Understanding Directions*	*Math Fluency*	*Calculation*	

Note. Italics indicates tests included in the 4-test batteries, and underlining indicates tests included in the 2-test batteries.

between the extracted general factor and the estimated true general factor, was calculated for each test battery.[2]

3.1. Independent batteries

For each analysis, Bartlett's test of sphericity was statistically significant ($p<.05$). For 8-test batteries and 4-test batteries, the Kaiser–Meyer–Olkin (KMO) measure of sampling adequacy was .60 or higher for each analysis, $M=.86$, $SD=.04$ and $M=.77$, $SD=.04$, respectively, but 10 analyses (29%) produced a KMO measure less than .60 using 2-test batteries ($M=.62$, $SD=.03$). The general factor accounted for the following percentages of the variance, on average, for each method: 44.70%, principal factors analysis; 45.03%, maximum-likelihood estimation; and 55.72%, principal components analysis. The general factor accounted for an average of 46.13% of variance across 8-test batteries, 48.65% across 4-test batteries, and 50.23% across 2-test batteries. The coefficients of congruence were .93 on average ($SD=.02$, range = .89 to .96) for 8-test batteries, .89 on average ($SD=.04$, range = .83 to 1.00) for 4-test batteries, and .82 on average ($SD=.06$, range = .66 to .91) for 2-test batteries.

Table 4 presents the general-factor loadings for the probe tests when entered into the five independent batteries of varying sizes and analyzed using the varying factor-extraction methods. In one instance, a general-factor loading was not obtained due to its communality exceeding 1.0. Following rules of thumb, general-factor loadings of .70 and higher indicate *high* measures of the general factor, those .69 to .50 indicate *medium* measures of the general factor, and those less than .50 indicate *low* measures of the general factor (McGrew & Flanagan, 1998, cf. Kaufman, 1994). Across conditions, the mean general-factor loadings for Verbal Comprehension and Concept Formation were generally high, whereas the mean general-factor loadings for Visual–Auditory Learning, Spatial Relations, Sound Blending, Visual Matching, and Numbers Reversed were medium.

Components for the main sources of variance in general-factor loadings, their interactions, and unexplained influences are presented in the second column of Table 5. In the primary analysis using the independent batteries, variance attributable to the probe tests constituted 33% of the total variance. When facets of error variance were considered, the main effect of factor-extraction method constituted 15% of variance. This effect is evident in the difference in mean values across the factor-extraction methods; principal components analysis typically produced higher general-factor loadings than both other methods. The interaction between the factor-extraction method and the probe tests constituted only 2% of variance, its interaction with the test battery constituted negligible variance, and its interaction with the number of tests in the test battery constituted only 2% of variance. The main effect of the test battery constituted only 3% of variance, and its interaction with the number of tests in the battery constituted only 3% of variance. However, its interaction with the probe tests constituted 18% of variance, which was the highest of any error variance component. This interaction is evident in the varying standard deviation values for each test within batteries of the same size. For example, Verbal Comprehension and Visual Matching demonstrated the greatest variation across these independent batteries, and Visual–Auditory Learning and Concept Formation seemed to demonstrate the least. The main effect of the number of tests in the battery constituted negligible variance, but its interaction with the probe tests contributed 5% of variance. The three-way interaction that contributed the most variance was the interaction between the probe test, the test battery, and the number of tests—totaling 12% of variance. Other three-way interactions constituted 2% of variance or less. The resulting dependability coefficient was .72, which indicates moderate dependability of general-factor loadings across different batteries of tests, different methods of factor analysis, and batteries of different sizes.

Based on consideration of the weaknesses inherent in principal components analysis and consideration of the range of reliability of the probe tests, we conducted a follow-up Generalizability theory analysis in which the general-factor loadings from this method were omitted. Thus, only general-factor loadings from principal factors analysis and maximum-likelihood estimation were included. The results are presented in the center column of Table 5. As is evident, when general-factor loadings from principal components analysis were excluded, the variance attributed to the factor-extraction method and its two-way interactions became negligible. As a

[2] The formula for this coefficient of congruence (r_{st}) is

$$r_{st} = \sqrt{\frac{n}{(n-1)}\left(1-\left(\frac{1}{\lambda}\right)\right)}$$

where n represents the number of tests included in a principal component analysis, and λ is the eigenvalue from the principal components analysis in which a single factor is extracted. The correlation r_{st} is the square root of the reliability of the factor as indicated by coefficient alpha in Harman's (1976) formula 11.29, p. 231. See also p. 236 of Jensen and Weng (1994). We appreciate John Kranzler for providing us information about this formula.

Author's personal copy

Table 3
Overlapping batteries selected randomly from the Woodcock–Johnson III.

1	2	3	4	5
Pair Cancellation	*Memory for Words*	*Memory for Words*	*Memory for Sentences*	Auditory Working Mem
Sound Awareness	Memory for Sentences	*Analysis–Synthesis*	Auditory Attention	*Analysis–Synthesis*
Incomplete Words	*Auditory Attention*	Decision Speed	*Numerical Reasoning*	Retrieval Fluency
Block Rotation	Numerical Reasoning	*Visual Closure*	Rapid Picture Naming	Cross Out
Understanding Directions	*Story Recall*	Block Rotation	Letter–Word Identification	Planning
Oral Comprehension	Academic Knowledge	Understanding Directions	Spelling of Sounds	*Picture Vocabulary*
Picture Vocabulary	Letter–Word Identification	*Word Attack*	*Calculation*	*Letter–Word Identification*
Passage Comprehension	*Editing*	Applied Problems	*Math Fluency*	*Spelling of Sounds*
6	**7**	**8**	**9**	**10**
Pair Cancellation	*Memory for Words*	*Auditory Working Mem*	*Auditory Attention*	*Auditory Attention*
Numerical Reasoning	Retrieval Fluency	Numerical Reasoning	Numerical Reasoning	*Story Recall*
Planning	Rapid Picture Naming	*Rapid Picture Naming*	*Analysis–Synthesis*	Picture Recognition
Academic Knowledge	*Cross Out*	Cross Out	Rapid Picture Naming	Oral Comprehension
Oral Comprehension	*Picture Recognition*	Picture Recognition	Decision Speed	*Writing Fluency*
Word Attack	Letter–Word Identification	*Word Attack*	*General Information*	Spelling of Sounds
Spelling of Sounds	Reading Fluency	*Spelling of Sounds*	Word Attack	*Quantitative Concepts*
Applied Problems	*Reading Vocabulary*	Math Fluency	*Calculation*	Math Fluency
11	**12**	**13**	**14**	**15**
Pair Cancellation	Memory for Sentences	Auditory Attention	*Pair Cancellation*	Memory for Words
Sound Awareness	*Incomplete Words*	*Memory for Names*	Sound Awareness	*Rapid Picture Naming*
Incomplete Words	Retrieval Fluency	Analysis–Synthesis	*Rapid Picture Naming*	*Cross Out*
General Information	*Cross Out*	Story Recall	*Planning*	*Picture Recognition*
Picture Vocabulary	Block Rotation	*Cross Out*	Understanding Directions	Block Rotation
Writing Fluency	*Oral Comprehension*	General Information	*Passage Comprehension*	Picture Vocabulary
Editing	*Letter–Word Identification*	*Picture Vocabulary*	Applied Problems	Passage Comprehension
Applied Problems	Writing Samples	Letter–Word Identification	Calculation	*Calculation*
16	**17**	**18**	**19**	**20**
Retrieval Fluency	Rapid Picture Naming	*Auditory Attention*	Sound Awareness	Memory for Sentences
Rapid Picture Naming	Picture Recognition	*Incomplete Words*	*Rapid Picture Naming*	Auditory Attention
Cross Out	*Academic Knowledge*	Memory for Names	*Understanding Directions*	Sound Awareness
Understanding Directions	Passage Comprehension	*Planning*	*Passage Comprehension*	*Story Recall*
Picture Vocabulary	Word Attack	Oral Comprehension	*Word Attack*	*Understanding Directions*
Letter–Word Identification	*Reading Fluency*	Picture Vocabulary	Reading Fluency	*Reading Fluency*
Word Attack	*Spelling of Sounds*	*Word Attack*	Reading Vocabulary	*Reading Vocabulary*
Reading Fluency	*Editing*	Spelling	Applied Problems	Spelling
21	**22**	**23**	**24**	**25**
Auditory Working Mem	*Auditory Working Mem*	Story Recall	Pair Cancellation	Memory for Words
Auditory Attention	Sound Patterns	Retrieval Fluency	Auditory Attention	*Retrieval Fluency*
Sound Awareness	Auditory Attention	Decision Speed	*Sound Awareness*	*Rapid Picture Naming*
Retrieval Fluency	*Numerical Reasoning*	*Academic Knowledge*	*Analysis–Synthesis*	Decision Speed
Cross Out	*Analysis–Synthesis*	*Understanding Directions*	*Visual Closure*	*Visual Closure*
Understanding Directions	Picture Vocabulary	*Oral Comprehension*	Block Rotation	Academic Knowledge
Oral Comprehension	*Reading Vocabulary*	Letter–Word Identification	Picture Vocabulary	General Information
Letter–Word Identification	Editing	*Spelling of Sounds*	*Calculation*	*Calculation*
26	**27**	**28**	**29**	**30**
Memory for Words	*Pair Cancellation*	Memory for Words	*Pair Cancellation*	*Memory for Words*
Sound Patterns	*Memory for Sentences*	*Memory for Sentences*	Sound Awareness	*Memory for Sentences*
Auditory Attention	Auditory Working Mem	Incomplete Words	*Numerical Reasoning*	Sound Patterns
Sound Awareness	*Incomplete Words*	Rapid Picture Naming	Cross Out	*Retrieval Fluency*
Memory for Names	Cross Out	*Oral Comprehension*	Writing Samples	Cross Out
Numerical Reasoning	Visual Closure	*Reading Vocabulary*	*Spelling of Sounds*	Passage Comprehension
Academic Knowledge	Understanding Directions	*Writing Samples*	*Editing*	*Reading Fluency*
Spelling of Sounds	*Spelling*	Quantitative Concepts	Math Fluency	Writing Samples

Note. Italics indicates tests included in the 4-test batteries, and underlining indicates tests included in the 2-test batteries.

result, the variance attributable to the probe tests rose to 42%, and the dependability coefficient rose to .84. Error variance from the interaction between the test battery and the probe tests rose to 22%, and the interaction between the probe test, the test battery, and the number of tests rose to 20%.

3.2. Overlapping batteries

For each analysis, Bartlett's test of sphericity was statistically significant ($p<.05$). For 8-test batteries and 4-test batteries, the Kaiser–Meyer–Olkin (KMO) measure of

Author's personal copy

Table 4
General-factor loadings and summary statistics for each probe test when tests inserted one at a time into each independent test battery and varying factor-extraction methods applied.

Test battery	Probe test																				
	VC			VAL			SR			SB			CF			VM			NR		
	PF	ML	PC	PF	ML	PC	PF	ML	PC	PF	ML	PC	PF	ML	PC	PF	ML	PC	PF	ML	PC
8-test batteries																					
1	.79	.81	.81	.62	.61	.67	.56	.56	.62	.69	.69	.74	.69	.69	.73	.70	.71	.74	.61	.61	.67
2	.84	.84	.84	.65	.63	.71	.57	.54	.65	.72	.70	.77	.73	.72	.77	.57	.53	.65	.67	.66	.73
3	.86	.87	.86	.63	.62	.69	.57	.55	.64	.67	.67	.73	.70	.69	.75	.61	.59	.68	.65	.65	.71
4	.91	.97	.91	.64	.62	.69	.57	.56	.63	.63	.62	.68	.71	.70	.75	.47	.47	.53	.59	.59	.64
5	.93	.94	.92	.64	.59	.70	.53	.50	.60	.65	.63	.71	.69	.66	.75	.45	.42	.53	.56	.53	.64
M	.87	.88	.87	.64	.61	.69	.56	.54	.63	.67	.66	.72	.70	.69	.75	.56	.54	.63	.62	.61	.68
SD	.06	.07	.04	.01	.02	.02	.02	.03	.02	.03	.04	.03	.02	.02	.01	.10	.11	.09	.04	.05	.04
4-test batteries																					
1	.78	.83	.81	.65	.65	.73	.62	.61	.71	.67	.68	.74	.72	.73	.78	.78	.79	.82	.67	.69	.74
2	.78	.79	.83	.63	.60	.74	.58	.53	.71	.65	.63	.75	.72	.70	.80	.68	.69	.76	.61	.58	.73
3	.84	.85	.86	.66	.65	.75	.65	.63	.75	.67	.69	.76	.76	.75	.81	.63	.60	.73	.71	.70	.78
4	.89	.93	.90	.61	.60	.71	.56	.56	.67	.59	.61	.69	.70	.70	.77	.47	.46	.59	.59	.59	.69
5	.93	.93	.90	.66	.67	.74	.53	.53	.64	.67	.67	.75	.72	.74	.78	.46	.46	.58	.59	.59	.69
M	.84	.87	.86	.64	.63	.73	.59	.57	.70	.65	.66	.74	.72	.72	.79	.60	.60	.70	.63	.63	.73
SD	.07	.06	.04	.02	.03	.02	.05	.05	.04	.03	.03	.03	.02	.02	.02	.14	.14	.11	.05	.06	.04
2-test batteries																					
1	.86	.87	.88	.70	.70	.80	.69	.69	.79	.72	.72	.81	.78	.78	.83	.80	.80	.84	.65	.65	.78
2	.53	.53	.72	.69	.70	.77	.84	.86	.81	.64	.64	.76	.76	.78	.79	.81	.84	.85	.66	.67	.77
3	.78	.78	.87	.62	.62	.79	.54	.54	.75	.67	.67	.83	.72	.72	.82	.73	.73	.83	.77	.77	.84
4	–	.66	.81	.54	.54	.74	.50	.50	.71	–	.43	.64	.62	.62	.79	.43	.43	.65	.53	.53	.73
5	.94	.95	.92	.58	.58	.75	.47	.47	.69	.73	.73	.83	.63	.63	.78	.44	.44	.67	.59	.59	.76
M	.78	.76	.84	.63	.63	.77	.61	.61	.75	.69	.64	.77	.70	.71	.80	.64	.65	.77	.64	.64	.78
SD	.18	.17	.08	.07	.07	.03	.15	.16	.05	.04	.12	.08	.07	.08	.02	.19	.20	.10	.09	.09	.04

Note. VC = Verbal Comprehension, VAL = Visual–Auditory Learning, SR = Spatial Relations, SB = Sound Blending, CF = Concept Formation, VM = Visual Matching, NR = Numbers Reversed. PF = principal factors analysis, ML = maximum-likelihood estimation, PC = principal components analysis.

sampling adequacy was .60 or higher for each analysis, $M = .88$, $SD = .02$ and $M = .78$, $SD = .04$, respectively, but 26 analyses (12%) produced a KMO measure less than .60 using 2-test batteries ($M = .64$, $SD = .04$). The general factor accounted for the following percentages of the variance, on average, for each method: 44.49%, principal factors analysis; 44.41%, maximum-likelihood estimation; and 55.34%, principal components analysis. The general factor accounted for an average of 45.43% of variance across 8-test batteries, 47.26% across 4-test batteries, and 51.56% across 2-test batteries. The coefficients of congruence were .93 on average ($SD = .01$, range = .90 to .95) for 8-test batteries, .88 on average ($SD = .03$, range = .80 to .95) for 4-test batteries, and .83 on average ($SD = .05$, range = .69 to .97) for 2-test batteries.

Tables 6–8 presents the general-factor loadings for the probe tests inserted one at a time into each of the 30 overlapping batteries of varying sizes and analyzed using each the factor-analytic method. As evident in Table 8, two general-factor loadings were not obtained due to their communality exceeding 1.0. Across the 8-test, 4-test, and 2-test batteries, the mean general-factor loadings for the probe tests were generally in the same ranges as those from the analysis using independent batteries.

As evident on the fourth column of Table 5, primary analysis revealed that variance attributable to the probe tests constituted 56% of the total variance. When facets of error variance were considered, the main effect of factor-extraction method constituted 13% of variance. Its interaction with the probe tests constituted 4% of variance, its interaction with the test battery constituted negligible variance, and its interaction with the number of tests in the test battery constituted 3% of variance. These results are consistent with the primary analysis of the independent batteries. The main effect of the test battery constituted only 1% of variance, and its interaction with the number of tests in the battery constituted 4% of variance. Because of the broad sampling of tests in the batteries (as well as their overlapping nature), the interaction between the test battery and the probe tests constituted only 7% of variance, which is approximately half of the proportion of variance attributed to this interaction in the primary analysis of the independent batteries. The main effect of the number of tests in the battery and its interaction with the probe tests constituted generally negligible variance. Again, the three-way interaction that contributed the most variance was the interaction between the probe test, the test battery, and the number of tests—now totaling only 7% of variance. Other three-way interactions constituted 2% of variance or less. The resulting dependability coefficient was .89, which indicates strong dependability.

Results from follow-up analyses omitting the general-factor loadings from principal components analysis are presented in the far right column of Table 5. In a manner consistent with the follow-up analysis of the independent batteries, the variance attributed to the factor-extraction method and its interactions was reduced to negligible levels. The variance attributable to

Author's personal copy

Table 5
Variance component estimates and dependability coefficients for general-factor loadings by type of battery.

Facet	Independent batteries	Independent batteries (without PC)	Overlapping batteries	Overlapping batteries (without PC)
Probe Test	0.00470 (33%)	0.00621 (42%)	0.00911 (56%)	0.01224 (71%)
Factor-Extraction Method	0.00221 (15%)	0.00000 (0%)	0.00214 (13%)	0.00000 (0%)
Probe Test * Factor-Extraction Method	0.00027 (2%)	0.00000 [a] (0%)	0.00062 (4%)	0.00003 (0%)
Test Battery	0.00046 (3%)	0.00048 (3%)	0.00015 (1%)	0.00017 (1%)
Probe Test * Test Battery	0.00263 (18%)	0.00329 (22%)	0.00109 (7%)	0.00140 (8%)
Number of Tests	0.00000 [a] (0%)	0.00000 [a] (0%)	0.00006 (0%)	0.00000 [a] (0%)
Probe Test * Number of Tests	0.00066 (5%)	0.00048 (3%)	0.00008 (1%)	0.00003 (0%)
Factor-Extraction Method * Test Battery	0.00000 [a] (0%)	0.00000 [a] (0%)	0.00000 (0%)	0.00000 (0%)
Factor-Extraction Method * Number of Tests	0.00028 (2%)	0.00000 [a] (0%)	0.00042 (3%)	0.00001 (0%)
Test Battery * Number of Tests	0.00047 (3%)	0.000048 (3%)	0.00069 (4%)	0.00127 (7%)
Probe Test * Factor-Extraction Method * Test Battery	0.00007 (1%)	0.00000 [a] (0%)	0.00006 (0%)	0.00001 (0%)
Probe Test * Factor-Extraction Method * Number of Tests	0.00000 [a] (0%)	0.00006 (0%)	0.00006 (0%)	0.00004 (0%)
Probe Test * Test Battery * Number of Tests	0.00171 (12%)	0.00296 (20%)	0.00117 (7%)	0.00180 (11%)
Factor-Extraction Method * Test Battery * Number of Tests	0.00034 (2%)	0.00010 (1%)	0.00025 (2%)	0.00005 (0%)
Residual	0.00049 (3%)	0.00007 (0%)	0.00026 (2%)	0.00007 (0%)
Total	0.01429	0.01462	0.01617	0.01712
ϕ	.72	.84	.89	.99

Note. Proportion of total variance in parentheses. PC = principal components analysis. ϕ = dependability coefficient.
[a] Negative estimated variance components were set to zero (Brennan, 2001; Cronbach et al., 1972).

the probe tests rose to 71%, and the dependability coefficient rose to .99, which indicates very strong dependability of general-factor loadings across different batteries of tests, batteries of different sizes, and different methods of factor analysis. Variance from the interaction between the test battery and the probe tests was 8%, which is consistent with the initial analysis of the overlapping batteries, but the interaction between the test battery and the number of tests in the battery rose slightly to 7%. The interaction between the probe test, the test battery, and the number of tests rose slightly to 11%.

4. Discussion

We examined the strength of and interactions between the effects of the factor-extraction method, the composition of the battery, and the number of tests in the battery on the identification of the general factor via testing a number of "worse-case scenarios" in which tests were randomly chosen to form batteries, decomposed into diminutive batteries, and analyzed using the most dubious factor-extraction method, principal components analysis. Results demonstrated that characteristics of the tests, per se, accounted for the greatest percentage of variance (i.e., at least one-third to more than two-thirds of variance) across the variety of influences on the general-factor loadings we investigated. Thus, the results are consistent with Thorndike's (1987) conclusion that "the *g*-loading of a type of test task has substantial stability and is to a considerable extent determined by the characteristics of the test itself, rather than the context in which is appears" (p. 586). The results support a sizeable body of research that indicates that the characteristics of tests that determine their general-factor loadings are (a) their reliability and (b) their degree of complexity (Jensen, 1982, 1998). For example, it is well known that unreliability in scores limits their correlations with other variables; consequentially, those tests with higher reliabilities tend to have higher general-factor loadings, as is demonstrated in this study (see Table 1). Furthermore, tests that require more conscious mental manipulation correlate more highly with the general factor. Thus, tests with the highest reliability and that require the greatest number of mental manipulations, such as the WJ III Verbal Comprehension and Concept Formation probe tests, yielded scores that consistently demonstrated the highest general-factor loadings.

4.1. Factor-extraction method

Some have argued that the construct validity of the general factor is diminished substantially due to the variation found across methods used to extract it. Our results indicate that the factor-analytic method employed does not contribute overwhelming error variance across general-factor loadings, but across our initial analyses of the independent batteries and overlapping batteries, the factor-extraction method contributed to more than 10% of the variance across general-factor loadings. When its interactions were considered, it contributed an additional 7% to 9% of variance. It is clear from reviewing (a) statistics from varied factor-extraction methods (e.g., the percentage of variance accounted for by the general factor) as well as (b) the patterns of general-factor loadings (and associated descriptive statistics) across these methods that principal components analysis produces results that seem to inflate the influence of the general factor on test scores. As a test of these effects, when the results from principal components analysis were omitted from the Generalizability theory analyses, variance due to the factor-extraction method (using principal factors analysis and maximum-likelihood estimation alone) and its interactions contributed negligible variance (i.e., 1% or less of total variance).

These results may seen contradictory with Ree and Earles (1991) and Jensen and Weng (1994), who demonstrated very strong correlations between component scores derived from principal components analysis, factor scores derived from principal factors analysis, and other measures of the general factor. Like this prior research, our results also yield high to very high correlations between the general-factor loadings from our three factor-extraction methods. For example, for the independent batteries analysis, the Spearman rho

Author's personal copy

Table 6
General-factor loadings and summary statistics for each probe test when tests inserted one at a time into each of the 30 8-test batteries and varying factor-extraction methods applied.

Test battery	Probe test																				
	VC			VAL			SR			SB			CF			VM			NR		
	PF	ML	PC	PF	ML	PC	PF	ML	PC	PF	ML	PC	PF	ML	PC	PF	ML	PC	PF	ML	PC
1	.93	.98	.92	.62	.60	.68	.59	.56	.65	.74	.72	.78	.71	.70	.75	.51	.48	.58	.60	.59	.66
2	.90	.92	.90	.62	.60	.67	.52	.51	.58	.66	.64	.71	.68	.67	.73	.47	.46	.54	.62	.60	.68
3	.86	.87	.86	.70	.70	.74	.68	.66	.72	.70	.69	.74	.80	.81	.82	.59	.57	.65	.66	.66	.71
4	.83	.85	.85	.61	.60	.68	.53	.52	.60	.62	.62	.68	.69	.67	.74	.60	.56	.66	.64	.62	.70
5	.92	.99	.90	.65	.63	.71	.58	.56	.65	.71	.70	.75	.73	.72	.77	.58	.54	.64	.64	.63	.69
6	.89	.91	.89	.63	.61	.69	.58	.56	.65	.64	.62	.70	.72	.71	.77	.54	.51	.61	.61	.60	.68
7	.86	.90	.86	.64	.62	.69	.52	.52	.59	.68	.68	.73	.66	.66	.71	.65	.65	.70	.61	.61	.67
8	.78	.80	.80	.65	.64	.70	.57	.56	.63	.64	.63	.69	.71	.71	.75	.69	.67	.74	.66	.66	.71
9	.87	.92	.87	.66	.66	.72	.61	.60	.67	.65	.63	.71	.75	.76	.79	.58	.54	.65	.62	.62	.68
10	.85	.87	.85	.65	.64	.70	.54	.54	.61	.63	.62	.68	.70	.70	.74	.62	.61	.68	.61	.61	.67
11	.95	.98	.93	.61	.57	.67	.56	.53	.63	.71	.68	.76	.70	.66	.74	.52	.44	.59	.61	.56	.67
12	.90	.91	.89	.64	.62	.69	.57	.55	.64	.73	.71	.76	.69	.68	.74	.56	.53	.63	.62	.61	.68
13	.95	.99	.93	.68	.60	.73	.58	.54	.65	.67	.65	.72	.72	.67	.76	.53	.44	.60	.58	.53	.64
14	.82	.86	.83	.62	.62	.68	.56	.55	.63	.69	.70	.73	.71	.71	.75	.65	.61	.71	.64	.65	.70
15	.92	1.0	.89	.70	.70	.74	.67	.66	.72	.72	.72	.75	.76	.75	.78	.63	.62	.69	.65	.64	.70
16	.89	.98	.88	.60	.60	.66	.53	.53	.59	.69	.69	.73	.68	.68	.72	.60	.58	.66	.57	.57	.63
17	.88	.91	.88	.60	.58	.66	.52	.51	.59	.68	.67	.73	.66	.66	.72	.54	.51	.61	.57	.56	.63
18	.94	.99	.91	.65	.62	.70	.56	.54	.63	.72	.71	.76	.67	.66	.72	.51	.50	.58	.57	.57	.64
19	.89	.90	.88	.61	.61	.66	.53	.54	.59	.69	.69	.73	.70	.70	.74	.53	.51	.59	.61	.62	.66
20	.91	.91	.90	.62	.62	.67	.50	.50	.57	.68	.69	.73	.68	.69	.73	.53	.52	.59	.62	.63	.68
21	.88	.89	.88	.63	.62	.68	.54	.54	.61	.71	.71	.75	.70	.70	.75	.58	.55	.64	.65	.65	.70
22	.93	.99	.91	.63	.62	.68	.54	.54	.61	.66	.65	.71	.72	.72	.76	.49	.48	.56	.62	.61	.67
23	.92	.93	.91	.61	.59	.67	.51	.51	.58	.67	.67	.72	.68	.68	.73	.52	.48	.58	.58	.57	.64
24	.92	1.0	.90	.69	.68	.74	.68	.66	.73	.71	.71	.75	.77	.77	.80	.56	.53	.63	.64	.65	.70
25	.92	.94	.90	.62	.56	.69	.56	.52	.63	.67	.62	.73	.69	.64	.74	.55	.43	.63	.59	.52	.66
26	.88	.91	.88	.68	.66	.73	.58	.57	.65	.73	.71	.77	.72	.72	.76	.50	.49	.57	.66	.66	.71
27	.83	.85	.84	.65	.64	.70	.58	.56	.65	.72	.71	.76	.72	.72	.76	.65	.62	.70	.66	.66	.71
28	.91	.92	.90	.62	.61	.67	.53	.53	.59	.71	.70	.75	.69	.68	.73	.49	.48	.56	.63	.62	.68
29	.81	.83	.82	.63	.62	.68	.58	.57	.64	.68	.69	.73	.71	.71	.75	.67	.63	.72	.67	.67	.72
30	.85	.86	.86	.63	.62	.69	.55	.54	.61	.70	.70	.74	.70	.69	.74	.60	.60	.66	.64	.63	.69
M	.89	.92	.88	.64	.62	.69	.57	.55	.63	.69	.68	.73	.71	.70	.75	.57	.54	.63	.62	.61	.68
SD	.04	.06	.03	.03	.03	.02	.05	.04	.04	.03	.03	.03	.03	.04	.02	.06	.07	.05	.03	.04	.03

Note. VC = Verbal Comprehension, VAL = Visual–Auditory Learning, SR = Spatial Relations, SB = Sound Blending, CF = Concept Formation, VM = Visual Matching, NR = Numbers Reversed. PF = principal factors analysis, ML = maximum-likelihood estimation, PC = principal components analysis.

correlation coefficients between general-factor loadings from each factor-extraction method ranged between .99 and 1.0 for the 8-test battery and the 4-test battery and .86 in both cases for the 2-test battery. However, because the Generalizability theory analysis takes into account both relative and absolute differences in general-factor loadings, and not relative differences alone, principal components analysis appears notably discrepant from the other two methods in our analysis due to its inflation of the general-factor loadings.

4.2. Test battery composition and test battery size

Some have argued that the general factor and its related scores are conglomerates, mixtures measures, and hodge-podges of distinct abilities that are determined primarily by the specific tests that contribute to such variables. Others have acknowledged that general factors estimates can be colored or flavored by the types of tests included in the analysis and tainted by psychometric sampling error. Our analysis of general-factor loadings from independent batteries and overlapping batteries revealed that variance attributable to the test battery used to form the general factor was minimal (consisting of 3% of total variance for independent batteries and only 1% for overlapping batteries). Thus, it was not apparent that some test batteries produced uniformly higher general-factor loadings and others produced uniformly lower general-factor loadings. It is in many ways counterintuitive that, although the test batteries may contain what appear to be tests that vary greatly in their content, operations, stimulus input, and modes of response, these variations seem to fade away as the essence underlying performance across all such tests, the general factor, is extracted in factor analysis.

Despite minimal evidence for a main effect of the test battery, our results demonstrate that the interaction between the test battery and the probe tests contributes the largest amount of error variance in general-factor loadings (18% to 22% for the independent batteries and 7% to 8% for the overlapping batteries). That is, general-factor loadings for some tests (e.g., Verbal Comprehension and Visual Matching probe tests) are inflated or deflated when there is such error, whereas general-factor loadings for other tests (e.g., Visual–Auditory Learning and Concept Formation probe tests) are more stable. For example, Visual Matching, a measure of the broad ability Processing Speed, demonstrated the greatest variability across general-factor loadings and its analysis using independent test batteries forming the general factor indicated clear evidence of psychometric sampling error.

Author's personal copy

Table 7
General-factor loadings and summary statistics for each probe test when tests inserted one at a time into each of the 30 4-test batteries and varying factor-extraction methods applied.

Test battery	Probe test																				
	VC			VAL			SR			SB			CF			VM			NR		
	PF	ML	PC	PF	ML	PC	PF	ML	PC	PF	ML	PC	PF	ML	PC	PF	ML	PC	PF	ML	PC
1	.90	.89	.90	.62	.59	.72	.59	.54	.70	.69	.69	.77	.74	.71	.80	.50	.48	.63	.59	.57	.70
2	.86	.86	.86	.62	.61	.72	.48	.46	.62	.68	.66	.76	.64	.63	.74	.51	.49	.64	.64	.62	.73
3	.87	.87	.87	.71	.72	.78	.65	.65	.74	.73	.73	.79	.79	.81	.82	.52	.52	.64	.67	.68	.75
4	.78	.76	.83	.61	.60	.72	.54	.54	.66	.55	.52	.67	.72	.71	.79	.60	.56	.70	.67	.64	.76
5	.97	1.0	.94	.62	.59	.71	.56	.53	.67	.70	.69	.78	.71	.69	.78	.47	.46	.59	.58	.57	.68
6	.91	.92	.91	.63	.61	.74	.61	.58	.72	.64	.62	.74	.73	.71	.80	.47	.45	.61	.59	.58	.71
7	.95	.98	.88	.79	.79	.81	.68	.69	.75	.76	.77	.80	.78	.78	.81	.74	.77	.78	.74	.74	.79
8	.81	.81	.84	.61	.59	.71	.50	.47	.63	.67	.65	.76	.66	.65	.75	.59	.56	.70	.63	.62	.73
9	.91	.94	.89	.66	.67	.75	.61	.61	.71	.63	.64	.73	.76	.77	.81	.52	.51	.64	.61	.61	.72
10	.64	.63	.73	.67	.67	.75	.56	.57	.68	.64	.63	.73	.73	.74	.79	.58	.57	.69	.60	.61	.71
11	.89	.89	.87	.64	.63	.75	.58	.55	.70	.80	.80	.84	.69	.68	.78	.61	.55	.73	.68	.67	.77
12	.90	.91	.90	.63	.61	.73	.58	.55	.69	.75	.73	.81	.68	.67	.77	.61	.54	.71	.61	.59	.71
13	.99	1.0	.92	.77	.79	.81	.62	.61	.71	.70	.70	.77	.74	.73	.79	.63	.73	.71	.59	.58	.69
14	.79	.84	.81	.64	.65	.73	.65	.66	.73	.71	.73	.77	.73	.76	.78	.74	.76	.79	.60	.62	.71
15	.72	.72	.78	.75	.76	.79	.67	.66	.75	.62	.61	.73	.75	.76	.79	.79	.79	.83	.63	.61	.73
16	.96	1.0	.92	.58	.58	.68	.48	.49	.60	.68	.68	.75	.68	.69	.76	.56	.56	.67	.56	.56	.67
17	.91	.93	.90	.58	.57	.68	.50	.49	.61	.68	.67	.76	.66	.65	.74	.56	.53	.67	.59	.57	.69
18	.84	.85	.84	.64	.64	.73	.63	.62	.73	.81	.82	.82	.66	.66	.75	.53	.52	.66	.59	.58	.70
19	.85	.87	.86	.60	.60	.70	.53	.54	.64	.70	.70	.77	.71	.72	.76	.52	.52	.64	.56	.57	.67
20	.92	.93	.90	.62	.61	.71	.48	.48	.60	.66	.66	.74	.69	.71	.77	.55	.55	.66	.58	.58	.68
21	.92	.92	.91	.64	.62	.74	.59	.56	.70	.73	.72	.80	.71	.69	.79	.60	.53	.71	.66	.64	.75
22	.88	.90	.89	.66	.66	.75	.56	.57	.67	.63	.61	.72	.76	.77	.82	.52	.51	.63	.67	.66	.75
23	.93	.94	.92	.59	.58	.69	.52	.52	.64	.70	.68	.77	.70	.69	.78	.47	.46	.59	.58	.56	.68
24	.84	.85	.86	.68	.68	.77	.63	.62	.73	.65	.68	.75	.77	.77	.82	.50	.50	.63	.65	.66	.75
25	.84	.84	.83	.64	.64	.74	.52	.51	.66	.65	.64	.74	.68	.69	.75	.58	.57	.70	.57	.55	.69
26	.83	.83	.86	.65	.64	.74	.58	.58	.70	.72	.70	.79	.72	.73	.79	.52	.50	.65	.72	.71	.79
27	.89	.90	.88	.65	.64	.75	.56	.54	.68	.76	.76	.81	.69	.68	.77	.63	.60	.73	.66	.66	.75
28	.94	.94	.92	.58	.58	.69	.49	.47	.60	.67	.67	.75	.66	.66	.74	.45	.45	.57	.59	.59	.69
29	.84	.85	.87	.63	.62	.74	.57	.53	.70	.69	.68	.77	.73	.72	.80	.62	.57	.74	.67	.66	.76
30	.83	.83	.85	.60	.59	.70	.47	.47	.59	.67	.67	.76	.65	.64	.74	.57	.59	.68	.66	.66	.74
M	.87	.88	.87	.64	.64	.73	.57	.56	.68	.69	.68	.77	.71	.71	.78	.57	.56	.68	.62	.62	.72
SD	.07	.08	.05	.05	.06	.03	.06	.06	.05	.05	.06	.03	.04	.05	.03	.08	.09	.06	.05	.05	.03

Note. VC = Verbal Comprehension, VAL = Visual–Auditory Learning, SR = Spatial Relations, SB = Sound Blending, CF = Concept Formation, VM = Visual Matching, NR = Numbers Reversed. PF = principal factors analysis, ML = maximum-likelihood estimation, PC = principal components analysis.

Observe the first independent test battery of 8 tests in Table 2. It contains four tests that measure Processing Speed: Decision Speed, Cross Out, Reading Fluency, and Writing Fluency. As a result of overrepresentation of measures of this broad ability in this test battery, the general-factor loadings from Visual Matching were high, whereas its general-factor loadings from analysis with the four other independent batteries were moderate or low (see first row of values in Table 4). In addition, other probe tests demonstrated some of their lowest general-factor loadings when analyzed with the first independent test battery that comprised so many tests of Processing Speed. These results clearly reflect the effects of psychometric sampling error.

Although decomposing the size of the test batteries decreased the magnitude of the results of Bartlett's test, the KMO measures, and the coefficients of congruence as well as increased the percentage of variance attributable to the general factor, the facet of the size of the test battery contributed miniscule variance in only one Generalizability theory analysis. However, the size of the test battery appeared to magnify the effects of psychometric sampling error through its interactions with the probe tests, with the factor-extraction method, and with the test battery as well as through its interactions with the probe tests and the test battery. Even when the effects of psychometric sampling error were reduced through the employment of numerous overlapping batteries and when the effects of principal components analysis were removed, interactions associated with the number of tests contributed to more than a quarter of total variance across general-factor loadings. Thus, as the number of test scores included in the analysis diminishes, the greater the effects psychometric sampling error has on the general factor and resulting scores.

4.3. Dependability of general-factor loadings

Our results, obtained from analysis of data from a large nationally representative sample of young adults, indicate the remarkable dependability of the general-factor loadings across different test batteries and factor-extraction methods when the effects of psychometric sampling error are controlled somewhat and when results from principal components analysis are removed. The resulting dependability coefficient was almost unity, .99. Even in our single worst-case scenario, which was our first analysis with the independent batteries that included principal components analysis, the resulting dependability coefficients was .72, which is below–but not far below–a reasonable standard for

Author's personal copy

Table 8
General-factor loadings and summary statistics for each probe test when tests inserted one at a time into each of the 30 2-test batteries and varying factor-extraction methods applied.

Test battery	Probe test																				
	VC			VAL			SR			SB			CF			VM			NR		
	PF	ML	PC	PF	ML	PC	PF	ML	PC	PF	ML	PC	PF	ML	PC	PF	ML	PC	PF	ML	PC
1	.91	.92	.91	.62	.62	.78	.55	.55	.74	.70	.70	.82	.76	.76	.85	.51	.51	.71	.61	.61	.77
2	.97	.98	.88	.79	.79	.82	.59	.59	.75	.79	.80	.83	.80	.80	.82	.53	.53	.72	.76	.77	.82
3	.88	.89	.88	.72	.72	.82	.65	.65	.79	.69	.69	.81	.81	.81	.86	.52	.52	.72	.62	.62	.77
4	.96	.96	.91	.71	.71	.82	.60	.60	.77	.70	.70	.81	.81	.82	.86	.51	.51	.72	.75	.75	.83
5	–	1.0	.95	.56	.56	.74	.53	.53	.73	.76	.76	.84	.65	.65	.79	.42	.42	.66	.55	.55	.74
6	.76	.76	.86	.68	.68	.80	.76	.76	.82	.61	.61	.77	.78	.79	.85	.56	.56	.75	.66	.65	.80
7	.98	.99	.92	.66	.67	.79	.54	.54	.73	.78	.78	.83	.71	.71	.81	.47	.47	.69	.73	.73	.82
8	.78	.78	.85	.62	.62	.77	.50	.50	.74	.78	.78	.83	.66	.66	.79	.62	.62	.76	.53	.53	.74
9	.89	.88	.92	.56	.56	.76	.54	.53	.75	.69	.69	.82	.61	.60	.80	.50	.50	.71	.55	.54	.75
10	.78	.78	.84	.59	.59	.77	.46	.46	.68	.62	.62	.76	.57	.57	.76	.47	.47	.69	.51	.51	.71
11	.89	.89	.89	.59	.59	.76	.49	.49	.70	.77	.77	.84	.61	.61	.77	.50	.50	.71	.59	.59	.76
12	.88	.89	.90	.58	.58	.75	.51	.51	.72	.77	.77	.85	.61	.61	.78	.46	.46	.68	.55	.55	.74
13	–	1.0	.96	.80	.80	.85	.63	.63	.77	.74	.74	.83	.75	.75	.82	.48	.48	.70	.57	.57	.75
14	.94	.98	.90	.81	.83	.82	.88	.91	.83	.82	.85	.84	.90	.95	.85	.67	.68	.78	.76	.77	.81
15	.65	.66	.78	.81	.81	.83	.74	.74	.81	.58	.57	.76	.73	.73	.81	.79	.79	.86	.59	.59	.76
16	.86	.87	.88	.60	.59	.76	.42	.41	.67	.64	.64	.79	.67	.66	.82	.52	.52	.72	.57	.57	.76
17	.91	.91	.93	.54	.54	.73	.45	.45	.68	.59	.59	.77	.61	.61	.78	.44	.44	.66	.53	.53	.73
18	.78	.79	.83	.56	.56	.75	.52	.51	.72	.79	.79	.84	.55	.55	.74	.52	.52	.72	.54	.54	.73
19	.91	.91	.91	.62	.62	.78	.55	.55	.74	.70	.70	.82	.76	.76	.85	.52	.52	.64	.61	.61	.77
20	.79	.79	.86	.56	.56	.75	.49	.48	.70	.65	.65	.79	.69	.69	.82	.64	.64	.80	.58	.58	.76
21	.87	.86	.89	.73	.73	.82	.70	.70	.81	.75	.75	.83	.79	.79	.84	.79	.79	.86	.70	.70	.81
22	.81	.81	.87	.67	.67	.80	.57	.57	.76	.64	.64	.79	.78	.78	.85	.58	.58	.76	.75	.75	.84
23	.95	.95	.94	.55	.55	.74	.50	.50	.71	.63	.63	.79	.65	.65	.80	.42	.42	.65	.52	.52	.72
24	.83	.83	.88	.66	.66	.80	.60	.60	.77	.62	.62	.81	.74	.74	.83	.51	.51	.71	.70	.70	.82
25	.97	1.0	.86	.82	.84	.82	.82	.84	.81	.74	.75	.79	.85	.88	.83	.59	.59	.76	.68	.69	.80
26	.83	.83	.89	.64	.64	.79	.48	.58	.76	.66	.66	.81	.73	.73	.84	.48	.48	.70	.68	.68	.82
27	.75	.72	.87	.75	.76	.81	.71	.72	.79	.76	.76	.82	.83	.83	.84	.78	.79	.83	.79	.79	.83
28	.92	.92	.91	.64	.63	.78	.54	.54	.73	.71	.71	.82	.71	.71	.82	.49	.49	.71	.66	.66	.80
29	.86	.87	.88	.57	.56	.75	.47	.47	.69	.71	.71	.82	.63	.63	.78	.52	.52	.72	.62	.61	.77
30	.86	.86	.88	.60	.60	.77	.49	.49	.70	.68	.68	.81	.68	.67	.80	.59	.58	.78	.64	.64	.79
M	.86	.87	.89	.65	.65	.78	.58	.58	.75	.70	.70	.81	.71	.72	.82	.55	.55	.73	.63	.63	.78
SD	.08	.08	.04	.09	.09	.03	.11	.12	.04	.07	.07	.02	.09	.10	.03	.10	.10	.06	.08	.09	.04

Note. Two general-factor loadings were not reported because their communality exceeded 1.0. VC = Verbal Comprehension, VAL = Visual–Auditory Learning, SR = Spatial Relations, SB = Sound Blending, CF = Concept Formation, VM = Visual Matching, NR = Numbers Reversed. PF = principal factors analysis, ML = maximum-likelihood estimation, PC = principal components analysis.

dependability coefficients (.80; R. J. Shavelson, personal communication, April 5, 2007). However, this coefficient is higher than any of the dependability coefficients stemming from analysis of teacher ratings of externalizing behaviors across time, rater, and instrument ($\phi = .47$ to .68; Bergeron, Floyd, McCormack, & Farmer, 2008) as well as higher than the vast majority of coefficients stemming from analysis of blood pressure metrics taken twice in the same day within a setting ($\phi = .30$ to .84; Llabre et al., 1988). Thus, we concur with Jensen (1998) who stated, "The *g* is always influenced, more or less, by both the nature and the variety of the tests from which it is extracted.… But the fact is that *g* remains quite invariant across many different collections of tests" (p. 85). We can now amend this quote by offering, "But the fact is that *g* remains quite invariant across many different collections of tests, across the factor-extraction methods of principal factors analysis and maximum-likelihood estimation, and across batteries of varying sizes." Because of this invariance, it appears that the relatively minimal differences across the hundreds of general factors we extracted (as represented in changes in general-factor loadings) can be attributable to experimental imprecision (Detterman, 2002; Jensen, 1998).

4.4. Limitations

Although this study drew data from large, nationally representative samples of young adults and enacted several innovations in design and analysis, it is not without its limitations. First, some may argue that our test batteries contained a preponderance of measures of the broad (stratum II) ability Comprehension–Knowledge. Some, such as Carroll (1993), have asserted that measures of reading and writing abilities are abilities subsumed by this ability, whereas others (e.g., McGrew & Woodcock, 2001; Woodcock, 1998) have asserted that they form a distinct ability at the broad (stratum II) ability level called Reading and Writing. Thus, based on our classifications presented in Table 1, 20 of the tests (43%) included across the test batteries we created could be considered as measures of Comprehension–Knowledge. Perhaps this preponderance of tests of this type led to the probe test, Verbal Comprehension, to have consistently high general-factor loadings, some values of 1.0, and some out-of-bound values across analyses (see also Gignac, 2006). In a related vein, some may argue that, despite our efforts to sample broadly from a wide range of cognitive ability measures, our sampling was insufficient

Author's personal copy

in breadth. For example, we included no measures targeting lower-order processes, such as those targeting kinesthetic abilities, reaction time, and decision speed (see, for example, Stankov, 2002 and McGrew, 2009). It is possible that the effects of the test battery and its interactions would have been more substantial with broader sampling, but it is unlikely the general findings from this study would have varied greatly.

Second, some may argue that our use of exploratory factor analysis is inappropriate when more controlled analyses, such as confirmatory factor analysis, can be used. We agree that control is needed for studies focusing on extracting general factors, and in fact, we believe that we used exploratory factor analysis methods in a confirmatory manner. Thus, we did not consider methods to extract multiple factors (see Thompson, 2004). However, we used the method for determining general-factor loadings that is commonly by structural equation modeling (SEM) software to determine parameter estimates—maximum-likelihood estimation. In fact, when a single-factor model was specified using SEM software and maximum-likelihood estimation was used, results matched ours exactly. In addition, based on our random selection of only 8 tests forming each independent and semi-independent battery, we are unable to specify consistently higher-order general factors from first-order broad (stratum II) factors that are adequately identified (i.e., with two or more indicators). Based on research by Ree and Earles (1991) and Jensen and Weng (1994), we do not believe that the general factor from a well-constructed hierarchical model would differ substantially from the general factors exacted in across our factor-extraction methods.

4.5. Conclusion

These results add to the body of evidence supporting the construct validity of the general factor, and they limit some criticisms of the general factor and related memes that pervade minds of many professionals and consumers of tests results. It is apparent that researchers, test authors and publishers, and other professionals involved in measuring the general factor should avoid using principal components analysis when computing general-factor loadings or when obtaining measures of the general factor. In addition, psychometric sampling error should be targeted as a problem and attempts at representative sampling of specific cognitive abilities should be made when constructing measures representing the general factor. Thoughtfully constructed batteries of cognitive ability tests should yield general factors and general-factor scores that are largely invariant across batteries.

Acknowledgements

The last author, Kevin S. McGrew, has a financial interest in the Woodcock-Johnson III because he is a co-author of that test battery.

We thank Richard Woodcock and the Woodcock–Muñoz Foundation for making data from the WJ III available for this research. We benefitted greatly from the reviews and comments of Joni Lakin. Katie McCloud, Danielle Murphy, and Danielle Steele are appreciated for providing computer, editing, and library support.

References

Bergeron, R., Floyd, R. G., McCormack, A. C., & Farmer, W. (2008). The generalizability of externalizing behavior composites and subscale scores across time, rater, and instrument. *School Psychology Review*, *37*, 91—108.

Bradley-Johnson, S., Morgan, S. K., & Nutkins, C. (2004). Test review. [Review of The Woodcock–Johnson III Tests of Achievement.]. *Journal of Psychoeducational Assessment*, *22*, 261—274.

Brennan, R. L. (2001). *Generalizability theory.* New York: Springer.

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

Cizek, G. J. (2003). [Review of the Woodcock–Johnson III]. In B. S. Plake & J.C. Impara (Eds.), *The fifteenth mental measurements yearbook* (pp. 1020—1024). Lincoln, NE: Buros Institute of Mental Measurements.

Cronbach, L. J., Gleser, G. C., Nanda, H., & Rajaratnam, N. (1972). *The dependability of behavioral measurements.* New York: Wiley.

Detterman, D. K. (2002). General intelligence: cognitive and biological explanations. In R. J. Sternberg & E.L. Grigorenko (Eds.), *The general factor of intelligence: How general is it?* (pp. 223—243). Mahwah, NJ: Erlbaum.

Floyd, R. G., Keith, T. Z., Taub, G. E., & McGrew, K. S. (2007). Cattell–Horn–Carroll cognitive abilities and their effects on reading decoding skills: g has indirect effects, more specific abilities have direct effects. *School Psychology Quarterly*, *22*, 200—233.

Floyd, R. G., McGrew, K. S., Barry, A., Rafael, F. A., & Rogers, J. (2009). General and specific effects on Cattell–Horn–Carroll broad ability composites: analysis of the Woodcock–Johnson III Normative Update CHC factor clusters across development. *School Psychology Review*, *38*(2).

Gould, S. J. (1981). *The mismeasure of man.* New York: Norton.

Gould, S. J. (1994). *The mismeasure of man*, Rev. ed. New York: Norton.

Gignac, G. E. (2006). Evaluating subtest "g" saturation levels via the single trait-correlated uniqueness (STCU) SEM approach: evidence in favor of crystallized subtests as the best indicators of "g". *Intelligence*, *34*, 29—46.

Harman, H. H. (1976). *Modern factor analysis*, revised 3rd ed. Chicago: University of Chicago Press.

Horn, J. L. (1985). Remodeling old models of intelligence. In B. B. Wolman (Ed.), *Handbook of intelligence: Theories, measurements, and applications* (pp. 267—300). New York: John Wiley & Sons.

Horn, J. L. (1989). Models of intelligence. In R. Linn (Ed.), *Intelligence: Measurement, theory and public policy. Proceedings of a symposium in honor of Lloyd G. Humphreys* (pp. 29—73).

Horn, J. L., & Blankson, N. (2005). Foundation for better understanding cognitive abilities. In D. P. Flanagan & P. Harrison (Eds.), *Contemporary intellectual assessment* (pp. 41—68)., 2nd ed. New York: Guilford Press.

Horn, J. L., & McArdle, J. J. (2007). Understanding human intelligence since Spearman. In R. Cudeck & R.C. MacCallum (Eds.), *Factor analysis at 100: Historical developments and future directions* (pp. 205—249). Mahwah, NJ: Erlbaum.

Jensen, A. R. (1982). *Bias in mental testing.* New York: Free Press.

Jensen, A. R. (1998). *The g factor: The science of mental ability.* Westport, CT: Preager.

Jensen, A. R., & Weng, L. J. (1994). What is a good g? *Intelligence*, *18*, 231—258.

Johnson, W., Bouchard, T. J., Jr., Krueger, R. F., McGue, M., & Gottesman, I. I. (2004). Just one g: consistent results from three test batteries. *Intelligence*, *32*, 95—107.

Johnson, W., te Nijenhuis, J., & Bouchard, T. J., Jr. (2008). Still just 1 g: consistent results from five test batteries. *Intelligence*, *36*, 81—95.

Kaufman, A. S. (1994). *Intelligent testing with the WISC-III.* New York: John Wiley & Sons.

Keith, T. Z., Kranzler, J. H., & Flanagan, D. P. (2001). What does the Cognitive Assessment System (CAS) measure? Joint confirmatory factor analysis of the CAS and the Woodcock–Johnson Tests of Cognitive Ability (3rd edition). *School Psychology Review*, *30*, 89—119.

Llabre, M. M., Ironson, G. H., Spitzer, S. B., Gellman, M. D., Weidler, D. J., & Schneiderman, N. (1988). How many blood pressure measurements are enough?: an application of generalizability theory to the study of blood pressure reliability. *Psychophysiology*, *25*, 97—106.

McGrew, K. S. (2009). CHC theory and the human cognitive abilities project: standing on the shoulders of the giants of psychometric intelligence research. *Intelligence*, *37*, 1—10.

McGrew, K. S., & Flanagan, D. P. (1998). *The intelligence test desk reference (ITDR): Gf-Gc cross-battery assessment.* Boston: Allyn & Bacon.

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

Phelps, L., McGrew, K. S., Knopik, S. N., & Ford, L. A. (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.

Ree, M. J., & Earles, J. A. (1991). The stability of g across different methods of estimation. *Intelligence*, *15*, 271—278.

Author's personal copy

Sares, T. (2005). [Review of the Woodcock–Johnson III Diagnostic Supplement to the Tests of Cognitive Abilities]. In R. A. Spies & B.S. Plake (Eds.), *The sixteenth mental measurements yearbook [Electronic version]* Retrieved July 29, 2005 from the Buros Institute's Test Reviews Online website: http://www.unl.edu/buros

Shavelson, R. J., & Webb, N. M. (1991). *Generalizability theory: A primer.* Newbury Park, NY: Sage.

Snook, S. C., & Gorsuch, R. L. (1989). Component analysis versus common factor analysis: a Monte Carlo study. *Psychological Bulletin*, *106*, 148—154.

Spearman, C. (1927). *The abilities of man: Their nature and measurement.* New York: Macmillan.

Stankov, L. (2002). g: a diminutive general. In R. J. Sternberg & E.L. Grigorenko (Eds.), *The general factor of intelligence: How general is it?* (pp. 19—37). Mahwah, NJ: Erlbaum.

Sternberg, R. J., & Grigorenko, E. L. (Eds.). (2002). *The general factor of intelligence: How general is it?* Mahwah, New Jersey: Erlbaum.

Taub, G. E., & McGrew, K. S. (2004). A confirmatory factor analysis of CHC theory and cross-age invariance of the Woodcock–Johnson Tests of Cognitive Abilities III. *School Psychology Quarterly*, *19*, 72—87.

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

Thompson, D. T. (2005). [Review of the Woodcock–Johnson III Diagnostic Supplement to the Tests of Cognitive Abilities]. In R. A. Spies & B.S. Plake (Eds.), *The sixteenth mental measurements yearbook [Electronic version]* Retrieved July 29, 2005 from the Buros Institute's Test Reviews Online website: http://www.unl.edu/buros

Thorndike, R. L. (1987). Stability of factor loadings. *Personality and Individual Differences*, *8*, 585—586.

Wilks, S. S. (1938). Weighting systems for linear functions of correlated variables when there is no dependent variable. *Psychometrica*, *3*, 23—40.

Woodcock, R. W. (1998). Extending Gf–Gc theory into practice. In J. J. McArdle & R.W. Woodcock (Eds.), *Human cognitive abilities in theory and practice* (pp. 137—156). Mahwah, NJ: Erlbaum.

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock–Johnson III Tests of Cognitive Abilities.* Itasca, IL: Riverside Publishing.

Woodcock, R. W., McGrew, K. S., Mather, N., & Schrank, F. A. (2003). *Woodcock–Johnson III Diagnostic Supplement to the Tests of Cognitive Abilities.* Itasca, IL: Riverside.