Special Issue: Woodcock-Johnson V

99 Ways to Obtain a General Factor of Intelligence—And They Are Not All Created Equal

Journal of Psychoeducational Assessment
2025, Vol. 0(0) 1–15

DOI: 10.1177/07342829251380405
journals.sagepub.com/home/jpa
Sage

Moritz Breit[1], Andrew R. A. Conway[2], and Kristof Kovacs[3]

Abstract

This paper examines whether general intelligence (g) factors derived from different test batteries are equivalent. There are three views regarding the equivalency of g-factors: (1) "indicator indifference" claims that test content is irrelevant as long as g loadings are identical and that single tests can be adequate indicators of g; (2) "complete dependence on test composition" claims that general factors are completely dependent on the tests from which they are extracted; and (3) an intermediate stance that emphasizes the importance of the diversity and comprehensiveness of cognitive test batteries from which g is obtained. The present study evaluates these competing views by analyzing g-factor correlations across all combinatorically possible combinations of subtests from the Woodcock–Johnson Tests of Cognitive Abilities V. Results showed strong correlations among g-factors across both one-factor and hierarchical models, increasing with the number of subtests or broad abilities included. Most g-factors closely matched the g-estimate obtained from all available subtests ($r > .9$). Low correlations were mainly tied to the overrepresentation of processing speed (Gs) in small test sets, highlighting the impact of content coverage. Overall, results support the intermediate view: reliable g-estimates require broad, balanced batteries covering at least three broad abilities.

Keywords

Woodcock–Johnson intelligence/cognition, intelligence models intelligence/cognition, factor analysis measurement, general factor of intelligence

Introduction

Are all general factors created equal? This is arguably the second longest-standing controversy in the history of intelligence research, following closely behind debates surrounding the interpretation of the general factor, *g*. Broadly speaking, there are three main perspectives on this issue.

[1]University of Trier, Trier, Germany
[2]New Mexico State University, Las Cruces, NM, USA
[3]ELTE Eotvos Lorand University, Budapest, Hungary

Corresponding Author:
Moritz Breit, University of Trier, Universitätsring 15, Trier 54286, Germany.
Email: breitm@uni-trier.de

Perspective 1: Indifference of the Indicator

The first perspective was offered by Charles Spearman, the discoverer of *g*. His two-factor theory differentiated *g*, a factor common to performance on all mental tests, from *s*-factors, which are test-specific. Spearman hypothesized that the general factor is "measuring something analogous to an 'energy'; that is to say, it is some force capable of being transferred from one mental operation to another different one" (Spearman, 1927, p. 414). If *g* truly functions like a transferable energy, then the measure used (i.e., the test battery) is no more significant than the type of thermometer used for body temperature. This conceptualization aligns with a realist interpretation of latent variables, which holds that a latent variable reflects a real entity independent of its measurement (Borsboom et al., 2003). Spearman called this notion the *indifference of the indicator*, proposing that the particular tests used to extract *g* are irrelevant, so long as they require mental effort. This assumption of indicator indifference is sometimes directly applied in scientific studies by using only a single, highly g-loaded test to approximate the general factor. In most cases, nonverbal reasoning tests such as Raven's Progressive Matrices are used, since they are assumed to be the best single estimates of *g* (e.g., Jensen, 1998), yet in other instances, vocabulary tests—reflecting comprehension-knowledge (Gc)—are used as a direct proxy for general intelligence (e.g., Li & Kanazawa, 2016).

Vernon (1989) examined the proposed indifference of the indicator based on test battery intercorrelations. He found that the *g*-factor extracted from a battery of reaction time tests correlated almost as highly with *g*-factors extracted from a psychometric battery measuring intelligence as the verbal and performance factor scores from the same intelligence battery correlate with one another (i.e., ranging from r = .26 to .69). According to Vernon, this provides "substantial support for the notion of the indifference of the indicator and for the generality of *g*" (p. 804). Because RT and IQ batteries do not overlap in content, he attributed the correlations among the resulting *g*-factors to the generality of *g* and as evidence for the indifference of the indicator. Moreover, he argued that the similarity in the magnitude of correlation across instruments supports their interchangeability as proxies of *g*. That is, under this approach, the correlation between two different indicators of *g* does not need to be perfect in order to establish that they are functionally equivalent indicators of *g*. Instead, it is sufficient that the correlation between A (e.g., reaction time) and B (IQ) is very similar in magnitude to the correlation between C (verbal IQ) and D (performance IQ). A more restrictive approach states that "ideally, equivalence between *g*-factors would be demonstrated by having *g*-factor scores or loadings that are statistically indistinguishable" (Major et al., 2011, p. 420).

Taken together, two statistical predictions follow from this perspective on *g*. First, any two batteries composed of mental tests should be substantially correlated, even if they do not overlap in their included tests or captured abilities. Second and relatedly, the generalizability of *g* is not a function of the number or breadth of tests used to obtain *g*, meaning that g-factors from small, narrow batteries will correlate just as highly as *g*-factors from large, comprehensive batteries.

Perspective 2: Complete Dependence on the Indicator

The second approach, originally advocated by Thurstone, is the direct antithesis of Spearman's view (Thurstone, 1947). It holds that the general factor of intelligence is not a real trait independent from its measurement but merely a statistical construct that is entirely dependent on the composition of the test battery from which it is extracted. This view was shared by Horn, who described *g* as a conglomerate of distinct abilities determined by the specific tests used (Horn, 1989). It was also popularized by Gould, who argued that *g* is a mathematical artifact with no psychological reality (Gould, 1996).

From this position, it follows that different batteries should yield completely different general factors—a prediction that has previously been falsified (Jensen & Weng, 1994; Reeve & Blacksmith, 2009a). Surprisingly, with regard to the number of indicators included in the batteries, the stance of "complete dependence on the indicator" is similar to the "indifference of the indicator" stance, since the number of tests used to estimate *g* is not relevant under either. If the indicators are indifferent, then an arbitrarily small and narrow set of tests is sufficient to estimate *g*. But if there is complete dependence on the indicators, then no matter how many tests are used, different batteries should always yield different *g*-factors.

Under the "complete dependence on the indicator" approach, *g* factors obtained from two different batteries are identical if and only if the composition of the two batteries is identical in terms of the narrow abilities they measure, that is, if each battery consists of subtests that measure the same exact narrow abilities with the same number of subtests and the subtests' factor loadings are equal. Since this is almost never the case, one cannot realistically expect very high correlations between *g*-factors obtained from different psychometric batteries. At the same time, if *g*-factors are estimated from batteries that measure completely different specific abilities, then this perspective predicts a 0 correlation between the *g*-factors obtained—a prediction that is the direct opposite of what the indifference of the indicator stance would predict. Similarly, the higher the extent of completely overlapping specific abilities, the higher the correlation, but with a limited number of such overlaps the correlations between general factors are expected to be low.

Perspective 3: The Intermediate View

An intermediate stance suggests that general factors derived from different test batteries are equivalent to the extent that the batteries broadly sample a wide range of abilities without overrepresenting any single ability. Jensen and Weng introduced the concept of psychometric sampling error: "Just as there is sampling error with respect to statistical parameters, there is psychometric sampling error with respect to g, because the universe of all possible mental tests is not perfectly sampled by any limited set of tests" (Jensen & Weng, 1994, p. 236). Consequently, single-test estimates of *g* are viewed as inadequate. General factors can be equivalent only if the test batteries are sufficiently large and diverse to average out specific processes. In contrast to the battery dependence perspective, this does not require an actual overlap in captured abilities, just sufficient diversity.

Two theories of the positive manifold (and thus psychometric g) closely align with this intermediary perspective. The Bonds Model (Bartholomew et al., 2009) explains the positive manifold of intelligence as the result of many small, simple processing units (or "bonds") in the brain that work collectively and are sampled by different mental tests in an overlapping fashion. In the Bonds Model, the correlations between any two tests reflect simply the number of processes they share relative to the total number of processes they collectively sample. Consequently, the Bonds Model predicts that general factors become increasingly similar as test batteries approach complete coverage and balanced representation of the brain's processing units. That is, the larger the number of tests from which *g* is obtained, the better the general factor.

Process Overlap Theory (Kovacs & Conway, 2016) is similar to the Bonds Model in that it posits that positive correlations among diverse mental tests result from overlapping sampling of cognitive processes. However, unlike in the Bonds Model, where all bonds have an equal probability of being sampled by any given test and therefore the correlation between any two tests is the direct, linear function of the number of bonds sampled by both tests in proportion to the processes uniquely sampled by each test, Process Overlap Theory identifies domains of achievement and ability and differentiates *generalist,* "multiple-demand" processes engaged across a wide range of tests, as opposed to *specific* processes sampled by a narrow range of tests only. Crucially, this theory proposes multiple generalist processes, with no single process shared by all tests. Thus, it predicts that general factors are identical only if they primarily reflect overlapping, generalist cognitive processes. For this to be achieved, a battery with a sufficient number and breadth of tests is needed. Conversely, any single test predominantly reflects the specific process related to the specific ability it measures and only secondarily the generalist processes.

Process Overlap Theory proposes that *g* does not reflect any psychological or neural construct; instead it is merely the factor analytic equivalent of the mathematical fact of the positive manifold itself. Additionally, the theory claims that the positive manifold and thus *g* are emergent properties resulting from an overlap of processes, which makes *g* a formative rather than a reflective construct—the common consequence, not the common cause of the positive correlations between tests that enable the extraction of a general factor (Kovacs & Conway, 2019). Nevertheless, Process Overlap Theory emphasizes that formative constructs can be useful for predicting outcomes. In particular, *g* can be useful in predicting *general* outcomes that are not more dependent on any of the specific abilities than on the others. For instance, *g* is more useful in predicting a grade point average than a reading score or a math score. Since formative constructs are dependent on their indicators by definition, in order for *g* to be a useful predictor of general outcomes, specific abilities need to be balanced in the batteries from which *g* factors are obtained.

Empirical Evidence on the Generality of g

Several studies have examined the generality of *g* by correlating general factors from different batteries. Thorndike (1987) reported an average correlation of .85 between *g* loadings across six batteries. Similarly, Ree and Earles (1991) found correlations ranging from .93 to .99. Keith, Kranzler, and Flanagan (2001) reported a .98 correlation between two batteries, Johnson et al. (2004) reported correlations as high as .99 and 1.0 across three batteries, and Johnson et al. (2008) found correlations ranging between .77 and 1 between five batteries. In a comprehensive analysis, Jensen and Weng (1994) found an average correlation of .98 between *g*-factors derived using different methods.

However, most of the aforementioned correlations were based on very comprehensive cognitive batteries (e.g., Wechsler Adult Intelligence Scale). Thus, the correlations are not very informative to distinguish between the presented perspectives on the *g*-factor, as all three perspectives would predict high correlations between comprehensive batteries that cover highly overlapping sets of broad abilities. Some indication that correlations may diminish in less comprehensive batteries can be found in the results by Johnson et al. (2008). Here, the *g*-factor from the shorter Cattell Culture Fair Test exhibited descriptively lower correlations with larger batteries than the larger batteries did with each other. Another study showed that *g*-factor loadings depend on the specific batteries used, concluding that "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" (Floyd et al., 2009, p. 464). A related study supported this conclusion, emphasizing that small, single-factor test samples do not adequately represent the general factor (Major et al., 2011).

The Present Study

In the present study, we present an analysis designed to distinguish between the predictions made by the three different perspectives on *g*. Specifically, we systematically varied the number of tests and abilities covered between test batteries. To do so, we utilized data from the recent U.S. standardization of the Woodcock–Johnson Tests of Cognitive Abilities V (WJ-COG-V) and examined all combinatorially possible combinations from which general factors could be extracted. We adopted the Cattell-Horn-Carroll (CHC) model of cognitive ability structure (McGrew, 2009) in selecting our higher-order model. To evaluate the predictions, we examined the magnitude and distribution of the correlations and their relationship with battery size. In accordance with the intermediate view, we expected substantial *g*-factor correlations that significantly increase with increasing numbers of tests included and abilities captured.

Methods

Test Instrument and Sample

The Woodcock–Johnson V (WJ V) is the latest edition of the Woodcock–Johnson test series (McGrew et al., 2025). It is an individually administered measure designed for ages 3 through adulthood. The WJ V comprises a cognitive test battery, an achievement test battery, and a virtual test library. The present investigation focuses on the cognitive test battery (WJ V COG). The WJ V COG contains 20 subtests, 14 of which form the standard set. These 14 subtests assess 7 CHC broad abilities (i.e., 2 subtests per broad ability): Comprehension-Knowledge (Gc), Fluid Reasoning (Gf), Cognitive Processing Speed (Gs), Visual Processing (Gv), Retrieval Fluency (Gr), Long-Term Storage (Gl), and Working Memory Capacity (Gwm). The subtests of the standard set with their corresponding CHC abilities are presented in Table 1.

The present study draws on the standardization sample of the WJ V. This sample was carefully recruited to closely represent the U.S. population in terms of geographic distribution, race/ethnicity, and educational levels. Detailed descriptions of the sampling procedures are provided in the technical manual. The sample comprises a total of $N = 5838$ participants aged 3 to 98. To ensure that all *g*-factor scores were based on the same sample, we excluded 506 participants with missing data on any subtest, resulting in a final sample size of $N = 5332$. Because the analyses in this study represent secondary analyses of data collected as part of WJ V standardization, institutional review board approval was not sought.

Analyses

We conducted two sets of analyses: the first based on the one-factor model, and the second based on the hierarchical (i.e., second-order) model of cognitive abilities. We used age-standardized subtest scores for all analyses. Prior to analysis, these scores were standardized as z-scores to minimize the risk of model-nonconvergence, which can sometimes be increased with large indicator values. All analyses were conducted in R version 4.2.2.

One-Factor Model. First, we generated all possible combinations of at least three WJ V COG subtests. Three subtests were the minimum required to identify the one-factor model. This procedure resulted in 16,278 unique combinations of subtests. Next, we estimated the one-factor model (top panel of Figure 1) based on each combination using the lavaan package (Rosseel, 2012) and exported the *g*-factor scores. This step produced a dataset of 5,332 participants, each with 16,278 *g*-factor scores. Third, we computed the full Pearson correlation matrix of all *g*-factor scores, resulting in 132,507,563 correlations in total. We extracted several statistics from this matrix: The minimum and maximum values, the median, mean, and standard deviation, and the proportion of correlations falling within 0.1-wide bands. In addition, we examined the subset of 16,277 correlations with the most comprehensive *g*-factor (*g*-complete) using all 14 subtests. In

Table 1. WJ V Standard Set Subtests

Name	Description	Broad ability
Analysis-synthesis	Analyze puzzles and determine missing components.	Gf
Block rotation	Identify two patterns of blocks that match a target pattern.	Gv
Letter-pattern matching	In a row of six-letter patterns, rapidly locate two identical patterns.	Gs
Matrices	Select the correct option to complete a figural matrix.	Gf
Number-pattern matching	In a set of numbers, rapidly locate two identical numbers.	Gs
Numbers reversed	Recall a sequence of aurally presented digits in reverse order.	Gwm
Oral vocabulary	Provide synonyms and antonyms of aurally presented words.	Gc
Phonemic word retrieval	Rapidly generate words based on phonemic cues.	Gr
Semantic word retrieval	Rapidly generate words based on semantic cues.	Gr
Spatial relations	Identify pieces that together form a target shape.	Gv
Story comprehension	Answer comprehension questions about aurally presented stories.	Gl
Story recall	Recall complex, aurally presented stories.	Gl
Verbal analogies	Complete logical word relationships.	Gc
Verbal attention	Answer questions about aurally presented series of animals and digits.	Gwm

this subset, we again extracted descriptive statistics and additionally estimated the correlation between the number of subtests included in the g-factor model and the magnitude of the correlation with *g*-complete.

Hierarchical Model. The analytic procedure for the hierarchical model was based on the same logic as the one-factor model. First, all possible combinations of at least three WJ V COG broad abilities were generated, resulting in 99 unique combinations. We then estimated the hierarchical model (bottom panel of Figure 1) based on each combination using the lavaan package (Rosseel, 2012), exported *g*-factor scores, and computed the correlation matrix. This resulted in a matrix of 4,851 correlations. The descriptive statistics were computed as described above. Again, we also examined the subset of 98 correlations with the most comprehensive *g*-factor (*g*-complete) that included all seven broad abilities. We computed the descriptive statistics and the correlation between the number of broad abilities included in the model and the magnitude of the correlation with *g*-complete.

Results

One-Factor Model

The correlations between the different *g*-factors in the one-factor model ranged from $r = .360$ to $r = 1{,}000$, with a median correlation of $r = .890$ (mean = .865; $SD = .087$). The distribution of correlations is visualized in the top panel of Figure 2, and the distribution within 0.1-wide bands is presented in Table 2. Almost half of the correlations exceeded $r = .9$ (44.1%), whereas 7.1% of correlations fell below $r = .8$.

The correlations with *g*-complete ranged from $r = .616$ to $r = .998$, with a median of $r = .949$ (mean = .929; $SD = .060$). This distribution is illustrated in the bottom panel of Figure 2. Table 2 presents the distribution within 0.1-wide bands. Most correlations exceeded $r = .9$ (82.5%), with

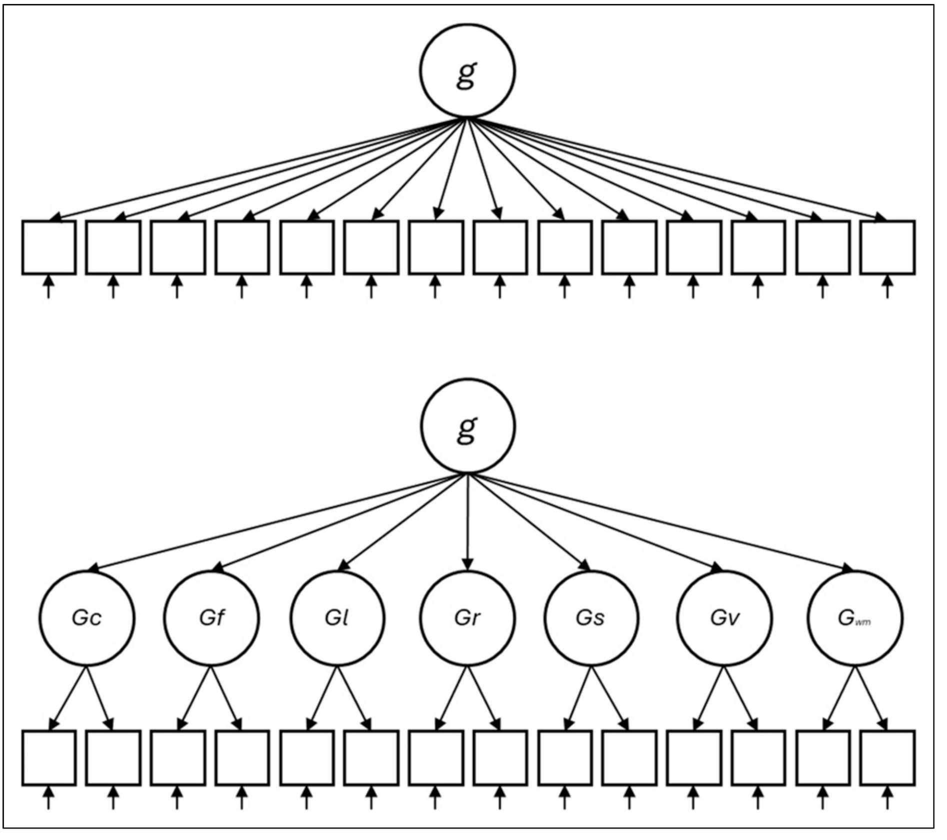

Figure 1. Illustrations of the One-Factor Model (Top) and Hierarchical Model (Bottom)

only 6% falling below r = .8. The correlation with g-complete increased significantly with an increasing number of included subtests (r = .602, p < .001). Figure 3 visualizes this relationship.

The histograms in Figure 3 show that for the g-factors based on smaller numbers of subtests (i.e., 3–6), there is a small group of correlations with g-complete that are much lower than the others. To understand this pattern, we conducted an exploratory analysis in which we separated these low correlations (r < .7 for 3 or 4 subtests, r < .8 for 5 or 6 subtests) from the others and examined the frequencies of all specific subtests underlying the g-factors of the different sets. The results are shown in Table 3, where a clear pattern emerges. Virtually all of the very low correlations are from g-factors that include both Gs subtests. In other words, when a g-factor is based on a low number of subtests, a strong representation of Gs subtests substantially lowers the correlation with g-complete.

Hierarchical Model

The correlations between the different g-factors in the hierarchical model ranged from r = .645 to r = .995, with a median correlation of r = .898 (mean = .886; SD = .067). The distribution of correlations is visualized in the top panel of Figure 4. Table 2 shows the distribution within 0.1-wide bands. Again, almost half of the correlations exceeded r = .9 (48.9%), whereas 12.5% of correlations fell below r = .8.

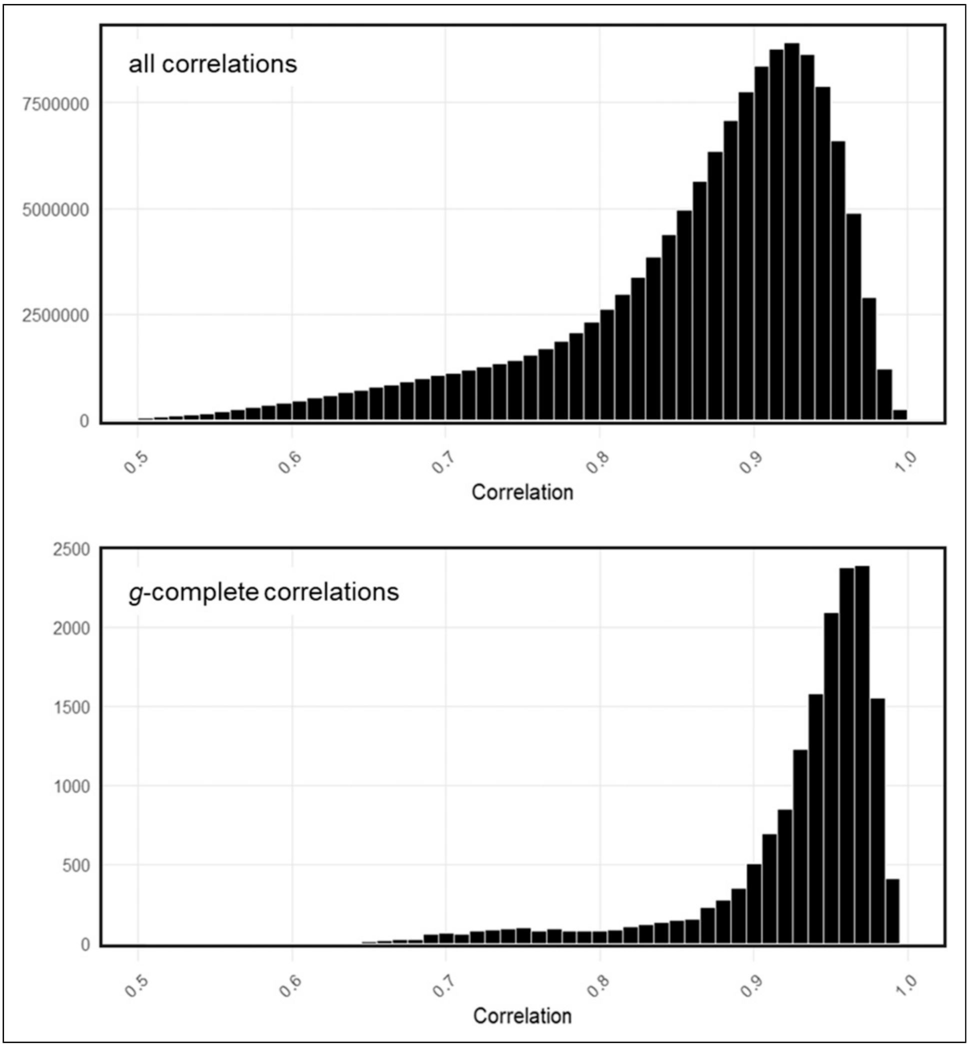

Figure 2. Distribution of the g-Factor Correlations Based on the One-Factor Model

The correlations with *g*-complete ranged from $r = .842$ to $r = .995$, with a median of $r = .947$ (mean = .941; $SD = .035$). The distribution of correlations is visualized in the bottom panel of Figure 4. Table 2 presents the distribution within 0.1-wide bands. A substantial majority of correlations exceeded $r = .9$ (83.7%), and no correlations fell below $r = .8$. Similar to the one-factor

Table 2. Distribution of g-Factor Correlations in Percentages

Magnitude of *r*	One factor model: All correlations	One factor model: *g*-complete correlations	Hierarchical model: All correlations	Hierarchical model: *g*-complete correlations
.90–1.00	44.09	82.52	48.90	83.67
.80–.90	36.95	11.48	38.57	16.33
.70–.80	11.85	4.98	11.87	
.60–.70	5.58	1.03	0.66	
.50–.60	1.45			
.40–.50	0.08			
.30–.40	<0.01			

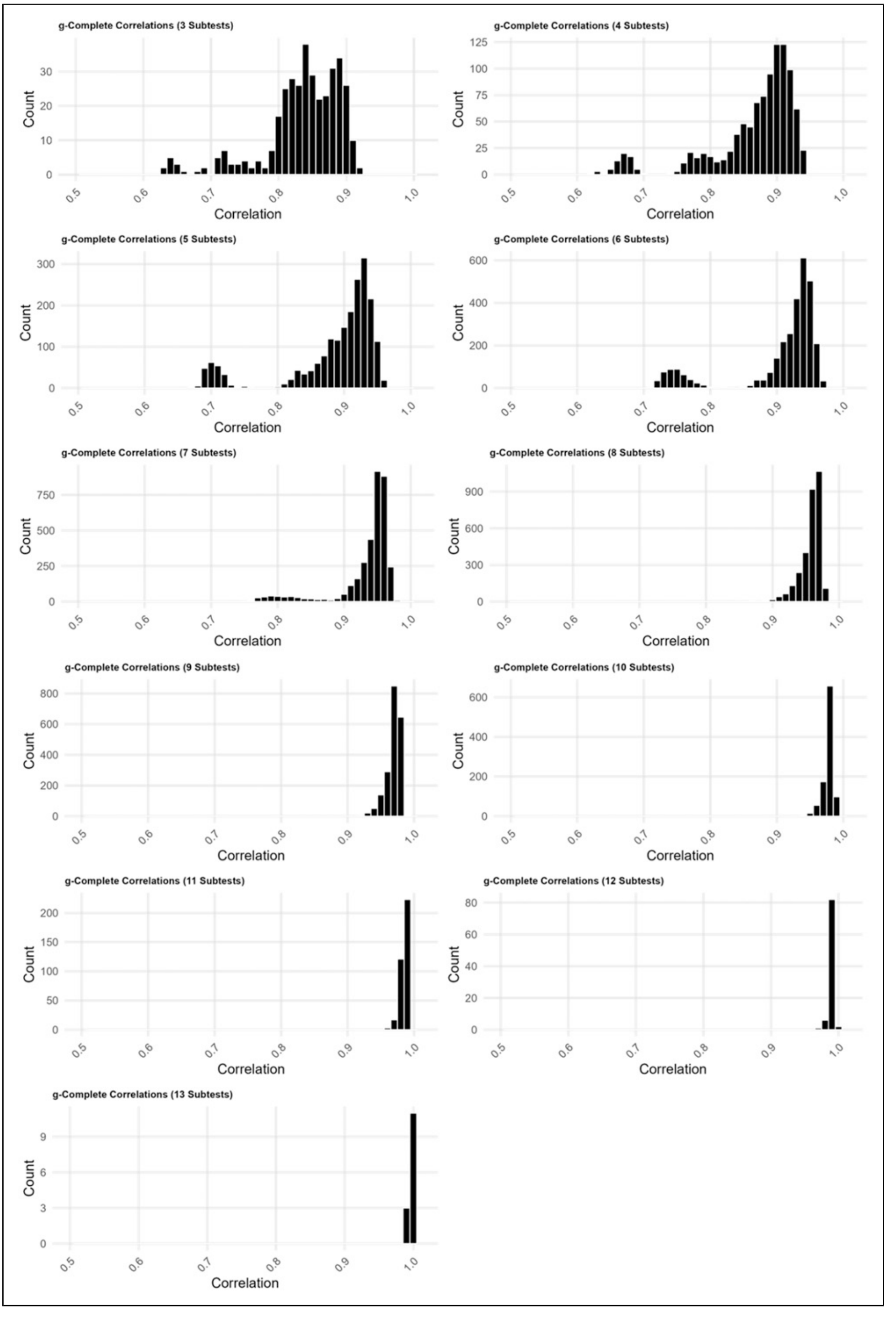

Figure 3. Distribution of the g-Complete Correlations Based on the One-Factor Model for Different Numbers of Included Subtests

Table 3. Frequency of Subtests Included in g-Factors that Correlate Low or High with g-Complete, for g-Factors Based on 3 to 6 Subtests

Subtest	Broad ability	% included low*	% included high	Difference
Oral vocabulary	Gc	24.38	38.81	−14.43
Verbal analogies	Gc	24.11	38.83	−14.72
Analysis-synthesis	Gf	29.18	38.19	−9.01
Matrices	Gf	30.27	38.03	−7.76
Story comprehension	Gl	29.45	38.15	−8.70
Story recall	Gl	28.77	38.22	−9.45
Phonemic word retrieval	Gr	30.96	37.96	−7.00
Semantic word retrieval	Gr	31.37	37.89	−6.52
Letter-pattern matching	Gs	99.45	29.07	70.38
Number-pattern matching	Gs	99.59	29.05	70.54
Block rotation	Gv	29.59	38.13	−8.54
Spatial relations	Gv	30.55	38.01	−7.46
Numbers reversed	Gwm	29.45	38.15	−8.70
Verbal attention	Gwm	29.18	38.19	−9.01

Note. *r < .7 for 3 or 4 subtests; r < .8 for 5 or 6 subtests.

model, the correlation with *g*-complete increased significantly with an increasing number of broad abilities included (r = .761, p < .001). This relationship is visualized in Figure 5.

Discussion

We investigated the generalizability of the general factor of intelligence (g) using the Woodcock–Johnson Tests of Cognitive Abilities administered to a large, nationally representative sample of the U.S. population. Two structural models were employed: a one-factor model and a higher-order (hierarchical) model aligned with the Cattell-Horn-Carroll (CHC) theory of cognitive abilities. To examine the robustness of the extracted *g*-factors, we systematically generated all possible combinations of subsets or abilities permitted by the model constraints (i.e., at least three indicators of *g*). We then analyzed (a) the correlations among *g*-factors derived from each of these combinations ("batteries") and (b) the correlation between each derived *g*-factor with the *g*-factor obtained from the full battery of all available tests or abilities ("*g*-complete"). If *g*-complete is interpreted as the most comprehensive and optimal estimate of the general factor of intelligence, then the correlation between the *g* extracted from each battery and *g*-complete provides a practical indicator of the quality or fidelity of that battery's *g*.

Our findings show that a greater number of subtests generally results in stronger correlations with *g*-complete. Reliably generating a high-fidelity estimate of *g* (which we define as a correlation greater than .90 with *g*-complete) required one of the following: (a) using at least six subtests in the one-factor model, with no more than one measuring Gs, or (b) using a hierarchical model with at least three, ideally four or more broad cognitive abilities. Both options imply that three or more broad abilities are present in the battery. These findings can inform applied assessments and scientific studies in which a high-quality *g*-factor estimate is of interest.

We also found that in certain circumstances, including Gs subtests in the battery resulted in a *g*-factor that correlated much more weakly with *g*-complete than other subtest combinations. Specifically, this occurred when six or fewer subtests were used in the one-factor model and both Gs subtests were included. This finding suggests that a strong representation of Gs in the set of

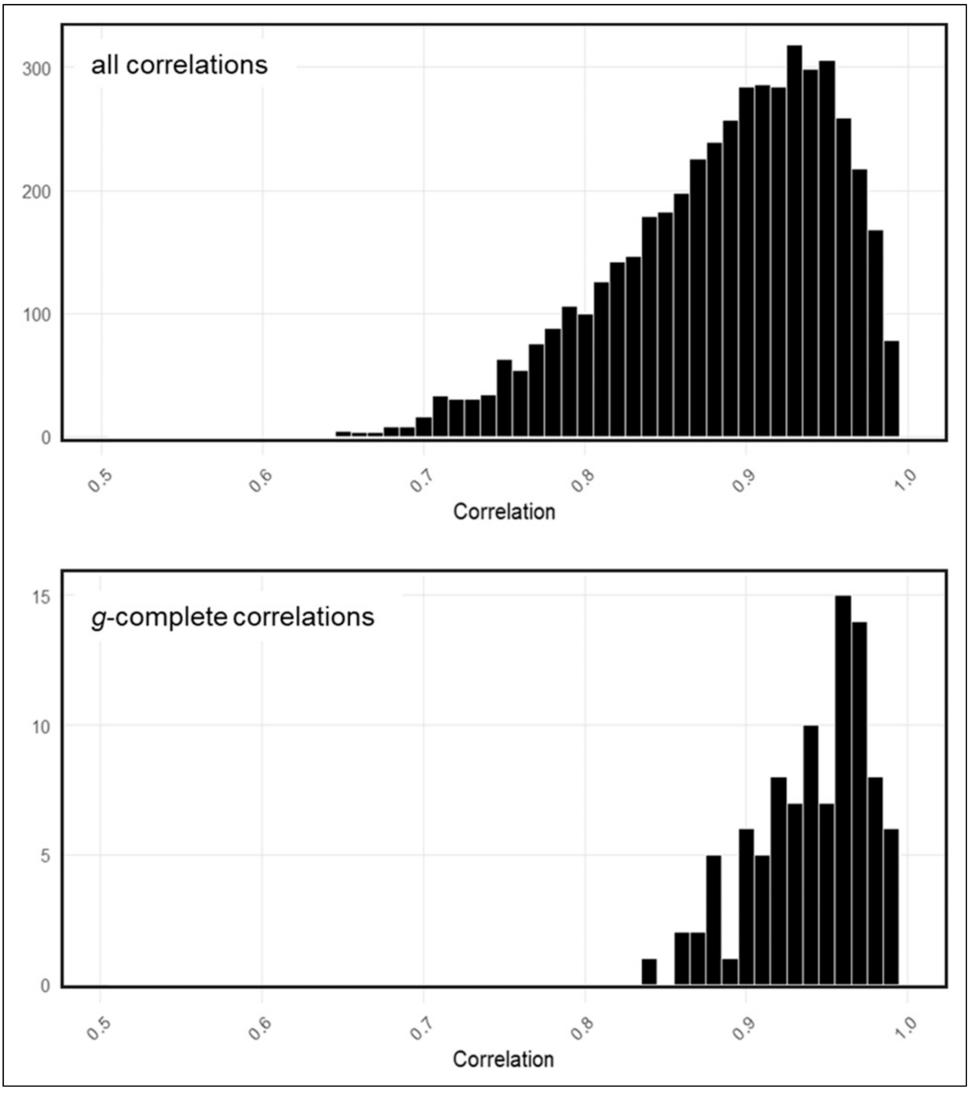

Figure 4. Distribution of the g-Factor Correlations Based on the Hierarchical Model

subtests used to estimate *g* leads to a *g*-factor that is heavily biased towards Gs, reducing the correlation with *g*-complete (although this correlation remains substantially positive). This phenomenon seems to be avoided by ensuring that Gs subtests represent less than a third of the indicators in the one-factor model or by using a hierarchical factor model.

Following this result, Gs should not be very strongly represented in cognitive batteries (i.e., less than one-third of the indicators in the one-factor model) if the purpose is to provide a general ability estimate. However, the visual search tasks used in the Woodcock–Johnson Tests to measure Gs represent only a subset of the tasks that can be used to meaningfully operationalize the processing speed factor. Thus, it is questionable whether our findings generalize to other operationalizations of Gs. In any case, this finding is certainly a warning that it is possible to select tasks and task combinations that lead to very different, and probably biased, *g*-estimates compared to most others. These results are particularly important because, according to a survey of factor

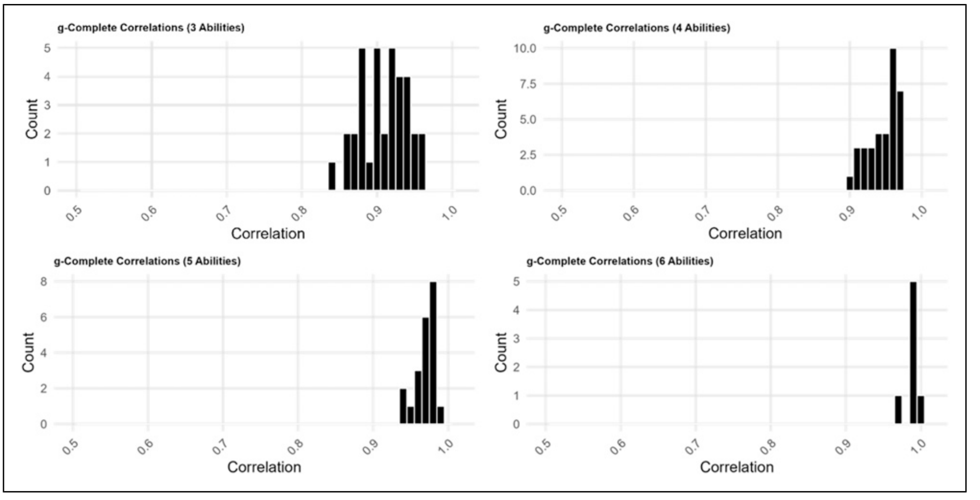

Figure 5. Distribution of the g-Complete Correlations Based on the Hierarchical Model for Different Numbers of Included Abilities

analytic practices, Gs is among the five most frequently assessed cognitive abilities when obtaining an estimate of *g*, alongside Gc, Gf, Gv, and Gq (Reeve & Blacksmith, 2009b). Generally, our findings underscore the importance of sampling broadly across the spectrum of mental abilities where at least three, but optimally more broad abilities, are represented.

Our results have direct implications for assessment with the WJ V in general and for the interpretation of test scores in particular. Overall indicators of general ability, such as the GIA and BIA, should be used with caution. In the GIA, each broad CHC ability is assessed with a single test. While a broader representation of the broad CHC-abilities may be desirable, this index meets the minimum requirements that we identified. Conversely, the BIA is a brief index that measures only three broad abilities (Gf, Gc, and Gwm) with one test each. This combination meets only the lowest limit of the requirements for g-factor estimates that we identified. Moreover, this set of abilities misses important abilities in general and broad visual ability (Gv in the CHC model) in particular, which is a very important yet frequently overlooked aspect of human cognition (Lubinski, 2010; Webb et al., 2007). Therefore, the BIA is a very crude indicator of overall intellectual functioning and should not be used for impactful diagnostic decisions or for designing interventions. However, the most problematic cases in practice are self-made short-form composite scores using only two or three, and sometimes just one, WJ subtests to estimate general ability. Most of these composites violate the minimum requirements for a g-factor estimate, leading to test scores that likely do not sufficiently correlate with a comprehensive g-factor estimate.

Theoretical Implications

Regarding the different perspectives on g, our results support the "intermediate stance," which states that general factors can be extracted from highly correlated batteries, but only if the batteries are large enough to cover a sufficient breadth of human cognitive abilities. That is, our results are in agreement with both the Bonds Model of intelligence and Process Overlap Theory, which predict that general factors obtained from different batteries are nearly identical as long as the batteries are sufficiently comprehensive and balanced.

Our results clearly do not align with a strong interpretation of the principle of the indifference of the indicator, which states that even a very narrow range of tests (maybe even just a single test) can be used as a proxy for *g*. Based on our results, we can safely advise against interpreting the result of any single test as an adequate proxy for *g*. If one administers the Raven's Progressive Matrices, they are assessing fluid reasoning; if one administers the Peabody Vocabulary Test, they are assessing comprehension-knowledge (Gc), etc. In order to estimate *g*, one needs a broad and large battery.

However, our results do not align with the stance that the general factor is a meaningless construct either—one that completely depends on the selection of tests used to obtain it. When *g* is estimated from a large and diverse battery where Gs is not overrepresented, general factors can indeed be generalized.

Limitations and Directions for Future Research

The present study has some limitations that need to be considered when interpreting the results. First, our models were constructed using a limited pool of 14 tests. Different operationalizations of the included CHC abilities and the inclusion of additional CHC abilities (e.g., Ga and Gq) may alter the results, raising questions about the generalizability of our findings and the need for replication studies.

Second, our hierarchical model was limited to two indicators per broad ability. Therefore, we could not examine the effect of varying the number of first-order factor indicators. This limits the practical implications we can derive from our results to the number of broad abilities included in the hierarchical model.

Third, we focused on the quantitative requirements that allow *g*-factors to be generalized. However, our results also required us to discuss substantive issues, such as including processing speed when obtaining *g* estimates. In a companion paper currently in preparation, we address these issues in greater detail. We investigate whether different *g*-factors result depending on the specific abilities included in the model; in particular, we examine whether it is possible to meaningfully differentiate verbal *g* and nonverbal *g*.

Finally, this paper focused solely on psychometric g: the general factors obtained from correlation matrices of cognitive tests. Although our results may have implications for a proposed psychological g—a hypothetical psychological or biological mechanism that psychometric g reflects—this paper does not address this directly.

ORCID iDs

Moritz Breit https://orcid.org/0000-0002-5983-2691
Andrew R. A. Conway https://orcid.org/0000-0002-6441-9930
Kristof Kovacs https://orcid.org/0000-0002-7452-067X

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: Kristof Kovacs received funding from the National Research, Development and Innovation Office of Hungary: Grant FK-21-138971.

Declaration of conflicting interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

References

Bartholomew, D. J., Deary, I. J., & Lawn, M. (2009). A new lease of life for Thomson's bonds model of intelligence. *Psychological Review*, *116*(3), 567–579. https://doi.org/10.1037/a0016262

Borsboom, D., Mellenbergh, G. J., & van Heerden, J. (2003). The theoretical status of latent variables. *Psychological Review*, *110*(2), 203–219. https://doi.org/10.1037/0033-295X.110.2.203

Floyd, R. G., Shands, E. I., Rafael, F. A., Bergeron, R., & McGrew, K. S. (2009). The dependability of general-factor loadings: The effects of factor-extraction methods, test battery composition, test battery size, and their interactions. *Intelligence*, *37*(5), 453–465. https://doi.org/10.1016/j.intell.2009.05.003

Gould, S. J. (1996). *The mismeasure of man*. WW Norton & Company.

Horn, J. L. (1989). Models of intelligence. In *Intelligence: Measurement, theory, and public policy* (pp. 29–73). University of Illinois Press.

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

Jensen, A. R., & Weng, L.-J. (1994). What is a good g? *Intelligence*, *18*(3), 231–258. https://doi.org/10.1016/0160-2896(94)90029-9

Johnson, W., Bouchard, T. J., Krueger, R. F., McGue, M., & Gottesman, I. I. (2004). Just one g: Consistent results from three test batteries. *Intelligence*, *32*(1), 95–107. https://doi.org/10.1016/S0160-2896(03)00062-X

Johnson, W., Nijenhuis, J. T., & Bouchard, T. J. (2008). Still just 1 g: Consistent results from five test batteries. *Intelligence*, *36*(1), 81–95. https://doi.org/10.1016/j.intell.2007.06.001

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*(1), 89–119. https://doi.org/10.1080/02796015.2001.12086102

Kovacs, K., & Conway, A. R. A. (2016). Process overlap theory: A unified account of the general factor of intelligence. *Psychological Inquiry*, *27*(3), 151–177. https://doi.org/10.1080/1047840X.2016.1153946

Kovacs, K., & Conway, A. R. A. (2019). What is IQ? Life beyond "general intelligence". *Current Directions in Psychological Science*, *28*(2), 189–194. https://doi.org/10.1177/0963721419827275

Li, N. P., & Kanazawa, S. (2016). Country roads, take me home. . . To my friends: How intelligence, population density, and friendship affect modern happiness. *British Journal of Psychology*, *107*(4), 675–697. https://doi.org/10.1111/bjop.12181

Lubinski, D. (2010). Spatial ability and STEM: A sleeping giant for talent identification and development. *Personality and Individual Differences*, *49*(4), 344–351. https://doi.org/10.1016/J.PAID.2010.03.022

Major, J. T., Johnson, W., & Bouchard, T. J. (2011). The dependability of the general factor of intelligence: Why small, single-factor models do not adequately represent g. *Intelligence*, *39*(5), 418–433. https://doi.org/10.1016/j.intell.2011.07.002

McGrew, K., Mather, N., & LaForte, E. M. (2025). *Woodcock-Johnson V tests of cognitive abilities*. Riverside Assessments, LLC.

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), 1–10. https://doi.org/10.1016/j.intell.2008.08.004

Ree, M. J., & Earles, J. A. (1991). The stability of g across different methods of estimation. *Intelligence*, *15*(3), 271–278. https://doi.org/10.1016/0160-2896(91)90036-D

Reeve, C. L., & Blacksmith, N. (2009a). Equivalency and reliability of vectors of g-loadings across different methods of estimation and sample sizes. *Personality and Individual Differences*, *47*(8), 968–972. https://doi.org/10.1016/j.paid.2009.07.025

Reeve, C. L., & Blacksmith, N. (2009b). Identifying g: A review of current factor analytic practices in the science of mental abilities. *Intelligence*, *37*(5), 487–494. https://doi.org/10.1016/j.intell.2009.06.002

Rosseel, Y. (2012). lavaan: An *R* package for structural equation modeling. *Journal of Statistical Software*, *48*(2). https://doi.org/10.18637/jss.v048.i02

Spearman, C. (1927). *The abilities of man: Their nature and measurement*. Macmillan. https://books.google.com/books?id=ws78nQEACAAJ&pgis=1

Thorndike, Robert L. (1987). Stability of factor loadings. *Personality and Individual Differences*, *8*(4), 585–586. https://doi.org/10.1016/0191-8869(87)90224-8

Thurstone, L. L. (1947). *Multiple-factor analysis*. University of Chicago Press.

Vernon, P. A. (1989). The generality of g. *Personality and Individual Differences*, *10*(7), 803–804. https://doi.org/10.1016/0191-8869(89)90129-3

Webb, R. M., Lubinski, D., & Benbow, C. P. (2007). Spatial ability: A neglected dimension in talent searches for intellectually precocious youth. *Journal of Educational Psychology*, *99*(2), 397–420. https://doi.org/10.1037/0022-0663.99.2.397