Intelligence 55 (2016) 69–78

Contents lists available at ScienceDirect

Intelligence

Residual group-level factor associations: Possibly negative implications for the mutualism theory of general intelligence

Gilles E. Gignac

School of Psychology, University of Western Australia, 35 Stirling Highway, Crawley, Western Australia 6009, Australia

ARTICLE INFO

Article history:
Received 2 July 2015
Received in revised form 15 January 2016
Accepted 18 January 2016
Available online xxxx

Keywords:
General intelligence
Mutualism
Factor analysis

ABSTRACT

The mutualism theory of general intelligence (*g*) posits that the positive manifold arises because of mutually beneficial interactions between originally orthogonal cognitive abilities, rather than because of a genuine general construct. In a recent investigation, Gignac (2014) reported that the strength of *g* was largely constant from the ages of 2.5 to 90 years, which was interpreted as an indirect failure to confirm the mutualism theory of *g*. In this investigation, a second indirect test of the mutualism theory of *g* was performed. Specifically, it was hypothesized that, if the extended mutualism theory of *g* is plausible, then there should be some consistent, positive associations between group-level factors, controlling for the effects of a general factor. To examine this possibility, the associations between cognitive ability group-level factors were estimated across a series of seven relatively large and relatively representative samples of intelligence battery data. Next, seven single-factor models were estimated against the seven group-level inter-correlation matrices. The corresponding single-factor residual correlation matrices were observed to yield an approximately equal number of positive and negative residual correlations and an overall mean of zero. Furthermore, the only moderately consistent residual effect was a negative association between crystallised intelligence and processing speed. Although only an indirect test, the results are interpreted to be more supportive of *g* factor theory than mutualism.

For many years, debates about the validity of the general factor of intelligence (*g*) revolved around statistical issues in the context of unrestricted factor analysis. For example, evidence in favor or against the *g* factor could be variously reported, contingent upon the decision to rotate a solution orthogonally, obliquely, or not all (Humphreys, 1979). With the advent of structural equation modeling (SEM), the thorny issue of rotation was circumvented, as the researcher was required to specify a particular structure: a structure which could then be tested against data and evaluated for plausibility, in part, via model fit indices (Bollen, 1989). Based on a number of empirical SEM investigations, there is now broad consensus on the plausibility of a psychometric *g* factor (Sternberg, 2003). However, the precise nature of *g* remains a contentious issue. In particular, some contend that *g* is a genuine construct representative of a meaningful and important human characteristic (Gottfredson, 2002; Jensen, 1998). By contrast, others have contended that the psychometric *g* factor is essentially an epiphenomenon that emerges during development: the mutualism theory of *g* (van der Maas et al., 2006). The purpose of this investigation was to test empirically these two competing theories via an indirect method: residual group-level factor inter-associations.

E-mail address: gilles.gignac@uwa.edu.au.

http://dx.doi.org/10.1016/j.intell.2016.01.007

1. *g* Factor theory vs. mutualism theory of *g*

The *g* factor was originally proposed as a causal explanation for the positive manifold, consistently positive correlations between a diverse collection of cognitive ability tests (Spearman, 1904). Contemporary representations of *g* factor theory do not assert that there is a single factor of intelligence, as originally contended by Spearman (1904, 1923). Instead, it is widely acknowledged that there are approximately 10 group-level factors of intelligence, in addition to *g* (Carroll, 2003). Although the group-level factors of intelligence are a source of active research (e.g., Brunner, 2008; Crawford, Deary, Allan, & Gustafsson, 1998; Frisby & Beaujean, 2015; Tommasi et al., 2015), the *g* factor has been observed to be the most dominant construct associated with a battery of cognitive ability tests, as well the most important with respect to predictive validity (Gottfredson, 2002; Jensen, 2006).

Spearman (1927) theorised that *g* arises due to individual differences in mental energy. Picking up form Spearman (1927); Lykken (2005) contended that *g* arises from individual differences in the ability for sustained concentration, which, presumably, would be aligned closely with the expenditure of mental energy. Based on a series of factor and multidimensional analyses, Marshalek, Lohman, and Snow (1983) argued that the fundamental basis of *g* was cognitive complexity. Jensen (1998, 2006) contended that *g* is probably best not conceptualised as a process. Instead, it is best considered a reflection of physical properties of the brain, which have yet to be identified.

Correspondingly, psychometric *g* has been reported to be approximately 60% heritable in young adulthood (Haworth et al., 2010). It has also been reported to correlate approximately .30 with brain volume (McDaniel, 2005; Ritchie et al., 2015). Finally, functional neural imaging studies have found that brains that can process information more efficiently tend to be associated with individuals with higher levels of *g* (Haier, 2011).

Several alternatives to *g* factor theory have been articulated over the years (e.g., Anderson, 1992; Bartholomew, Allerhand, & Deary, 2013; Gardner, 1987; Sternberg, 1985; Thomson, 1951). Perhaps the most compelling alternative is the mutualism theory of *g* (van der Maas et al., 2006). The mutualism theory of *g* postulates that the positive manifold arises not because of some general process, but, instead, due to the mutually beneficial interactions that emerge during development between originally orthogonal cognitive processes. Consequently, as the capability of one cognitive process grows, so do others in a mutually beneficial manner through reciprocal causation.

The empirical establishment of the mutualism theory of *g* would have profound implications for the manner in which intelligence is conceptualised and modeled, as it disputes the notion that the positive manifold is caused by a construct (van der Maas et al., 2006). Consequently, a general factor latent variable is considered an inappropriate representation of the positive manifold, according to the mutualism theory of *g*. Instead, a formative model is considered a more acceptable representation of the positive manifold (van der Maas, Kan, & Borsboom, 2014). A formative model does not posit the presence of an entity theorised to explain the observation of a pattern of correlations between subtests (i.e., a construct). Instead, in a formative model, an optimally weighted sum of score based on a battery of subtests is simply a convenient method to create an index. Thus, the observation of a factor with positive loadings of varying magnitudes across several cognitive ability subtests is considered completely meaningless, according to the mutualism theory of *g*. Consequently, an overall intelligence index score based on an optimally weighted sum of cognitive ability subtest scores should be viewed as no more theoretically meaningful than the composite value associated with the Dow Jones Industrial Average index, for example (van der Maas et al., 2014).

Evidently, *g* factor theory and the mutualism theory of *g* are at odds with each other in a fundamental manner. One of the limitations associated with the mutualism theory of *g* is that there are no known methods to test the theory empirically, at least not directly (van der Maas et al., 2006). However, one implication of the mutualism theory of *g* is that the strength of the positive manifold should increase across development, particularly early development, as the mutually beneficial interactions begin to emerge (van der Maas et al., 2006). To test this possibility, Gignac (2014) estimated the strength of the *g* factor (omega hierarchical; Ω_h) across the Wechsler batteries from the ages of 2.5 to 90 years ($N = 5200$). Gignac (2014) reported a largely equally strong *g* factor across age, although the Ω_h index did suggest some increase in the strength of the *g* factor from the ages of 2.5 to 7 years. However, when controlling for the number of subtests within a battery (a characteristic which impacts Ω_h positively), the reduction in the strength of the positive manifold was no longer observed. In fact, the strength of the positive manifold was observed to decrease from about 2.5 to 7 years, which would be inconsistent with the mutualism theory of *g*. Gignac (2014) interpreted the results of his investigation to suggest that the indirect empirical test failed to support the mutualism theory of *g*. Gignac (2014) encouraged others to generate additional approaches to test the mutualism theory of *g*, even if indirectly, in order to evaluate the theory from alternative perspectives.

2. A proposed second indirect test

As a fundamental feature associated with the dynamic mutualism theory of *g* pertains to the mutually beneficial inter-actions between cognitive abilities, it may be suggested that, in order to help evaluate the plausibility of the theory, some focus should be placed upon an examination of the inter-associations between the group-level factors. More specifically, if the mutualism theory of *g* is plausible, then it would seem very unlikely that the postulated pattern of mutually beneficial interactions between group-level factors occur in a manner that is captured perfectly, or nearly so, by a single factor. Instead, one would expect *some* level of shared variance between one or more pairs of group-level factors, independently of a general factor. Furthermore, most of the residual group-level factor associations would be expected to be positive in direction, rather than negative, if the group-level factor associations are the result of mutually beneficial interactions. Admittedly, the conventional correlated factor model is not an entirely accurate representation of mutualism (van der Maas et al., 2006), as only the associations between the group-level factors are estimated in a correlated factor model, rather than their interactions. If there were a SEM model that could be specified to represent the mutualism theory of *g*, the theory could be tested directly. It is in this context that the proposed test described here is considered to be an indirect test of the extended mutualism theory of *g* (see Fig. 1c, van der Maas et al., 2006).

Somewhat surprisingly, very little systematic research relevant to the inter-associations between group-level factors, independent of *g*, has been conducted. Without providing references, Humphreys (1979) asserted that when tested against a heterogeneous battery of cognitive ability tests, a general factor model yields residual correlations mostly near zero. There are, of course, many higher-order model confirmatory factor analytic investigations which have included some group-level factor residual correlations in their models of intelligence. For example, in Gustafsson's (1984) third-order factor model of intelligence, four correlated residuals were included between the lower-order factor residuals. Invariably, however, past studies have included only a small percentage of the total possible correlated group-level factor residual correlations in their higher-order models of intelligence. Moreover, the added residual correlations are typically included in a relatively ad-hoc manner, i.e., in order to simply achieve satisfactory model fit. Consequently, it is difficult, if not impossible, to evaluate the current literature with respect to the associations between group-level factors, independent of *g*, in a comprehensive or systematic manner.

In a recent investigation, Gignac and Watkins (2015) were specifically interested in the unique latent variable association between fluid intelligence and working memory capacity, independently of *g*. In particular, Gignac and Watkins (2015) hypothesized that some level of positive shared variance should be observed between the fluid intelligence and working memory capacity lower-order factor residuals, if the association between these two constructs is special, as commonly asserted (e.g., Carpenter, Just, & Shell, 1990; Fry & Hale, 1996; Oberauer, Su, Wilhelm, & Sander, 2007). Based on a higher-order model and the Wechsler Intelligence Scale for Children — V normative sample (Wechsler, 2014), Gignac and Watkins (2015) reported a non-significant fluid intelligence and working memory capacity residual correlation of $-.10$ ($p = .152$). Thus, the hypothesis of a special or unique association between fluid intelligence and working memory capacity was not considered supported. From the perspective of the mutualism theory of *g*, the absence of a unique association between fluid intelligence and working memory capacity may be considered at least somewhat inconsistent, as it may be suggested that these two constructs, in particular, would be expected to develop some level of uniquely beneficial interactions over the course of human development. As it was not the focus of their investigation, Gignac and Watkins (2015) did not report any results relevant to the other group-level factor inter-associations, independently of *g*.

Perhaps part of the reason why the unique associations between group-level factors, independent of *g*, have not been examined previously in a comprehensive manner is because it is not possible to identify a higher-order model which includes correlated terms between all of the lower-order factor residuals (Schmiedek & Li, 2004). Such a model

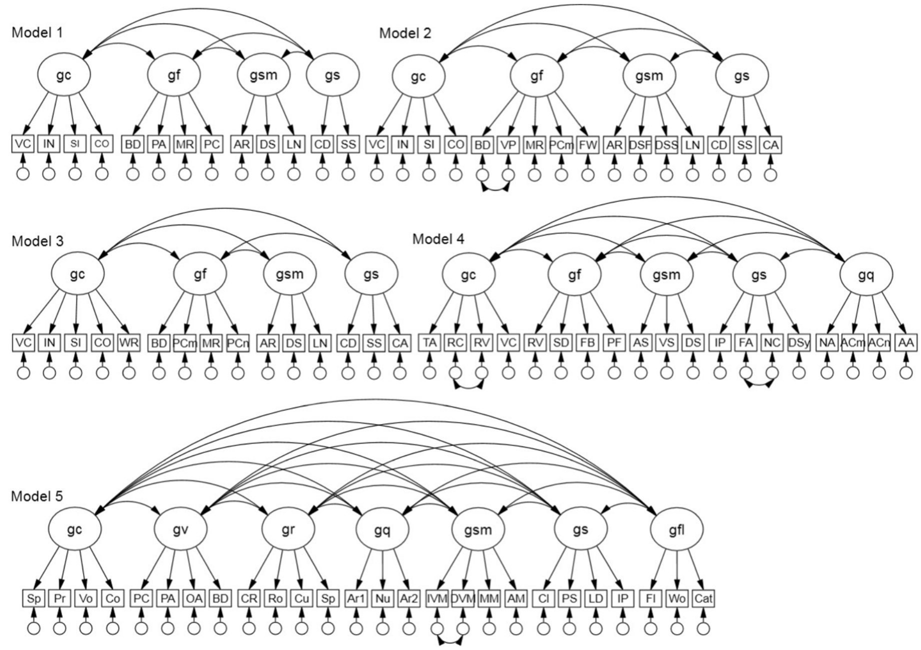

Fig. 1. Correlated factor models: Model 1 = WAIS-III; Model 2 = WAIS-IV & WISC-V; Model 3 = WISC-IV; Model 4 = Marshalek et al.; Model 5 = MISTRA; Model 6 = WJ-R; g_c = crystallised; g_f = fluid; g_{sm} = short-term memory; g_s = processing speed; g_q = quantitative reasoning; g_v = visual; g_r = mental rotation; g_{fl} = fluency.

would also not be identified in the bifactor case, where correlated terms are added between all of the nested factors (Jennrich & Bentler, 2012).

A possible solution to the identification problem, in this context, would be to estimate all of the inter-latent variable associations between the group-level factors based on a conventional correlated factor model. Then, use those inter-latent variable correlations as input to estimate (1) a general factor, and (2) the residual correlations between the group-level indicators associated with the general factor solution. Such a procedure is equivalent to the method used to estimate a higher-order model solution via unrestricted factor analysis (Gorsuch, 1974). However, for the purposes of addressing the question posed in this investigation, particular focus would be placed upon an examination of the residual correlation matrix. More specifically, if *g* factor theory is more plausible, then the residual correlation matrix would be expected to be associated with relatively small values and a mean of zero (Humphreys, 1979). Such an observation would imply that the pattern of inter-latent variable correlations between group-level factors is accounted for completely by a single phenomenon (i.e., *g*). By contrast, if the extended mutualism theory of *g* is more plausible, then at least *some* consistent, positive associations between the group-level indicators within the residual correlation matrix should be observed across samples. Such an observation would imply that a single factor fails to capture all of the covariance between group-level factors, possibly because that covariance arises from the complex, mutually beneficial interplay between the group-level factors during development, as described to occur by the extended mutualism theory of *g* (van der Maas et al., 2006). Stated alternatively, it would seem very unlikely that the pattern of mutually beneficial interactions between group-level cognitive ability factors arise *precisely* in a manner that their latent variable inter-associations can be accounted for by a single latent variable. Instead, some residual effects should remain, if the extended mutualism theory of *g* is valid. Furthermore, those residual correlations would be expected to be positive in direction, rather than negative, if the interactions are beneficial in nature.

To a large degree, the procedure described here is essentially akin to Spearman's vanishing tetrad differences test (Spearman & Holzinger, 1925). However, whereas Spearman and others attempted to prove the plausibility of *g*, and only *g*, based on the application of the tetrad differences test on the inter-*subtest* correlation matrix, a method that ultimately led to some acknowledgement of the presence of group factors (Spearman, 1927), the focus in this investigation is upon the associations between group-level factors, which is the level at which the interactions are theorised to occur in the extended mutualism model (van der Maas et al., 2006).

In light of the above, the purpose of this investigation was to estimate and test for statistical significance the residual correlations between group-level factors, controlling for the effects of a single factor, across several batteries of cognitive ability and relatively large and representative samples. Such an investigation was considered a second indirect test of the mutualism theory of *g*, alongside Gignac's (2014) indirect test. In the event that little or no positive residual correlations between group-level factors are observed, then greater support for *g* factor theory would be implied. By contrast, if some statistically significant positive residual correlations are observed consistently between some group-level factors, independently of the variance captured by a single factor, then support for the extended mutualism theory of *g* would be suggested.

3. Method

3.1. Samples & measures

All analyses were performed on seven previously published sample correlation matrices. The samples were selected based principally on the following five characteristics: (1) relatively large samples (i.e., $N = 200+$); (2) relatively representative of the population (e.g., not university students); (3) relatively well-known factor structures based on previously published investigations; (4) associated with a minimum of four group-level factors,[1] and (5) could be used to create correlated factor models with group-level factors defined by, typically, four good quality subtests.

Four of the seven correlation matrices were derived from the Wechsler scale normative samples: the Wechsler Adult Intelligence Scale — IV (WAIS-IV; Wechsler, 2008; $N = 1800$), the Wechsler Adult Intelligence Scale — III (WAIS-III; Wechsler, 1997; $N = 2450$), the

[1] A general factor based on only three indicators necessarily yields a residual correlation matrix with values of .00 exclusively.

Wechsler Intelligence Scale for Children — V (WISC-V; Wechsler, 2014; $N = 2200$), and the Wechsler Intelligence Scale for Children — IV (WISC-IV; Wechsler, 2003; $N = 2200$). All four of the Wechsler normative samples were obtained based on a stratified sampling strategy to reflect the US census results relevant to gender, age, race/ethnicity, education, and geographic location. Although the Wechsler scale technical manuals use a different nomenclature (e.g., Verbal Comprehension, Perceptual Reasoning), for the purposes of this investigation, the four Wechsler scales included in the current investigation were broadly considered to be measures of four group-level factors: crystallised intelligence, g_c (Vocabulary, Information, Similarities, Comprehension); fluid intelligence, g_f (Matrix Reasoning, Figure Weights, Picture Completion, Block Design, Visual Puzzles); short-term memory, g_{sm} (Digit Span Forward, Digit Span Backward, Digit Span Sequencing, Letter-Number Sequencing, Arithmetic); and processing speed, g_s (Symbol Search, Coding, Cancelation).[2]

Next, the correlation matrix associated with Marshalek et al.'s (1983) investigation was used ($N = 241$).[3] Unfortunately, there are few published details associated with the Marshalek et al. (1983) sample. However, it is known that the sample was comprised of high school students, which suggests that it was relatively diverse in nature. The Marshalek et al. cognitive ability test battery consisted of a total of 34 tests. However, for the purposes of this investigation, only 19 of the subtests were selected for the analysis, as the subtest battery was reported to measure a total of only five group-level factors (Marshalek et al., 1983): four subtests were selected as indicators of g_c (Terman Concept Mastery, Reading Comprehension, Reading Vocabulary, Vocabulary), four subtests were selected as indicators of g_f (Advanced Progressive Matrices, Surface Development, Paper Form Board, Paper Folding), three subtests as indicators of g_{sm} (Auditory Letter Span, Visual Number Span, Digit Span), four subtests as indicators of g_s (Identical Pictures, Find As, Number Comparison, Digit Symbol), and, finally, four subtests as indicators of g_q (Necessary Arithmetic Operations, Arithmetic Computation, Arithmetic Concepts, Arithmetic Applications).

Next, the Minnesota Study of Twins Reared Apart (MISTRA; Johnson & Bouchard, 2011) sample correlation matrix based on 433 adults (combination of twin pairs, family members, and friends; age range: 18–79) was used. Participants in the MISTRA study were administered a total of three intelligence batteries (42 subtests in total). For the purposes of this investigation, 26 of the subtests were selected to represent the seven group-level factors associated with the correlation matrix: g_c (Spelling, Proverbs, Vocabulary, Comprehension), visual intelligence, g_v (Picture Completion, Picture Arrangement, Object Assembly, Block Design), g_r (Card Rotation, Rotation, Cubes, Spatial), quantitative reasoning, g_q (Arithmetic I, Arithmetic II, Numerical), g_{sm}, (Immediate Visual Memory, Delayed Visual Memory, Meaningful Memory, Associative Memory), g_s (Closure, Perceptual Speed, Lines and Dots, Identification of Pictures), and verbal fluency, g_{fl} (Fluency, Words, Categories). It will be noted that four subtests (Raven's, Inductive, Pedigrees, Uses) were originally selected from the MISTRA data (thus, 30 out of 42) to represent an eighth group-level factor (g_f). However, the correlated factor model which included the g_f latent variable yielded a non-positive definite matrix (i.e., several of the standardised covariances with g_f were very large). Consequently, the g_f latent variable was omitted from the correlated factor model and subsequent analyses.

Finally, the age corrected normative sample ($N = 1425$) correlation matrix associated with the Woodcock-Johnson-Revised (WJ-R; McGrew, Werder, & Woodcock, 1991) published in Carroll (2003) was used. The WJ-R consists of 29 subtests designed to measure a total of eight group-level factors: g_c (Picture Vocabulary, Oral Vocabulary, Listening Comprehension, Science, Social Studies, Humanities), g_f (Analysis-Synthesis, Concept Formation, Verbal Analogies), g_{sm} (Memory for Sentences, Memory for Words, Numbers Reversed), g_s (Visual Matching, Cross Out, Writing Fluency), long-term retrieval, g_{lr} (Memory for Names, Visual–Auditory Learning, Memory for Names Delayed Recall, Visual–Auditory Learning Delayed Recall), g_v (Visual Closure, Picture Recognition, Spatial Relations), auditory processing, g_a (Word Attack, Sound Blending, Sound Patterns, Incomplete Words), and g_q (Calculation, Applied Problems, Quantitative Concepts).

3.2. Data analysis

First, a series of seven correlated factor models were specified, in order to estimate the correlations between the group-level factors (see Figs. 1 and 2). Then, the inter-correlations between the group-level factors were used as input for the purposes of testing a series of seven single-factor models (see Fig. 3). Particular focus was placed upon estimating the residual correlation matrices associated with the single-factor models. If the mutualism theory of g is plausible, it was expected that some positive residual correlations would be observed between two or more of the group-level factors across the samples. The residual correlations were estimated within the SPSS factor analysis utility by extracting a single factor (maximum likelihood estimation). In order to test the statistical significance of the residual correlations, the corresponding normalised residual correlations were estimated in Amos 21 (Arbuckle, 2012).[4] Normalised residual correlations correspond to the z-distribution. Consequently, values greater than $|1.96|$ imply statistical significance (Arbuckle, 2012).

4. Results

As can be seen in Table 1 (top half), all of the correlated factor models fit reasonably well. However, there was a greater level of model misfit associated with the three larger correlated factors models (Fig. 1, Models 4 and 5; and Fig. 2, Model 6). For the purposes of estimating the associations between the group-level latent variables, the level of fit associated with the correlated factors models was considered sufficient. Next, the single-factor models (see Fig. 3) were fitted against the group-level factor inter-latent variable correlations obtained from the correlated factor model solutions (available upon request). As can be seen in Table 1 (bottom half), none of the single factor models fit perfectly (i.e., statistically significant χ^2 values). Even from a close-fit perspective, there was a notable amount of unmodeled covariance associated with some of the models, which suggested the possibility of appreciable correlated residuals between the group-level factor indicators.[5] As can be seen in Table 2, all of the single-factor model solutions were associated with group-level indicator loadings that were positive, appreciable in magnitude, and statistically significant ($p < .001$).

With respect to the four Wechsler scales, it can be seen that there were six statistically significant residual correlations within the four residual correlation matrices (see Table 3). The residual correlation between g_c and g_s was negative and statistically significant across all four samples (WAIS-IV: $r = -.06$, $p = .038$; WAIS-III: $r = -.07$, $p = .005$; WISC-IV: $r = -.08$, $p = .002$; WISC-V: $r = -.08$, $p = .010$). Thus, controlling for individual differences in g, higher levels of processing speed were associated with lower levels of crystallised intelligence.

With respect to the Marshalek et al. sample, none of the residual correlations were found to be statistically significant (see Table 4).

[2] There are minor subtest compositional differences across the four Wechsler scales included in this investigation. For example, the WAIS-III does not include the Visual Puzzles subtest.

[3] The correlation matrix used in this investigation was obtained from a report prepared by Meng (2005).

[4] Amos 21 estimates only residual covariances and normalised residual correlations, consequently, the residual correlations were estimated in SPSS.

[5] Some of the RMSEA values appear to be rather large given the corresponding incremental fit index values. However, it has been shown that RMSEA is an invalid indicator of model fit when the degrees of freedom are very low (<5; Kenny, Kaniskan, & McCoach, 2014).

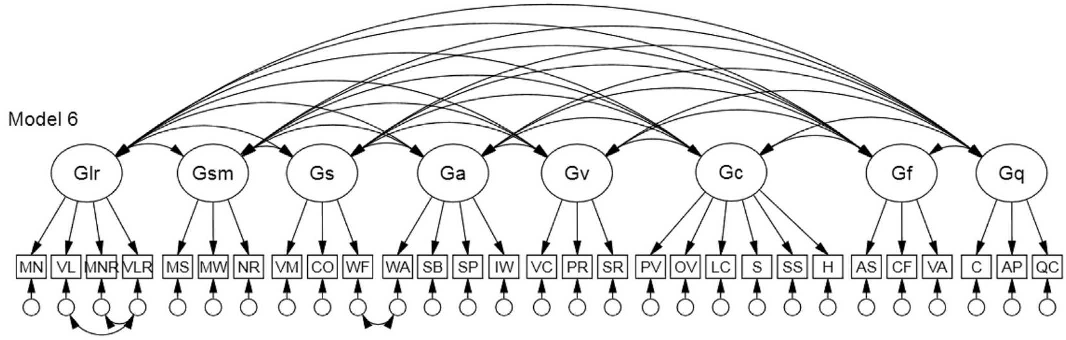

Fig. 2. Correlated factor model tested on the WJ-R data; g_a = auditory processing; g_{lr} = long-term retrieval; see Fig. 1 caption for remaining group-level factor acronym spellings.

However, it was noted that the estimated residual correlation between g_c and g_s was, again, negative in direction ($r = -.07$, $p = .308$). With respect to the MISTRA sample, three of the residual correlations were statistically significant (one negative and two positive; see Table 5). The numerically largest residual correlation at .27 ($p < .001$) was observed between two very similar group-level factors, g_v and g_r. Thus, higher levels of visual intelligence were associated with higher levels of mental rotation ability. Although not significant statistically, it was noted again that the estimated residual association between g_c and g_s was negative in direction ($r = -.09$, $p = .150$).

Finally, with respect to the WJ-R sample, six of the residual correlations were observed to be statistically significant (see Table 6). Four of those correlations were positive in direction and two were negative in direction. The numerically largest residual correlation was observed between g_{sm} and g_a at .18 ($p < .001$). Thus, higher levels of auditory processing were associated with higher levels of short-term memory and processing speed, controlling for individual differences in g. However, one of the larger statistically significant residual correlations was also negative in direction (g_{sm}/g_v $r = -.11$, $p = .013$). Thus, higher levels of short-term memory were associated with lower levels of visual/spatial intelligence, controlling for g. Additionally, the g_c and g_s residual correlation was observed to be negative and statistically significant ($-.06$, $p = .044$).

Across all seven samples (i.e., 83 residual correlations), the mean and median residual correlation was equal to exactly .00 (SD = .06). Thus, an approximately equal number of residual correlations were positive and negative in direction (33 = negative, 11 = zero, and 39 = positive). The range in residual correlations was −.14 to .27. As can be seen in Fig. 4, the distribution of residual correlations approximately reasonably closely a normal distribution. However, from a statistical perspective, the distribution was both positively skewed (skew = .97, $z = 3.73$, $p < .001$) and positively kurtotitic (kurtosis = 3.32, $z = 6.38$, $p < .001$). Excluding the relatively large .27 residual correlation between g_v and g_r, the distribution was not observed to be skewed (skew = .27, $z = 1.03$, $p = .303$).

Finally, the issue of model misfit associated with the Marshalek et al., MISTRA, and WJ-R correlated factor models was revisited with some supplementary models. Specifically, in addition to the correlated factor models reported above, corresponding higher-order models and bifactor models were estimated, in order to test the possibility that such models would fit the data better. As can be seen in Table 7, none of the higher-order models were observed to be associated with improvements in model fit, as it might be expected. However, for two out of three of the samples (Marshalek and WJ-R), the bifactor model was found to fit better than the corresponding correlated factor models, even when a penalty for model

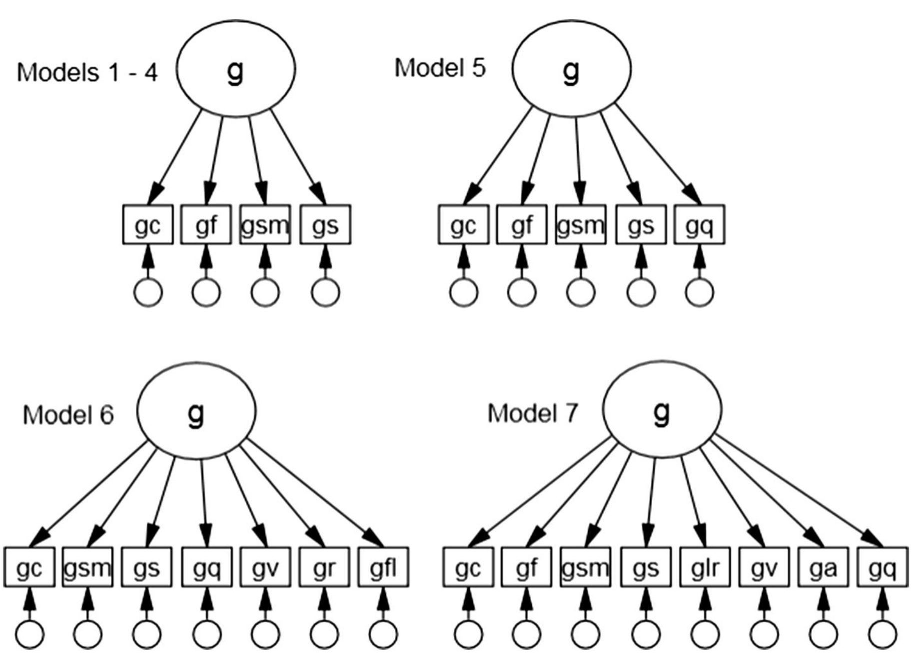

Fig. 3. Single factor models tested in this investigation (Models 1–4 = Wechsler normative samples; Model 5 = Marshalek sample; Model 6 = MISTRA sample; Model 7 = WJ-R normative sample; see Figs. 1 and 2 captions for acronym spellings.

Table 1
Correlated factor model fit statistics and indices.

	χ^2	*df*	SRMR	RMSEA	CFI	TLI
Correlated factor models						
WAIS-IV	965.93	112	.042	.065	.949	.938
WAIS-III	592.35	59	.029	.061	.972	.962
WISC-IV	625.69	84	.034	.054	.967	.958
WISC-V	534.80	112	.029	.041	.973	.968
Marshalek et al.	342.88	141	.057	.077	.931	.916
MISTRA	811.23	277	.057	.067	.904	.888
WJ-R	1804.62	346	.056	.054	.927	.915
Single-factor models						
WAIS-IV	56.88	2	.020	.124	.989	.913
WAIS-III	240.94	2	.025	.224	.971	.913
WISC-IV	228.10	2	.031	.227	.905	.968
WISC-V	85.35	2	.025	.138	.958	.986
Marshalek et al.	24.04	5	.040	.126	.970	.939
MISTRA	584.62	14	.080	.307	.798	.697
WJ-R	1246.45	20	.053	.208	.837	.883

Note. SRMR = standardised root mean residual and RMSEA = Root Mean Square Error of Approximation (<.080 indicative of acceptable fit); CFI = comparative fit index and TLI = Tucker–Lewis index (values of .950 or greater indicative of good fit); AIC = Akaike Information Criterion (small values indicate better fit); both AIC and TLI incorporate a penalty for model complexity.

complexity was applied: Marshalek, ΔAIC = −66.16, ΔTLI = .028; WJ-R, ΔAIC = −326.67, ΔTLI = .019.[6]

5. Discussion

Across the seven single-factor models tested in this investigation, there were 16 statistically significant residual correlations (out of 83) within the corresponding residual correlation matrices. However, there were approximately an equal number of positively and negatively directed residual correlations. Furthermore, the 83 residual correlations were associated with a mean of .00. Finally, the only moderately consistent residual correlation was a negative association between g_c and g_s. Thus, overall, the results of this investigation may be suggested to be supportive of *g* factor theory (Humphreys, 1979; Jensen, 1998) and unsupportive of the extended mutualism theory of *g* (van der Maas et al., 2006).

It will be noted that, although 16 of the 83 residual correlations were reported to be statistically significant, no adjustment was implemented to manage the increased likelihood of Type I errors due to the relatively large number of estimated correlations. It was considered relatively harsh to implement a Bonferroni correction, in light of the fact that some researchers implement the adjustment in their investigations, while many others do not. Based on a Bonferroni adjustment, only six of the 83 residual correlations (three negative and three positive) were identified as statistically significant ($p < .05$). Thus, from a more conservative perspective, the lack of support for the mutualism theory of *g* may be underscored by the Bonferroni adjusted results, as very few of the residual correlations may be expected to have arisen for reasons other than sampling fluctuations.

Although there was one relatively large positive residual correlation ($r = .27$) between two group-level factors (g_v and g_r; MISTRA sample), it was observed between two very similar lower-order cognitive ability constructs: visual intelligence and mental rotation. Although there is some evidence to suggest that visual intelligence and mental rotation are, to some degree, distinct constructs (e.g., Johnson & Bouchard, 2005; Kaufman, 2007), they are probably much more closely aligned psychometrically than other group-level factors. Consequently, the relatively large residual correlation of .27 may have been observed because g_v and g_r are fundamentally similar in nature, rather than because of mutually beneficial interactions that occur during development.

Excluding the g_v and g_r residual correlation, there were no appreciable and/or consistently positive residual correlations observed between any of the group-level factors. Thus, the results of this investigation may be viewed as an extension of Gignac and Watkins (2015) who failed to observe a unique association between g_f and g_{sm}, based on a hybrid higher-order modeling strategy. However, whereas Gignac and Watkins (2015) interpreted their results to suggest that there may be nothing special about the association between g_f and g_{sm}, a commonly asserted position (see Conway & Kovacks, 2013), the more comprehensive results of this investigation suggest that the mutualism theory of *g* (van der Maas et al., 2006) may not be an accurate account of the positive manifold.

It should be made clear that the results of this investigation do not suggest that the positive correlations between cognitive ability *tests* can be accounted for by a single factor. There is previously reported evidence for the presence of group-level factors, independent of *g*, associated with the data analyzed in this investigation (see Carroll, 2003; Gignac, 2006; Gignac & Watkins, 2013, 2015; Johnson, Bouchard, Krueger, McGue, & Gottesman, 2004; Marshalek et al., 1983). Instead, the results suggest that the positive correlations between the group-level factors can be (largely) accounted for by a single factor, which is, arguably, not what would be predicted by the extended mutualism theory of *g*. That is, one would expect some consistently positive associations between group-level factors, independently of what can be accounted for by a single-factor, if these group-level factors are in fact interacting mutually beneficially during development.

Table 2
Completely standardised factor loadings associated with the single-factor models.

	WAIS-IV	WAIS-III	WISC-V	WISC-IV	Marshelak	MISTRA	WJ-R
g_c	.82	.87	.88	.88	.81	.87	.86
g_f	.93	.93	.95	.91	.80	–	.98
g_{sm}	.88	.90	.86	.94	.63	.80	.74
g_s	.72	.80	.52	.70	.56	.86	.65
g_q	–	–	–	–	.94	.86	.89
g_v	–	–	–	–	–	.79	.87
g_r	–	–	–	–	–	.68	–
g_{fl}	–	–	–	–	–	.87	–
g_a	–	–	–	–	–	–	.75
g_{lr}	–	–	–	–	–	–	.60

Note. All of the factor loadings were statistically significant ($p < .05$); a value of – indicates the absence of the group-level factor within that battery.

Table 3
Residual correlation (lower) and normalised residual correlation (upper) matrices — Wechsler Scales.

	WAIS-IV ($N = 1800$)				WAIS-III ($N = 2450$)			
	g_c	g_f	g_{sm}	g_s	g_c	g_f	g_{sm}	g_s
g_c		.22	.32	−2.08		.35	.69	−2.84
g_f	.01		−.25	.30	.01		−.58	.80
g_{sm}	.01	−.01		.79	.02	−.02		.84
g_s	**−.06**	.01	.02		**−.07**	.02	.02	
	WISC-V ($N = 2200$)				WISC-IV ($N = 2200$)			
	g_c	g_f	g_{sm}	g_s	g_c	g_f	g_{sm}	g_s
g_c		.12	.14	−2.57		−.01	.49	−3.08
g_f	.00		−.19	.33	.00		−.35	2.27
g_{sm}	.00	−.01		2.07	.01	−.01		−.03
g_s	**−.06**	.01	**.05**		**−.08**	**.06**	.00	

Note. Residual correlations in bold were statistically significant ($p < .05$).

[6] According to Raftery (1995), smaller AIC values equal to 10 or greater may be interpreted as an indication of 'very strong' evidence for a superior fitting model. However, as the AIC is very heavily influenced by sample size (Burnham & Anderson, 2004), the TLI difference of .01 or greater recommended by Gignac (2007) may be considered a more useful indicator of practical improvement in model fit, in this context.

Table 4
Residual correlation (lower) and normalised residual correlation (upper) matrix — Marshalek et al.

	g_c	g_f	g_{sm}	g_s	g_q
g_c		−.33	.74	−1.02	.09
g_f	−.03		−.33	.82	.05
g_{sm}	.05	−.02		1.55	−.25
g_s	−.07	.06	.11		−.12
g_q	.01	.00	−.02	−.01	

Note. $N = 241$; none of the residual correlations were statistically significant ($p < .05$).

5.1. What are the residual correlations?

How to interpret the substantive meaningfulness of the small residual correlations between cognitive abilities, controlling for *g*, may be a useful line of enquiry. It is not difficult to generate a plausible hypothesis that may be able to help understand the nature of the negative association between g_c and g_s. For example, Luciano, Leisser, Wright, and Martin (2004) reported a correlation of −.12 between extraversion and g_c. Thus, it is possible that individuals who are more introverted tend to spend more time engaged in solitary activities such as literary reading (e.g., Lau & Cheung, 1988), which would be expected to help increase their vocabulary and knowledge of worldly facts, independent of *g*, to some degree.

Additionally, with respect to processing speed, Doucet and Stelmack (1997) reported correlations of approximately −.30 to −.40 between extraversion and movement time within the conventional reaction time paradigm. Importantly, Doucet and Stelmack (1997) failed to observe an association between extraversion and reaction time. Thus, the results were interpreted to suggest that there was no benefit to being extraverted with respect to the core cognitive elements involved with processing speed: stimulus evaluation and response selection processes. Instead, the effect was restricted to only the physical execution of a response. In light of the results reported by Doucet and Stelmack (1997) and those reported by Luciano et al. (2004), it is possible that processing speed type subtests and crystallised intelligence type tests are, to a small degree, confounded by individual differences in extraversion. Thus, the negative association between g_c and g_s, controlling for individual differences in *g*, may be mediated by extraversion — a hypothesis that could be tested empirically relatively easily in a future investigation. As all of the residual correlations reported in this investigation were small (excepting the g_v and g_r correlation), it may be speculated that many of them, even the positive ones, are the result of various non-intellective factors. Such a possibility established empirically would arguably be further supportive of *g* factor theory, rather than the mutualism theory of *g*.

5.2. The bifactor model and *g*

It will be discussed here only briefly that amongst the three largest correlation matrices examined in this investigation (Marshalek, MISTRA, WJ-R), neither the correlated factor models nor the higher-order models were observed to be particularly well-fitting models, based on conventional standards of model fit evaluation (Marsh, Hau, & Wen, 2004). However, in two out of three of the samples (Marshalek and WJ-R), the bifactor model was associated with improvements in model fit based on the AIC and TLI values, both of which incorporate penalties for model complexity (i.e., fewer degrees of freedom). Many other samples of cognitive ability test batteries have been found to be associated with better fitting bifactor models than correlated factor and/or higher-order models (see Canivez, in press, for review; Gignac, 2008). Precisely why the bifactor model tends to fit better than the competing correlated factor and higher-order models is a contentious issue (Morgan, Hodge, Wells, & Watkins, 2015; Murray & Johnson, 2013). However, if the bifactor model of intelligence is ever demonstrated to be a more accurate representation of the shared variance between cognitive ability tests, on grounds more than just model fit, then it may have negative implications for the extended mutualism theory of *g*, as the bifactor model specifies group-level factors which are completely orthogonal to each other (i.e., no interactions). Analytical and theoretical work which deals explicitly with the implications of the bifactor model in relation to the mutualism theory of *g* (van der Maas et al., 2006, Fig. 1b) is encouraged.

Some may contend that the bifactor model is inconsistent with *g* theory. However, such a notion would only be true for those who theorise that the correlations between group-factors are necessary for the plausibility of the *g* factor (e.g., Reynolds & Keith, 2013). Spearman (1904) theorised the existence of a single intelligence factor, which would necessarily imply a first-order *g* factor. Although Spearman later acknowledged the presence of group-factors, they were interpreted as method factors, not substantive factors which were inter-correlated with each other (Spearman, 1939). Additionally,

Table 5
Residual correlation (lower) and normalised residual correlation (upper) matrix — MISTRA.

	g_c	g_{sm}	g_s	g_q	g_v	g_r	g_{fl}
g_c		1.04	−1.44	.55	−.40	−2.54	1.27
g_{sm}	.06		−.62	−.49	.60	.23	−.44
g_s	−.09	−.04		.68	1.41	2.68	−.51
g_q	.03	−.03	.04		−1.89	−1.47	.68
g_v	−.02	.03	.08	−.11		4.89	−1.18
g_r	**−.14**	.01	**.15**	−.08	**.27**		−1.53
g_{fl}	.08	−.03	−.03	.04	−.07	−.09	

Note. $N = 433$; residual correlations in bold were statistically significant ($p < .05$).

Table 6
Residual correlation (lower) and normalised residual correlation (upper) Matrix — WJ-R.

	g_c	g_f	g_{sm}	g_s	g_{lr}	g_a	g_v	g_q
g_c		−.05	1.52	−2.01	1.63	−2.09	1.16	1.29
g_f	.00		−.12	−.60	−.05	−.58	.53	.01
g_{sm}	.05	.00		2.15	−.11	5.89	−3.50	−.15
g_s	**−.06**	−.02	.06		−2.10	4.03	1.93	2.41
g_{lr}	.05	.00	.00	**−.06**		1.76	.96	−1.78
g_a	**.04**	−.02	**.18**	**.12**	.05		−.51	.37
g_v	−.07	.02	**−.11**	.06	.03	−.02		−1.60
g_q	.04	.00	.00	**.07**	−.05	.01	−.05	

Note. $N = 1425$; residual correlations in bold were statistically significant ($p < .05$; after Bonferonni adjustment).

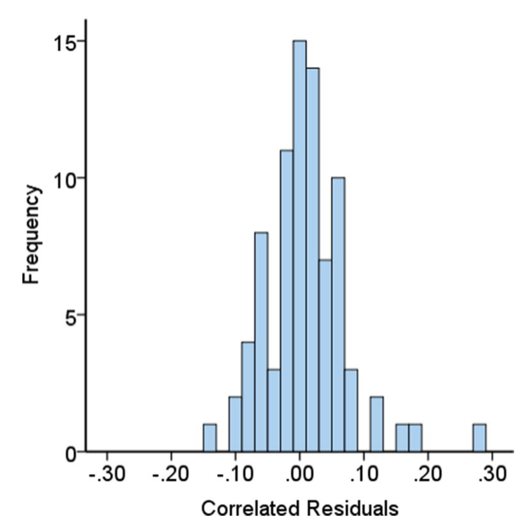

Fig. 4. The 83 group-level factor correlated residuals across all seven single-factor models tested in this investigation; mean = .00.

Table 7
Model fit statistics and indices associated with the Marshalek et al., MISTRA, and WJ-R samples: correlated factor, higher-order, and bifactor models.

	χ^2	*df*	SRMR	RMSEA	CFI	TLI	AIC
Correlated factor models							
Marshalek et al.	342.88	141	.057	.077	.931	.916	440.88
MISTRA	811.23	277	.057	.067	.904	.888	959.23
WJ-R	1804.62	346	.056	.054	.927	.915	1982.62
Higher-order models							
Marshalek et al.	354.82	146	.062	.077	.929	.916	442.82
MISTRA	926.20	291	.067	.071	.886	.873	1046.20
WJ-R	2040.73	366	.061	.057	.917	.907	2178.73
Bifactor models							
Marshalek et al.	258.72	132	.047	.063	.957	.944	374.72
MISTRA	863.04	272	.065	.071	.894	.873	1021.04
WJ-R	1475.95	345	.039	.048	.944	.934	1655.95

Note. SRMR = standardised root mean residual and RMSEA = Root Mean Square Error of Approximation (<.080 indicative of acceptable fit); CFI = comparative fit index and TLI = Tucker–Lewis index (values of .950 or greater indicative of good fit); AIC = Akaike Information Criterion (small values indicate better fit); both AIC and TLI incorporate a penalty for model complexity.

Spearman (1939, p. 11) endorsed the bifactor method as a useful procedure for estimating a *g* factor in addition to group-factors (see also Spearman, 1946, p. 120–121). Thus, it may be argued that the bifactor model is consistent with *g* theory as developed by Spearman. Jensen and Weng (1994) recommended the bifactor model as an attractive method to estimate *g*. Finally, J. B. Carroll, a major contributor to the Cattell–Horn–Carroll model of intelligence (Carroll, 1993, 2003), was very open to modeling intelligence within the context of a bifactor model and did so on a number of occasions (see Beaujean, 2015). Thus, evidence in favor of the bifactor model is arguably supportive of the plausibility of a *g* factor. By contrast, evidence in favor of the bifactor model should probably be viewed as evidence against the extended mutualism theory of *g*.

5.3. Limitations

It should be acknowledged that the approach used in this investigation to test the mutualism theory of *g* rests upon the premise that it is unlikely that the pattern of interactions between group-level factors described by the extended mutualism theory of *g*, as represented by inter-latent variable associations, occur in such a way as to be accounted for perfectly (or nearly so) by a single-factor. Some may disagree with such a premise. In fact, van der Maas et al. (2006) provided an idealised example of simulated data consistent with mutualism and a general factor without corrected residuals. However, considerations relevant to plausibility and simplicity should be entertained in the evaluation of a scientific theory (Beshers, 1957). That is, had consistent, positive correlated residuals been observed in this investigation, an account of the data more complex than *g* factor theory would be suggested. However, given that none were observed, *g* factor theory may be argued to be an acceptable representation of the data that is much simpler theoretically than mutualism. Thus, based Occam's razor, *g* factor theory may be suggested to be more plausible than mutualism, at this stage.

Ultimately, clearly articulated and empirically testable hypotheses which can yield unambiguous results relevant to the fundamental nature of the positive manifold would be welcome from both *g* factor theorists and mutualism theorists. As decades of research would attest, such hypotheses may not be forthcoming anytime soon. Although the theme of this investigation was framed within the context of an indirect test of the extended mutualism theory of *g*, those who do not accept the premise may nonetheless find the results of this investigation of interest, as there has been very little research relevant to the nature of the associations between group-level factors, independent of *g*.

Additionally, it should be acknowledged that the extended mutualism model was tested indirectly in this investigation. The extended mutualism model predicts that the mutually beneficial interactions take place between group-level factors such as g_f, g_c, g_{sm}, and g_s, etc. There is an alternative mutualism model that predicts the mutually beneficial interactions take place at the observed score level (i.e., subtests; see Fig. 1b, van der Maas et al., 2006). Consequently, the implications of the results reported in this investigation are limited to the extended mutualism model. Future research may consider adopting a similar residual correlation approach to that used in this investigation for the purposes of testing the alternative mutualism model at the observed score level. The challenge would be to find a sufficient number of samples with very large sample sizes to potentially identify the residual effects as statistically significant and replicable. That is, the observed score residual effects would be expected to be even smaller in magnitude than the negligible effects reported in this investigation, as they would be based on observed scores, rather than latent variables.

The results reported in this investigation were not segregated across age, because such data were not available in most cases. Consequently, a comprehensive examination of the possibility that there may have been some age differentiation effects was not examined. However, based on supplementary analyses, the possibility of age effects was explored with the WISC-V normative sample correlation matrices, which were available in age segregated samples (Wechsler, 2014). The small, negative residual correlation between g_s and g_c was largely consistent across age: $r = -.05$ (6 to 7 years), $r = -.03$ (8 to 9 years), $r = -.06$ (10 to 11 years), $r = -.06$ (12 to 13 years), $r = -.07$ (14 to 16 years). As the corresponding adult correlated residual based on the WAIS-IV and WAIS-III was estimated at −.06 and −.07, respectively, there was arguably no clear evidence to suggest the effect was moderated by age. No other consistent residual correlation age-based patterns were observed.

In addition to age, sex may also play a role (moderator or mediator) in the observation of the residual effects, as gender differences have been observed for specific cognitive abilities. For example, Keith, Reynolds, Roberts, Winter, and Austin (2011) found that female children/adolescents outperformed males on a g_s first-order factor residual, based on a higher-order model of the Differential Ability Scales (Elliott, 2007). Thus, the negative residual correlation between g_s and g_c observed in this investigation may be impacted by gender. Future research is encouraged in this area based on data for which gender is available.

The selections of subtests used to define the latent variables were based principally upon the factor analysis results associated with the publications within which the correlation matrices were published. Additionally, the latent variables were considered specified in such a way as to be largely consistent with CHC theory (McGrew, 2009). Nonetheless, it remains a possibility that alternative correlated factor models with equal levels of model fit and theoretical plausibility could be specified. Furthermore, such models may possibly yield residual group-factor correlations that are substantively different to those reported in this investigation. Thus, the results reported in this investigation should be considered valid to the degree that such models are not present within the correlation matrices.

Finally, the seven samples included in this investigation do not represent all of the potentially relevant correlation matrices in the literature. They were chosen based on the five criteria described in the method section, as well as the fact that they are well known. It remains a possibility that additional samples could be identified which yield some consistently positive unique associations between group-level factors. It will also be noted that most group-level factors were defined by only four indicators. It remains a possibility that group-level factors defined by more indicators may yield more substantial residual correlations. In two of the samples (Marshalek and MISTRA), not all of the available subtests were included in the models. However, this was only to help define clearly interpretable model solutions with reasonably acceptable levels of model fit, and for no other reason.

6. Conclusion

Spearman's (1927) *g* is a remarkable postulated construct. On the one hand, *g* has been described as "…one of the great discoveries of psychology" (Matarazzo, 1972, p. 82). By contrast, it has also been contended that "…the rejection of general ability theories of intelligence … may eventually be viewed as one of Psychology's greatest achievements" Conway and Kovacks (2013, p .252). Obviously, both of these positions cannot be correct. Based on the results of this investigation, in addition to those reported by Gignac (2014), it would appear, at this stage, that there is more empirical support, albeit indirect, for interpreting the positive manifold as due to a genuine construct, rather than an epiphenomenon as described by van der Maas et al. (2006).

References

Anderson, M. (1992). *Intelligence and development: A cognitive theory.* Malden, MA: Blackwell.

Arbuckle, J. L. (2012). *Amos (Version 21.0).* Chicago: IBM SPSS.

Bartholomew, D. J., Allerhand, M., & Deary, I. J. (2013). Measuring mental capacity: Thomson's Bonds model and Spearman's *g*-model compared. *Intelligence*, *41*(4), 222–233.

Beaujean, A. A. (2015). John Carroll's views on intelligence: bi-factor vs. Higher-order models. *Journal of Intelligence*, *3*(4), 121–136.

Beshers, J. M. (1957). Models and theory construction. *American Sociological Review*, *32–38*.

Bollen, K. A. (1989). *Structural equations with latent variables.* New York: John Wiley & Sons.

Brunner, M. (2008). No *g* in education? *Learning and Individual Differences*, *18*(2), 152–165.

Burnham, K. P., & Anderson, D. R. (2004). Multimodel inference understanding AIC and BIC in model selection. *Sociological Methods & Research*, *33*(2), 261–304.

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

Carpenter, P. A., Just, M. A., & Shell, P. (1990). What one intelligence test measures: a theoretical account of the processing in the Raven Progressive Matrices Test. *Psychological Review*, *97*, 404–431.

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

Carroll, J. B. (2003). The higher-stratum structure of cognitive abilities: Current evidence supports *g* and about ten broad factors. In H. Nyborg (Ed.), *The scientific study of general intelligence: Tribute to Arthur R. Jensen* (pp. 5–21). New York: Pergamon Press.

Conway, A. R. A., & Kovacks, K. (2013). Individual differences in intelligence and working memory: A review of latent variable models. In B. H. Ross (Ed.), *The psychology of learning and motivation* (pp. 233–270). San Diego, CA: Academic Press.

Crawford, J. R., Deary, I. J., Allan, K. M., & Gustafsson, J. E. (1998). Evaluating competing models of the relationship between inspection time and psychometric intelligence. *Intelligence*, *26*(1), 27–42.

Doucet, C., & Stelmack, R. M. (1997). Movement time differentiates extraverts from introverts. *Personality and Individual Differences*, *23*(5), 775–786.

Elliott, C. D. (2007). *Differential Ability Scales—Second edition: Introductory and technical manual.* San Antonio, TX: Harcourt Assessment.

Frisby, C. L., & Beaujean, A. A. (2015). Testing Spearman's hypotheses using a bi-factor model with WAIS-IV/WMS-IV standardization data. *Intelligence*, *51*, 79–97.

Fry, A. F., & Hale, S. (1996). Processing speed, working memory, and fluid intelligence: Evidence for a developmental cascade. *Psychological Science*, *7*, 237–241.

Gardner, H. (1987). The theory of multiple intelligences. *Annals of Dyslexia*, *37*(1), 19–35.

Gignac, G. E. (2006). The WAIS-III as a nested factors model: A useful alternative to the more conventional oblique and higher-order models. *Journal of Individual Differences*, *27*(2), 73–86.

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

Gignac, G. E. (2008). Higher-order models versus bifactor modes: *g* as superordinate or breadth factor? *Psychology Science*, *50*(1), 21–43.

Gignac, G. E. (2014). Dynamic mutualism versus *g* factor theory: An empirical test. *Intelligence*, *42*, 89–97.

Gignac, G. E., & Watkins, M. W. (2013). Bifactor modeling and the estimation of model-based reliability in the WAIS-IV. *Multivariate Behavioral Research*, *48*(5), 639–662.

Gignac, G. E., & Watkins, M. W. (2015). There may be nothing special about the association between working memory capacity and fluid intelligence. *Intelligence*, *52*, 18–23.

Gorsuch, R. L. (1974). *Factor analysis.* Philadelphia: W. B. Saunders.

Gottfredson, L. S. (2002). *g*: Highly general and highly practical. In R. J. Sternberg, & E. L. Grigorenko (Eds.), *The general intelligence factor: How general is it?* (pp. 331–380). Mahwah, NJ: Erlbaum.

Gustafsson, J. E. (1984). A unifying model for the structure of intellectual abilities. *Intelligence*, *8*(3), 179–203.

Haier, R. J. (2011). Biological basis of intelligence. In R. J. Sternberg, & S. B. Kaufman (Eds.), *Cambridge handbook of intelligence* (pp. 351–370). New York: Cambridge University Press.

Haworth, C. M. A., Wright, M. J., Luciano, M., Martin, N. G., De Geus, E. J. C., Van Beijsterveldt, C. E. M., ... Plomin, R. (2010). The heritability of general cognitive ability increases linearly from childhood to young adulthood. *Molecular Psychiatry*, *15*(11), 1112–1120.

Humphreys, L. G. (1979). The construct of general intelligence. *Intelligence*, *3*(2), 105–120.

Jennrich, R. I., & Bentler, P. M. (2012). Exploratory bi-factor analysis: The oblique case. *Psychometrika*, *77*(3), 442–454.

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

Jensen, A. R. (2006). *Clocking the mind: Mental chronometry and individual differences.* Amsterdam: Elsevier.

Jensen, A. R., & Weng, L. J. (1994). What is a good g? *Intelligence*, *18*(3), 231–258.

Johnson, W., & Bouchard, T. J. (2005). The structure of human intelligence: It is verbal, perceptual, and image rotation (VPR), not fluid and crystallized. *Intelligence*, *33*(4), 393–416.

Johnson, W., & Bouchard, T. J. (2011). The MISTRA data: Forty-two mental ability tests in three batteries. *Intelligence, 2011*, *39*(2), 82–88.

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.

Kaufman, S. B. (2007). Sex differences in mental rotation and spatial visualization ability: Can they be accounted for by differences in working memory capacity? *Intelligence*, *35*(3), 211–223.

Keith, T. Z., Reynolds, M. R., Roberts, L. G., Winter, A. L., & Austin, C. A. (2011). Sex differences in latent cognitive abilities ages 5 to 17: Evidence from the Differential Ability Scales—Second Edition. *Intelligence*, *39*(5), 389–404.

Kenny, D. A., Kaniskan, B., & McCoach, D. B. (2014). The performance of RMSEA in models with small degrees of freedom. *Sociological Methods & Research*, *0049124114543236*.

Lau, S., & Cheung, S. M. (1988). Reading interests of Chinese adolescents: Effects of personal and social factors. *International Journal of Psychology*, *23*(1–6), 695–705.

Luciano, M., Leisser, R., Wright, M. J., & Martin, N. G. (2004). Personality, arousal theory and the relationship to cognitive ability as measured by inspection time and IQ. *Personality and Individual Differences*, *37*(5), 1081–1089.

Lykken, D. T. (2005). Mental energy. *Intelligence*, *33*(4), 331–335.

van der Maas, H. L., Dolan, C. V., Grasman, R. P., Wicherts, J. M., Huizenga, H. M., & Raijmakers, M. E. (2006). A dynamical model of general intelligence: The positive manifold of intelligence by mutualism. *Psychological Review*, *113*(4), 842–861.

van der Maas, H. L., Kan, K. J., & Borsboom, D. (2014). Intelligence is what the intelligence test measures. Seriously. *Journal of Intelligence*, *2*(1), 12–15.

Marsh, H. W., Hau, K. T., & Wen, Z. (2004). In search of golden rules: Comment on hypothesis testing approaches to setting cutoff values for fit indexes and dangers in overgeneralising Hu & Bentler's (1999) findings. *Structural Equation Modeling*, *11*, 320–341.

Marshalek, B., Lohman, D. F., & Snow, R. E. (1983). The complexity continuum in the radex and hierarchical models of intelligence. *Intelligence*, *7*(2), 107–127.

Matarazzo, J. D. (1972). *Wechsler's measurement and appraisal of adult intelligence* (5th ed.). New York: Oxford University Press.

McDaniel, M. A. (2005). Big-brained people are smarter: A meta-analysis of the relationship between in vivo brain volume and intelligence. *Intelligence*, *33*(4), 337–346.

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.

McGrew, K. S., Werder, J. K., & Woodcock, R. W. (1991). *The WJ-R technical manual.* Allen, TX: DLM.

Meng, H. (2005). Hierarchical model of human intelligence factors on 37 ability tests. Unpublished manuscript.

Morgan, G. B., Hodge, K. J., Wells, K. E., & Watkins, M. W. (2015). Are fit indices biased in favor of bi-factor models in cognitive ability research?: A comparison of fit in correlated factors, higher-order, and bi-factor models via Monte Carlo simulations. *Journal of Intelligence*, *3*(1), 2–20.

Murray, A. L., & Johnson, W. (2013). The limitations of model fit in comparing the bi-factor versus higher-order models of human cognitive ability structure. *Intelligence*, *41*(5), 407–422.

Oberauer, K., Su, H. -M., Wilhelm, O., & Sander, N. (2007). Individual differences in working memory capacity and reasoning ability. In A. R. A. Conway, C. Jarrold, M. J. Kane, A. Miyake, & J. N. Towse (Eds.), *Variation in working memory* (pp. 21–48). New York, NY: Oxford University Press.

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

Reynolds, M. R., & Keith, T. Z. (2013). Measurement and statistical issues in child psychological assessment. In D. H. Soklofske, V. L. Schwean, & C. R. Reynolds (Eds.), *Oxford handbook of child and adolescent assessment* (pp. 48–83). New York: Oxford University Press.

Ritchie, S. J., Booth, T., Hernández, M. D. C. V., Corley, J., Maniega, S. M., Gow, A. J., ... Deary, I. J. (2015). Beyond a bigger brain: Multivariable structural brain imaging and intelligence. *Intelligence*, *51*, 47–56.

Schmiedek, F., & Li, S. C. (2004). Toward an alternative representation for disentangling age- associated differences in general and specific cognitive abilities. *Psychology and Aging*, *19*(1), 40–56.

Spearman, C. (1904). General intelligence, objectively determined and measured. *The American Journal of Psychology*, *15*(2), 201–292.

Spearman, C. (1923). *The nature of "intelligence" and the principles of cognition.* London: Macmillan.

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

Spearman, C. (1939). Thurstone's work re-worked. *Journal of Educational Psychology*, *30*(1), 1–16.

Spearman, C. (1946). Theory of general factor. *British Journal of Psychology: General Section*, *36*(3), 117–131.

Spearman, C., & Holzinger, K. (1925). Note on the sampling error of tetrad differences. *British Journal of Psychology. General Section*, *16*(2), 86–88.

Sternberg, R. J. (1985). *Beyond IQ: A triarchic theory of human intelligence*. New York: Cambridge University Press.

Sternberg, R. J. (2003). My house is a very very very fine house—but it is not the only house. In H. Nyborg (Ed.), *The scientific study of general intelligence: Tribute to Arthur R. Jensen* (pp. 373–395). New York: Pergamon.

Thomson, G. H. (1951). *The factorial analysis of human ability* (5th ed.). London: University of London Press.

Tommasi, M., Pezzuti, L., Colom, R., Abad, F. J., Saggino, A., & Orsini, A. (2015). Increased educational level is related with higher IQ scores but lower g-variance: *Evidence from the standardization of the WAIS-R for Italy*. *Intelligence*, *50*, 68–74.

Wechsler, D. (1997). *WAIS-III WMS-III technical manual*. San Antonio, TX: Psychological Corporation.

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

Wechsler, D. (2008). *Wechsler Adult Intelligence Scale—Fourth Edition: Technical and interpretive manual*. San Antonio, TX: Pearson Assessment.

Wechsler, D. (2014). *Wechsler Intelligence Scale for Children—Fifth Edition: Technical and interpretive manual*. San Antonio, TX: Pearson Assessment.