Intelligence 52 (2015) 71–79

Contents lists available at ScienceDirect

Intelligence

Raven's is not a pure measure of general intelligence: Implications for g factor theory and the brief measurement of g

Gilles E. Gignac

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

ARTICLE INFO

Article history:
Received 23 May 2015
Received in revised form 19 July 2015
Available online 12 August 2015

Keywords:
Raven's Progressive Matrices
General intelligence
Fluid intelligence

ABSTRACT

It has been claimed that Raven's Progressive Matrices is a pure indicator of general intelligence (*g*). Such a claim implies three observations: (1) Raven's has a remarkably high association with *g*; (2) Raven's does not share variance with a group-level factor; and (3) Raven's is associated with virtually no test specificity. The existing factor analytic research relevant to Raven's and *g* is very mixed, likely because of the variety of factor analytic techniques employed, as well as the small sample sizes upon which the analyses have been performed. Consequently, the purpose of this investigation was to estimate the association between Raven's and g, Raven's and a theoretically congruent group-level factor, and Raven's test specificity within the context of a bifactor model. Across several large samples, it was observed that Raven's (1) shared approximately 50% of its variance with *g*; (2) shared approximately 10% of its variance with a fluid intelligence group-level factor orthogonal to *g*; and (3) was associated with approximately 25% test specific reliable variance. Overall, the results are interpreted to suggest that Raven's is not a particularly remarkable test with respect to *g*. Potential implications relevant to the commonly articulated central role of Raven's in *g* factor theory, as well as the Flynn effect, are discussed. Finally, researchers are discouraged to include only Raven's in an investigation, if a valid estimate of *g* is sought. Instead, as just one example, a four-subtest combination from the Wechsler scales with a *g* validity coefficient of .93 and 14 min administration time is suggested.

1. Introduction

Raven's Progressive Matrices (RPM; Raven, 1940; Raven & Court, 1998) have been contended in the intelligence literature to be a pure (or nearly so) indicator of Spearman's (1927) general intelligence (*g*; Deary & Smith, 2004; Eysenck, 1998; Jensen, 1998; Llabre, 1984; Neisser, 1998; Thorndike, 1986; Vernon, 1947). Correspondingly, researchers across various disciplines within psychology commonly include a single measure of intelligence in their design, Raven's, and interpret their results (or lack thereof) under the pretence that they have measured *g* (e.g., Basso, De Renzi, Faglioni, Scotti, & Spinnler, 1973; Corben et al., 2006; Day et al., 2005; Schellenberg & Moreno, 2010; Walker, Pierre, Christie, & Chang, 2013). The purpose of this investigation was to review critically the literature upon which claims of "pure *g*" have been made. Additionally, the question will be examined empirically across several data sets, in conjunction with a rigorous, direct, factor analytic approach, the bifactor model (Gustafsson & Balke, 1993).

2. Raven's and g: background

Although there are several versions of RPM that have been published over the years (Raven, 1940; Raven, 1966; Raven, Court, & Raven, 1994; Raven, Raven, & Court, 1962), they all have in common the inclusion of items that consist of visually presented figural matrices within which one piece is missing. In order to complete a particular matrix, the participant must choose one piece amongst several alternatives which are presented below the matrix. In more practical terms, this would involve the successful identification of one or more figural patterns associated with the displayed pieces within the matrix.

It is commonly stated that Raven's may be considered the purest expression of *g* (Martinez, 2013). Spearman (1946) considered matrices type tests to be the nearest representation of *g*, as they contained items that required the eduction of relations and correlates to solve, a view shared by others (e.g., Holyoak, 2012; Jensen, 1998; Thorndike, 1986). Some researchers view the *g* factor saturation associated with Raven's to be so substantial that the item-total correlations may be considered an accurate estimate of each items *g* loading (Rushton & Skuy, 2001). However, few of these sources provide primary references, or, in many cases, any references at all, to support the contention that Raven's is a pure measure of *g*.

Perhaps one of the researchers who has most frequently asserted Raven's to be a pure measure of *g* is Arthur Jensen (e.g., Jensen, 1973; Jensen, 1980a; Jensen, 1987; Jensen, 1998). However, again, only occasionally has a reference been provided to support such a claim. In a review of Raven's, Jensen (1980a, p. 646) wrote:

> "Factorially the Progressive Matrices apparently measures *g* and little else (Burke, 1958). The loadings that are occasionally found on other

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

http://dx.doi.org/10.1016/j.intell.2015.07.006
0160-2896/

> "perceptual" and "performance"-type factors, independently of *g*, are usually so trivial and inconsistent from one analysis to another as to suggest that the RPM does not reliably measure anything but *g* in the general population."

Thus, Jensen (1980a) relied upon Burke (1958) for evidence to support his assertion. Although it is the case that Burke (1958) did review some factor analytic research to suggest that Raven's may be a relatively pure measure of *g*, it is also the case that Burke (1958) reviewed a substantial amount of factor analytic research to suggest that Raven's was not a relatively pure measure of *g*. Importantly, Burke (1958, p. 212) concluded the following on the matter:

> "The evidence is not convincing that [Progressive Matrices] has validity as a pure measure of the Spearman construct of *g*; and doubt may be raised whether such a construct can be measured independently of the modality through which it is expressed, the selectivity of the subjects and their sex, and possibly the presuppositions of the factor analyst."

In another investigation, Jensen (1987) cited Vernon (1983) as support for the contention that Raven's may be considered a relatively pure indicator of *g*. Vernon (1983) tested 100 university students on reaction time measures, as well as the Wechsler Adult Intelligence Scale (WAIS; Wechsler, 1955) and the Advanced Progressive Matrices (APM; Raven, 1966). Based on the extraction of a single principal component, Vernon (1983) reported the APM to be associated with a *g* loading of .797, which was numerically larger than the next largest *g* component loading (Block Design $g = .692$). Thus, based on the reported APM component loading, it would appear that Raven's is a relatively substantial indicator of *g*. However, as Vernon (1983) reported the results for a single component only, it was not possible to evaluate the possibility of a secondary Raven's loading. Finally, the sample size of 100 would not be considered large for the purposes of estimating a component loading in a precise manner.

In contrast to the numerous authors cited above, there are some who have expressed a more uncertain view in relation to the possibility of Raven's as a pure indicator of *g*. For example, Carroll (1993) acknowledged Raven's as a good measure of *g*, but it was not identified as clearly the single best measure of *g*. Furthermore, Carroll (1993) suggested that there was some evidence that Raven's loaded on a separate induction intelligence factor or possibly a separate spatial ability factor. Brody (1987) was also unconvinced by the position that Raven's was a pure measure of *g*. Instead, Brody (1987) suggested that previously reported factor loadings of intelligence test batteries appear to be too highly variable to allow any meaningful conclusions in that respect. Such a position suggests that no test has an intrinsic *g* loading, as it would be expected to fluctuate based on the other tests included in the analysis. Also, the fact that the Flynn effect is most pronounced on Raven's may be suggested to be evidence contrary to the notion that it is a pure measure of *g* (Johnson, 2012). Finally, in a discussion relevant to test development, Sidney Irvine opined that Raven's could not possibly be a pure measure of *g*, as it would be expected to be associated with a large amount of test specific variance (Wainer, 2002).

3. Some empirical research since Burke (1958)

Brody's (1987) point in relation to the substantial amount of variability in the factor analytic results is underscored by a review of some of the factor analytic literature published since Burke (1958). That is, some researchers report Raven's to be the most substantial indicator of *g*, while others do not. Similarly, Raven's is occasionally observed to load onto a secondary group-factor, and on other occasions it is not.

For example, Rogers, Fisk, and Hertzog (1994) administered a battery of 20 cognitive ability tests, including the Standard Progressive Matrices (SPM; Raven, Court, & Raven, 1977), to a sample of 140 participants (mix of university students and members of the community) in an investigation relevant to intelligence and visual search performance across age. The battery consisted of four semantic knowledge tests (Vocabulary, Analogies, Information, Controlled Associations), three induction tests (Mathematic Reasoning, SPM, Letter Sets), three working memory tests (Computation Span, Listening Span, Alphabet Span), four perceptual speed tests (Digit Symbol, Identical Pictures, Number Comparison, Finding *A*s), three semantic memory tests (Semantic Matching, Lexical Access, Synonym Matching), and three psychomotor speed tests (Simple Reaction Time, Making *X*s, Crossing Lines). Based on a higher-order model, the SPM was observed to have an association of .76 with *g* (Schmid–Leiman transformed) and no other group-level factor, which would suggest that Raven's may be exclusively associated with *g*.

In another investigation, DeYoung, Peterson, and Higgins (2005) administered Vocabulary, Similarities, Arithmetic, Digit Symbol, and Block Design from the WAIS-III, as well as the APM (Raven, Raven & Court, 1998), to a sample of 174 university students. Based on the extraction of a single factor, the APM was reported to be associated with the second largest *g* loading (.72). The largest *g* loading was associated with Block Design (.74). The inter-subtest correlation between the APM and Block Design was the largest at $r = .65$, however, there were no other appropriate subtests with which to model a possible a group-level spatial reasoning factor.

Kranzler and Jensen (1991) administered the WAIS and the APM (Raven, 1966) to a sample of 101 university students. Based on a Schmid–Leiman transformation of the higher-order model solution, the APM was reported to have a rather small *g* loading of .44, which was numerically lower than the mean subtest *g* loading of .48. Furthermore, the APM was also observed to load (.48) onto a Perceptual Organisation group-level factor. These results would suggest that Raven's is neither a remarkable *g* loading test, nor exclusively associated with *g*.

Finally, Ackerman (1988) administered a battery of 22 cognitive ability tests (including the SPM) to a sample of 65 undergraduate students and performed a higher-order model factor analysis with a corresponding Schmid–Leiman transformation. The battery consisted of five tests of perceptual speed (Name Comparison, Clerical Ability, Perceptual Speed, Letter/Number Substitution, Scattered Xs), three tests of movement speed (Circle Tapping, Square Marking, Pursuit Aiming), three tests of memory (Object–Number, Picture Number, First and Last Names), three tests of verbal ability (Analogies, Vocabulary, Word Beginnings), five tests of reasoning (SPM, Patterns, Number Series, Figure Classification, Letter Sets), and three indicators of reaction time (simple, two-choice, and four-choice). The SPM was observed to be associated with a *g* loading of .57, which was numerically smaller than several other tests (Letter Sets $g = .62$; Letter/Number Substitution $g = .60$). Thus, Raven's was not observed to be associated with the largest *g* loading in this investigation. Furthermore, the SPM was reported to be associated with a secondary loading of .45 on the group-level reasoning factor, which was similar in magnitude to the secondary loadings associated with the other four reasoning tests. Thus, Raven's was neither observed to be associated with a remarkably large *g* loading, nor was it observed to be exclusively associated with *g*.

Although the above is not an exhaustive review of the relevant factor analytic literature published since Burke (1958), it may nonetheless be plausible to suggest that the published results in relation to Raven's *g* factor saturation are highly variable. Arguably, the very mixed results may be due to the rather small sample sizes (often $N \leq 100$) upon which the factor analyses have been performed. Additionally, the different factor analytic techniques that were employed may be expected to yield variation in the results, not to mention results of questionable validity, in some cases. For example, a non-negligible percentage of the investigations simply extracted a single component to estimate the loading of Raven's on *g*, which is arguably an inadequate approach to the estimation a non-biased general factor solution (Ashton, Lee, & Vernon, 2001; Gignac, 2006a). Also, the extraction of a single component necessarily precludes the possibility of a secondary Raven's group-level