Intelligence 44 (2014) 1–10

Contents lists available at ScienceDirect

Intelligence

The *g* beyond Spearman's *g*: Flynn's paradoxes resolved using four exploratory meta-analyses

James R. Flynn [a,*,1], Jan te Nijenhuis [b,1], Daniel Metzen [b]

[a] Departments of Political Studies and Psychology, University of Otago, Dunedin, New Zealand
[b] Work and Organizational Psychology, University of Amsterdam, Amsterdam, The Netherlands

ARTICLE INFO

Article history:
Received 20 October 2013
Received in revised form 29 January 2014
Accepted 30 January 2014
Available online 25 February 2014

Keywords:
Flynn effect
Jensen effect
Method of correlated vectors
Significance of IQ gains over time
Significance of black IQ gains on whites

ABSTRACT

When IQ tests are ranked by the magnitude of their score gains over time, this hierarchy lacks a positive correlation with the same tests ranked by their *g* loadings. Therefore, Jensen declared IQ gains "hollow" and, by implication, extended this judgment to score gains that indicated that blacks had made IQ gains on whites. We offer four exploratory meta-analyses that apply Jensen's method to the subtest score differences between normal subjects and those suffering from certain afflictions: iodine deficiency ($K = 6$, $N = 196$), prenatal cocaine exposure ($K = 2$, $N = 215$), fetal alcohol syndrome and degree of fetal alcohol syndrome (respectively, $K = 1$, $N = 110$; and $K = 3$, $N = 125$), and traumatic brain injury ($K = 14$, $N = 629$). All of these create a substantial cognitive deficit in those afflicted. However, the correlations between subtest score differences and *g* loadings run from −0.23 to +0.12, with an unweighted average of 0.00.

1. Introduction

Jensen used what he called the method of correlated vectors to test whether a "Flynn effect" was a "Jensen effect". The application of this method implies that IQ gains over time do not signal significant cognitive progress. In this paper, we will: (1) describe the method of correlated vectors; (2) state two paradoxes that its application entails; (3) use data from four exploratory meta-analyses to sharpen those paradoxes; (4) propose a solution based on limiting the method's application.

1.1. The method

Jensen recommended the method as a criterion for evaluating the significance of IQ gains over time. He ranked IQ tests into two hierarchies, best exemplified by ranking the 10 or 11 subtests of the WISC (Wechsler Intelligence Scale for Children) or the WAIS (Wechsler Adult Intelligence Scale). The first hierarchy ran from the subtest that had the greatest *g* loading down to the one that had the least; the second hierarchy ran from the subtest on which there had been the largest gains over time down to the one on which there had been the least. If the correlation was positive, the IQ gains were *g* gains; if negative, they could not be intelligence gains with all that entailed. A recent meta-analysis based on a large total *N* shows the meta-analytical correlation is *rho* = −.38 (te Nijenhuis & van der Flier, 2013).

The appeal of the method rests on the fact that cognitive tasks become more complex as their *g* loading rises. For example, the *g* loading of Digit Span forward, a simple task of repeating a series of random numbers in the order in which they are read out, has a low *g* loading. Digit Span backward, a more complex task of saying numbers in reverse of the order in which they are read out, has a much higher *g* loading. Scrambling eggs has a lower *g* loading than making a soufflé. Speed of shoe tying would have a *g* loading of close to zero. Most of us believe that the more cognitively complex a task the more it measures intelligence.

* Corresponding author at: Department of Political Studies, University of Otago, PO Box 56, Dunedin, New Zealand.
[1] The two first authors contributed equally to this study.

0160-2896/$ – see front matter © 2014 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.intell.2014.01.009

1.2. Applications and paradoxes

When Jensen applied the method of correlated vectors, he asserted that its results would determine whether IQ gains were cognitively significant or "hollow". By significant, he meant that it would test whether group differences were *g* differences. If the result was negative, groups could still differ on the acquisition of task-specific skills (Jensen, 1998, 320–321 & 332). Had he expanded on this and emphasized that these skills could be cognitively significant, this would have been felicitous. But he did not. Moreover, there would still have been a serious problem. Because Jensen identified "intelligence differences" with "*g* differences", his position meant that that having flunked the test of correlated vectors, group IQ differences could never be equated with intelligence differences. We will show that they can: *g* aside, they still discriminate between groups concerning whom there is virtually universal acceptance that they differ in intelligence.

Further, the identification of *g* with intelligence led to deceptive conclusions about the significance of black versus white trends over time, trends that reduced the racial IQ gap. Rushton and Jensen (2006, p. 922) exemplify this. First, they state that the best estimate of black/white convergence is between 0 and 3.44 IQ points. Second, they state that "While secular increases on various tests cluster together, they do so independently of Black–White differences, which cluster with the *g* factor and genetic indices such as inbreeding depression and twin differences" (Rushton & Jensen, 2006, p. 922). The implication is that black gains on whites are somehow devalued because they are not on the *g* factor.

As the above exemplifies, the method can be applied to groups that are separated, not by time, but purely by the fact that they have different subtest score hierarchies. As Jensen says, at any given time, American blacks and whites differ; and the magnitude of their subtest differences correlates with *g* loadings. These results convinced Jensen that race differences are significant. If the correlation had been negative, his method would have shown that blacks and whites were separated not by intelligence differences but merely by hollow differences. He did not, of course, simply assume that intelligence differences were genetic in origin but stated many arguments to that effect.

The application to black/white comparisons leads directly to a paradox. Its advocates must assert three propositions, which cannot be reconciled: the black/white IQ gap in 1972 was significant and real because it correlated with *g*; the black/white IQ gap of 2002 (a lesser gap) was also significant and real because it correlated with *g*; the IQ gains that reduced the gap were hollow and unreal because they did not correlate with *g*. Collectively these propositions posit that a hollow trend can make a real-world difference. Rindermann and Thompson (2013) show that this real world difference is not hypothetical (and larger than Rushton and Jensen concede). Between 1971 and 2008, averaging scores for reading and mathematics from the NAEP (National Assessment of Educational Progress), blacks gained the equivalent of 6.39 IQ points on whites, leaving the final gap for all ages at 9.94 points. As Flynn (2013b) notes, this is almost identical to black IQ gains on whites between 1972 and 2002: a gain of 5.5 points, leaving the final gap at 10.00 points for ages 9 to 17. Yet, the IQ gains, whatever their magnitude, fail the test of correlated vectors. On this, Jensen and Ruston are correct.

The application of the method of correlated vectors to generational comparisons poses another paradox, one relevant to the general significance of IQ gains over time. People over time have made huge gains on subtests every one of which poses problems of cognitive complexity. However, the fact that the gains do not rank tasks according to the *magnitude* of their cognitive complexity is posited as a rationale for denying that significant gains have occurred. This has an unstated assumption: large gains on simple tasks show that lesser gains on complex tasks are not significant. For example, WISC trends over 54 years show that gains on Coding (low *g* loading) are much greater than gains on Vocabulary (high *g* loading). Nonetheless, Vocabulary gains amount to 0.30 SDs, which means that school children can communicate better than they could in the past. There is no other area in which progress on a complex task is discounted because of greater progress on a simple task. Pole-vaulting is more complex than sprinting. But we do not consider how much performance in the latter has improved to assess the significance of improved performance in the former.

The paradox may be stated as: one group betters another on ten tasks that all involve cognitive complexity; it does not do so according to the magnitude of their cognitive complexity; therefore, the gains on each and every task have only trivial significance. That implies a hypothesis: under these conditions, evidence of the real-world significance of the gains will be weak or non-existent. Research over the last five years demonstrates the contrary:

- When you deduct *g* from performance on the SAT, the scores still predict grades (Coyle & Pillow, 2008).
- Education promotes autonomous and diverse skills. Its effects are not mediated by *g* but by direct links to specific subtests (Ritchie, Bates, & Deary, 2014; Woodley, 2012a).
- IQ gains over time (which of course do not correlate with *g*) parallel and predict growth in GDP per capita (Woodley, 2012b).
- Autonomous skills allow one to adapt cognitively to modernity and thereby promote a better life (Woodley, Figueredo, Ross, & Brown, 2013)
- Modernity in general encourages greater sensitivity to a whole range of rules, operating independently in a complex web of social situations, rather than collectively as assumed by *g* (Armstrong & Woodley, 2014).

A basketball analogy is inevitable. Imagine a team that bettered another team on every basketball task. They can make layups better, shoot fouls better, make set shots better, do fade ways jumps better, pass better, guard better. But when these tasks are ranked in terms of difficulty, the degree of their superiority does not correlate with their complexity, that is, their advantage on foul shots (more difficult) was less than their advantage on lay ups (less difficult). If someone objects that their advantage is only trivial because it has been "hollowed-out" by the failed correlation, evidence must decide: it shows that the first team beats the second team easily.

1.3. Seeking additional evidence

The conclusion that score differences between two groups on Wechsler subtests need *not* correlate with *g* in order to be cognitively significant will be reinforced by presenting