Personality and Individual Differences 55 (2013) 224–228

Contents lists available at SciVerse ScienceDirect

Personality and Individual Differences

journal homepage: www.elsevier.com/locate/paid

The Flynn effect, group differences, and *g* loadings

Jan te Nijenhuis *

Work and Organizational Psychology, University of Amsterdam, Amsterdam, The Netherlands

ARTICLE INFO

Article history:
Available online 14 January 2012

Keywords:
Flynn effect
Secular score gains
g Score
g Loadings
IQ tests
Intelligence
Method of correlated vectors
Group differences

ABSTRACT

Flynn effect gains are predominantly driven by environmental factors. Might these factors also be responsible for group differences in intelligence? Group differences in intelligence have been clearly shown to strongly correlate with *g* loadings. The empirical studies on whether the pattern of Flynn effect gains is the same as the pattern of group differences yield conflicting findings. We present new evidence on the topic using a number of datasets from the US and the Netherlands. Score gains and *g* loadings showed a small negative average correlation. The general picture is now that there is a small, negative correlation between *g* loadings and Flynn effect gains. It appears that the Flynn effect and group differences have different causes.

It is always a great pleasure to meet Phil Rushton at conferences. Although he has been relentlessly attacked since the early 1980s, it has never lessened his passion for research or slowed him down. From early in the morning until late at night he shows a contagious enthusiasm for professional scientific discussions, showing an encyclopedic knowledge of the literature.

In his long and outstanding career, Phil Rushton has made important contributions to many scientific discussions. His research program is truly progressive (Lakatos & Musgrave, 1974): every couple of years he comes up with innovative ideas. This is in contrast to many other researchers who come up with only one good idea in their whole career. Science owes much to Phil Rushton.

1. Introduction

Secular gains in IQ test scores are among the most intriguing and controversial findings in psychology. Flynn (1984) was the first to show that average scores on intelligence tests are rising substantially and consistently, all over the world. Between 1930 and 1990 the gain on standard broad-spectrum IQ tests averaged three IQ points per decade. For verbal tests, or more precisely, tests with a content that most reflects the traditional classroom subject matter, the gain is 2 IQ points per decade, and for non-verbal (Fluid and Visual) tests 4 IQ points per decade. Gains on specific measures, such as the Raven's Progressive Matrices when used for the assessment of military recruits average about 7 IQ points per decade.

Group differences in mean IQ are more the rule than the exception (Lynn & Vanhanen, 2002). Large-scale research shows that, on average, the adult American Black population scores below the White population by about 1.2 standard deviations, equivalent to about 18 IQ points (Jensen, 1998). There is some discussion about the gap diminishing (see Dickens & Flynn 2006a,b; see Rushton & Jensen, 2005). Dutch-first-generation non-Western immigrants differences are about the same size as the US Black/White differences, but become substantially smaller for the second generation of non-Western immigrants (te Nijenhuis, de Jong, Evers, & van der Flier, 2004). Jensen (1998, pp. 380–383) has shown that *g* loadings correlate about .60 with Black/White IQ test score differences. te Nijenhuis and van der Flier (2003) showed that Dutch-non-Western immigrant differences are also strongly predicted by *g* loadings.

What, then, are the causes of these differences? Some argue there is a strong genetic component to group differences (Rushton & Jensen, 2005), whereas others argue group differences are wholly caused by the environment (Nisbett, 2009). The secular gains are massive and the time period too short for large genetic changes in the population, so therefore the changes must be largely environmental – although reduced inbreeding has been suggested to play a role (Mingroni, 2007; see also Woodley, 2011). It is an empirical question whether the strong environmental forces causing the scores over generations to rise are the same as the forces causing the group differences.

1.1. Rushton's contribution to the area

Rushton (1989) joined the fray on this topic showing that inbreeding depression scores from Japan predicted the magnitude

* Address: Gouden Leeuw 746, 1103 KR Amsterdam, The Netherlands. Tel.: +31 20 6002067.
E-mail address: JanteNijenhuis@planet.nl

0191-8869/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.paid.2011.12.023

of the Black/White differences on the same subtests in the US. Inbreeding depression is an established genetic phenomenon that occurs when people who are genetically related have children together, thereby producing in their offspring, on average, a lower score on IQ than would otherwise have been the case.

Subsequently, in an exchange with Flynn (1999), Rushton (1999) showed that secular gains from the US, Germany, Austria, and Scotland had modest to small negative correlations with *g* loadings. This is an important result given that *g* loadings correlate substantially with group differences, as shown by Rushton in a series of articles. In South Africa *g* loadings of items of the Raven Matrices predicted mean differences on the items between White, South Asian, and Black students (Rushton, Skuy, & Bons, 2004; Rushton, Skuy, & Fridjohn, 2002; Rushton, Skuy, & Fridjohn, 2003). In Zimbabwe *g* was a strong predictor of the score differences between African and White 12- to 14-year-olds on the WISC-R (Rushton & Jensen, 2003). In Serbia item *g* loadings from the Raven Matrices predicted mean differences between Roma and Whites (Rushton, Čvorovič, & Bons, 2007).

These findings increase in importance when one takes into consideration that most studies show that *g* loadings of tests correlate highly with their heritabilities (Jensen, 1987; Pedersen, Plomin, Nesselroade, & McClearn, 1992; Rijsdijk, Vernon, & Boomsma, 2002; Spitz, 1988). Moreover, Rushton, Bons, Vernon, and Čvorovič (2007) computed heritabilities for items from the Raven's Standard Progressive Matrices and showed that they could predict various group differences.

Rushton (1999) and Rushton and Jensen (2010) argues that if the gains over generations had been similar to the B/W differences they would have been in line with environmental explanations of group differences; when the pattern in secular score gains is not similar to the pattern in B/W differences, the latter is in line with a genetic component in group differences. However, it should not be forgotten that these findings are at best indirect evidence.

Rushton's (1999) finding has been challenged by Flynn (2000) and Nisbett (2009) who claim that there actually is a substantial positive correlation between secular score gains and *g* loadings. If this were a fact, it would indeed jeopardize Rushton's position; it would mean that *g* loadings correlate highly with both environmental and genetic effects, making them useless. Since Rushton's study suggesting secular trends are not related to *g*, various other studies have been carried out (Colom, Juan-Espinosa, & García, 2001; Flynn, 1999, 2000; Must, Must, & Raudik, 2003; Wicherts et al., 2004) yielding conflicting findings.

Flynn (2007, 2010) states that even if there is a small negative correlation between secular score gains and *g* loadings it sheds no light on the race and IQ debate. Flynn (2010) accepts the empirical findings that Black/White score differences on subtests of IQ batteries rise as their *g* loadings, cognitive complexity, heritability, and inbreeding sensitivity rise. However, he argues that the fact that the performance gap is larger on more complex tasks than on easier tasks does not necessarily tell us something about genes versus environment. For instance, he hypothesizes that when one group has better genes for height and keen reflexes, but finds itself in a less rich basketball environment – less incentive, low-quality coaches, less play – the environmental disadvantage will expand the between-group performance gap as complexity rises, just as much as a genetic deficit would. The skill gap between challenged and unchallenged players is hypothesized to be more pronounced the more difficult the task. So, someone exposed to an inferior environment will hit a "complexity ceiling" and this ceiling does not differentiate whether the phenotypic gap is due to genes or environment. Elsewhere Flynn (2007; see also Dickens & Flynn, 2001) has argued that Blacks tend to be systematically underexposed to cognitive complexity throughout their life-courses. Flynn argues that the correlations reported by Rushton do not decide the causal question.

1.2. Research question

We used a number of datasets to see whether the method of correlated vectors yields a modest positive or negative correlation between score gains and *g* loadings.

2. Method

2.1. Test

The GATB (United States Department of Labor, 1970; van der Flier & Boomsma-Suerink, 1994) is a test of general intelligence with eight subtests: *Three-Dimensional Space* measures Visualization (g_v), *Vocabulary* measures Induction (g_{fl}) and Lexical Knowledge (g_{cr}), *Arithmetic Reason* measures Quantitative Reasoning (g_{fl}), *Computation* measures Numerical Ability (g_{cr}), *Tool Matching* measures Perceptual Speed (g_v), *Form Matching* measures Spatial Relations (g_v), *Name Comparison* measures Perceptual Speed (g_v) and Numerical Ability (g_{cr}), and *Mark Making* measures Aiming (General Psychomotor Speed). There are also two additional tests for finger dexterity: *Assemble* and *Disassemble*; as well as two more tests for manual dexterity: *Place* and *Turn*.

2.2. Samples

Sample 1: workers representative of the general working population from 1947: The first general working population norms for the GATB were based on 519 employed workers (US GATB manual, 1970). It was recognized that the sample probably was not truly representative of the general working population, but since it did include a wide range of occupational classifications, it was believed to yield a reasonably close approximation to test performance typical of the general working population. The date for the first sample is not explicitly given in the GATB manual, but close reading of the text suggests the year 1947. The mean age of the sample is 30.4 years (SD = 10.9 years) and the mean education is 11.0 years (SD = 2.4 years). This resulted in means and SDs for the GATB subtests, including the four tests of finger and manual dexterity.

Sample 2: workers representative for the general working population from 1952: In 1952, general working population norms were established on the basis of a selected sample of 4000 which was stratified to obtain proportional occupational representation of the general working population. The mean age of the sample is 30.4 years (SD = 9.9 years) and the mean education is 11.0 years (SD = 2.6 years). Means and SDs are reported for all 12 GATB subtests.

Samples 3 + 4: Dutch applicant bus drivers from 1975 to 1976 (sample 3), and 1983–1985 (sample 4): The Dutch GATB manual reports that for a specific time period a random sample (N = 110 for sample 3, and N = 1091 for sample 4) was taken from all persons who applied for positions of bus driver at regional bus companies and were tested at Dutch Railways selection centers.

Sample 5: Dutch applicant bus drivers from 1988 to 1992: For a specific time period, a random sample (N = 221) was taken from all persons who applied for positions of bus driver at regional bus companies and were tested at Dutch Railways selection centers (data are taken from te Nijenhuis, 1997).

Samples 6–8: Dutch 16-year-old students in higher general secondary education representative for the years 1975, 1985, and 2005, respectively: The Dutch GATB manual (van der Flier &