Intelligence 41 (2013) 802–807

Contents lists available at ScienceDirect

Intelligence

Is the Flynn effect on g?: A meta-analysis

Jan te Nijenhuis [a,*], Henk van der Flier [b]

[a] Work and Organizational Psychology, University of Amsterdam, Amsterdam, The Netherlands
[b] Social and Organizational Psychology, Free University, Amsterdam, The Netherlands

ARTICLE INFO

Article history:
Received 30 November 2012
Received in revised form 31 January 2013
Accepted 5 March 2013
Available online 27 March 2013

Keywords:
Flynn effect
Secular score gains
Group differences in intelligence
Method of correlated vectors

ABSTRACT

Black/White differences in mean IQ have been clearly shown to strongly correlate with *g* loadings, so large group differences on subtests of high cognitive complexity and small group differences on subtests of low cognitive complexity. IQ scores have been increasing over the last half century, a phenomenon known as the Flynn effect. Flynn effect gains are predominantly driven by environmental factors. Might these factors also be responsible for group differences in intelligence? The empirical studies on whether the pattern of Flynn effect gains is the same as the pattern of group differences yield conflicting findings. A psychometric meta-analysis on all studies with seven or more subtests reporting correlations between *g* loadings and standardized score gains was carried out, based on 5 papers, yielding 11 data points (total $N = 16{,}663$). It yielded a true correlation of −.38, and none of the variance between the studies could be attributed to moderators. It appears that the Flynn effect and group differences have different causes. Suggestions for future research are discussed.

1. Introduction

A first finding is that IQ scores are the best predictors of a large number of important life outcomes (Jensen, 1998; Schmidt & Hunter, 1998) and that some groups show substantial differences in their mean IQ scores, such as Blacks and Whites in the US, and Europeans and non-Western immigrants in Europe (Hunt, 2011). These group differences have been clearly shown to strongly correlate with *g* loadings, so large group differences on subtests of high cognitive complexity and small group differences on subtests of low cognitive complexity.

A second finding is that IQ scores have been increasing over the last half century, a phenomenon known as the Flynn effect. Flynn effect gains are predominantly driven by environmental factors.

We combine these two findings and ask the question whether the Flynn effect and group differences have the same causes. The empirical studies on whether the pattern of Flynn effect gains is the same as the pattern of group differences yield conflicting findings. We carried out a psychometric meta-analysis on all published studies reporting correlations between *g* loadings and standardized score gains attempting to estimate the true correlation.

1.1. Group differences in IQ

Lynn and Vanhanen (2002, 2012) have shown that there are large group differences in mean IQ scores. In the US, for instance, the Black/White difference is about one S.D. (Jensen, 1998).

The method of correlated vectors is a means of identifying variables that are associated with the *g* factor of mental ability. This method involves calculating the correlation between (a) the column vector of the *g* factor loadings of the subtests of an intelligence test or similar battery, and (b) the column vector of those same subtests' relations with the variable in question. When the latter variable is dichotomous the relations are usually calculated in terms of an effect size statistic. When the

* Corresponding author at: Gouden Leeuw 746, 1103 KR Amsterdam, the Netherlands.
E-mail address: JanteNijenhuis@planet.nl (J. te Nijenhuis).

0160-2896/$ – see front matter
http://dx.doi.org/10.1016/j.intell.2013.03.001

latter variable is continuous (or nearly so), the relations are usually calculated in terms of a correlation coefficient (Ashton & Lee, 2005). Jensen (1998, ch.10) quite convincingly argues that it is plausible that the true correlation between *g* loadings and B/W IQ test score differences is close to .90. Virtually all other group differences that have been studied using correlations between *g* loadings and group differences show substantial to high positive correlations (Rushton, Čvorovič, & Bons, 2007; Rushton & Jensen, 2003; te Nijenhuis & van der Flier, 2003). What makes these findings so important is that all studies computing correlations between heritabilities of subtests of an IQ battery and the *g* loadings of these same subtests show substantial to strong values of *r* (Jensen, 1987; Pedersen, Plomin, Nesselroade, & McClearn, 1992; Rijsdijk, Vernon, & Boomsma, 2002; Spitz, 1988).

1.2. The Flynn effect

Many studies have shown that IQ scores have been increasing over the last half century (Flynn, 2012). The average gain on standard broad-spectrum IQ tests between 1930 and 1990 was three IQ points per decade. Recently, however, studies from Scandinavia suggest the secular gains may have stopped in Western, industrialized countries, although the gains are still in progress in Estonia (Must, te Nijenhuis, Must, & van Vianen, 2009). The secular gains are massive and the time period too short for large positive genetic changes in the population, so there is strong consensus that the changes must be largely environmental. There may, however, be a quite modest role for a genetic effect called heterosis, meaning that genetically unrelated parents have children with IQs that are slightly higher than the mean IQ of the general population (see Mingroni, 2007; Woodley, 2011).

Are the strong environmental forces causing the scores over generations to rise the same as the forces causing the group differences? Rushton (1999) showed that secular gains from four Western countries had modest to small negative correlations with *g* loadings. Rushton's (1999) finding has been challenged by Flynn (1999a, 1999b, 2000) and Nisbett (2009) claiming there actually is a substantial positive correlation between secular score gains and *g* loadings. If *g* loadings indeed did correlate highly with both environmental and genetic effects, it would make 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, 1999a, 1999b, 2000; Must, Must, & Raudik, 2003; te Nijenhuis, in press; te Nijenhuis & van der Flier, 2007; Wicherts et al., 2004) yielding correlations ranging from substantial and negative to large and positive. The present paper aims to reduce the uncertainty regarding the question how strongly the Flynn effect is on the *g* factor by carrying out a psychometric meta-analysis of all published studies on this subject.

2. Method

To test the size of the true correlation between *g* loadings of tests and secular score gains (*d*), we carried out a meta-analysis of all published studies reporting correlations between *g* loading of tests and secular score gains. We identified all studies for the meta-analysis by manual search of Jensen (1998, ch. 10) and the journals Personality and Individual Differences, Intelligence, Psychology in the Schools, Journal of School Psychology, and Journal of Clinical Psychology. Additional search strategies were manual searches at four Dutch Universities, and a computer search of library databases available to us, including ERIC, PsycINFO, MEDLINE, PiCarta, Academic search premier, Web of Science, PubMed, Education-line, SSRN, Cogprints, ROAR and Open DOAR. We used the following keywords: Flynn effect, secular score gains, Jensen effects, and method of correlated vectors. The reference sections of the articles and book chapters obtained were checked, and researchers contributing to this specialist discussion were contacted.

To be included in the present review the following criteria had to be met. First, studies had to report secular gains on well-validated tests. Second, to get reliable correlations between *g* and *d* batteries had to be comprised of at least seven subtests. This choice is based on our experience in a psychometric meta-analysis of the correlation between retest effects and *g* loadings (te Nijenhuis, van Vianen, & van der Flier, 2007), where including datasets with less than seven subtests gave inconsistent results. One could choose to counter with extra strong corrections for unreliability, but we decided to increase reliability by simply dropping the small datasets. There is also a practical consideration that limiting oneself to batteries with, for instance, at least 12 subtests would mean that there are virtually no datasets that could be analyzed with the Method of Correlated Vectors. Application of these inclusion rules yielded five papers with 11 correlations between *g* and *d*.

Psychometric meta-analysis (Hunter & Schmidt, 2004) aims to estimate what the results of studies would have been if all studies had been conducted without methodological limitations or flaws. One of the goals of the present meta-analysis is to have a reliable estimate of the true correlation between *g* loadings and secular score gains (*d*). The collected studies were heterogeneous across various possible moderators.

Standardized secular score gains were computed by dividing the raw score gain by the S.D. of the earlier sample. In general, *g* loadings were computed by submitting a correlation matrix to a principal axis factor analysis and using the loadings of the subtests on the first unrotated factor. In some cases *g* loadings were taken from studies where other procedures were followed; these procedures have been shown empirically to lead to highly comparable results (Jensen & Weng, 1994). Pearson correlations between the standardized score gains and the *g* loadings were computed.

Psychometric meta-analytical techniques (Hunter & Schmidt, 2004) were applied to the resulting 11 r_{gd}s using the software package developed by Schmidt and Le (2004). Psychometric meta-analysis is based on the principle that there are artifacts in every dataset and that most of these artifacts can be corrected for. In the present study we corrected for five artifacts that alter the value of outcome measures listed by Hunter and Schmidt (2004): (1) sampling error, (2) reliability of the vector of *g* loadings, (3) reliability of the vector of score gains, (4) restriction of range of *g* loadings, and (5) deviation from perfect construct validity. We present the outcomes step by step.

2.1. Correction for sampling error

In many cases sampling error explains the majority of the variation between studies, so the first step in a psychometric meta-analysis is to correct the collection of effect sizes for