Personality and Individual Differences 55 (2013) 387–390

Contents lists available at SciVerse ScienceDirect

Personality and Individual Differences

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

Establishing an association between the Flynn effect and ability differentiation

Michael A. Woodley [a,b,*], Guy Madison [a]

[a] Umeå University, Department of Psychology, Umeå, Sweden
[b] Vrije Universiteit Brussel, Center Leo Apostel for Interdisciplinary Studies, Brussels, Belgium

ARTICLE INFO

Article history:
Received 23 February 2013
Received in revised form 19 March 2013
Accepted 25 March 2013
Available online 1 May 2013

Keywords:
Ability differentiation
CD-IE
Flynn effect
g Loading

ABSTRACT

The relationship between the Flynn effect and ability differentiation is investigated in a reanalysis of published data on Estonian student cohorts tested in 1933/36, 1997/98 and 2006 on the National Intelligence Test (Must, te Nijenhuis, Must, & van Vianen, 2009). To determine whether there was a relationship we computed the vector correlation between the Flynn effects (d) and the change in the g loading (Δg) between measurement occasions for each of the 10 NIT subtests and for each of the seven cohort comparisons, giving a total N of 70 effect sizes. The association between d and Δg was robustly negative (indicating that the Flynn effects were negatively associated with changes in the g loading of subtests) for all cohort comparisons, with values of r ranging from $-.100$ to $-.461$ ($N = 10$). When all effect sizes were analyzed together, the vector correlation was found to be $-.281$ ($p \leqslant .05$, $N = 70$). This indicates a significant association between the Flynn effect and ability differentiation. Possible causes of this association are discussed.

1. Introduction

The Flynn effect is an aggregate increase in pen-and-paper IQ test performance of about three points per decade (Flynn, 1987, 2009). The effect appears to occur unequally across measures of different cognitive abilities, however. The largest Flynn effects tend to be reported for measures of fluid intelligence such as the Raven's Progressive Matrices (between 5–7 points per decade, Flynn, 2009), and the smallest ones for measures of crystallized intelligence (around 2–3 points per decade; Flynn, 2009). The Flynn effect also does not appear to be a Jensen effect in that the preponderance of studies indicate that its effect magnitude is typically either unrelated or negatively related to the *g*-loadings of the subtests on which it occurs (Jensen, 1998; Must, Must, & Raudik, 2003a, 2003b; Rushton, 1999; te Nijenhuis, 2013; te Nijenhuis & van der Flier, 2006, 2007, in press; Woodley & Meisenberg, 2013). Despite this some studies have reported Jensen effects on the effect in cases when it is considered in the context of fluid rather than crystallized measures of intelligence (Colom, Juan-Espinosa, & García, 2001; Flynn, 1999a, 1999b, 2000).

The Flynn effect's differential affinity for test batteries, coupled with the disputed idea that it might be associated with an actual increase in *g*, has led some researchers to attempt to connect it with Spearman's Law of Diminishing Returns (SLODR), or the *ability differentiation hypothesis* (e.g. Juan-Espinosa, Cuevas, Escorial, & García, 2006; Lynn & Cooper, 1993, 1994). This is the observation that as the level of *g* increases in a population, the diversity of more specialized cognitive abilities also increases. In other words the overall variance attributable to the *g* factor decreases relative to the subtest or group-factor-specific variance (*s*), which increases as a function of increasing level of ability (Jensen, 2003; Spearman, 1927).

A handful of studies have demonstrated that the Flynn effect is associated with a lack of *factorial invariance* between times of measurement, that is, across the dimension on which the Flynn effect occurs (Must et al., 2009; Wicherts et al., 2004). In other words, what IQ tests measure over time appears to change as a result of the Flynn effect, which could be taken to support ability differentiation in the sense that subtest-specific or group factors influence performance on IQ tests to a greater extent in higher-ability relative to lower-ability cohorts.

Several studies have explicitly investigated the relationship between the Flynn effect and the ability differentiation hypothesis and have found evidence for an association between the two phenomena in countries such as France (Lynn & Cooper, 1993), Japan (Lynn & Cooper, 1994), the US (Juan-Espinosa et al., 2006; Kane, 2000; Kane & Oakland, 2000), Norway (Sundet, Barlaug, & Torjussen, 2004) and Estonia (Must et al., 2003a, 2009). These studies typically infer a relationship between the Flynn effect and ability differentiation on the basis that Flynn effects appear to be occurring concomitantly with a negative change in the variance proportion accounted for by the g factor over time. It is important to note that these studies

* Corresponding author at: Umeå University, Department of Psychology, Umeå, Sweden.
E-mail address: Michael.Woodley@psy.umu.se (M.A. Woodley).

0191-8869/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.paid.2013.03.016

do not attempt to establish the *degree* to which the Flynn effect is associated with this process instead they merely demonstrate that the two phenomena are occurring concurrently. Indeed this would be difficult to do given the relatively small number of data points typically collected in such studies, and also the relatively small differences in *g* saturation that are observed over time.

Here, we directly test for an association between the two phenomena utilizing a variant of the method of correlated vectors. This method has traditionally been used in exploring the affinity of an association between IQ and another variable (such as time in the case of the Flynn effect) and the *g* loading of the tests on which the association is established. This parameter can be determined by simply correlating the vector of the association or effect size (*d*) with the vector of the subtest *g* loadings. Positive correlations indicate the presence of Jensen effects, i.e. that the change in the strength of the IQ/variable association is positively mediated by *g* saturation (Jensen, 1998; Rushton, 1998).

2. Methods

As the reported change in the variance proportion of the *g* factor over time across studies is typically small, a relatively large number of subtests will be needed in order to determine if there is any kind of correlation between this tendency and the magnitude of the Flynn effect (*d*). To that end we chose to reanalyze the data presented in the study of Must et al. (2009). This study provides evidence for the continuity of the Flynn effect in Estonia between the years 1933/36, 1997/98 and 2006 using the Estonian National Intelligence Test, which is divisible into ten subtests administered to (1) a sample of 12–13 year old students in 1933/36 ($N = 270$) and again in 2006 ($N = 243$); (2) a sample of 13–14 year old students in 1933/36 ($N = 222$), again in 1997/98 ($N = 224$) and again in 2006 ($N = 343$); (3) a sample of 14–15 year old students in 1933/36 ($N = 407$), again in 1997/98 ($N = 137$) and again in 2006 ($N = 327$).

Flynn effects are computed by comparing all testing year permutations (i.e. 1933/36 – 1997/98; 1933/36 – 2006; 1997/98 – 2006) yielding a total of 70 subtest-specific Flynn effects for the whole sample. The study also reports the *g* loadings for each subtest at each testing occasion, in addition to demonstrating that the Flynn effect across cohorts appears to be concomitant with respect to a small overall decrease in the average *g* saturation of the NIT (1933/36 – 1997/98 = −.02; 1933/36 – 2006 = −.02), suggesting a potential role for ability differentiation. Here the difference in the *g* loadings (Δg) for each subtest is calculated by subtracting the *g* loading at the later time-point from the value at the earlier time point, hence a negative value indicates a decrease in the size of the subtest *g* loading with time and a positive value indicates the opposite tendency.

All of these values are rescaled to measure decadal change, taking the median year in cases where the measurements take place over multiple years (i.e. 1934.5 in the case of the 1933/36 measurement occasion), so that they are directly comparable. Finally each set of ten effect sizes is *z*-transformed based on the means and standard deviations of each set so as to reduce the error of measurement associated with comparing between cohorts using effect sizes standardized based on the whole sample of effect sizes.

To compute the vector correlation between *d* and Δg the Pearson's product moment correlation (*r*) is employed. Two sets of findings are reported, namely (1) the vector correlations within each of the seven sets of ten subtests for each cohort comparison, and (2) the overall vector correlation for all effect sizes. In the case of the latter the bi-directional significance level is computed on the basis of the subtest/effect size number (70), rather than the combined sample *N* (2,173) so as to reduce the potential for Type I error.

Table 1
Vector correlations computed for the seven paired cohort comparisons and all effect sizes combined.

Cohort comparison (*N* subtests)	Vector correlation (*r*); *d* vs. Δg
12–13 year olds 1934.5 vs. 2006 (10)	−.220
13–14 year olds 1934.5 vs. 1997.5 (10)	−.461
13–14 year olds 1997.5 vs. 2006 (10)	−.250
13–14 year olds 1934.5 vs. 2006 (10)	−.132
14–15 year olds 1934.5 vs. 1997.5 (10)	−.413
14–15 year olds 1997.5 vs. 2006 (10)	−.391
14–15 year olds 1934.5 vs. 2006 (10)	−.100
All effect sizes (70)	−.281*

* $P \leqslant .05_{\text{(bi-directional)}}$.

3. Results

Table 1 presents the vector correlations computed for all seven sets of ten subtests representing all paired cohort comparisons, along with the vector correlation for the full set of effect sizes.

Figure 1 presents the results of a scatter plot illustrating the distributions of all *z*-transformed values of *d* and Δg.

4. Discussion

Based on the use of a variant of the method of correlated vectors we can demonstrate that there is a significant and negative association between the size of the Flynn effects and the change in the *g* loadings of the subtests on which they occur. This analysis is furthermore the first to quantify this relationship rather than simply relying on inference. The results presented here indicate that a change in the *g* loadings of subtests accounts for nearly 10% of the variance in the Flynn effect, which suggests a modest association between these two variables. Furthermore, although the *N*'s were too small to establish significance in the case of the seven paired cohort comparisons, in all cases the signs on the resultant vector correlations were in the expected negative direction (i.e. the Flynn effect is associated with an overall negative change in direction of the subtest *g* loadings). Despite their non-significance the consistency of the direction of the effects across different cohort comparisons suggests some validity via the criterion of inductive generalization. This interpretation is bolstered by the observation that these vector correlations are non-heterogeneous in magnitude when subjected to a chi-square test ($\chi^2 = 1.03$, df = 6, $P > .05$).

The results of this analysis can therefore be considered as directly confirmatory of the hypothesis that the Flynn effect is associated with ability differentiation, however the *source* of this ability differentiation is less clear.

One hypothesis (as was discussed in the introduction) is that this result indicates a parallel between the Flynn effect and SLODR, in that the former might be associated with an increase in the population level of *g*, and hence will replicate the phenomenology of the latter effect (e.g. Juan-Espinosa et al., 2006; Lynn & Cooper, 1993, 1994). Compelling evidence for a dysgenic effect concentrated on *g* speaks against this interpretation: the magnitude of the gradient of dysgenic fertility (i.e. the size of the negative correlation between subtest performance and completed fertility) is clearly a strong Jensen effect (Woodley & Meisenberg, in press). Recall that the Flynn effect has the opposite phenomenology: the preponderance of studies along with the results of psychometric meta-analysis clearly indicate that it tends to be less pronounced on the more *g* loaded subtests (te Nijenhuis & van der Flier, in press). Given that subtest heritabilities are positively monotonically related to subtest *g* loadings (Rushton & Jensen, 2010), this suggests a strongly genetic basis for the effects of dysgenics