Personality and Individual Differences 88 (2016) 280–284

Contents lists available at ScienceDirect

Personality and Individual Differences

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

Showing their true colours: Possible secular declines and a Jensen effect on colour acuity — More evidence for the weaker variant of Spearman's Other Hypothesis

Michael A. Woodley of Menie [a,b,*], Heitor B.F. Fernandes [c]

[a] Department of Psychology, Technische Universität Chemnitz, Chemnitz, Germany
[b] Center Leo Apostel for Interdisciplinary Studies, Vrije Universiteit Brussel, Brussels, Belgium
[c] Department of Psychology, Federal University of Rio Grande do Sul, Porto Alegre, Brazil

ARTICLE INFO

Article history:
Received 24 April 2015
Received in revised form 14 August 2015
Accepted 3 September 2015

Keywords:
100 Hue
Colour acuity
Co-occurrence model
Jensen effect
Secular decline
Spearman's Other Hypothesis

ABSTRACT

Spearman's Other Hypothesis predicts that the common factor amongst sensory discrimination measures corresponds to general intelligence (*g*). The co-occurrence model predicts that low-complexity physiological information-processing indicators reliably measure *g* across cohorts, and should therefore decline with time due to genetic changes in the broader population. As strong relations exist between general sensory discrimination and *g*, such measures should show evidence of secular declines. This is tested using *N*-weighted temporal regression of square-root Total Error Scores (√TES), obtained from four Western normative samples published in the 1980s, 90s and 2000s (combined $N = 752$) evaluated using the Farnsworth–Munsell 100-Hue colour acuity test (disattenuated *g* loading = .78). A significant temporal β value of .37 was found (controlling for national IQ), suggesting a decline in colour acuity equating to a reduction in *g* of −3.15 points per decade. Analysis of the subset of the cohorts aged 20–29 years, in which colour acuity is maximized, reveals a larger secular decline ($\beta = .67$, $N = 199$, −5.85 points per decade). The small number of studies employed in these analyses makes these findings tentative however. Also consistent with a weaker variant of the Other Hypothesis is the finding that 100-Hue acuity-IQ correlations are associated with the Jensen effect. The aggregate vector correlation across two studies is .63 ($N = 932.5$, $p < .05$).

1. Introduction

Sir Francis Galton (1883) was the first to propose that fine sensory discrimination might be associated with higher general intelligence (i.e. the cognitive processes common to the solving of diverse tests of mental ability; Jensen, 1998; Spearman, 1904), however the claim was not investigated empirically until Spearman (1904) conducted an analysis in which student performance on various sensory discrimination tasks (lightness, sound and weight) was found to correlate with teacher-ratings of their general intelligence (*g*). Spearman further posited the existence of a *general factor of discriminative ability* existing in the common factor variance amongst various sensory discrimination tasks. He argued that this should correlate nearly perfectly with *g*.

After many decades of neglect, in the 1990s and 2000s a series of papers by Ian Deary revisited what came to be termed *Spearman's Other Hypothesis* (Deary, 1994; Deary, 2000a, 2000b). The first direct test of the Other Hypothesis was conducted in 2004, when Deary and co-workers collected data on various sensory discrimination tasks amongst a sample of 62 Scottish secondary school students, along with various measures of IQ. Utilizing structural equations modelling (SEM) to estimate the common factor variance amongst the sensory discrimination and the cognitive ability measures, the latent general discrimination and *g* factors were found to correlate at .92, making them virtually isomorphic — consistent with the prediction of the Other Hypothesis. In a second analysis, Deary, Bell, Bell, Campbell, and Fazal (2004) reanalysed a much larger dataset (899 individuals) for which measures of both cognitive and sensory discrimination ability had been collected and analysed in a previous publication (Acton & Schroeder, 2001). Using the same SEM-based method it was found that *g* correlated with the general discrimination factor at .676 for the male and .681 for the female cohort, which indicated some divergence between the two common factors, but also demonstrated considerable shared variance, consistent with a weaker form of the Other Hypothesis.

Subsequent studies have demonstrated convergence between sensory discrimination and *g* using different modalities of sensory discrimination (e.g. Meyer, Hagmann-von Arx, Lemola, & Grob, 2010).

1.1. The co-occurrence model

IQ is a highly heterogeneous measure. Applying the basic bi-factor variance components model first proposed by Spearman (1904), IQ is

* Corresponding author at: Department of Psychology, Technische Universität Chemnitz, Chemnitz, Germany.
E-mail address: Michael.Woodley@vub.ac.be (M.A. Woodley of Menie).

http://dx.doi.org/10.1016/j.paid.2015.09.009

broadly composed of general and specialized ability variances (*g* and *s*). As was mentioned previously, *g* variance relates to the cognitive processes common to the solving of diverse tests of mental ability whereas *s* variances relate to uncorrelated and specialized sets of processes that are specific to the domain of each ability.

It has been found that the magnitude of the negative correlation between IQ and fertility relates positively to the *g* loading of the test on which it is established, making it a *Jensen effect* (Woodley & Meisenberg, 2013). The *g* loadings of tests also positively relate to the strength of their association with phenotypic indicators of mutation load (i.e. the complement of deleterious mutations), such as measures of fluctuating asymmetry and factors that increase mutation load, such as inbreeding depression (Prokosch, Yeo, & Miller, 2005; Rushton & Jensen, 2010). IQ subtests that are more *g*-loaded are also more heritable on average (see: Woodley of Menie, Fernandes & Hopkins, 2015 for a review of this literature), indicating that genetic factors influence individual differences in *g* to a greater extent than *s*.

At the opposite end of the continuum are variables that influence IQ primarily via environmental routes, such as educational interventions (te Nijenhuis, Jongeneel-Grimen & Kirkegaard, 2014), IQ gains amongst adopted children (te Nijenhuis, Jongeneel-Grimen & Armstrong, 2015), IQ gains via retesting effects (te Nijenhuis, van Vianen & van der Flier, 2007), and the Flynn effect (te Nijenhuis & van der Flier, 2013). In all cases, the effects of these environmental factors are bigger when the sub-scales are *g*-loaded to a lesser (and *s* loaded to a correspondingly greater) extent.

This observation has led to the development of the *co-occurrence model* which posits that genetic changes occurring within modern populations should be reducing the level of *g*, whereas environmental improvements of one sort or another should be increasing performance with respect to numerous *s*-variances associated with cognitive abilities simultaneously.

Whilst the gain in *s* is well-attested on the basis of ubiquitously rising IQ scores, there are several lines of evidence supporting co-occurrent declines in indicators of *g*, such as historical declines in per capita macro-innovation and genius, suggesting falling creativity (Huebner, 2005, Murray, 2003), diminishing backwards digit span (a working memory measure; Woodley of Menie & Fernandes, 2015), and decreased usage of high-difficulty words over 155 years, across a representative corpus of the English language, indicating declining crystallized ability, specifically vocabulary in direct response to genetic selection (Woodley of Menie, Fernandes, Figueredo & Meisenberg, 2015).

More germane to the issue of sensory discrimination is the finding of an apparent decline in performance on measures of simple reaction time (SRT) since the 19th century (Silverman, 2010; Woodley, te Nijenhuis & Murphy, 2013). This finding was not received uncritically (e.g. Dodonova & Dodonov, 2013). Reanalysis of a more closely matched subset of the original data to which various adjustments for methods variance had been made revealed declines of between −.57 and −1.21 points per decade, based on different applications of the corrections (Woodley, te Nijenhuis & Murphy, 2014). Tentative indications of generational slowing of SRT have also been found by comparing the extrapolated longitudinal to the observed cross-sectional ageing trend amongst various studies (Verhaeghen, 2014). A more detailed analysis of three Scottish birth cohorts utilizing a variant of this method, found that amongst the female cohorts the secular slowing in SRT performance equates to a decrease in *g* of −1.8 points per decade (Woodley, Madison & Charlton, 2014).

Even though SRT does not correlate strongly with IQ (−.31, rising to −.54 when corrected for range restriction, reliability and validity; Woodley et al., 2013), it is nonetheless likely a stable measure of a phenotype fundamental to *g* (i.e. information processing speed) across cohorts, as its simplicity reduces its sensitivity to training effects (Jensen, 2006), which may actually be increasing performance on more complex measures of processing speed (Verhaeghen, 2014). Pencil and paper tests on the other hand are sensitive to training effects and changes in test taking habits, therefore they frequently fail to measure the same parameter across cohorts (Wicherts et al. 2004). This lack of measurement invariance across cohorts is the reason why tests can sometimes be highly *g*-loaded within cohorts (such as the Raven's Progressive Matrices) but may nonetheless yield large Flynn effects between cohorts, tracking instead the development of specialized skills and abilities (e.g. Fox & Mitchum, 2013).

Sensory discrimination tasks are therefore potentially ideal for tracking secular trends in *g* over time, as like SRT, they exhibit low-complexity and are potentially resistant to training, therefore they should be relatively stable measures of a fundamental psychophysical indicator of *g* across generations.

1.2. The present study

Here the focus will be on colour discrimination ability evaluated using the Farnsworth–Munsell 100-Hue colour perception test (Farnsworth, 1943), which was first found to correlate with IQ in the 1960s (Lakowski, 1970). This test evaluates colour acuity by having the participants physically arrange a series of 85 caps, each of subtly different hue, along a spectrum defined by two end caps (e.g. blue to green, pink to purple etc.). Normative performance data have been collected from four Western populations between the 1980s and 2000s, which will here be reanalysed for the presence of possible secular trends.

In order to validate potential declines in colour acuity with respect to potential underlying declines in *g*, an analysis involving the method of correlated vectors will be performed by reanalysing data from Acton and Schroeder (2001) and Deary and co-workers (2004), which provide information on subtest-100-Hue acuity correlations, along with *g* loadings. This is an important relationship to establish as single indicators can contribute to a *g*-loaded aggregate even when, taken on their own, they relate more to specialized abilities (e.g. Inspection Time; Deary & Crawford, 1998).

2. Methods

2.1. Secular trend estimation

Secular trend data on 100-Hue acuity were obtained using four normative studies published in the 1980's (Verriest, Van Laetham, & Uvijls, 1982), 90s (Roy, Podgor, Collier, & Gunkel, 1991), and 2000s (Kinnear & Sahraie, 2002; Mäntyjärvi, 2001). These studies were conducted on mixed-sex samples of 232 Belgian, 112 American, 160 Finnish and 286 British subjects with the greatest age range extending from five to >80 years. Each of these studies employed the physical 'hardware' variant (as opposed to a more recent electronic variant) of the 100-Hue test. Very few studies have attempted to examine the normative characteristics of the 100-Hue test. Earlier normalizations were carried out on a sample of Americans in the 1950s (Farnsworth, 1957) and also amongst a Belgian sample in the 1960s (Verriest, Vandevyvere, & Vanderdonck, 1962), however owing to differences in the instrumentation used and also the statistical procedures employed in analysing the data, only the results of the more recent four studies are directly comparable to one another at the methodological level (Kinnear & Sahraie, 2002). There also exists biocultural heterogeneity between the countries, which potentially restricts comparability. A major source of this is national IQ, which ranges from 97.5 (USA) to 100.9 (Finland) (Lynn & Vanhanen, 2012).

100-Hue test performance is not associated with sex differences (Kinnear & Sahraie, 2002), however it is strongly age-dependent, with the highest Total Error Scores (the sum of the cap arrangement errors — these values are typically transformed by taking the square-root in order to normalize their distributions [$\sqrt{\text{TES}}$]) being found amongst very young samples (i.e. <10 years of age) and those in old age. The lowest $\sqrt{\text{TES}}$ values are typically found amongst young adults in their 20s. In their comparative analysis of three normative samples, Kinnear and Sahraie (2002) compared the different age groups in each study with