Intelligence 57 (2016) 1–6

Contents lists available at ScienceDirect

Intelligence

A NIT-picking analysis: Abstractness dependence of subtests correlated to their Flynn effect magnitudes

Elijah L. Armstrong [a,*], Jan te Nijenhuis [b], Michael A. Woodley of Menie [c,d], Heitor B.F. Fernandes [e], Olev Must [f], Aasa Must [g]

[a] Washington University in St. Louis, United States
[b] National Research Center for Dementia, Chosun University, Gwangju, Republic of Korea
[c] Department of Psychology, Technische Universität Chemnitz, Germany
[d] Center Leo Apostel for Interdisciplinary Research, Vrije Universiteit Brussel, Belgium
[e] Department of Psychology, Federal University of Rio Grande do Sul Department of Psychology, Brazil
[f] Department of Psychology, University of Tartu, Estonia
[g] Estonian National Defense College, Estonia

ARTICLE INFO

Article history:
Received 21 December 2015
Received in revised form 24 February 2016
Accepted 24 February 2016
Available online 28 March 2016

Keywords:
Abstract thinking
Flynn effect
Intelligence
National Intelligence Test
Estonia
g loading

ABSTRACT

We examine the association between the strength of the Flynn effect in Estonia and highly convergent panel-ratings of the 'abstractness' of nine subtests on the National Intelligence Test, in order to test the theory that the Flynn effect results in part from an increase in the use of abstract reference frames in solving cognitive problems. The vectors of abstractness ratings and Flynn effect gains, controlled for guessing) exhibit a near-zero correlation ($r = -.02$); however, abstractness correlates positively with (and is therefore confounded by) *g*-loadings ($r = .61$). A General Linear Model is used to determine the degree to which the abstractness vector predicts the Flynn effect vector, independently of subtest *g*-loadings and the portion of the secular IQ gain due to guessing (the Brand effect). Consistent with the abstract reasoning model of the Flynn effect, abstractness positively predicts Flynn effect magnitudes, once controlled for confounds ($sr = .44$), which indicates an increasing tendency to utilize factors external to the items in order to abstract their solutions.

1. Introduction

The Flynn effect describes the tendency for IQ scores to rise across tests at a rate of approximately three points per decade (Flynn, 2009; Pietschnig & Voracek, 2015). The causes of this effect are unknown, although many factors have been postulated, including reduced inbreeding, better education, improved nutrition, lower parasite prevalence, and slower life history speed (see Williams, 2013, and Pietschnig & Voracek, 2015, for reviews of possible causes). To better understand the effect's etiology, it is helpful to understand the profile of tests on which it is most pronounced (e.g., Lynn, 1990; Rushton, 1999; Pietschnig & Voracek, 2015; see also Rushton & Jensen, 2005). Previous research has documented that the Flynn effect is more prominent on tests with lower *g* loadings, i.e., that correlate less strongly with the set of other tests (te Nijenhuis & van der Flier, 2013); stronger on fluid, as opposed to crystallized, tests (e.g., Pietschnig & Voracek, 2015); and stronger on tests of mathematical achievement, as opposed to verbal achievement (e.g., Herrnstein & Murray, 1994; Rindermann & Thompson, 2013; Wai & Putallaz, 2011). The present study investigates one further proposed determinant of Flynn effect strength — namely *abstract thinking ability*: the capacity to infer general properties when solving problems, and to ignore irrelevant concrete facts (e.g., Flynn, 2009; Jensen, 1998; Pinker, 2011; Terman, 1921, 1922; see Flynn, 1998 for criticism).[1] Some tests, which rely heavily on this ability, such as the

[1] Jensen (1998), for instance, defines a similar concept: "In almost every subject in the school curriculum, pupils learn to discover the general rule that applies to a highly specific situation and to apply a general rule in a wide variety of different contexts. The use of symbols to stand for things in reading (and musical notation); basic arithmetic operations; consistencies in spelling, grammar, and punctuation; regularities and generalizations in history; categorizing, serializing, enumerating, and inferring in science, and so on. Learning to do these things, which are all part of the school curriculum, instills cognitive habits that can be called *decontextualization* of cognitive skills" (p. 325). This definition, however, does not exhaust abstract thinking as we define it: we include taking false or unknown hypotheticals seriously, and having absorbed, and being able to apply, scientific concepts, in our definition (ref. Terman, 1956). Our definition includes one analogical or "decontextualization"-based item, one "scientific spectacles" item that requires answering based on abstract rather than concrete similarities, and one syllogism with a bizarre premise that requires taking false hypotheticals seriously. The second and third are from Flynn (2009) and Luria (1976), respectively; the first is from Flynn (2012) summarizing Fox and Mitchum (2013). These different aspects of abstract reasoning are theoretically separable, but the high correlation between ratings suggests they were related in this dataset.

* Corresponding author.
E-mail address: armstrong357@wustl.edu (E.L. Armstrong).

http://dx.doi.org/10.1016/j.intell.2016.02.009
0160-2896/

Wechsler Similarities (Flynn, 2009) and the Raven's Matrices (Flynn, 2012; Fox & Mitchum, 2013), have shown large Flynn effects.

Flynn (2009, 2012) has proposed that improvements in abstract thinking ability explain much of the Flynn effect. He reasons that over time, people have become better able to use scientific classifications in examining the world. Along similar lines, Fox and Mitchum (2013) have demonstrated greater gains in RPM items that are more dependent on analogical reasoning, a related construct to abstract thinking; Jensen (1998) hypothesizes that improvements in decontextualized problem solving, as taught in schools, may contribute to the Flynn effect; and Armstrong and Woodley (2014) suggest an improvement over time in the ability to induce and apply rules. Sowell (1978) provides some evidence that culturally deprived groups show particularly poor abstract reasoning capacity.

The present paper represents a quantitative test of the hypothesis that abstract thinking ability and Flynn effects are related. In the present study, we test the abstract thinking hypothesis using the Estonian translation of the National Intelligence Test. We employ guessing-corrected secular IQ gain data for nine subtests, independently rated for abstract thinking dependence by twenty-seven raters. We further control for potential confounds of the Flynn effect magnitudes per subtest based on other theoretical predictors.

2. Methods

Flynn effects were derived from the difference in scores between 1933/36 and 2006 administrations of the National Intelligence Test to samples of Estonian schoolchildren ($N = 890$ for the older sample, 913 for the more recent sample; Must & Must, 2013). For examples of subtest items, see Must and Must (2013).

The Method of Correlated Vectors (MCV) was utilized to determine the effect of abstractness on the Flynn effect independent of both subtest *g* loadings and the Brand effect — or the portion of the secular gain in IQ that is due purely to the results of guessing.

We employed the following techniques in the study.

2.1. Ratings of abstractness

We had 28 raters independently assess the abstract thinking dependence of the ten subtests of the British version of the *National Intelligence Test*, an early group test of intelligence (Haggerty, Terman, Thorndike, Whipple, & Yerkes, 1920; see Whipple, 1921 for some remarks on its development); we sent all raters a .pdf file with a complete copy of the test coupled with a short text on abstract thinking dependence (Supplementary Material 1). The National Intelligence Test was translated into Estonian by Tork (1940) and administered to large samples of children across seventy years. Thus, while our raters assessed the English-language version of the NIT, the data on Flynn effects and *g* loadings are derived from the Estonian translation. The discrepancy, if any, produced by this inconsistency has not been examined. None of the raters were authors on the paper.

The 28 raters used to obtain the *abstract thinking dependencies* for each subtest were classified into the following categories: nonprofessionals (without degrees in psychology), graduate students, or professionals ($N = 10$ nonprofessionals, 5 graduate students, 13 professionals). Each rater rated the abstract thinking dependency of each subtest on a scale from 0–100, using a text vignette defining abstract thinking (Supplement 1) as a rating criterion. The text gave examples of three hypothetical test items heavily dependent on abstract thinking; one was drawn from Luria (1976), one from Flynn (2009), and one from Flynn (2012) in a discussion of Fox and Mitchum (2013). The raters used Form 2 of the British National Intelligence Test to rate the abstract thinking dependence of each subtest.

Impressionistic ratings of test criteria have been utilized previously in the literature (e.g., Helms-Lorenz, Van de Vijver, & Poortinga, 2003; Kan, Wicherts, Dolan & van der Maas, 2013; McGurk, 1953). Ceci (1996; ch. 9) discusses the concept of 'abstractness' as related to IQ tests; he reviews two studies in which the abstractness of tests was rated, in both cases finding minimal consistency between raters (and zero correlations with ethnic differences on the test in question).

Amongst the nine subtests employed in the present study, the Cronbach's alpha values for the abstract thinking dependencies, calculated using all 28 raters, was .95, which is considered "excellent" (Cohen, 1988). (With subtest B4 excluded as a potential outlier, the reliability dropped to .67.) We sampled professional intelligence researchers whereas Ceci did not, perhaps accounting in part for the much higher interrater reliability we found. We believe our drastically different results from Ceci may also in part result from our more definite characterization of abstract reasoning in terms of three specific items (representing syllogisms with false premises, analogical items, and "scientific" similarities).

We performed a supplementary analysis to determine whether the abstractness ratings were indeed reliable. Test A5 and B5 are highly similar, loading on processing speed, perceptual speed, or clerical ability (Flanagan et al, 1962; Carroll, 1993; Wechsler et al., 2008, but see p. 472 in Carroll, 1993), and seem to have similar levels of abstractness (although the former test involves abstract stimuli, symbols, more heavily, while the latter test often includes names, which are more concrete). However, the abstractness ratings correlated at only .48, about average for the correlations between two pairs of tests. To supplement this result, we also examined the correlations for the pair of subtests A1 and B1 (both involving numerical computation; .83) and A4 and B3 (both testing vocabulary; .39). This supplementary analysis indicates a somewhat low level of validity, despite the high level of reliability, for our abstractness ratings.

Finally, note that many of our raters did not speak English as their first language, which may have hurt their ability to assess abstractness based on the text of the subtests (the NIT is highly verbal). However, we did not assess how much this biased the ratings. The exception was one non-native English speaker who warned us about his/her ratings, which we therefore checked (as it turned out, his/hers were very concordant with the others, so we kept them). All ratings positively loaded on the first principal component except those of another non-native speaker, but we did not discard his/hers.

2.2. Confounds

We also used two other theoretical predictors of Flynn effect magnitudes on distinct tests in a regression analysis (Type III Sum of Squares implemented in SPSS v.21):

1) The *g-loading of tests*. Most IQ subtests (and many other cognitive tests) correlate positively with one another (e.g., Jensen, 1998); the *g* loading of a test refers to the size of its correlation with the set of other tests (the higher the *g* loading, the stronger the correlation). Studies have demonstrated that the *g* loadings of tests correlate moderately inversely with the magnitude of the Flynn effect on those tests (te Nijenhuis & van der Flier, 2013) and measurement invariance does not hold for cohort comparisons (Must, te Nijenhuis, Must, & van Vianen, 2009; Wicherts et al., 2004), suggesting that the causes of *g* only in small part (if at all) overlap with the causes of the Flynn effect (e.g., Jensen, 1998; see Meisenberg, 2015).
   It is tempting to say that as *g* represents "general" intelligence, since the Flynn effect is not on *g*, it must be a gain in specialized abilities (e.g., Woodley, 2012). Indeed, if the Flynn effect overlapped strongly with *g*, it would probably be a quite general effect, not just a gain in narrow, test-related or "academic" abilities; this is because *g* is known to be quite general, extending into many domains (e.g., Gottfredson, 1997; but see Kanazawa, 2004; Stanovich, 2009; Sternberg et al., 2000, for some classes of cognitive tasks that may be relatively minimally *g* loaded). The converse inference, however,