Available online at www.sciencedirect.com
ScienceDirect
Personality and Individual Differences 43 (2007) 1259–1265
PERSONALITY AND INDIVIDUAL DIFFERENCES
www.elsevier.com/locate/paid

The secular rise in IQs in the Netherlands: Is the Flynn effect on *g*?

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

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

Received 27 October 2006; received in revised form 22 February 2007; accepted 22 March 2007
Available online 8 June 2007

Abstract

IQ scores have been increasing over the last half century, a phenomenon known as the Flynn effect. In this study, we focused on the question to what extent these secular gains are on the *g* factor. Two IQ batteries: the Interest-School achievement-Intelligence Test (ISI) and the Groningen Final Examination Primary Education (GALO) yielded small and modest negative correlations between standardized gains and *g* loadings. As these studies employ large samples this suggests that the combined literature now shows a modest negative relationship between *d* (the secular change in test score) and *g*.
© 2007 Elsevier Ltd. All rights reserved.

Keywords: Flynn effect; Secular score gains; *g*; *g* score; *g* loadings; IQ tests; Intelligence; Method of correlated vectors

1. Introduction

Many studies have shown that IQ scores have been increasing over the last half century. These increases have been reported in countries on all continents (Flynn, 2006). The most essential research question is to what extent these empirical gains are on the *g* factor and therefore reflect a functional increase of real-life problem solving ability rather than simply an increase in familiarity

* Corresponding author. Tel.: +31 20 6002067.
E-mail address: JanteNijenhuis@planet.nl (J. te Nijenhuis).

0191-8869/$ - see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.paid.2007.03.016

with taking tests or some other less cognitive function. In this paper, the correlation between secular gains and *g* loadings is tested using two datasets on tests that have not been analyzed before.

Jensen (1998, pp. 320, 321) was the first to ask the question whether secular score gains are correlated with *g* loadings. He reported data on four test batteries and concluded that these test's *g* loadings are not highly correlated with the amount of secular change in scores. Rushton (1999) showed that secular test score gains from the US, Germany, Austria, and Scotland had modest to small negative correlations with *g* loadings. He was the first to expand the discussion on the link between secular gains and the general mental ability factor by focusing on patterns in correlations with theoretically linked variables. He carried out a principal components analysis of the secular gains in IQ, along with, amongst others, inbreeding depression scores from cousin-marriages in Japan, and *g* loadings from the standardization samples of the WISC-R and WISC-III. The relevant findings were: (a) the IQ gains on the WISC, WISC-R, and WISC-III form a cluster, showing that the secular trend in overall test scores is a reliable phenomenon; and (b) this cluster is independent of a second cluster formed by inbreeding depression scores (a purely genetic effect), and *g* factor loadings (a largely genetic effect). It should be noted, however, that the factor solution does not show simple structure: there are substantial secondary loadings for many of the variables. Since Rushton's report suggesting secular trends are not strongly related to *g*, various other studies have been carried out (Colom, Juan-Espinosa, & García, 2001; Flynn, 1999a; Flynn, 1999b; Flynn, 2000; Must, Must, & Raudik, 2003; te Nijenhuis et al., submitted; Wicherts et al., 2004) which have produced conflicting results. The present paper aims to reduce the uncertainty regarding the question how strongly the Flynn effect is on the *g* factor by adding two new datasets to the discussion.

2. Method

2.1. Tests

2.1.1. Interests-school achievement-intelligence tests

The Interesse-Schoolvorderingen-Intelligentietests (Interests-School achievement-Intelligence tests) Vorm II (ISI Form II; Snijders & Welten, 1968) is a collection of tests, that measures interests, school achievement, and intelligence. Only the six intelligence tests are used in the present study. Using Carroll's (1993) terminology synonyms measures lexical knowledge (g_{cr}), opposites measures induction (g_{fl}), sorting words measures induction (g_{fl}), cut figures measures visualization (g_{v}), turning figures measures visualization (g_{v}), and sorting figures measures induction (g_{fl}).

2.1.2. Groningen final examination primary education

The Groninger Afsluitingsonderzoek Lager Onderwijs (Groningen Final examination Primary Education) (Kema & Kouwer, 1958) is a test of general intelligence, consisting of nine subtests. Synonyms measures lexical knowledge (g_{cr}), numbers measures quantitative reasoning (g_{fl}) and numerical ability (g_{cr}), verbal analogies measures induction (g_{fl}), figure analogies measures induction (g_{fl}), filling in signs measures quantitative reasoning (g_{fl}) and numerical ability (g_{cr}), filling in words measures cloze ability (g_{cr}), unfolding measures spatial relations (g_{v}), categories measures induction (g_{fl}), and sketch in figures measures visualization (g_{v}).

2.2. Samples

2.2.1. Sample 1: Dutch primary-school children from 1975, 1977, 1979, and 1981

Dr. Theo Nijsse of Groningen University collected scores on the ISI and the GALO in the city of Groningen – the largest town in the north of the Netherlands – and the small towns and villages in its surroundings in 1975, 1977, 1979, and 1981. Great care was taken to collect representative samples (Nijsse, personal communication 2006). First, using the classification of the Institute of Applied Sociology of the University of Nijmegen (van Westerlaak, Kropman, & Collaris, 1975), job level of the father of each child was gathered. Table 1 shows the percentages of children with a father with a low-, middle-, and high-level job for 1975, 1977, 1979, and 1981, respectively. Second, data on low-, middle-, and high-level jobs gathered by the Dutch Central Bureau of Statistics for the complete Dutch male population from the same four years are reported in Table 2 (Dutch Central Bureau of Statistics, 1977; Dutch Central Bureau of Statistics, 1980; Dutch Central Bureau of Statistics, 1982; Dutch Central Bureau of Statistics, 1985). A comparison of the percentages in Tables 1 and 2 clearly shows that the distributions are reasonably similar, but that there is an oversampling of high-SES children and an undersampling of low-SES children. The maximum differences between the cumulative distributions for 1975, 1977, 1979 and 1981 are 4.8%, 14.4%, 9.5% and 12.7%, respectively. Third, to make sure that the mean differences on the tests are not due to selection differences weighing procedures were applied using the percentages from Table 2.

Table 1
Percentages of children taking the ISI and the GALO with fathers with low-, middle-, and high-level jobs in 1975, 1977, 1979, and 1981

Job level father	Year			
	1975	1977	1979	1981
Low	49.4	47.4	46.8	37.0
Middle	25.5	19.2	23.5	35.4
High	25.1	33.5	29.7	27.6

Table 2
Percentages of 15–65 year-old males in the Dutch population with low-, middle-, and high-level jobs in 1975, 1977, 1979, and 1981

Job level father	Year			
	1975	1977	1979	1981
Low	52.0	51.9	50.9	49.7
Middle	27.7	29.1	28.9	28.6
High	20.3	19.0	20.3	21.6

Note. Source for 1975 data: Pocket book statistics 1977, p. 101. Source for 1977 data: Pocket book statistics 1980, p. 122. Source for 1979 data: Pocket book statistics 1982, p. 122. Source for 1981 data: Counting the labor force 1981, p. 27.

Fourth, the data of the four samples from 1975, 1977, 1979, and 1981 were combined, because aggregate data are more reliable. So, Nijsse's combined datasets provide a good approximation of national norms. The ISI ($N = 1227$) was taken in fifth grade and the GALO ($N = 1140$) was taken half a year later, in sixth grade.

2.2.2. Samples 2 and 3: Dutch norm samples of the ISI and GALO

The samples collected by Nijsse were compared against the validation samples of the ISI (Snijders & Welten, 1968) and the GALO (Kema & Kouwer, 1958; Kouwer, 1961). The norm sample reported in the ISI manual ($N = 2000$) was collected in 1966 with great care. Children were selected from various counties varying in degree of urbanization in such a way that they matched the distribution of the degree of urbanization in the whole country as reported by the Dutch Central Bureau of Statistics. For grade five for each of the five categories of degree of urbanization the number of boys and girls was exactly the same. A comparison between the cumulative distribution over urbanization categories in the sample and the nationally representative distribution, using Snijders and Welten's (1968, p. 37) Table 4.5, showed a maximum difference of 0.3%. So, the sample is nationally representative. The influential Dutch Testing Committee (1974) awarded the ISI the highest possible score, meaning its quality, including the representative norm group, is excellent.

The validation of the GALO was carried out around 1954, using a large sample of sixth-grade children ($N = 2218$), in counties differing strongly in number of inhabitants, spread over various parts of the Netherlands (Kouwer, 1961). The sample did not perfectly match the national distributions of various socio-economic variables, but the use of a weighing system turned it into a nationally representative sample (Kouwer, 1961, p. 409).

2.3. Statistical analyses

2.3.1. Gain scores

The mean scores of various cohorts were compared. ISI scores from 1966 were compared against weighed scores from 1975 to 1981. Dutch GALO scores from 1954 were compared against weighed scores from 1975 to 1981. Standardized gain scores were computed by subtracting the score of the earlier sample on the test from the score of the later sample on the same test, and dividing the difference by the standard deviation of the earlier sample.

2.4. g loadings

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.

2.5. Method of correlated vectors

The method of correlated vectors requires the computation of a vector of *g* loadings and a vector of gain scores that are subsequently correlated. Pearson correlations between the standardized score gains and the *g* loadings were computed.

3. Results

3.1. Gains

Table 3 shows there are gains on virtually all the ISI and the GALO subtests. The standardized gains per decade are 0.47 SD (7.05 IQ points) on the ISI's three fluid/visual subtests, and 0.32 SD (4.80 IQ points) on the three verbal/scholastic tests. The standardized gains per decade are 0.21

Table 3
Means and standard deviations on the ISI and the GALO for the two standardization samples and the sample from 1975 to 1981, standardized score gains, and *g* loadings

Tests	Norm sample		1975–1981		*d*	*g*
	m	SD	*m*	SD		
ISI						
Synonyms	8.9	3.3	9.75	3.00	0.26	.63
Opposites	9.5	3.7	11.44	2.98	0.52	.69
Sorting words	11.4	3.5	12.67	3.27	0.36	.68
Cut figures	7.9	3.0	9.27	3.62	0.46	.55
Turning figures	10.7	5.1	12.67	4.61	0.40	.55
Sorting figures	10.7	4.8	14.72	3.97	0.84	.58
GALO						
Synonyms	8.5	4.83	8.94	4.11	0.09	.73
Numbers	10.5	5.33	12.58	5.19	0.39	.67
Verbal analogies	17.5	6.83	20.82	5.76	0.49	.80
Figure analogies	9	4.33	11.92	4.12	0.67	.69
Filling in signs	9.5	4.83	8.90	4.73	−0.12	.67
Filling in words	14	3.67	13.82	3.77	−0.05	.67
Unfolding	8	5.67	12.86	4.00	0.86	.60
Categories	16.5	6.67	16.94	6.17	0.07	.79
Sketch in figures	6	3.67	7.73	3.63	0.47	.60
ISI IQ	100	14.31	106.86	14.98	0.48	
GALO IQ	100	15.00	107.15	13.87	0.48	

Note. Means and SDs for the ISI standardization sample taken from Table 6.10 from the manual, and for the GALO standardization sample estimated from Table 1 from the manual. SDs for GALO estimated by taking the raw scores corresponding with the values 1.5 SD above and 1.5 SD below the average, respectively, and dividing the difference between the raw scores by three. This method is based on the principles used in the method of estimating the dollar value of job performance in costing human resources (see Schmidt et al., 1979). Since no data are reported for the SD of GALO IQ, the common value of 15.00 was used.
ISI *g* loadings based on standardization sample from 1966, computed from Table B1.2, and GALO *g* loadings based on standardization sample from 1954 (Kouwer, 1961), respectively.
1975–1981 data: mean based on weighted sample and SD based on unweighted sample.
Samples from 1975–1981: ISI: $N = 1227$; GALO: $N = 1140$.
Standardized gains on the ISI correlated .999 with standardized gains on the ISI after weighing for SES in Dutch population; standardized gains on the GALO correlated .992 with standardized gains on the GALO after weighing for SES in Dutch population.

SD (3.15 IQ points) for the GALO's five fluid/visual subtests, and 0.03 SD (0.45 IQ point) for the four verbal/scholastic subtests. Again, there are stronger gains on the fluid/visual tests.

3.2. Method of correlated vectors

For the ISI and the GALO the correlations between score gains and *g* loadings were −.23 ($p = .33$) and −.32 ($p = .20$), respectively. So, overall, there are only small or modest, non-significant negative correlations.

4. Discussion

Previous studies of the Flynn effect made extensive use of the Wechsler tests and the various versions of Raven's Progressive Matrices. This study has now also shown clear Flynn effects for the ISI and the GALO.

Fifteen studies on whether secular trends are related to *g* have been carried out, but have produced conflicting results. The two additional studies in the present paper yielded small and modest negative correlations between standardized gains and *g* loadings, and these studies employed large samples. A simple count of the findings from all seventeen independent studies yields six positive correlations and eleven negative correlations between *d* and *g*. When we only look at batteries with at least seven subtests a simple count yields three positive correlations and eight negative correlations. These findings suggests that the combined literature now shows a modest negative relationship between *d* (the secular change in test score) and *g*. However, a psychometric meta-analysis (Hunter & Schmidt, 1990) is required to determine the true correlation between *d* and *g* and to check the influence of both statistical artifacts and moderators.

Acknowledgements

We like to thank Theo Nijsse for permission to use his datasets and his enthusiasm in answering detailed questions.

Thanks to Roberto Colom, Conor Dolan, Jelte Wicherts, Raegan Murphy, Dimitri van der Linden, and Richard Lynn for feedback on a previous version of this paper. Thanks to Wim Hofstee, Arne Evers, and Mart-Jan de Jong for discussions.

References

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. Cambridge University Press.

Colom, R., Juan-Espinosa, M., & García, L. F. (2001). The secular increase in test scores is a "Jensen effect". *Personality and Individual Differences, 30*, 553–559.

Dutch Central Bureau of Statistics (1977). *Statistisch zakboek 1977* (Statistical pocket book 1977). The Netherlands, the Hague: Central Bureau of Statistics.

Dutch Central Bureau of Statistics (1980). *Statistisch zakboek 1980* (Statistical pocket book 1980). The Netherlands, the Hague: Central Bureau of Statistics.

Dutch Central Bureau of Statistics (1982). *Statistisch zakboek 1982* (Statistical pocket book 1982). The Netherlands, the Hague: Central Bureau of Statistics.

Dutch Central Bureau of Statistics (1985). *Arbeidskrachtentelling 1981. Deel 1: Methode; bevolking en beroepsbevolking; werkzame personen en werklozen* (Counting the labor force 1981. Part 1: Method; population and labor force; working persons and unemployed). The Netherlands, the Hague: Central Bureau of Statistics.

Dutch Institute of Psychologists (1974). *Documentatie van tests en testresearch in Nederland 1974* (Documentation of tests and test research in the Netherlands 1974). The Netherlands, Zaandijk: Nederlands Insituut van Psychologen together with Heijnis.

Flynn, J. R. (1999a). Evidence against Rushton: The genetic loading of WISC-R subtests and the causes of between-group IQ differences. *Personality and Individual Differences, 26*, 373–379.

Flynn, J. R. (1999b). Reply to Rushton. A gang of *g*s overpowers factor analysis. *Personality and Individual Differences, 26*, 391–393.

Flynn, J. R. (2000). IQ gains, WISC subtests and fluid *g*: *g* Theory and the relevance of Spearman's hypothesis to race. In G. R. B. J. Goode (Ed.), *The nature of intelligence* (pp. 202–227). New York: Wiley.

Flynn, J. R. (2006). Efeito Flynn: Repensando a inteligência e seus efeitos (The Flynn effect: Rethinking intelligence and what effects it). In C. F.-M. R. Colom (Ed.), *Introdução à Psicologia das Diferenças Individuais* (pp. 387–411). Brazil: Porto Alegre: ArtMed.

Hunter, J. E., & Schmidt, F. L. (1990). *Methods of meta-analysis: Correcting error and bias in research findings*. London: Sage.

Jensen, A. R. (1998). *The g factor: The science of mental ability*. London: Praeger.

Kema, G. N., & Kouwer, B. J. (1958). *Handleiding G.A.L.O.: Groninger Afsluitingsonderzoek Lager Onderwijs* (G.A.L.O. manual: Groningen Final Examination Primary Education). The Netherlands, Groningen: Groninger Instituut voor Toegepaste Psychologie en Psychotechniek.

Kouwer, B. J. (1961). De factorstructuur van de G.A.L.O. (The factor structure of the G.A.L.O.). *Paedagogische Studiën, 38*, 408–417.

Must, O., Must, A., & Raudik, V. (2003). The secular rise in IQs: In Estonia the Flynn effect is not a Jensen effect. *Intelligence, 167*, 1–11.

Rushton, J. P. (1999). Secular gains in IQ not related to the g factor and inbreeding depression – unlike Black-White differences: A reply to Flynn. *Personality and Individual Differences, 26*, 381–389.

Schmidt, F. L., Hunter, J. E., McKenzie, R. C., & Muldrow, T. W. (1979). Impact of valid selection procedures on work-force productivity. *Journal of Applied Psychology, 64*, 609–626.

Snijders, J. T., & Welten, V. J. (1968). *I.S.I.-publikaties: Interesse-Schoolvorderingen-Intelligentie. 5. De I.S.I-schoolvorderingen- en intelligentietests vorm I en II* ([I.S.I. publications: Interests-School achievement-Intelligence. 5. The I.S.I. school achievement and intelligence tests form I and II]). The Netherlands, Groningen: Wolters-Noordhoff.

te Nijenhuis, J., & van der Flier, H. (submitted). The Flynn effect: Gains in *g* and empty gains.

van Westerlaak, J. M., Kropman, J. A., & Collaris, J. W. M. (1975). *Beroepenklapper* (Occupation file). The Netherlands: University of Nijmegen, Institute of Applied Sociology.

Wicherts, J. W., Dolan, C. V., Oosterveld, P., van Baal, G. C. V., Boomsma, D. I., & Span, M. M. (2004). Are intelligence tests measurement invariant over time? Investigating the nature of the Flynn effect. *Intelligence, 32*, 509–537.