Personality and Individual Differences 55 (2013) 224–228

Contents lists available at SciVerse ScienceDirect

Personality and Individual Differences

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

The Flynn effect, group differences, and g loadings

Jan te Nijenhuis *

Work and Organizational Psychology, University of Amsterdam, Amsterdam, The Netherlands

ARTICLE INFO

Article history:
Available online 14 January 2012

Keywords:
Flynn effect
Secular score gains
g Score
g Loadings
IQ tests
Intelligence
Method of correlated vectors
Group differences

ABSTRACT

Flynn effect gains are predominantly driven by environmental factors. Might these factors also be responsible for group differences in intelligence? Group differences in intelligence have been clearly shown to strongly correlate with *g* loadings. The empirical studies on whether the pattern of Flynn effect gains is the same as the pattern of group differences yield conflicting findings. We present new evidence on the topic using a number of datasets from the US and the Netherlands. Score gains and *g* loadings showed a small negative average correlation. The general picture is now that there is a small, negative correlation between *g* loadings and Flynn effect gains. It appears that the Flynn effect and group differences have different causes.

It is always a great pleasure to meet Phil Rushton at conferences. Although he has been relentlessly attacked since the early 1980s, it has never lessened his passion for research or slowed him down. From early in the morning until late at night he shows a contagious enthusiasm for professional scientific discussions, showing an encyclopedic knowledge of the literature.

In his long and outstanding career, Phil Rushton has made important contributions to many scientific discussions. His research program is truly progressive (Lakatos & Musgrave, 1974): every couple of years he comes up with innovative ideas. This is in contrast to many other researchers who come up with only one good idea in their whole career. Science owes much to Phil Rushton.

1. Introduction

Secular gains in IQ test scores are among the most intriguing and controversial findings in psychology. Flynn (1984) was the first to show that average scores on intelligence tests are rising substantially and consistently, all over the world. Between 1930 and 1990 the gain on standard broad-spectrum IQ tests averaged three IQ points per decade. For verbal tests, or more precisely, tests with a content that most reflects the traditional classroom subject matter, the gain is 2 IQ points per decade, and for non-verbal (Fluid and Visual) tests 4 IQ points per decade. Gains on specific measures, such as the Raven's Progressive Matrices when used for the assessment of military recruits average about 7 IQ points per decade.

Group differences in mean IQ are more the rule than the exception (Lynn & Vanhanen, 2002). Large-scale research shows that, on average, the adult American Black population scores below the White population by about 1.2 standard deviations, equivalent to about 18 IQ points (Jensen, 1998). There is some discussion about the gap diminishing (see Dickens & Flynn 2006a,b; see Rushton & Jensen, 2005). Dutch-first-generation non-Western immigrants differences are about the same size as the US Black/White differences, but become substantially smaller for the second generation of non-Western immigrants (te Nijenhuis, de Jong, Evers, & van der Flier, 2004). Jensen (1998, pp. 380–383) has shown that *g* loadings correlate about .60 with Black/White IQ test score differences. te Nijenhuis and van der Flier (2003) showed that Dutch-non-Western immigrant differences are also strongly predicted by *g* loadings.

What, then, are the causes of these differences? Some argue there is a strong genetic component to group differences (Rushton & Jensen, 2005), whereas others argue group differences are wholly caused by the environment (Nisbett, 2009). The secular gains are massive and the time period too short for large genetic changes in the population, so therefore the changes must be largely environmental – although reduced inbreeding has been suggested to play a role (Mingroni, 2007; see also Woodley, 2011). It is an empirical question whether the strong environmental forces causing the scores over generations to rise are the same as the forces causing the group differences.

1.1. Rushton's contribution to the area

Rushton (1989) joined the fray on this topic showing that inbreeding depression scores from Japan predicted the magnitude

* Address: Gouden Leeuw 746, 1103 KR Amsterdam, The Netherlands. Tel.: +31 20 6002067.

E-mail address: JanteNijenhuis@planet.nl

0191-8869/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.paid.2011.12.023

of the Black/White differences on the same subtests in the US. Inbreeding depression is an established genetic phenomenon that occurs when people who are genetically related have children together, thereby producing in their offspring, on average, a lower score on IQ than would otherwise have been the case.

Subsequently, in an exchange with Flynn (1999), Rushton (1999) showed that secular gains from the US, Germany, Austria, and Scotland had modest to small negative correlations with *g* loadings. This is an important result given that *g* loadings correlate substantially with group differences, as shown by Rushton in a series of articles. In South Africa *g* loadings of items of the Raven Matrices predicted mean differences on the items between White, South Asian, and Black students (Rushton, Skuy, & Bons, 2004; Rushton, Skuy, & Fridjohn, 2002; Rushton, Skuy, & Fridjohn, 2003). In Zimbabwe *g* was a strong predictor of the score differences between African and White 12- to 14-year-olds on the WISC-R (Rushton & Jensen, 2003). In Serbia item *g* loadings from the Raven Matrices predicted mean differences between Roma and Whites (Rushton, Čvorovič, & Bons, 2007).

These findings increase in importance when one takes into consideration that most studies show that *g* loadings of tests correlate highly with their heritabilities (Jensen, 1987; Pedersen, Plomin, Nesselroade, & McClearn, 1992; Rijsdijk, Vernon, & Boomsma, 2002; Spitz, 1988). Moreover, Rushton, Bons, Vernon, and Čvorovič (2007) computed heritabilities for items from the Raven's Standard Progressive Matrices and showed that they could predict various group differences.

Rushton (1999) and Rushton and Jensen (2010) argues that if the gains over generations had been similar to the B/W differences they would have been in line with environmental explanations of group differences; when the pattern in secular score gains is not similar to the pattern in B/W differences, the latter is in line with a genetic component in group differences. However, it should not be forgotten that these findings are at best indirect evidence.

Rushton's (1999) finding has been challenged by Flynn (2000) and Nisbett (2009) who claim that there actually is a substantial positive correlation between secular score gains and *g* loadings. If this were a fact, it would indeed jeopardize Rushton's position; it would mean that *g* loadings correlate highly with both environmental and genetic effects, making 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, 1999, 2000; Must, Must, & Raudik, 2003; Wicherts et al., 2004) yielding conflicting findings.

Flynn (2007, 2010) states that even if there is a small negative correlation between secular score gains and *g* loadings it sheds no light on the race and IQ debate. Flynn (2010) accepts the empirical findings that Black/White score differences on subtests of IQ batteries rise as their *g* loadings, cognitive complexity, heritability, and inbreeding sensitivity rise. However, he argues that the fact that the performance gap is larger on more complex tasks than on easier tasks does not necessarily tell us something about genes versus environment. For instance, he hypothesizes that when one group has better genes for height and keen reflexes, but finds itself in a less rich basketball environment – less incentive, low-quality coaches, less play – the environmental disadvantage will expand the between-group performance gap as complexity rises, just as much as a genetic deficit would. The skill gap between challenged and unchallenged players is hypothesized to be more pronounced the more difficult the task. So, someone exposed to an inferior environment will hit a "complexity ceiling" and this ceiling does not differentiate whether the phenotypic gap is due to genes or environment. Elsewhere Flynn (2007; see also Dickens & Flynn, 2001) has argued that Blacks tend to be systematically underexposed to cognitive complexity throughout their life-courses. Flynn argues that the correlations reported by Rushton do not decide the causal question.

1.2. Research question

We used a number of datasets to see whether the method of correlated vectors yields a modest positive or negative correlation between score gains and *g* loadings.

2. Method

2.1. Test

The GATB (United States Department of Labor, 1970; van der Flier & Boomsma-Suerink, 1994) is a test of general intelligence with eight subtests: *Three-Dimensional Space* measures Visualization (g_v), *Vocabulary* measures Induction (g_{fl}) and Lexical Knowledge (g_{cr}), *Arithmetic Reason* measures Quantitative Reasoning (g_{fl}), *Computation* measures Numerical Ability (g_{cr}), *Tool Matching* measures Perceptual Speed (g_v), *Form Matching* measures Spatial Relations (g_v), *Name Comparison* measures Perceptual Speed (g_v) and Numerical Ability (g_{cr}), and *Mark Making* measures Aiming (General Psychomotor Speed). There are also two additional tests for finger dexterity: *Assemble* and *Disassemble*; as well as two more tests for manual dexterity: *Place* and *Turn*.

2.2. Samples

Sample 1: workers representative of the general working population from 1947: The first general working population norms for the GATB were based on 519 employed workers (US GATB manual, 1970). It was recognized that the sample probably was not truly representative of the general working population, but since it did include a wide range of occupational classifications, it was believed to yield a reasonably close approximation to test performance typical of the general working population. The date for the first sample is not explicitly given in the GATB manual, but close reading of the text suggests the year 1947. The mean age of the sample is 30.4 years (*SD* = 10.9 years) and the mean education is 11.0 years (*SD* = 2.4 years). This resulted in means and *SD*s for the GATB subtests, including the four tests of finger and manual dexterity.

Sample 2: workers representative for the general working population from 1952: In 1952, general working population norms were established on the basis of a selected sample of 4000 which was stratified to obtain proportional occupational representation of the general working population. The mean age of the sample is 30.4 years (*SD* = 9.9 years) and the mean education is 11.0 years (*SD* = 2.6 years). Means and *SD*s are reported for all 12 GATB subtests.

Samples 3 + 4: Dutch applicant bus drivers from 1975 to 1976 (sample 3), and 1983–1985 (sample 4): The Dutch GATB manual reports that for a specific time period a random sample (*N* = 110 for sample 3, and *N* = 1091 for sample 4) was taken from all persons who applied for positions of bus driver at regional bus companies and were tested at Dutch Railways selection centers.

Sample 5: Dutch applicant bus drivers from 1988 to 1992: For a specific time period, a random sample (*N* = 221) was taken from all persons who applied for positions of bus driver at regional bus companies and were tested at Dutch Railways selection centers (data are taken from te Nijenhuis, 1997).

Samples 6–8: Dutch 16-year-old students in higher general secondary education representative for the years 1975, 1985, and 2005, respectively: The Dutch GATB manual (van der Flier &

Boomsma-Suerink, 1994, pp. 148–153) and an update of the manual (Akkerman, 2011) give large representative samples for various school types from 1975, 1985, and 2005. The sample sizes for students from higher general secondary education were $N = 130$ for 1975, $N = 270$ for 1985, and $N = 498$ for 2005.

2.3. Comparison of samples

The mean scores of various cohorts were compared. (1) GATB scores from workers representative for the general working population from, respectively, 1947 and 1952 were compared. (2) Dutch GATB scores of three groups of applicants for the position of bus driver from 1975 to 1976, 1983 to 1985, and 1988 to 1992, respectively, were compared. (3) Dutch GATB scores of 16-year-old students in higher general secondary education in 1975, 1985, and 2005, respectively, were compared.

2.4. Statistical analyses

2.4.1. Gain scores

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.

2.4.2. g loadings

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.

2.4.3. 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. The method of correlated vectors was only employed on samples in which the vast majority of subtests showed score gains.

3. Results

3.1. Gains

Tables 1–3 show the gains on the GATB for various samples. The general picture in Tables 1 and 2 is an increase while Table 3 shows both an increase and a decrease. The aggregated scores show a much clearer pattern than do the scores on the individual tests. The subtests closest to traditional classroom subjects are Vocabulary, Arithmetic Reason, and Computation; and the tests of Broad Visual Perception are Three Dimensional Space, Tool Matching, and Form Matching. Name Comparison has both a visual and a scholastic component and therefore does not fit unambiguously into one of the two previous clusters. Mark Making and the four dexterity tests have low *g* loadings and low correlations with the other, more highly *g*-loaded tests. Table 4 shows the standardized gains per decade for the combination of visual tests and scholastic tests, respectively.

Table 1
Means and *SD*s on the General Aptitude Test Battery for the two general working population samples from 1947 ($N = 519$) to 1952 ($N = 4000$), Standardized score gains, and *g* loadings.

Test	1947		1952		*d*	*g*
	M	*SD*	*M*	*SD*		
Three-dimensional space	17.5	6.9	19.0	7.0	0.17	.64
Vocabulary	20.9	9.2	21.5	9.4	0.07	.68
Arithmetic reason	8.7	3.4	9.9	3.8	0.35	.69
Computation	26.6	7.6	28.1	8.1	0.20	.72
Tool matching	21.2	5.7	21.8	6.1	0.11	.70
Form matching	26.3	7.8	26.9	8.1	0.08	.72
Name comparison	71.3	20.1	70.7	22.1	−0.03	.79
Mark Making	71.0	9.7	69.5	10.3	−0.15	.67
Place	88.0	8.7	89.8	8.6	0.21	.40
Turn	101.0	8.7	100.8	9.6	−0.03	.41
Assemble	27.6	4.6	28.3	4.6	0.15	.40
Disassemble	28.8	3.7	29.5	3.7	0.19	.46

Note: *g* loadings were computed from the intercorrelations for the 1952 sample, as reported in the GATB manual (1970, p. 30); $N = 4000$.

It appears that the gains on the visual cluster have remained roughly the same between 1917 and 1989. Further, between 1917 and about 1960 the gains on the scholastic cluster were roughly comparable, and then the scores started to drop dramatically. The pattern in Table 3 is very clear: between 1975 and 1985 there is a gain in the scores on visual tests and a large drop in the scores on scholastic tests; the differences between 1975 and 2005 show the effects are in the same direction and that they become much stronger. Flynn (2001) describes a similar shift in the US data around 1948, but the Dutch data suggest a shift around 1960 for the Netherlands. Flynn hypothesizes that this is because in the middle of the 20th century schools shifted their emphasis from the traditional classroom subjects of reading, writing, and arithmetic to reasoning with de-contextualized problems.

Table 1 shows a decrease in Mark Making, but an average increase for the four dexterity tests. Table 2 shows both a decrease of −0.03 *SD* per decade and an increase of 0.07 *SD* per decade on Mark Making. In sum, there is a modest increase on these subtests.

3.2. Method of correlated vectors

We used the method of correlated vectors only on samples in which the vast majority of subtests showed score gains. Since the samples of students in higher general secondary education showed dramatic decreases on half the GATB subtests, they were excluded from the analyses (see above for explanation). For the general working population samples from 1947 to 1952 the score gains correlated $r = -.19$ with *g* loadings. For the applicant bus drivers from 1975 to 1976 and 1988 to 1992 the score gains correlated $r = .04$ with *g* loadings. For the Dutch applicant bus drivers from 1975 to 1976 and 1983 to 1985 the score gains correlated $r = .35$ with *g* loadings. For all three studies this results in an *N*-weighted $r = -.07$.

4. Discussion

Secular score gains are predominantly driven by environmental factors and might these factors also be responsible for group differences in intelligence? Is the pattern of secular score gains the same as the pattern of group differences? Group differences are strongly linked to *g* loadings, but the literature up to now yields conflicting findings concerning the link between *g* loadings and score gains, so additional studies are required. We used a number of datasets, which showed that the method of correlated vectors yields a small negative average correlation between score gains and *g* loadings. So, the general picture is now that of a small negative correlation between *g* loadings and gains. It appears that the Flynn effect and group differences have different correlations with *g* loadings, which suggests they have different causes.

Table 2
Means and *SD*s on the Dutch General Aptitude Test Battery for Applicant bus drivers from 1975 to 1976, 1983 to 1985, and 1988 to 1992, standardized score gains, and *g* loadings.

Test	1975–1976		1983–1985		1988–1992		d_1	d_2	*g*
	M	*SD*	*M*	*SD*	*M*	*SD*			
Three dimensional space	20.02	6.18	19.80	5.90	21.17	5.70	−0.04	0.19	.58
Vocabulary	20.01	6.33	20.32	6.77	24.50	7.18	0.05	0.71	.68
Arithmetic Reason	12.72	3.39	12.30	3.16	13.39	3.58	−0.12	0.20	.68
Computation	21.49	5.61	22.06	4.71	21.61	4.93	0.10	0.02	.67
Tool matching	27.47	5.90	28.13	5.55	28.42	5.69	0.11	0.16	.49
Form matching	25.16	5.71	26.68	5.85	28.45	5.98	0.27	0.58	.53
Name comparison	54.81	12.83	55.39	12.01	60.35	13.87	0.05	0.43	.62
Mark Making	67.14	7.66	66.91	8.52	67.68	9.75	−0.03	0.07	.14

Note: 1975–1976: $N = 110$; 1983–1985: $N = 1091$; 1988–1992: $N = 212$.
d_1: Gain from 1975–1976 to 1983–1985; d_2: gain from 1975–1976 to 1988–1992.
g loadings from de Wolff and Buiten (1963).

Table 3
Means and *SD*s on the Dutch General Aptitude Test Battery for 16-year-old students in higher general secondary education in 1975, 1985, and 2005, standardized score gains, and *g* loadings.

Test	1975		1985		2005		d_1	d_2	*g*
	M	*SD*	*M*	*SD*	*M*	*SD*			
Three dimensional space	21.55	5.32	21.76	5.46	24.03	5.49	0.04	0.47	.58
Vocabulary	29.72	5.15	27.70	4.94	23.65	4.84	−0.39	−1.18	.68
Arithmetic reason	14.92	3.05	14.29	3.27	11.42	3.02	−0.21	−1.15	.68
Computation	25.11	4.64	22.21	4.04	19.50	4.18	−0.63	−1.21	.67
Tool matching	33.56	5.04	34.44	5.39	36.14	5.67	0.18	0.51	.49
Form matching	30.28	5.79	30.90	6.51	32.37	6.54	0.11	0.36	.53
Name comparison	75.24	11.82	66.06	10.48	65.02	11.56	−0.78	−0.87	.62

Note. 1975: $N = 130$; 1985: $N = 270$; 2005: $N = 498$.
d_1: Gain from 1975 to 1985; d_2: gain from 1975 to 2005.
g loadings from de Wolff and Buiten (1963).

Table 4
Standardized gains per decade (in *SD*s) in GATB scores from Tables 1–3.

Born	Sample	Gain per decade	
		Visual	Scholastic
1917–1922	Working population	0.24	0.42
1945–1954	Applicant bus drivers	0.13	0.01
1945–1960	Applicant bus drivers	0.21	0.21
1959–1969	16-year-old students	0.11	−0.41
1959–1989	16-year-old students	0.15	−0.39

Note: Visual subtests: three dimensional space, tool matching, and form matching; scholastic subtests: vocabulary, arithmetic reason, and computation.
The average age for the applicant bus drivers from 1988 to 1992 was 29.91 years (data from te Nijenhuis (1997)) and we take this value as the average of the other two samples of applicant bus drivers.

As expected, we found overall gains on the large majority of measures. 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 General Aptitude Test Battery. Flynn (2006) states that there are few datasets on adults, and most of these are on military conscripts. We added data from samples of applicant bus drivers and workers representative for the general working population and found clear secular score gains.

Secular gains of about one *SD* have been shown for infants on tests for motor development, such as the Griffiths scale and the Bayley scales (see Lynn, 2009). We found Flynn effects for adults on perceptual motor tests, such as the four GATB tests which measure Finger and Manual dexterity as well as the GATB tests Mark Making, which measures Aiming. So, secular increases in perceptual motor ability and attention occur not only for toddlers and young children, but also for adults. The gains on motor tests seem to last into adulthood.

The method of correlated vectors yielded small to modest positive and negative correlations between score gains and *g* loadings in all cases where there were Flynn effects on the large majority of subtests, with an *N*-weighted $r = -.07$. The combined literature is now suggestive of a modest negative relationship between *g* and *d*.

5. Conclusion

The research literature overwhelmingly showed that group differences and *g* loadings are strongly correlated, but the literature on the Flynn effect and *g* loadings up to now showed conflicting findings. However, all three additional studies in the present study taken together show a small negative correlation between *g* loadings and secular gains. The strong positive correlation of Flynn (2000) and Nisbett (2009) appears to be an outlier. There are strong differences of opinion about the meaning of these findings, so more research is clearly needed.

5.1. Thoughts about the future of the area

The nomological net of group differences, secular score gains, and *g* loadings needs to be explored more fully, using the method of correlated vectors. For instance, Jensen (1998) showed that variables such as head size, brain volume, brain's gray matter, brain's evoked potential, brain glucose metabolic rate, peripheral nerve conduction velocity, brain pH, body symmetry, inbreeding depression, and hybrid vigor yield high positive correlations with *g* scores. Moreover, the gains resulting from test–retest and test training on IQ tests correlate perfectly negatively with *g* scores (te Nijenhuis, van Vianen, & van der Flier, 2007; te Nijenhuis, Voskuijl, & Schijve, 2001). For instance, various causes have been hypothesized for the Flynn effect, including schooling, nutrition, health care, heterosis, GDP, urbanization, smaller families, increased exposure to movies and optical displays, the dissemination of visual-spatial toys, increased environmental complexity, and teacher to student ratio. Usable data can be tested with the method of correlated vectors, to see whether they have a pattern that is more comparable to the pattern of secular score gains or to the pattern of group differences.

The method of correlated vectors is not a strong statistic; it's simply the correlation between a small number of observations. However, it is a golden combination with one of the strongest methods available in science, namely psychometric meta-analysis (Hunter & Schmidt, 2004), which estimates what the results of studies would have been if all studies had been conducted without methodological limitations or flaws. What is badly needed is a psychometric meta-analysis of all the studies correlating *g* loadings and secular score gains. The hypothesis that the large variety in effect sizes can be fully explained by statistical artifacts can be tested in a psychometric meta-analysis.

References

Akkerman, A. E. (2011). *GATB handleiding [GATB manual]*. The Netherlands, Utrecht: Stichting GATB-Research.

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.

de Wolff, C. J., & Buiten, B. (1963). Een factoranalyse van vier testbatterijen [A factor analysis of four test batteries]. *Nederlands Tijdschrift voor de Psychologie, 18*, 220–239.

Dickens, W. T., & Flynn, J. R. (2001). Heritability estimates versus large environmental effects: The IQ paradox resolved. *Psychological Review, 198*, 346–369.

Dickens, W. T., & Flynn, J. R. (2006a). Black Americans reduce the racial IQ gap: Evidence from standardization samples. *Psychological Science, 17*, 913–920.

Dickens, W. T., & Flynn, J. R. (2006b). Common ground and difference. *Psychological Science, 17*, 923–924.

Flynn, J. R. (1984). The mean of IQ of Americans: Massive gains 1932–1978. *Psychological Bulletin, 95*, 29–51.

Flynn, J. R. (1999). 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. (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. (2001). *The history of the American mind in the 20th century*. Sydney, Australia: Paper presented at Spearman Seminar.

Flynn, J. R. (2006). Efeito Flynn: Repensando a inteligência e seus efeitos [the Flynn effect: Rethinking intelligence and what effects it]. In C. Flores-Mendoza & R. Colom (Eds.), *Introdução à Psicologia das Diferenças Individuais* (pp. 387–411). Brazil, Porto Alegre: ArtMed.

Flynn, J. R. (2007). *What is intelligence?* Cambridge University Press.

Flynn, J. R. (2010). The spectacles through which I see the race and IQ debate. *Intelligence, 38*, 363–366.

Hunter, J. E., & Schmidt, F. L. (2004). *Methods of meta-analysis: Correcting error and bias in research findings* (second ed.). London: Sage.

Jensen, A. R. (1987). Individual differences in mental ability. In J. A. Glover & R. R. Ronning (Eds.), *Historical foundations of educational psychology* (pp. 61–88). New York: Plenum.

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

Lakatos, I., & Musgrave, A. (1974). *Criticism and the growth of knowledge*. Cambridge University Press.

Lynn, R. (2009). What has caused the Flynn effect?: Secular increases in the development quotients of infants. *Intelligence, 37*, 16–24.

Lynn, R., & Vanhanen, T. (2002). *IQ and the wealth of nations*. London: Praeger.

Mingroni, M. A. (2007). Resolving the IQ paradox: Heterosis as a cause of the Flynn effect and other trends. *Psychological Review, 114*, 806–829.

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.

Nisbett, R. E. (2009). *Intelligence and how to get it*. New York: Norton.

Pedersen, N. L., Plomin, R., Nesselroade, J. R., & McClearn, G. E. (1992). A quantitative genetic analysis of cognitive abilities during the second half of the life span. *Psychological Science, 3*, 346–353.

Rijsdijk, F. V., Vernon, P. A., & Boomsma, D. I. (2002). Application of hierarchical genetic models to Raven and WAIS subtests: A Dutch twin study. *Behavior Genetics, 32*, 199–210.

Rushton, J. P. (1989). Japanese inbreeding depression scores: Predictors of cognitive differences between blacks and whites. *Intelligence, 13*, 43–51.

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.

Rushton, J. P., Bons, T. A., Vernon, P. A., & Čvorovič, J. (2007). Genetic and environmental contributions to population group differences on the Raven's Progressive Matrices estimated from twins reared together and apart. *Proceedings of the Royal Society of London. Series B: Biological Sciences, 274*, 1773–1777.

Rushton, J. P., Čvorovič, J., & Bons, T. A. (2007). General mental ability in South Asians: Data from three Roma (Gypsy) communities in Serbia. *Intelligence, 35*, 1–12.

Rushton, J. P., & Jensen, A. R. (2003). African-White IQ differences from Zimbabwe on the Wechsler Intelligence Scale for Children-Revised are mainly on the *g* factor. *Personality and Individual Differences, 2003*, 177–183.

Rushton, J. P., & Jensen, A. R. (2005). Thirty years of research on race differences in cognitive ability. *Psychology, Public Policy, and Law, 11*, 235–294.

Rushton, J. P., & Jensen, A. R. (2010). The rise and fall of the Flynn effect as a reason to expect a narrowing of the Black–White IQ gap. *Intelligence, 38*, 213–219.

Rushton, J. P., Skuy, M., & Bons, T. A. (2004). Construct validity of Raven's advanced progressive matrices for African and non-African engineering students in South Africa. *International Journal of Selection and Assessment, 12*, 220–229.

Rushton, J. P., Skuy, M., & Fridjohn, P. (2002). Jensen Effects among African, Indian, and White engineering students in South Africa on Raven's standard progressive matrices. *Intelligence, 30*, 409–423.

Rushton, J. P., Skuy, M., & Fridjohn, P. (2003). Performance on Raven's Advanced Progressive Matrices by African, East Indian, and White engineering students in South Africa. *Intelligence, 31*, 123–137.

Spitz, H. H. (1988). Wechsler subtest patterns of mentally retarded groups: Relationship to *g* and to estimates of heritability. *Intelligence, 12*, 279–297.

te Nijenhuis, J. (1997). *Comparability of test scores for immigrants and majority group members in the Netherlands* Unpublished doctoral dissertation. Amsterdam: Vrije Universiteit.

te Nijenhuis, J., de Jong, M.-J., Evers, A., & van der Flier, H. (2004). Are cognitive differences between immigrant and majority groups diminishing? *European Journal of Personality, 405*, 434.

te Nijenhuis, J., & van der Flier, H. (2003). Immigrant-majority group differences in cognitive performance: Jensen effects, cultural effects, or both? *Intelligence, 31*, 443–459.

te Nijenhuis, J., van Vianen, A. E. M., & van der Flier, H. (2007). Score gains on *g*-loaded tests: No *g*. *Intelligence, 35*, 283–300.

te Nijenhuis, J., Voskuijl, O. F., & Schijve, N. B. (2001). Practice and coaching on IQ tests: Quite a lot of *g*. *International Journal of Selection and Assessment, 9*, 302–308.

United States Department of Labor (1970). *Manual for the USTES General Aptitude Test Battery*. Washington, DC: United States Department of Labor, Section III. Development.

van der Flier, H., & Boomsma-Suerink, J. L. (1994). *Handbook GATB [GATB manual]*. The Netherlands, Amsterdam: Stichting GATB-Research.

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.

Woodley, M. A. (2011). Heterosis doesn't cause the Flynn effect: A critical examination of Mingroni (2007). *Psychological Review, 118*, 689–693.