PERGAMON

Personality and Individual Differences 30 (2001) 553–559

PERSONALITY AND INDIVIDUAL DIFFERENCES

www.elsevier.com/locate/paid

The secular increase in test scores is a "Jensen effect"

Roberto Colom *, Manuel Juan-Espinosa, Luís F. García

Facultad de Psicología, Universidad Autónoma de Madrid (UAM), 28049, Madrid, Spain

Received 19 November 1999; received in revised form 1 February 2000; accepted 1 March 2000

Abstract

The "Jensen effect" results from the correlation between the *g* factor and a huge number of psychological and biological variables. Although Rushton (Rushton, J. P. (1999). Secular gains in IQ not related to the *g* factor and inbreding depression — unlike Black–White differences: a reply to Flynn. *Personality and Individual Differences, 26*, 381–389) proposed that the secular increase in test scores is not a "Jensen effect", the present study demonstrates that this is true for *crystallized* tests but not for *fluid* tests. A fluid *g* factor is correlated with the generational changes observed in two successive Spanish standardizations of the DAT battery. Contrary to Rushton's (1999) findings — based on a crystallised *g* — there is a positive correlation between a fluid *g* and the generational cognitive change. There is one strong implication of the generational cognitive difference observed in the present study for the comparison of contemporary human populations: an environmental explanation of the current cognitive gap between some populations need only posit that the *current* average environment for population A (with a lower average score) matches the quality of the average environment for population B (with a higher average score) a generation ago.

Keywords: Flynn effect; Jensen effect; Generational IQ changes; Crystallized intelligence; Fluid intelligence; Differential aptitude test (DAT); Population differences

1. Introduction

The "Jensen effect" results after the finding of significant correlations occurring between the *g* factor and other variables. Jensen (1998) showed that the *g* vector is responsible for the observed correlation between a variety of cognitive tests and variables such as scholastic achievement, work-place performance, brain size, brain glucose metabolic rate, average evoked potential, reaction time, and so forth. The Jensen effect can be seen whenever there is a significant correlation

* Corresponding author. Tel.: +34-91-397-4114; fax: +34-91-397-5215.
E-mail address: roberto.colom@uam.es (R. Colom).

0191-8869/01/$ - see front matter
PII: S0191-8869(00)00054-4

between the vector of tests' *g* loadings and the vector of the same tests' loadings on variable X. Rushton (1998) proposed that when a significant correlation occurs between those vectors, the result be called a "Jensen effect".

It is recognized that the "Jensen effects" are not omnipresent. Thus, for instance, Rushton (1999) suggested that the secular increase in test scores is not a "Jensen effect":

> "these results clearly imply that environmental effects on IQ affect the non-*g* component" (p. 387).

Therefore, the gains in IQ over time are unrelated to *g*:

> "the secular gains are little, if at all related to *g*. Perhaps, then, they are mainly due to gains in the subtest's group factors and test specificity" (Rushton, 1999, p. 388).

However, Flynn (1999b) has pointed out that the zero correlation reported by Rushton (1999) could be a product of two related things:

1. the crystallized nature of the WISC battery; and
2. the fact that a crystallized *g* is biased against IQ gains.

Flynn (1999b) wrote:

> "if you have a battery with a crystallized bias, it will give higher *g* loadings to verbal than performance subtests. Therefore, you will get a negative correlation between IQ gains and *g*-loadings on its subtests. The WISC battery has a crystallized bias" (Flynn, 1999b, p. 392).

Reanalyzing Rushton's (1999) data, Flynn (1999b) ranked the WISC subtests in terms of their correlations with Raven's *g*, thus injecting a fluid bias. The result is that the correlation between *g* loadings and IQ gains turn out to be positive. However, it is recognized by Flynn (1999b) that the new values are not pure (fluid) *g* loadings. Therefore, Flynn (1999a) has invited others to add datasets germane to this debate (see Flynn, 1999a, p. 374).

The purpose of the present study is twofold:

1. to satisfy Flynn's invitation; and
2. to contribute to the demonstration that generational IQ changes challenge some conclusions extracted after the comparison of contemporary human populations.[1]

2. Method

2.1. Participants

The participants were 4177 high school students (2611 males and 1566 females) with a mean age of 18 years, considered in the 1979 and the 1995 Spanish standardizations of the Differential

[1] We are grateful to James Flynn for reviewing a draft of the present paper.

Aptitude Test (DAT — see Colom, Andrés-Pueyo & Juan-Espinosa, 1998; Colom, Quiroga & Juan-Espinosa, 1999).

2.2. Measures

The Spanish translation of the DAT is a battery of reliable group-administered paper-and-pencil tests. The DAT subscales measure Verbal Reasoning (VR), Space Relations (SR), Numerical Ability (NA), Mechanical Reasoning (MR), and Abstract Reasoning (AR). Each DAT ability is measured by a single type of item, and item types have remained constant over the Spanish standardizations of the DAT conducted in 1979 and 1995. Because the DAT standardizations span a 16-year-period and the means and standard deviations are obtained by sex, the DAT's norms constitute an ideal data base for the assessment of generational differences in fluid intelligence (Lynn, 1994).

2.3. Analyses

Colom et al. (1998) analyzed the 1979 and 1995 Spanish standardizations of the DAT, computing the generational cognitive change in *d* units. This produced a column vector representing the generational change for every DAT subtest. The *d* vectors corresponding to the male and the female subsamples considered in the DAT standardizations are shown in Table 1.

g was extracted from the correlation matrix presented in Table 2. The DAT intersubtests' correlations are presented separately for males and females.

The *g* factor was represented by the first principal unrotated factor. Jensen and Weng (1994) demonstrated that it seldom makes any difference whether *g* is represented by the highest order factor in a hierarchical factor analysis or by the first unrotated principal factor in a principal factor analysis. Table 3 shows the *g* vector separately for males and females.[2]

The vector of *g* loadings was correlated with the vector of generational changes in the test scores. We analyzed the data separately for males and females. The Pearson *r* and Spearman's rank-order correlations, r_s, are suitable measures of the degree of relationship between the two vectors. The test of significance of r_s is a stringent statistical test, because the *n* (number of tests) is typically small (Jensen, 1998).

Table 1
Generational changes (*d*) for the male and female samples

DAT subjects	(*d*) Males	(*d*) Females
Verbal Reasoning (VR)	−0.020	−0.020
Numerical Ability (NA)	−0.160	−0.140
Abstract Reasoning (AR)	0.090	0.330
Spatial Relations (SR)	0.100	0.180
Mechanical Reasoning (MR)	−0.020	0.100

[2] The congruence coefficient between the *g* vectors corresponding to the male and to the female samples was +0.995. Therefore, the *g* factor is the same irrespective of sex.

Table 2
Intersubtests' correlations. Above the diagonal are the correlations for the male sample, and below the diagonal (shown in italic) are the correlations for the female sample. Subtests' reliabilities are shown at the diagonal

Correlations	VR	NA	AR	SR	MR
Verbal Reasoning (VR)	0.89	0.5	0.47	0.45	0.32
Numerical Ability (NA)	*0.32*	0.9	0.48	0.33	0.26
Abstract Reasoning (AR)	*0.47*	*0.46*	0.73	0.55	0.41
Spatial Relations (SR)	*0.42*	*0.3*	*0.54*	0.94	0.58
Mechanical Reasoning (MR)	*0.38*	*0.29*	*0.43*	*0.59*	0.82

Using five tests in a correlated vector analysis is relatively weak and risks a Type II error, because the statistical test is based on the number of variables, which is small here. To surpass this problem, the vectors for males and females were combined to yield a statistical test based on g and d with an $n = 10$. This gives the null hypothesis (e.g. absence of correlation) a fair chance of rejection. Note that there is no trouble with this combination, since there are no sex differences in g (see Aluja, Colom, Abad & Juan-Espinosa, 2000; Colom, Juan-Espinosa, Abad & García, 2000; Jensen, 1998).

Finally, test unreliability is considered, because a correlation between the g vector and the d vector could be entirely an artifact of differences in the subtests' reliabilities (Jensen, 1998). Thus, the partial correlation controlling for subtests' reliabilities was also computed. Subtests' reliabilities — Split-Half Method — are presented at the diagonal of the Table 2.

3. Results

The Spearman rank-order correlations between g and d are $+0.821$ ($p = 0.089$) and $+0.9$ ($p = 0.037$), Pearson correlations are $+0.634$ ($p = 0.251$) and $+0.925$ ($p = 0.024$), and the partial correlations controlling for r_{xx} are $+0.794$ ($p = 0.206$) and $+0.9891$ ($p = 0.011$), for the male and the female sample, respectively.

Combining the male and the female g and d vectors, the results are as follows: Pearson $r = +0.707$ ($p = 0.022$), Rho $= +0.695$ ($p = 0.026$), and the partial correlation controlling for $r_{xx} = +0.7811$ ($p = 0.013$). Therefore, the correlation is strongly positive.

Table 3
g loadings for the male and female samples. The g factor is represented by the first principal unrotated factor

DAT subsets	(g) Males	(g) Females
Verbal Reasoning (VR)	0.664	0.620
Numerical Ability (NA)	0.649	0.517
Abstract Reasoning (AR)	0.702	0.730
Spatial Relations (SR)	0.786	0.743
Mechanical Reasoning (MR)	0.586	0.643
% Variance	46.4	43

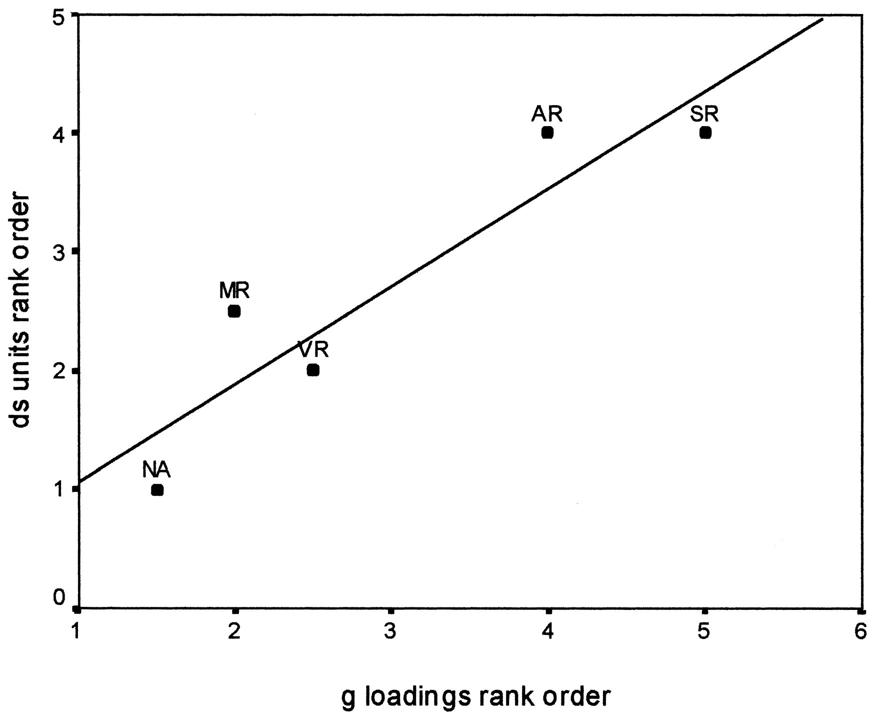

Fig. 1. DAT subtests ordered according to their *g* loadings and the generational changes (*d*). NA = Numerical Ability, VR = Verbal Reasoning, MR = Mechanical Reasoning, AR = Abstract Reasoning, SR = Spatial Relations.

Overall, the pattern of correlations between the subtests' *g* loadings and the subtests' generational changes (*d*), indicate that the changes in fluid test scores are *g* changes. Therefore, the secular change in fluid test scores is a "Jensen effect": the higher the *g* loading, the larger the generational change observed (Fig. 1).

4. Discussion

The results in the present study are indicative of the fact that secular changes in fluid test scores are within the domain of the so-called "Jensen effects" (Rushton, 1998). Clearly, there is a positive correlation between a fluid *g* and the secular changes in several cognitive tests. More specifically, a fluid *g* is positively correlated with the secular change in the observed test scores at the successive Spanish standardizations of the DAT battery.

Given that the cognitive tests analyzed in the present study are a fine-grained estimate of a fluid *g* (contrary to the Wechsler scales analyzed by Rushton, 1999) we must conclude that the association between *g* and the secular changes in test scores is highly positive for fluid cognitive batteries, but zero for crystallized cognitive batteries.

A theoretical derivation of the results reported in the present study can be developed as follows:

1. given that IQ differences between generations are clearly environmental in origin;
2. given that there is a positive correlation between fluid *g* and secular IQ changes — and, ipso facto, IQ gains are *g* changes, then;
3. *g* changes can be environmentally caused;
4. *Mutatis mutandis*, there is nothing conclusive at the average phenotypic difference in standardized IQ tests between contemporary human populations directing to a genetic explanation.

The results in the present study, as well as its main conclusion, are in line with those of Flynn (1999 a,c): IQ gains are *g* gains, which is a demonstration of the plausible environmental nature of the average phenotypic IQ difference between some contemporary human populations.

Harris's (1995) Group Socialization Theory of Development (GS Theory) has offered an explanation of the nature of this environmental causation. Note that GS Theory has received some empirical support:

"a number of the results were consistent with Harris's theory" (Loehlin, 1997, p. 1197).

According to the GS Theory, intra- and inter-group processes are responsible for the transmission of culture and for environmental modification of children's characteristics. On this ground, separate peer group cultures for populations A and B could change some of their psychological traits (intelligence, personality, and so forth) in the direction pointed out by their reference groups. Moreover, this kind of pressure can be as homogeneous as could be expected from a plausible "X Factor".

In summary, there are enough grounds for the assumption based on the idea of separate socialization cultures for contemporary populations. This agrees with Flynn's (1999c) statement that an environmental explanation of an IQ gap between contemporary human populations need only posit that the average environment for population A (with a contemporary lower average IQ score) in, say, 1995, matches the quality of the average environment for population B (with a contemporary higher average IQ score) in, say, 1945.

References

Aluja, A., Colom, R., Abad, F. J., & Juan-Espinosa, M. (2000). Sex differences in general intelligence defined as *g* among young adolescents. *Personality and Individual Differences, 28*(4), 813–820

Colom, R., Andrés, Pueyo A., & Juan-Espinosa, M. (1998). Generational IQ gains: Spanish data. *Personality and Individual Differences, 25*(5), 927–935.

Colom, R., Juan-Espinosa, M., Abad, F. J., & García, L. F. (2000). Negligible sex differences in general intelligence. *Intelligence, 28*(1), 57–68

Colom, R., Quiroga, MªA, & Juan-Espinosa, M. (1999). Are cognitive sex differences disappearing Evidence from Spanish populations? *Personality and Individual Differences, 27*(6), 1189–1195.

Flynn, J. (1999a). Searching for justice. *American Psychologist, 54*(1), 5–20.

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

Harris, J. R. (1995). Where is the child's environment A group socialization theory of development? *Psychological Review, 102*(3), 458–489.

Jensen, A. R. (1998). *The* g *factor*. New York: Praeger.

Jensen, A., & Weng, L. (1994). What is a good *g*? *Intelligence, 18*, 231–258.

Loehlin, J. (1997). A test of J. R. Harris's theory of peer influences on personality. *Journal of Personality and Social Psychology, 72*(5), 1197–1201.

Lynn, R. (1994). Sex differences in intelligence and brain size: a paradox resolved. *Personality and Individual Differences, 17*(2), 257–271.

Rushton, J. P. (1998). The "Jensen effect" and the "Spearman–Jensen Hypothesis" of Black–White IQ Differences. *Intelligence, 26*(3), 217–225.

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.