Available online at www.sciencedirect.com
ScienceDirect
Intelligence 35 (2007) 451 – 456
INTELLIGENCE

Brother–sister differences in the *g* factor in intelligence: Analysis of full, opposite-sex siblings from the NLSY1979

Ian J. Deary [a,*], Paul Irwing [b], Geoff Der [c], Timothy C. Bates [a]

[a] Department of Psychology, University of Edinburgh, 7 George Square, Edinburgh EH8 9JZ, UK
[b] Manchester Business School, University of Manchester, UK
[c] MRC Social and Public Health Sciences Unit, University of Glasgow, UK

Received 23 August 2006; received in revised form 14 September 2006; accepted 14 September 2006
Available online 24 October 2006

Abstract

There is scientific and popular dispute about whether there are sex differences in cognitive abilities and whether they are relevant to the proportions of men and women who attain high-level achievements, such as Nobel Prizes. A recent meta-analysis (Lynn, R., and Irwing, P. (2004). Sex differences on the progressive matrices: a meta-analysis. Intelligence, 32, 481–498.), which suggested that males have higher mean scores on the general factor in intelligence (*g*), proved especially contentious. Here we use a novel design, comparing 1292 pairs of opposite-sex siblings who participated in the US National Longitudinal Survey of Youth 1979 (NLSY1979). The mental test applied was the Armed Services Vocational Aptitude Battery (ASVAB), from which the briefer Armed Forces Qualification Test (AFQT) scores can also be derived. Males have only a marginal advantage in mean levels of *g* (less than 7% of a standard deviation) from the ASVAB and AFQT, but substantially greater variance. Among the top 2% AFQT scores, there were almost twice as many males as females. These differences could provide a partial basis for sex differences in intellectual eminence.

Keywords: Intelligence; *g*; Sex; Gender

1. Introduction

Whereas a large body of research supports the existence of sex differences in specific cognitive abilities, some favouring men, and some favouring women (Hyde, 2005; Kimura, 2002), the existence of an overall difference in general cognitive ability has recently been strongly debated (Blinkhorn, 2005, 2006; Irwing & Lynn, 2006, Lawrence, 2006; Muller et al., 2005, Spelke, 2005). One particular aspect of the debate receiving attention is the relation of any sex differences to the proportion of men and women apt to attain the highest level achievements, such as Nobel Prizes.

There is certainly an effect to be explained: for example, in terms of indices of scientific achievement, men were awarded 545 out of the 557 Nobel prizes awarded for science. However, there is also scope for widely divergent explanations of this effect. For example, Lawrence Summers, then President of Harvard University, was reported as suggesting that a male advantage in high school tests of maths and science supported a genetic explanation for the paucity of female success in these fields (Muller et al., 2005). In reply, Muller et al. (2005) contended that current evidence clearly supports an explanation in terms of

* Corresponding author.
E-mail address: i.deary@ed.ac.uk (I.J. Deary).

0160-2896/$ - see front matter © 2006 Elsevier Inc. All rights reserved.
doi:10.1016/j.intell.2006.09.003

socialisation. They also suggested that scores in the top of the range on standardised tests may not even be related to subsequent success in science careers.

Here, before approaching the topic of sex differences in intelligence with a novel research design, we recap on some of the principal difficulties that beset attempts to answer the question. First, there is the dependence on age. Based on research primarily on infants, Spelke concluded that there is, "no evidence for sex differences in overall aptitude for mathematics or science at any point in development" (Spelke, 2005, p. 950). Supporting this at the later age of 11 years, a large study of almost all children in Scotland born in 1921 and attending school on June 1st 1932 showed that mean levels of *g* did not differ (Deary et al., 2003). A similar result was found in a more recent, massive study of UK children of the same age (Strand, Deary, & Smith, 2006). However, Lynn (1994) has suggested that, like other sexually dimorphic traits, sex differences in cognition should show female advantage pre-puberty, with males gaining from about age 15 into adulthood, so that data from after puberty would be needed.

Another important methodological issue concerns the choice of intelligence measure. Arguably, the strongest evidence for a sex difference in *g* consists of the meta-analysis of the Raven's Matrices (Lynn & Irwing, 2004, 2005) but, although the Progressive Matrices is typically highly *g*-loaded, its item content is heterogeneous (Lynn, Allick, & Irwing, 2004; Mackintosh & Bennett, 2005; Van der Ven & Ellis, 2000). There is evidence suggesting that this item-specific variance cancels out to produce a pure or near pure measure of *g* (Carroll, 1993; Snow, Kyllonen, & Marshalek, 1984), but there is as yet no definitive test of this. Some have suggested, but not found definitive evidence for, the possibility that any male advantage on Raven's Matrices could be caused by items' visuo-spatial nature (Abad, Colom, Rebollo, & Escorial, 2004). In short, a measure is needed that is appropriately and fairly constructed and which unambiguously measures *g* in adults (see Jensen & Weng, 1994). In fact, there are comparatively few studies of adult sex differences in *g* and those are questionable (Dolan et al., 2006; Van der Sluis, et al., 2006). The principal difficulty is that of ensuring the samples of men and women are directly comparable, and do not differ in terms of, for example, genetic or socio-economic factors. A high proportion of the samples reviewed by Lynn and Irwing (2004) are standardisation samples and, amongst adults, heterogeneity tests establish that point estimates are largely identical across samples. However, standardisation samples employ exclusionary criteria, either by accident or design, opening the question of selection bias (Blinkhorn, 2005). That is, the female samples might be less selective, thus accounting for the observed sex difference in *g*.

One way to control for these important background factors and to maximise comparability is to compare opposite-sex, full siblings. This implicitly and fully controls for any factor, like SES, which differs between families but is common to both siblings. Here we apply this novel approach to the study of sex differences in intelligence, using a large sample of opposite-sex, full siblings; a large, well-validated mental test battery; and computation of a general factor using structural equation modelling.

2. Methods

2.1. Participants

We analysed the data available in the US National Longitudinal Survey of Youth (NLSY1979; http://www.bls.gov/nls/nlsy79.htm; accessed 13 September 2006). The subjects for the full USA survey were 12,686 young people aged between 14 and 22 years on January 1st, 1979. The survey comprised three subsamples: the cross-sectional sample ($n=6111$) designed to be representative of non-institutionalised civilian young people; the supplementary sample ($n=5295$) designed to oversample Hispanic, black and economically disadvantaged youth; and a military sample ($n=1280$) representing those serving in the military. We selected opposite-sex, full siblings from the civilian samples. Because it was sampled differently, there were no sibling pairs in the military sample. Where sibships exceeded two, we chose the oldest two opposite-sex siblings. There were 1292 sibling pairs for analysis, well matched for age: male mean (SD) = 18.43 (2.07) years; female = 18.38 (2.08). The proportion of pairs derived from the cross-sectional sample ($n=722$, 56%) closely matched that in the survey as a whole (54%).

2.2. Mental test battery

Participants were administered the Armed Services Vocational Aptitude Battery (ASVAB) which has 10 subtests: science, arithmetic, word knowledge, paragraph comprehension, numerical operations, coding speed, auto and shop information, mathematics knowledge, mechanical comprehension, and electronics information. Four of the subtests comprise the Armed Forces Qualification Test (AFQT) which includes only the more general, less vocationally-specific tests: arithmetic, word knowledge, paragraph comprehension, and mathematics knowledge. The concurrent validity of the ASVAB general cognitive (*g*) factor is supported by almost perfect (>0.99) correlation with the *g* factor from Kyllonen's (1993) large, and very different–with respect to form, content, and method of administration–mental test battery (Stauffer, Ree, & Carretta, 1996). Whereas the ASVAB has 10 subtests which are standard paper-and-pencil tests, Kyllonen's (1993) Cognitive Abilities Assessment battery has 25 computer-based subtests, derived from information processing (cognitive components) principles. We record this observation as evidence that individual differences in the *g* factor from the ASVAB are not peculiar to the contents of the ASVAB's subtests.

2.2.1. Computing AFQT-standardised scores

An AFQT percentile score was available, already computed, from the NLSY database. We also computed a standardised score for the AFQT. These new AFQT-standardised scores were computed from the raw ASVAB subtest scores (Table 1). The procedure is: a 'Verbal' score computed as ASVAB3 + ASVAB4; Verbal, Mathematics knowledge, and Arithmetic reasoning are converted to *z* scores (mean = 0; SD = 1); multiply the Verbal standard score by two; sum the standard scores for Verbal, Mathematics knowledge, and Arithmetic reasoning.

2.2.2. Estimates of *g* from the ASVAB and AFQT

In order to extract estimates of *g* from the ASVAB and AFQT, we applied robust maximum likelihood estimation to a series of confirmatory factor models using LISREL 8.72 software. For the ASVAB, it was first necessary to establish the number of factors. An exploratory factor analysis using PRELIS tested the fit of models with 1 to 5 factors, with the Root Mean Square Error of Approximation, supporting a four-factor model as providing the best fit. This conforms with a number of previous analyses which have reported a four-factor solution to the ASVAB (Ree & Carretta, 1995). We tentatively concluded that a hierarchical factor model with one general factor and four group factors (Fig. 1), as indicated by the exploratory factor analysis, best describes the data. This model provided a good fit to the data with a Non-Normed Fit Index (NNFI) of .99 and a Standardised Root Mean Square Residual (SRMR) of .039, meeting the criteria as specified for the two-index presentation strategy of Hu and

Table 1
Sex differences in means and standard deviations for Armed Services Vocational Aptitude Battery (ASVAB) subtests and total scores and for Armed Forces Qualification Test (AFQT) scores

Test name	Meaning	Male mean	Female mean	*p* for mean difference	Cohen's *d* for sex difference in means	Male SD	Female SD	*p* for equality of variances	Male–female standard deviation ratio
Armed services vocational aptitude battery									
ASVAB1	Science	14.3	13.2	<.0001	0.21	5.7	4.8	<.0001	1.19
ASVAB2	Arithmetic	15.9	14.7	<.0001	0.17	7.5	6.7	<.0001	1.12
ASVAB3	Word knowledge	22.3	22.9	.005	−0.07	9.0	8.2	.0010	1.10
ASVAB4	Paragraph comprehension	9.2	10.0	<.0001	−0.21	3.9	3.6	.0027	1.08
ASVAB5	Numerical operations	29.3	32.4	<.0001	−0.27	11.5	11.5	.99	1.00
ASVAB6	Coding speed	36.8	44.1	<.0001	−0.48	15.8	16.8	.028	0.94
ASVAB7	Auto and shop information	14.1	9.7	<.0001	0.89	5.9	3.7	<.0001	1.59
ASVAB8	Mathematics knowledge	11.9	11.9	.90	0.00	6.5	6.1	.013	1.07
ASVAB9	Mechanical comprehension	13.6	10.7	<.0001	0.58	5.8	4.1	<.0001	1.41
ASVAB10	Electronics information	10.8	8.5	<.0001	0.56	4.7	3.4	<.0001	1.38
ASVAB-g[c]	Score extracted from higher order confirmatory factor analysis using robust maximum likelihood	8.66	8.45	.003	.068	3.2	2.8	<.001	1.16
Armed forces qualification test									
AFQT89[a]	Revised AFQT percentile score	38.7	38.2	.53	0.02	30.1	27.7	.0025	1.09
AFQT-standardised[b]	Sum of standardised scores for the 4 AFQT subtests	−0.034	0.034	.44	−0.02	3.9	3.5	.0003	1.11
AFQT-*g*[c]	Scores from confirmatory factor analysis of the 4 AFQT subtests	15.08	14.66	.009	.064	6.7	6.0	<.001	1.11

[a] These scores were obtained directly from the NLSY79 database.
[b] These scores were computed from the raw ASVAB subtest scores (see Methods).
[c] To extract estimates of *g* from the ASVAB and AFQT, we applied robust maximum likelihood estimation to a series of confirmatory factor models using LISREL 8.72 software (see Methods).

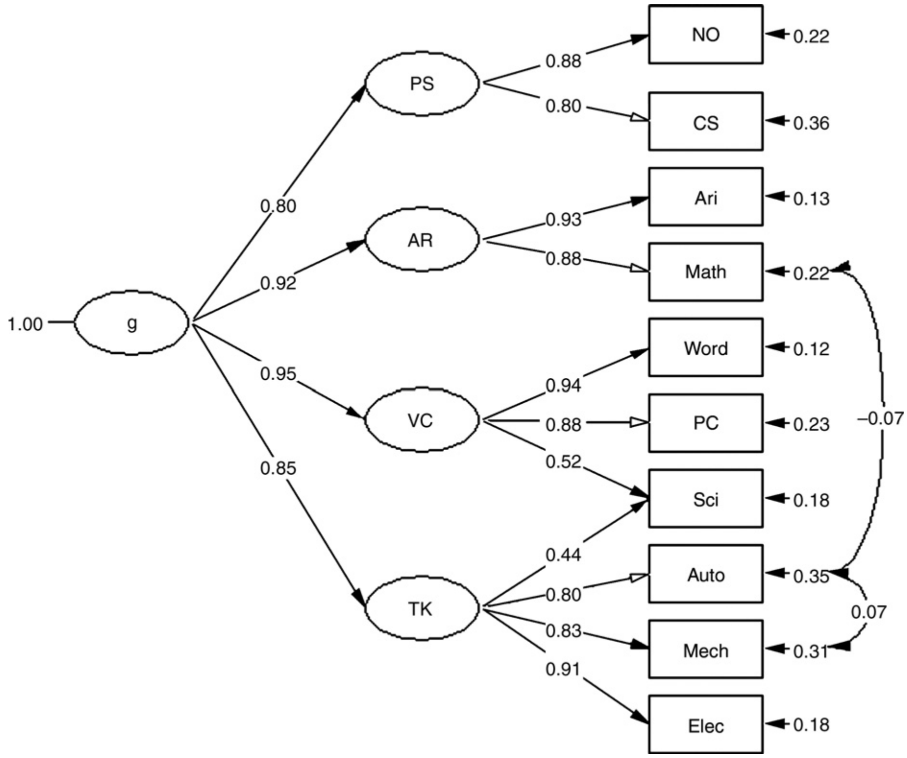

Fig. 1. Hierarchical factor model of the ASVAB: standardised solution. (*g* = general cognitive ability, PS = perceptual speed, AR = mathematical reasoning, VC = verbal reasoning, TK = technical knowledge, NO = Numerical operations, CS = Coding speed, Ari = Arithmetic reasoning, Math = Mathematics knowledge, Word = Word knowledge, PC = Paragraph comprehension, Sci = General science, Auto = Auto and shop information, Mech = Mechanical comprehension and Elec = Electrical information).

Bentler (1999). Based on factor score coefficients generated by LISREL, *g* factor scores were calculated as follows: $g = \text{ASVAB1} * .0810 + \text{ASVAB2} * .0159 + \text{ASVAB3} * .0982 + \text{ASVAB4} * .1049 + \text{ASVAB5} * .0254 + \text{ASVAB6} * .0103 + \text{ASVAB7} * .0320 + \text{ASVAB8} * .0578 + \text{ASVAB9} * .0186 + \text{ASVAB10} * .0565$. For the AFQT we first tested a one-factor solution, which the fit indices showed to be mis-specified (NNFI = .76, SRMR = .043). After allowing a correlated error between Word knowledge and Paragraph comprehension, representing a verbal specific, the fit indices were NNFI = 1.00 and SRMR = .00096, indicating near-perfect fit. The factor score coefficients based on this model were used to estimate *g* from the AFQT scales according to the formula: $g = \text{ASVAB2} * .4927 + \text{ASVAB3} * .0844 + \text{ASVAB4} * .1503 + \text{ASVAB8} * .3347$.

3. Results

3.1. ASVAB subtest comparisons

There was no significant male–female difference in Mathematics knowledge, but *p* values for all other ASVAB subtests were $< .001$ (Table 1). Females scored significantly higher on the Word knowledge, Paragraph comprehension, Numerical operations, and Coding speed subtests of the ASVAB. Males score higher on Science, Arithmetic, Auto and shop information, Mechanical comprehension, and Electronics information. Effect sizes range from small to large (Table 1).

Males have significantly greater variance in all ASVAB subtests except Numerical operations and Coding speed, and on all the AFQT summary scores (Table 1). The Standard deviation ratios range from 1.07 to 1.38.

3.2. General factor comparisons

The principal results in this study are the male–female comparisons on the general factor of intelligence. Males show a very small (Cohen's $d = 0.064$) but significant advantage on the *g* factor extracted from the AFQT (Table 1). Males score significantly higher on the *g* factor from the ASVAB, though the effect size is again very small (Cohen's $d = 0.068$). The strongest finding is for significantly greater variance in male scores. The standard deviations of the *g* factors from the ASVAB and the AFQT have male:female standard deviation ratios of 1.16 and 1.11, respectively. Among the people in our sample with the top 50 scores on the *g* factor from the AFQT (roughly, the top 2%), 33 were male and 17 were female.

4. Discussion

There were small male advantages on the means of the *g* scores from the AFQT and the ASVAB, but there was perhaps a more important difference on the variability. In a sample of young adults, this replicates the findings from population-level studies of 11 year olds (Deary et al., 2003; Strand et al., 2006) of the "more so" nature of males (Heim, 1970). This demonstration in older subjects is important, because Lynn's (Lynn, 1994, 1998, 1999) theory suggests that male–female differences emerge only after puberty.

Individual subtest means showed a range of male and female strengths. For example females score better on Word knowledge, Paragraph comprehension, Numerical operations, and Coding speed, which may reflect a female advantage in phonological coding (Hecht, Torgesen, Wagner, & Rashotte, 2001; Majeres, 1999). Males score better on arithmetic, in accordance with other findings. The higher male mean scores on Auto and shop information, Mechanical comprehension, and Electronics information probably reflect spheres of greater male interest (Ackerman, Bowen, Beier, & Kanfer, 2001) combined with a male advantage in semantic memory (Ackerman et al., 2001; Lynn, Irwing, & Cammock, 2002).This is why more attention should be given to the AFQT.

The previous consensus that there was a cut-off beyond which further increments in IQ do not increase levels of attainment has been refuted (Wai, Lubinski, & Benbow, 2005). Participation will therefore reflect differences in the numbers of very high scorers. The possible role of mean ability differences in explaining the male excess in certain occupational groups, especially science, has been discussed (Lynn & Irwing, 2004). The present data suggest that a stronger difference in mental ability variance between the sexes could give rise to excess males at higher levels, balanced by an excess at lower levels. Differences in variance in the general factor in intelligence, therefore, have potentially important implications for the proportions of high achieving males in areas such as science.

A strength of this study is the sibling design, which controls for possible confounders associated with family background, such as genetic and socio-economic factors. The large numbers and population-representativeness of a large part of the sample are also valuable.

It is useful to consider the advantages and disadvantages of different samples in studying this topic. Children toward the end of primary/elementary education offer the advantages that they can be tested with little biasing in the samples, or in sex-biased educational choices and experiences (Deary et al., 2003; Hedges & Nowell, 1995). However, the findings might not be relevant to adult sex differences in abilities. On the other hand, more mature subjects are affected by marked dimorphism of educational subject choices, which in turn might affect development and performance on mental tests. Of course, these sex differences in school and college subject choices are likely to be influenced by differences in original ability, or by socialisation, or by a combination of these and other factors. Thus, it will remain methodologically difficult to ascertain the nature and causes of any sex differences in abilities among adults. These considerations, though, probably apply more to means than to variances. The present study at least contributes by controlling for a range of genetic and social background factors by including only subjects matched for family background.

References

Abad, F. J., Colom, R., Rebollo, I., & Escorial, S. (2004). Sex differential item functioning in the Raven's Advanced Progressive Matrices: Evidence for bias. *Personality and Individual Differences*, *36*, 1459−1470.

Ackerman, P. L., Bowen, K. R., Beier, M. E., & Kanfer, R. (2001). Determinants of individual differences and gender differences in knowledge. *Journal of Educational Psychology*, *93*, 797−825.

Blinkhorn, S. (2005). Intelligence: A gender bender. *Nature*, *438*, 31−32.

Blinkhorn, S. (2006). Blinkhorn replies. *Nature*, *442*, doi:10.1036/nature04967.

Carroll, J. B. (1993). *Human cognitive abilities.* Cambridge: Cambridge University Press.

Deary, I. J., Thorpe, G., Wilson, V., Starr, J. M., & Whalley, L. J. (2003). Population sex differences in IQ at age 11: The Scottish Mental Survey 1932. *Intelligence*, *31*, 533−542.

Dolan, C. V., Colom, R., Abad, F. J., Wicherts, J. M., Hessen, D. J., & van der Sluis, S. (2006). Multi-group covariance and mean structure modelling of the relationship between the WAIS-III common factors and sex and educational attainment in Spain. *Intelligence*, *34*, 193−210.

Hecht, S. A., Torgesen, J. K., Wagner, R. K., & Rashotte, C. A. (2001). The relations between phonological processing abilities and emerging individual differences in mathematical computation skills: A longitudinal study from second to fourth grades. *Journal of Experimental Child Psychology*, *79*, 192−227.

Hedges, L. V., & Nowell, A. (1995). Sex differences in mental test scores, variability, and numbers of high-scoring individuals. *Science*, *269*, 41−45.

Heim, A. (1970). *Intelligence and personality.* Harmondsworth, UK: Penguin.

Hu, L. -T., & Bentler, P. M. (1999). Cutoff criteria for fit indices in covariance structure modelling: Conventional criteria versus new alternatives. *Structural Equation Modelling*, *6*, 1−55.

Hyde, J. S. (2005). The gender similarities hypothesis. *American Psychologist*, *60*, 581−592.

Irwing, P., & Lynn, R. (2005). Sex differences in means and variability on the progressive matrices in university students: A meta-analysis. *British Journal of Psychology*, *96*, 505−524.

Irwing, P., & Lynn, R. (2006). Is there a sex difference in IQ scores? *Nature*, *442*, doi:10.1038/nature04966.

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

Kimura, D. (2002). Sex hormones influence human cognitive pattern. *Neuroendocrinology Letters*, *23*(Suppl. 4), 67−77.

Lawrence, P. A. (2006). Men, women, and ghosts in science. *Plos Biology*, *4*, doi:10.1371/journal.pbio.0040019.

Kyllonen, P. C. (1993). Aptitude testing inspired by information processing: A test of the four-sources model. *Journal of General Psychology*, *120*, 375−405.

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

Lynn, R. (1998). Sex differences in intelligence: A rejoinder to Mackintosh. *Journal of Biosocial Science*, *30*, 529−532.

Lynn, R. (1999). Sex differences in intelligence and brain size: A developmental theory. *Intelligence*, *27*, 1−12.

Lynn, R., Allick, J., & Irwing, P. (2004). Sex differences on three factors identified in Raven's Standard Progressive Matrices. *Intelligence*, *32*, 411−424.

Lynn, R., & Irwing, P. (2004). Sex differences on the progressive matrices: A meta-analysis. *Intelligence*, *32*, 481−498.

Lynn, R., Irwing, P., & Cammock, T. (2002). *Intelligence*, *30*, 27−39.

Mackintosh, N. J., & Bennett, E. S. (2005). What do Raven's Matrices measure? An analysis in terms of sex differences. *Intelligence*, *33*, 6663−6674.

Majeres, R. L. (1999). Sex differences in phonological processes: Speeded matching and word reading. *Memory and Cognition*, *27*, 246−253.

Muller, C. B., Ride, S. M., Fouke, J., Whitney, F., Denton, D. D., Cantor, N., et al. (2005). Gender differences and performance in science. *Science*, *307*, 1043.

Ree, M. J., & Carretta, T. R. (1995). Group differences in aptitude factor structure on the ASVAB. *Educational and Psychological Measurement*, *55*, 268−277.

Snow, R. E., Kyllonen, P. C., & Marshalek, B. (1984). The topography of ability and learning correlations. In R. J. Sternberg (Ed.), *Advances in the Psychology of Human Intelligence*, *vol. 2*. Hillsdale, NJ: Erlbaum.

Spelke, E. S. (2005). Sex differences in intrinsic aptitude for mathematics and science. *American Psychologist*, *60*, 950−958.

Stauffer, J. M., Ree, M. J., & Carretta, T. R. (1996). Cognitive components tests are not much more than g: An extension of Kyllonen's analyses. *Journal of General Psychology*, *123*, 193−205.

Strand, S., Deary, I. J., & Smith, P. (2006). Sex differences in cognitive ability test score: A UK national picture. *British Journal of Educational Psychology*, *76*, 463−480.

Van der Sluis, S., Posthuma, D., Dolan, D. V., de Geus, E. J. C., Colom, R., & Boomsma, D. I. (2006). Sex differences on the Dutch WAIS-III. *Intelligence*, *34*, 273−289.

Van der Ven, A. H. G. S., & Ellis, J. L. (2000). A Rasch analysis of Raven's Standard Progressive Matrices. *Personality and Individual Differences*, *29*, 45−64.

Wai, J., Lubinski, D., & Benbow, C. (2005). Creativity and occupational accomplishments among intellectually precocious youths: An age 13 to age 33 longitudinal study. *Journal of Educational Psychology*, *97*, 449−484.