Personality and Individual Differences 73 (2015) 56–60

Contents lists available at ScienceDirect

Personality and Individual Differences

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

Are adoption gains on the g factor? A meta-analysis

Jan te Nijenhuis [a,*], Birthe Jongeneel-Grimen [b], Elijah L. Armstrong [c]

[a] University of Amsterdam, Work and Organizational Psychology, The Netherlands
[b] University of Amsterdam, Amsterdam Medical Center, The Netherlands
[c] Fairfax, CA 94390, United States

ARTICLE INFO

Article history:
Received 3 July 2014
Received in revised form 12 September 2014
Accepted 13 September 2014

Keywords:
g loadings
Adoption gains
Method of correlated vectors
Psychometric meta-analysis

ABSTRACT

When children are adopted into prosperous families they generally show IQ gains. The meta-analytic correlation between subtest *g* loadings and adoption gains is examined ($K = 4$, combined $N = 3018$). A number of meta-analytic corrections are applied to the estimate, yielding a correlation of −1. The results are discussed.

1. Introduction

Nearly all mental tests, no matter how diverse, correlate to some extent (Jensen, 1969, 1998a; Spearman, 1927). This means that nearly all tests share common variance, and the cross-test common variance is called *g*, or 'general intelligence'. When a test is a good measure of *g* (i.e., it predicts the results of other tests better, or to use a more technical definition, has a stronger loading on the first principal component extracted from different correlated tests), it is said to have a high *g* loading. Analogously, when a test poorly measures *g*, its *g* loading is low. Some tests have much higher *g* loadings than others (e.g., Jensen, 1969, 1998a; Johnson, Bouchard, Krueger, McGue, & Gottesman, 2004). For example, in the widely used Wechsler Intelligence Scale for Children (WISC) IQ battery, the most *g*-loaded subtest (Vocabulary) loads on the first principal component of the WISC at .904, whereas the least *g*-loaded subtest (Coding) loads on the first principal component of the WISC at .605 (Kan, 2011). *g* loadings are best understood as a measure of cognitive complexity: tests demanding higher cognitive complexity have high *g* loadings, whereas tests demanding lower cognitive complexity have low *g* loadings (Gottfredson, 1997).

The *g* loadings of different tests affect their properties. When a test is more *g*-loaded, scores on the test are damaged more strongly by inbreeding (Rushton, 1999; Rushton & Jensen, 2010), it shows higher differences between ethnic groups (Jensen, 1985), and it correlates more strongly with brain size and reaction time (Jensen, 1998a; Rushton & Ankney, 2009). When a test is less *g*-loaded, the practice effect on that test is larger; likewise the Flynn effect is larger (te Nijenhuis & van der Flier, 2013; te Nijenhuis, van Vianen, & van der Flier, 2007). Other correlates of *g* loadings are given in Armstrong and Woodley (2014) and Jensen (1998a). It appears, therefore, that when a test is more *g*-loaded, it has a stronger relationship to biological factors, and when it is less *g*-loaded, it is more influenced by the environment and by social factors. A few exceptions to this trend are listed in Flynn, te Nijenhuis, and Metzen (2014), who list some environmental sources of IQ decrements (including foetal alcohol syndrome and Traumatic Brain Injury) that, despite being environmental, show correlations of 0 with the *g* loadings of different tests.

One common method of determining the relationship between *g* loadings and variables is called the "method of correlated vectors" (Jensen, 1998a), hereafter MCV. This method entails correlating the *g* loadings of different tests with changes on test scores produced by given variables. When this relationship is positive, the IQ differences are more likely to be concentrated on *g*; when it is negative, they are more likely to be concentrated on specific abilities (e.g., Woodley, 2011). However, MCV is best combined with psychometric corrections for unreliability of tests and restricted range of *g* loadings (e.g., Jensen, 1998a; Woodley, te Nijenhuis, Must, & Must, 2014). Moreover, small sample sizes can produce anomalous results when analyzed with MCV (e.g., Dolan,

* Corresponding author at: Gouden Leeuw 746, 1103 KR Amsterdam, The Netherlands.
E-mail address: JanteNijenhuis@planet.nl (J. te Nijenhuis).

http://dx.doi.org/10.1016/j.paid.2014.09.022
0191-8869/

2000; Jensen, 1998a). Meta-analytic corrections such as those described above (Hunter & Schmidt, 2004), strongly help in producing more sensible results when using MCV (Woodley et al., 2014).

When children are adopted into prosperous families (from less prosperous families), they generally show IQ gains relative to their parents (e.g., Capron & Duyme, 1996; Scarr & Weinberg, 1976; Skodak & Skeels, 1949), though these gains are not always present in adulthood (e.g., Jensen, 1998a; Weinberg, Scarr, & Waldman, 1992). In some cases, the IQ gains are high (10 points; Flynn, 1993; Scarr & Weinberg, 1976). It is not surprising that children adopted into prosperous families show IQ gains, since they generally receive better and longer education and better nutrition, are more cognitively stimulated by their environment, become more test-wise, and are less likely to be exposed to neurotoxins. These sources of IQ gain, of course, are environmental. Therefore, since environmental IQ gains tend not to be on *g*, it stands to reason that IQ gains from adoption would not be *g*-loaded either. In the present study, we test this prediction by using MCV, coupled with psychometric meta-analytic techniques, to examine the relationship between adoption gains and test *g* loadings.

2. Method

Psychometric meta-analysis (Hunter & Schmidt, 2004) aims to quantify underlying construct-level relationships without attenuation by statistical artifacts (Schmidt & Hunter, 1999). The goal of the present psychometric meta-analysis is to provide reliable estimates of the true correlation between adoption gains and the magnitude of *g* loadings. As the techniques we use are relatively unknown to the majority of our readers, we choose to give a detailed description of the techniques. However, highly similar descriptions have also been used in other recent publications.

2.1. Searching and screening studies

To identify studies for inclusion in the meta-analysis, both electronic and manual searches for studies that contained IQ data for adopted children or for adults adopted as children were conducted in 2007. Four methods were used to obtain adoption gains from both published and unpublished studies for the present meta-analysis. First, an electronic search for published work was conducted, using PsycINFO, ERIC, MEDLINE, PiCarta, Academic search premier, Web of science, and PubMed. The following keyword combinations were used to conduct searches: adopt∗ (where an asterisk indicates that the search contained but was not limited to that word or word fragment), adopted children, and adoption in combination with the keywords IQ, intelligence, intellectual development, *g*, GMA, general mental ability, cognitive development, cognitive ability, and general cognitive ability. Second, we browsed the content tables of several major research journals of development, genetics, and of intelligence, such as *Behavior Genetics* 1970–2007, *Intelligence* 1977–2007, *Psychological Science* 1990–2007, *Child Development* 1930–2007, and *Developmental Psychology* 1969–2007. Third, several well-known researchers who have conducted IQ research on the adopted were contacted in order to obtain any additional articles or supplementary information. Finally, we checked the reference list of all currently included empirical studies to identify any potential articles that may have been missed by earlier search methods.

2.2. Inclusion rules

For a study to be included in the meta-analysis two criteria had to be met: First, to get a reliable estimate of the true correlation between adoption gains and the *g* loadings the cognitive batteries had to have a minimum of seven subtests; second, well-validated tests had to be used. The general inclusion rules were applied and yielded three papers resulting in four correlations between *g* and *d*.

2.3. Computation of adoption gains

One of the goals of the present meta-analysis is to have a reliable estimate of the true correlation between adoption gains (d) and *g*. All studies except that of Frydman and Lynn (1989) reported results using a comparison group. To be able to compute d (adoption gains) from that study we needed to compare the results of the intervention group against the results of a comparison group. We therefore decided to compare the mean of the scaled scores of adopted children on the French WISC reported in Frydman and Lynn (1989) with the mean of the scaled scores of the standardization group of 10-year-old children of the French WISC manual (Wechsler, 1965). Adoption gains (d) were computed by subtracting the mean of the comparison group from the mean of the intervention group, and then dividing the result by the (mean) *SD* of the standardization group(s) of the particular test in question.

2.4. Computation of 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 (Jensen & Weng, 1994). Finally, Pearson correlations between adoption gains and the *g* loadings were computed.

2.5. Corrections for artifacts

Psychometric meta-analytical techniques (Hunter & Schmidt, 2004) were applied using the software package developed by Schmidt and Le (2004). As stated in the introduction, there are many statistical artifacts that affect correlations, and psychometric meta-analysis is used to correct for these artifacts. We corrected for several statistical artifacts (Hunter & Schmidt, 2004): sampling error, unreliability (of *g* loadings and adoption gains), restricted range of *g* loadings, and suboptimal construct validity.

2.5.1. Correction for sampling error

In many cases sampling error explains the majority of the variation between studies, so the first step in a psychometric meta-analysis is to correct the collection of effect sizes for differences in sample size between the studies.

2.5.2. Correction for reliability of the vector of g loadings

The values of r ($g \times$ adoption gains) are attenuated by the reliability of the vector of *g* loadings for a given battery. When two samples have a comparable N, the average correlation between vectors is an estimate of the reliability of each vector. Several samples were compared that differed little on background variables. For the comparisons using children we chose samples that were highly comparable with regard to age. Samples of children in the age of 3–5 years were compared against other samples of children who did not differ more than 0.5 year of age. Samples of children in the age of 6–17 years were compared against other samples of children who did not differ more than 1.5 year of age. For the comparisons of adults we compared samples in the age of 18–95 years.

We collected correlation matrices from test manuals, books, articles, and technical reports. The large majority came from North

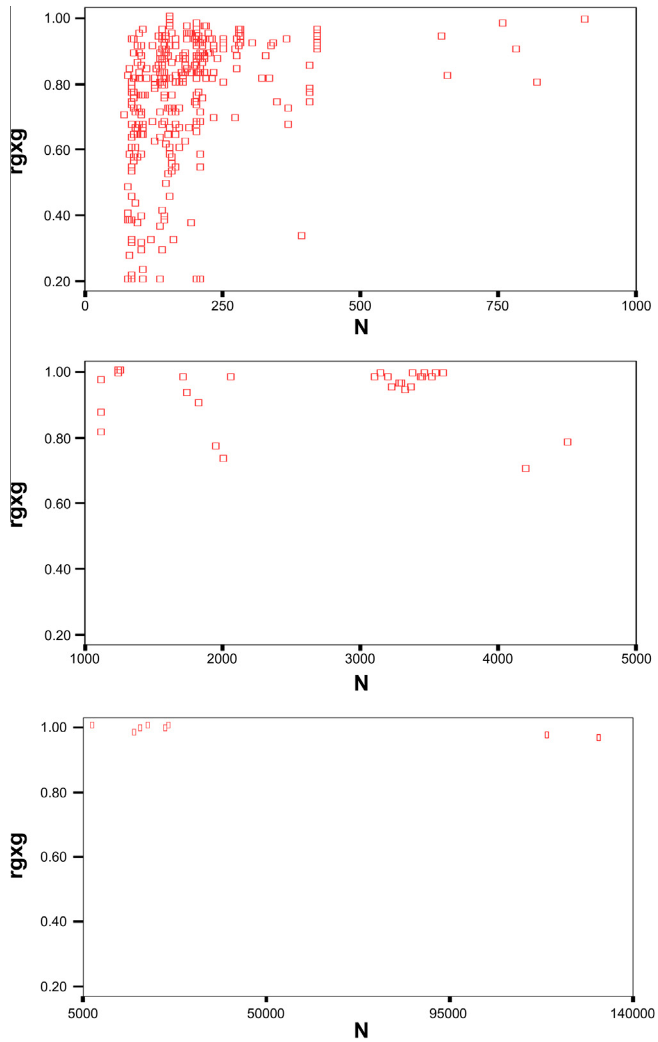

Fig. 1. Three scatter plot of reliability of the vector of g loadings and sample size each for a different range of N.

America, with a large number of European countries, and also a substantial number from Korea, China, Hong Kong, and Australia. This resulted in about 700 data points, which led to 385 comparisons of g loadings of comparable groups which provided an indication of the reliability for that group.

A scatter plot of reliabilities against Ns should show that the larger N becomes, the higher the value of the reliability coefficients, with an asymptotic function between r $(g \times g)$ and N expected. We checked to see which curve gave the best fit to the expected asymptotic function. The logarithmic regression line resembled quite well the expected asymptotic distribution for reliabilities. However, because the extreme range on the x-axis resulted in a picture that is not informative, the regression line for r $(g \times g)$ and N is not reported. For the same reason we divided Fig. 1 into three parts, each showing the scatter plot of reliability of the vector of g loadings and sample size for a specific range of N.

2.5.3. Correction for reliability of the vector of adoption gains (d)

The value of r $(g \times d)$ is attenuated by the reliability of the vector of adoption gains for a given battery. The reliability of the vector of adoption gains was estimated using the present datasets and by comparing the samples that took the same test and that were comparable in regard to age and sample size. As an illustration of the procedure, consider the vectors of adoption gains from datasets on the WISC. Frydman and Lynn (1989) tested Belgian children ($N = 19$) with an average age of 10 years (age range 6.04–13.11 years) and (Colombo, de la Parra, and Lopez (1992)) tested Chilean children ($N = 27$) with an average age of 8.8 years (age range 5.7–11.2 years). This yielded a correlation of −.097 between the two vectors (total $N = 46$; average $N = 23$). To limit the risk of over-correction, we set the reliability at .20, which is the lowest positive reliability we found in any of the three datasets. We decided to set the reliability at .20 for adoption datasets with a reliability of less than .20.

An asymptotic function between r $(d \times d)$ and N is expected. We checked to see which curve gave the best fit to the expected asymptotic function. Figure 2 shows the scatter plot of reliability of the vector of adoption gains and sample size, and the curve that fitted optimally.

2.5.4. Correction for restriction of range of g loadings

The values of r ($g \times$ adoption gains) are attenuated by the restriction of range of g loadings in many of the standard test batteries. The most highly g-loaded batteries tend to have the smallest range of variation in the subtests' g loadings. Jensen (1998a, pp. 381–382) showed that restriction in the magnitude of g loadings strongly attenuates the correlation between g loadings and standardized group differences. Hunter and Schmidt (1990, pp. 47–49) state that the solution to range variation is to define a reference population and express all correlations in terms of that reference population. The Hunter and Schmidt meta-analytical program computes what the correlation in a given population would be if the standard deviation were the same as in the reference population. The standard deviations can be compared by dividing the standard deviation of the study population by the standard deviation of the reference group, that is $u = SD_{study}/SD_{ref}$. As references we used tests that are broadly regarded as exemplary for the measurement of the intelligence domain, namely the

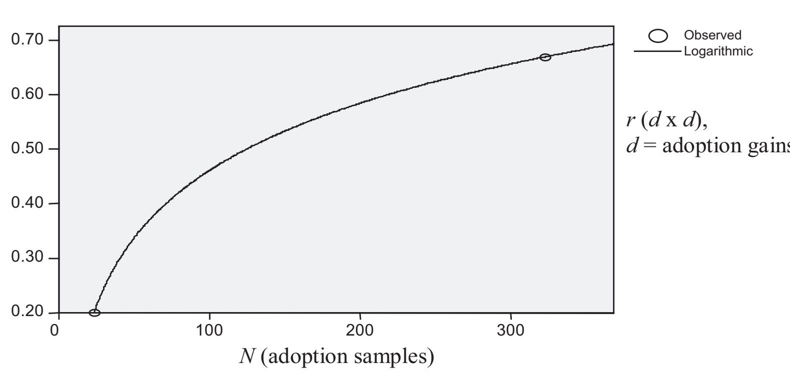

Fig. 2. Scatter plot of reliability of the vector of adoption gains and sample size and regression line.

Table 1
Studies of correlations between *g* loadings and adoption gains.

Reference	Test	*r*	*N*	Age mean (range)
Frydman and Lynn (1989)	"French adaptation of" WISC	−.329	19	10 (6.04–13.11)
Colombo et al., 1992	"Chilean adaptation of" WISC	.131	27	8.9 (5.7–11.2)
Wadsworth, DeFries, and Fulker (1993)	WISC-R	−.640	415	7.4
Wadsworth et al. (1993)	WISC-R	−.728	230	12.3

Note. In general, the *g* loadings were based on the correlation matrix taken from test manuals or from the correlation matrix based on the largest sample size we could find.

various versions of the Wechsler tests for children and adults. The average standard deviation of *g* loadings of the various versions of the Wechsler-Bellevue (W-B), Wechsler Preschool and Primary Scale of Intelligence (WPPSI), Wechsler Intelligence Scale for Children (WISC), Wechsler Intelligence Scale for Children-Revised (WISC-R), Wechsler Intelligence Scale for Children-Third Edition (WISC-III), and the Wechsler Intelligence Scale for Children-Fourth Edition (WISC-IV) from datasets from countries all over the world was 0.132. We used this value as our reference in the studies with children. The average standard deviation of *g* loadings of the various versions of the Wechsler Adult Intelligence Scale (WAIS), Wechsler Adult Intelligence Scale-Revised (WAIS-R), and the Wechsler Adult Intelligence Scale-Third Edition (WAIS-III) from datasets from countries all over the world was 0.107. This was used as the reference value in the studies with adults. In so doing, the SD of *g* loadings of all test batteries was compared to the average *SD* in *g* loadings in the Wechsler tests for, respectively, children and adults.

2.5.5. Correction for deviation from perfect construct validity

The deviation from perfect construct validity in *g* attenuates the values of *r* ($g \times$ adoption gains). The sample values of *g* are affected by psychometric sampling error, since two *g* factors extracted from different tests will not be perfect measures of *g* (Jensen, 1998a), but the fact that *g* is very substantially correlated across different test batteries implies that the differing obtained values of *g* can all be interpreted as estimates of a "true" *g* (Johnson et al., 2004). The values of *r* ($g \times$ adoption gains) are attenuated by psychometric sampling error in each of the batteries from which a *g* factor has been extracted. As in other studies (e.g., Woodley et al., 2014), we used the conservative value of .90 for the correlation between the measured and the true *g* score.

3. Results

The results of the studies on the correlation between *g* loadings and adoption gains are presented in Table 1. The table gives data derived from four studies, with participants numbering a total of 691. The table lists the reference for the study, the cognitive ability test used, the correlation between *g* loadings and adoption gains, the sample size, and the mean age (and range of age). It is clear that most correlations are negative, and half of them quite strongly so. Table 2 shows the results of the psychometric meta-analysis of the four data points.

The estimated true correlation has a value of −.92, with only 37% of the variance in the observed correlations explained by artifactual errors. However, Hunter and Schmidt (1990) state that extreme outliers should be left out of the analyses, because they are most likely the result of errors in the data. They also argue that strong outliers artificially inflate the *SD* of effect sizes and thereby reduce the amount of variance that artifacts can explain. We chose to leave out one outlier – a value of *r* more than 7.5 *SD* above the average *r* of the sample of three final data points. This resulted in a small change in the value of the true correlation, a very large decrease in the *SD* of rho with 100%, and a very large increase in the amount of variance explained in the observed correlations by artifacts. Therefore, when one extreme outlier is excluded artifacts explain all of the variance in the observed correlations. Finally, a correction for deviation from perfect construct validity in *g* took place, using a conservative value of .90. This resulted in a value of −1.06 for the final estimated true correlation between *g* loadings and adoption gains. This value is obviously too high, since −1.00 is the lowest possible correlation, so we adopted −1.00 as our final estimate.

Table 2
Meta-analytical results for correlation between adoption gains and *g* loadings after corrections for reliability, restriction of range, and imperfectly measuring the construct.

K	*N*	*r*	*SDr*	Rho-4	SDrho-4	Rho-5	%VE	80% CI
4	691	−.64	.17	−.92	.14	−1.01	37	−1.10 to −.74
3	664	−.66	.06	−.95	.00	−1.06	211	−.95 to −.95

Note. K = number of correlations; *N* = total sample size; *r* = mean observed correlation (sample size weighted); *SDr* = Standard deviation of observed correlation; rho-4 = observed correlation corrected for unreliability and range restriction; *SD*rho = Standard deviation of true correlation; rho-5 = observed correlation corrected for unreliability, range restriction, and imperfect measurement of the construct;%VE = Percentage of variance accounted for by artifactual errors; 80% CI = 80% credibility interval.

The outcome of any meta-analysis based on a limited number of studies depends to some extent on study properties that vary randomly across studies. This phenomenon is called "second-order sampling error". It results from the sampling of studies in a meta-analysis. Percentages of variance explained having values greater than 100% are not uncommon when only a limited number of studies are included in the analysis. The proper conclusion is that all the variance is explained by statistical artifacts (see Hunter & Schmidt, 2004, pp. 399–401 for an extensive discussion).

4. Discussion

Our meta-analysis of four studies (with a combined *N* of 3018 individuals) yielded a perfectly inverse correlation between the *g* loadings of different IQ subtests and the degree to which adoption gains increased scores on those tests. This replicated Jensen's (Jensen, 1998b) finding that adoption gains were not on *g*. Therefore, it is now fairly well-established that adoption gains are not on *g*.

The present meta-analysis, however, has a few weaknesses. For example, it includes only four studies. Moreover, it includes only one IQ test, the Wechsler. It would be preferable to have a larger number of studies, and a more diverse collection of tests (Jensen, 1998a; te Nijenhuis, David, Metzen, & Armstrong, 2014). However, it should not be forgotten that the present findings are based on the full world literature and that studies reporting scores on all subtests of an IQ battery are quite rare. Future research on adoption gains and the *g* factor ought to use a more diverse sample of tests and studies, if possible.

It appears that adoption gains are like environmental sources of IQ differences in that they are not on *g*. By now, therefore, it is very well established that environmental IQ differences are not usually *g* loaded, while genetic ones are. Therefore, whether a source of IQ differences is *g* loaded can provide a good indication of whether that source of IQ differences is environmental or genetic.

It must be noted, however, that while adoption gains are not *g* loaded, that does not mean they are insignificant. Some extremely significant sources of IQ differences, like TBI, are not *g*-loaded (Flynn et al., 2014). Therefore, IQ gains caused by adoption, or by other cultural factors (like education), may still have desirable real-world effects, despite not being on *g*.

4.1. Implications for the study of education

It is known that IQ gains may be induced by formal education (e.g., Ceci, 1991). Adopted children probably receive superior education relative to most children, so it seems plausible that their IQ advantages are partly due to education. Hence, the presence of adoption gains confirms others' studies in suggesting that education can raise IQ.

Educational gains are environmental, so it stands to reason that, like other environmental sources of IQ variance, they are not *g* loaded, although to the best of the authors' knowledge, no published study has ever directly measured the *g* loadedness of educationally induced gains. The present study suggests that educational gains are indeed not *g* loaded, since adopted children do not demonstrate any gain in *g*.

However, it remains yet to be seen whether educational gains are "significant", that is to say, whether they have real-world consequences. Flynn (2009, 2012) has argued that in fact educational gains *are* "significant", in that they influence one's ability to think abstractly and to categorize in scientific rather than practical terms. Flynn et al. (2014), in support of this hypothesis, demonstrate that some environmental, non-*g* loaded changes in cognitive ability that are very significant (such as TBI, prenatal cocaine exposure and foetal alcohol syndrome) are not *g* loaded, which counteracts Jensen's (1998) assertion that only *g* loaded gains are strongly significant. It is therefore an open question, and beyond the capacity of the present study to answer, whether educational (or other adoption-related) gains, despite not being gains in *g*, are strongly related to real-world functioning.

Acknowledgements

We would like to thank Dr. Marinus van Ijzendoorn, Dr. Jon Sundet, Dr. Henk van der Flier, Dr. Sally Wadsworth, and Dr. Panagiota Vorria for their enthusiastic support of this project.

References

References marked with an asterisk indicate studies included in the meta-analysis

Armstrong, E. L., & Woodley, M. A. (2014). The rule-dependence model explains the commonalities between the Flynn effect and IQ gains via retesting. *Learning and Individual Differences, 29*, 41–49.

Capron, C., & Duyme, M. (1996). Effect of socioeconomic status of biological and adoptive parents on WISC-R subtests scores of their French adoptive children. *Intelligence, 22*, 259–275.

Ceci, S. J. (1991). How much does schooling influence general intelligence and its cognitive components? A reassessment of the evidence. *Developmental Psychology, 27*, 703–722.

*Colombo, M., de la Parra, A., & Lopez, I. (1992). Intellectual and physical outcome of children undernourished in early life is influenced by later environmental conditions. *Developmental Medicine and Child Neurology, 34*, 611–622.

Dolan, C. V. (2000). Investigating Spearman's hypothesis by means of multi-group confirmatory factor analysis. *Multivariate Behavioral Research, 35*, 21–50.

Flynn, J. R. (1993). Skodak and Skeels: The inflated mother–child IQ gap. *Intelligence, 17*, 557–561.

Flynn, J. R. (2009). *What is intelligence?* (2nd ed). New York: Cambridge University Press.

Flynn, J. R. (2012). *Are we getting smarter?* New York: Cambridge University Press.

Flynn, J. R., te Nijenhuis, J., & Metzen, D. (2014). The g beyond Spearman's *g*: Flynn's paradoxes resolved using four exploratory meta-analyses. *Intelligence, 44*, 1–10.

*Frydman, M., & Lynn, R. (1989). The intelligence of Korean children adopted in Belgium. *Personality and Individual Differences, 10*, 1323–1325.

Gottfredson, L. S. (1997). Why *g* matters: The complexity of everyday life. *Intelligence, 24*, 79–132.

Hunter, J. E., & Schmidt, F. L. (1990). *Methods of meta-analysis*. London: Sage.

Hunter, J. E., & Schmidt, F. L. (2004). *Methods of meta-analysis: Correcting error and bias in research findings* (2nd ed.). London: Sage.

Jensen, A. R. (1969). How much can we boost IQ and scholastic achievement? *Harvard Educational Review, 33*, 1–123.

Jensen, A. R. (1985). The nature of the black–white difference on various psychometric tests. Spearman's hypothesis. *The Behavioral and Brain Sciences, 8*, 193–263.

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

Jensen, A. R. (1998). Adoption data and two *g*-related hypotheses. *Intelligence, 25*, 1–6.

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

Johnson, W., Bouchard, T. J., Jr., Krueger, R. F., McGue, M., & Gottesman, I. I. (2004). Just one g: Consistent results from three test batteries. *Intelligence, 32*, 95–107.

Kan, K-J. (2011). *The nature of nurture: The role of gene–environment interplay in the development of intelligence* (Ph.D. Dissertation). University of Amsterdam.

Rushton, J. P. (1999). Secular gains 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., & Ankney, C. D. (2009). Whole brain size and general mental ability: A review. *International Journal of Neuroscience, 119*, 691–731.

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.

Scarr, S., & Weinberg, R. A. (1976). IQ test performance of black children adopted by white families. *American Psychologist, 31*, 726–739.

Schmidt, F. L., & Hunter, J. E. (1999). Theory testing and measurement error. *Intelligence, 27*, 183–198.

Schmidt, F. L., & Le, H. (2004). *Software for the Hunter–Schmidt meta-analysis methods*. Iowa city, IQ, 42242: University of Iowa, Department of Management & Organization.

Skodak, M., & Skeels, H. M. (1949). A final follow-up study of one hundred adopted children. *Journal of Genetic Psychology, 75*, 85–125.

Spearman, C. (1927). *The abilities of man: Their nature and measurement*. New York: Macmillan.

te Nijenhuis, J., & van der Flier, H. (2013). Is the Flynn effect on g? A meta-analysis. *Intelligence, 41*, 802–807.

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., David, H., Metzen, D., & Armstrong, E. L. (2014). Spearman's hypothesis tested on European Jews vs non-Jewish Whites and vs Oriental Jews: Two meta-analyses. *Intelligence, 44*, 15–18.

*Wadsworth, S. J., DeFries, J. C., & Fulker, D. W. (1993). Cognitive abilities of children at 7 and 12 years of age in the Colorado adoption project. *Journal of Learning Disabilities, 26*, 611–615.

Wechsler, D. (1965). *Echelle d'Intelligence de Wechsler pour enfants (WISC). Étalonnages Français: Manuel. [Wechsler Scale of Intelligence for Children WISC. French calibration: Manual]*. Paris: Centre de Psychologie Appliquée.

Weinberg, R. A., Scarr, S., & Waldman, I. D. (1992). The Minnesota Transracial Adoption Study: A follow-up of IQ test performance at adolescence. *Intelligence, 16*, 117–135.

Woodley, M. A. (2011). The cognitive differentiation–integration effort hypothesis: A synthesis between the fitness indicator and life history models of human intelligence. *Review of General Psychology, 15*, 228–245.

Woodley, M. A., te Nijenhuis, J., Must, O., & Must, A. (2014). Controlling for increased guessing enhances the independence of the Flynn effect from g: The return of the Brand effect. *Intelligence, 43*, 27–34.