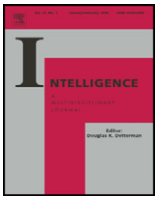
Intelligence 44 (2014) 1–10

Contents lists available at ScienceDirect

Intelligence



The *g* beyond Spearman's *g*: Flynn's paradoxes resolved using four exploratory meta-analyses

James R. Flynn[a,*,1], Jan te Nijenhuis[b,1], Daniel Metzen[b]

[a] *Departments of Political Studies and Psychology, University of Otago, Dunedin, New Zealand*
[b] *Work and Organizational Psychology, University of Amsterdam, Amsterdam, The Netherlands*

ARTICLE INFO

Article history:
Received 20 October 2013
Received in revised form 29 January 2014
Accepted 30 January 2014
Available online 25 February 2014

Keywords:
Flynn effect
Jensen effect
Method of correlated vectors
Significance of IQ gains over time
Significance of black IQ gains on whites

ABSTRACT

When IQ tests are ranked by the magnitude of their score gains over time, this hierarchy lacks a positive correlation with the same tests ranked by their *g* loadings. Therefore, Jensen declared IQ gains "hollow" and, by implication, extended this judgment to score gains that indicated that blacks had made IQ gains on whites. We offer four exploratory meta-analyses that apply Jensen's method to the subtest score differences between normal subjects and those suffering from certain afflictions: iodine deficiency ($K = 6$, $N = 196$), prenatal cocaine exposure ($K = 2$, $N = 215$), fetal alcohol syndrome and degree of fetal alcohol syndrome (respectively, $K = 1$, $N = 110$; and $K = 3$, $N = 125$), and traumatic brain injury ($K = 14$, $N = 629$). All of these create a substantial cognitive deficit in those afflicted. However, the correlations between subtest score differences and *g* loadings run from -0.23 to $+0.12$, with an unweighted average of 0.00.

1. Introduction

Jensen used what he called the method of correlated vectors to test whether a "Flynn effect" was a "Jensen effect". The application of this method implies that IQ gains over time do not signal significant cognitive progress. In this paper, we will: (1) describe the method of correlated vectors; (2) state two paradoxes that its application entails; (3) use data from four exploratory meta-analyses to sharpen those paradoxes; (4) propose a solution based on limiting the method's application.

1.1. The method

Jensen recommended the method as a criterion for evaluating the significance of IQ gains over time. He ranked IQ tests into two hierarchies, best exemplified by ranking the 10 or 11 subtests of the WISC (Wechsler Intelligence Scale for Children) or the WAIS (Wechsler Adult Intelligence Scale). The first hierarchy ran from the subtest that had the greatest *g* loading down to the one that had the least; the second hierarchy ran from the subtest on which there had been the largest gains over time down to the one on which there had been the least. If the correlation was positive, the IQ gains were *g* gains; if negative, they could not be intelligence gains with all that entailed. A recent meta-analysis based on a large total *N* shows the meta-analytical correlation is $rho = -.38$ (te Nijenhuis & van der Flier, 2013).

The appeal of the method rests on the fact that cognitive tasks become more complex as their *g* loading rises. For example, the *g* loading of Digit Span forward, a simple task of repeating a series of random numbers in the order in which they are read out, has a low *g* loading. Digit Span backward, a more complex task of saying numbers in reverse of the order in which they are read out, has a much higher *g* loading. Scrambling eggs has a lower *g* loading than making a soufflé. Speed of shoe tying would have a *g* loading of close to zero. Most of us believe that the more cognitively complex a task the more it measures intelligence.

---

[*] Corresponding author at: Department of Political Studies, University of Otago, PO Box 56, Dunedin, New Zealand.
[1] The two first authors contributed equally to this study.

0160-2896/$ – see front matter © 2014 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.intell.2014.01.009

1.2. Applications and paradoxes

When Jensen applied the method of correlated vectors, he asserted that its results would determine whether IQ gains were cognitively significant or "hollow". By significant, he meant that it would test whether group differences were *g* differences. If the result was negative, groups could still differ on the acquisition of task-specific skills (Jensen, 1998, 320–321 & 332). Had he expanded on this and emphasized that these skills could be cognitively significant, this would have been felicitous. But he did not. Moreover, there would still have been a serious problem. Because Jensen identified "intelligence differences" with "*g* differences", his position meant that that having flunked the test of correlated vectors, group IQ differences could never be equated with intelligence differences. We will show that they can: *g* aside, they still discriminate between groups concerning whom there is virtually universal acceptance that they differ in intelligence.

Further, the identification of *g* with intelligence led to deceptive conclusions about the significance of black versus white trends over time, trends that reduced the racial IQ gap. Rushton and Jensen (2006, p. 922) exemplify this. First, they state that the best estimate of black/white convergence is between 0 and 3.44 IQ points. Second, they state that "While secular increases on various tests cluster together, they do so independently of Black–White differences, which cluster with the *g* factor and genetic indices such as inbreeding depression and twin differences" (Rushton & Jensen, 2006, p. 922). The implication is that black gains on whites are somehow devalued because they are not on the *g* factor.

As the above exemplifies, the method can be applied to groups that are separated, not by time, but purely by the fact that they have different subtest score hierarchies. As Jensen says, at any given time, American blacks and whites differ; and the magnitude of their subtest differences correlates with *g* loadings. These results convinced Jensen that race differences are significant. If the correlation had been negative, his method would have shown that blacks and whites were separated not by intelligence differences but merely by hollow differences. He did not, of course, simply assume that intelligence differences were genetic in origin but stated many arguments to that effect.

The application to black/white comparisons leads directly to a paradox. Its advocates must assert three propositions, which cannot be reconciled: the black/white IQ gap in 1972 was significant and real because it correlated with *g*; the black/white IQ gap of 2002 (a lesser gap) was also significant and real because it correlated with *g*; the IQ gains that reduced the gap were hollow and unreal because they did not correlate with *g*. Collectively these propositions posit that a hollow trend can make a real-world difference. Rindermann and Thompson (2013) show that this real world difference is not hypothetical (and larger than Rushton and Jensen concede). Between 1971 and 2008, averaging scores for reading and mathematics from the NAEP (National Assessment of Educational Progress), blacks gained the equivalent of 6.39 IQ points on whites, leaving the final gap for all ages at 9.94 points. As Flynn (2013b) notes, this is almost identical to black IQ gains on whites between 1972 and 2002: a gain of 5.5 points, leaving the final gap at 10.00 points for ages 9 to 17. Yet, the IQ gains, whatever their magnitude, fail the test of correlated vectors. On this, Jensen and Ruston are correct.

The application of the method of correlated vectors to generational comparisons poses another paradox, one relevant to the general significance of IQ gains over time. People over time have made huge gains on subtests every one of which poses problems of cognitive complexity. However, the fact that the gains do not rank tasks according to the *magnitude* of their cognitive complexity is posited as a rationale for denying that significant gains have occurred. This has an unstated assumption: large gains on simple tasks show that lesser gains on complex tasks are not significant. For example, WISC trends over 54 years show that gains on Coding (low *g* loading) are much greater than gains on Vocabulary (high *g* loading). Nonetheless, Vocabulary gains amount to 0.30 SDs, which means that school children can communicate better than they could in the past. There is no other area in which progress on a complex task is discounted because of greater progress on a simple task. Pole-vaulting is more complex than sprinting. But we do not consider how much performance in the latter has improved to assess the significance of improved performance in the former.

The paradox may be stated as: one group betters another on ten tasks that all involve cognitive complexity; it does not do so according to the magnitude of their cognitive complexity; therefore, the gains on each and every task have only trivial significance. That implies a hypothesis: under these conditions, evidence of the real-world significance of the gains will be weak or non-existent. Research over the last five years demonstrates the contrary:

☐ When you deduct *g* from performance on the SAT, the scores still predict grades (Coyle & Pillow, 2008).
☐ Education promotes autonomous and diverse skills. Its effects are not mediated by *g* but by direct links to specific subtests (Ritchie, Bates, & Deary, 2014; Woodley, 2012a).
☐ IQ gains over time (which of course do not correlate with *g*) parallel and predict growth in GDP per capita (Woodley, 2012b).
☐ Autonomous skills allow one to adapt cognitively to modernity and thereby promote a better life (Woodley, Figueredo, Ross, & Brown, 2013)
☐ Modernity in general encourages greater sensitivity to a whole range of rules, operating independently in a complex web of social situations, rather than collectively as assumed by *g* (Armstrong & Woodley, 2014).

A basketball analogy is inevitable. Imagine a team that bettered another team on every basketball task. They can make layups better, shoot fouls better, make set shots better, do fade ways jumps better, pass better, guard better. But when these tasks are ranked in terms of difficulty, the degree of their superiority does not correlate with their complexity, that is, their advantage on foul shots (more difficult) was less than their advantage on lay ups (less difficult). If someone objects that their advantage is only trivial because it has been "hollowed-out" by the failed correlation, evidence must decide: it shows that the first team beats the second team easily.

1.3. Seeking additional evidence

The conclusion that score differences between two groups on Wechsler subtests need *not* correlate with *g* in order to be cognitively significant will be reinforced by presenting analogous cases. The task is to find two groups so that: (1) Group B has a significant advantage over Group A for Full Scale IQ on the same test; (2) that advantage arises out of a score advantage on most or all of its subtests; (3) the advantages create a subtest hierarchy that does not correlate positively with the subtest *g*-loading hierarchy. This is analogous to IQ gains over time in that time in itself means nothing. Imagine that Group A is the last generation, that Group B is the present generation, and that Group B's advantage on subtests represents score gains over time. In other words, IQ gains become a special case of a generalization: the absence of the correlation with *g* loadings does not show that the difference between the two groups lacks cognitive significance

1.4. The analogous cases

We applied the method of correlated vectors in four exploratory meta-analyses of, respectively, iodine deficiency, prenatal cocaine exposure, fetal alcohol syndrome, and traumatic brain injury. Although we made an effort to find a substantial number of studies, studies reporting the scores on all subtests of an IQ battery are quite rare. We will first describe the afflictions.

1.5. Iodine deficiency

According to the WHO, iodine deficiency is the single most preventable cause for mental retardation in the world. A diet that has a strong deficiency in the trace element iodine can lead to endemic goiter and endemic cretinism. The former refers to a swelling of the thyroid glance that leads to a swollen neck. The latter refers to a restriction in physical and mental growth due to a lack of thyroid hormones (Bleichrodt, Drenth, & Querido, 1980). Additional abnormalities in endemic cretinism include bilateral hearing loss or deaf-mutism and neurological abnormalities such as paralysis (Bleichrodt, Garcia, Rubio, de Escobar, & del Rey, 1987). In a meta-analysis on IQ differences between iodine deficient and control groups, an IQ difference of 14 IQ points has been reported (Bleichrodt & Born, 1994). Iodine deficiency has also been shown to have a detrimental effect on psycho-motor development (Bleichrodt et al., 1987). The supplementation with iodine of individuals with iodine deficiency has been shown to lead to an increase in IQ (van den Briel et al., 2000).

1.6. Prenatal cocaine exposure

Cocaine is a stimulant drug that has powerful effects on the central nervous system. If consumed during pregnancy, cocaine can have adverse effects on the developing brain through alterations of the central monoamine systems as well as through maternal vascular disruptions (Arendt et al., 2004). Prenatal cocaine exposure has been found to have detrimental effects on a wide range of cognitive abilities, such as general knowledge, arithmetic skills, visual–spatial skills (Arendt et al., 2004), attention span (Bandstra, Morrow, Anthony, Accornero, & Fried, 2001), and verbal comprehension (Lewis et al., 2004; Morrow et al., 2004).

1.7. Fetal alcohol syndrome

The term fetal alcohol syndrome (FAS) comprises all neurological, intellectual, and behavioral abnormalities in an individual that can be attributed to the exposure of the toxin alcohol through the use or abuse of alcohol by the individual's mother during pregnancy (Juretko, 2006). Since even relatively small amounts of alcohol can cause devastating effects in the children of mothers with a low neurological resistance to the toxin alcohol, mothers of children with FAS are not necessarily heavy drinkers (Löser, 1995). However, if the neurological or organic tolerance of mother and/or embryo is high, sustained alcohol intake of mothers does not necessarily lead to FAS in their children (Dehaene, 1995).

There are three different degrees of FAS (FAS 1st°, FAS 2nd°, and FAS 3rd°) whereby a higher degree corresponds to more severe damage to the brain. A less severe form of FAS is known as fetal alcohol effects (FAE). Although individuals are diagnosed with FAE and FAS 1st°, FAS 2nd°, or FAS 3rd° based on intellectual and behavioral functioning, there is no clear neurological definition of what constitutes a FAS or FAE. Neuropsychological damage can have different causes and intellectual and behavioral abnormalities can also be caused by environmental influence and heritage. Gottfredson and Deary (2004) suggest that individuals with a lower IQ are more likely to fail to manage the challenges of maintaining good health and are more prone to expose themselves to unhealthy environments and lifestyles. Therefore it is reasonable to assume that heavy-drinking mothers, that are obviously either indifferent to or uninformed about the hurtful effects of chronic or excessive alcohol consumption during pregnancy, have on average a lower IQ, than not alcohol consuming mothers. Since IQ has been found to be highly heritable, deficits in intellectual performance of individuals with FAS or FAE might to some extent be due to genetic predisposition rather than toxic damage caused by alcohol.

1.8. Traumatic brain injury

Traumatic brain injury (TBI) refers to severe damage to the brain resulting from external force. Since severity of damage can vary extremely from case to case, we classify severity of damage following the criteria of the Glasgow Coma Scale (GCS), a widely used instrument to classify TBI. Classification is based on a patient's verbal, motor, and eye movement reactions to various stimuli. Scores of this classification range from 3 to 15, whereby a score higher than or equal to 13 leads to a diagnosis of mild TBI. A score lower than 13 and higher than or equal to 8 leads to a diagnosis of moderate TBI, and a score lower than 8 leads to a diagnosis of severe TBI.

Since TBI also often leads to memory impairment (Dikmen, Machamer, Winn, & Temkin, 1995), a second classification is based on the time a patient suffers from post-traumatic amnesia (PTA). PTA refers to a state of confusion immediately after suffering a TBI. PTA can include retrograde amnesia as well as anterograde amnesia. The former refers to the loss of memories before the accident. For example, a patient might not be able to recall his name, address, or other autobiographic memories. The latter refers to the inability to store memories of events that happened after the injury in memory. Patients who suffer from PTA less than a day are diagnosed with mild TBI. A PTA of one to seven days indicates a moderate TBI. If a patient suffers more than seven days from PTA, a diagnosis of severe TBI is made.

A third classification is based on the time a patient suffered from loss of consciousness (LOC). If a patient had an LOC of less than 30 min, this indicates a mild TBI. LOC of 30 min to 24 h points to a moderate TBI. A LOC longer than 24 h indicates a severe TBI. Since scores on the Glasgow Coma Scale, the duration of post traumatic amnesia, and the duration of loss of consciousness do not necessarily lead to the same classification with regard to severity of TBI, the classification of severity depends to some extent on the judgment of the medical staff in question.

TBI results in brain damage that is most often both focal and diffuse in nature. Focal damage refers to the brain area directly beneath the location where the external force was exposed to the head. Diffuse damage refers to damage to all other areas of the brain following the exposure to an external force. Since TBI can be the result of a hit to virtually any part of the head and diffuse damage can occur throughout the brain we can assume that each TBI is unique to some extent.

Patients who suffered from TBI are often found to have lower IQ scores. Batty, Deary, and Gottfredson (2007) convincingly argue that people with a lower IQ are more likely than people with a higher IQ to suffer from accidents. The data from Roma appear to fit nicely into this pattern. Roma IQ has been estimated to be at least one *SD* below the West-European mean of 100 (see, for instance, Rushton, Čvorović, & Bons, 2007) and the percentage of Roma dying in traffic accidents is very large.

A test of the correlation between TBI and *g* is most informative when the study includes a comparison involving a carefully matched control group. A control group should have an intellectual level comparable to that of the TBI group before the accident. As many studies did not include a control group we also based our comparisons on data from nationally representative samples. This comparison, however, is less meaningful, because the TBI group did not only suffer neurological damage, but might have been less intelligent in the first place. It will be checked whether the two types of comparisons yield comparable outcomes.

1.9. Statistical analyses

We carried out four exploratory meta-analyses correcting only for sampling error. So, bare-bones meta-analytical techniques (Hunter & Schmidt, 1990, 2004) were applied to the r ($g \times d$)s using the software package developed by Schmidt and Le (2004). We chose to use a random-effects model as they generally apply in the large majority of cases (Borenstein, Hedges, Higgins, & Rothstein, 2009). Heterogeneity was assessed by measuring the percentage variance between the data points explained by sampling error. We did not check for publication bias (Rothstein, Sutton, & Borenstein, 2005) because none of the studies tested the method of correlated vectors; moreover, all data points in our meta-analyses took effect sizes (d) and g loadings from separate sources.

In general, *g* loadings were computed by conducting a principal-axis factor analysis on the correlation matrix of a test battery's subtest scores. The subtests' loadings on the first unrotated factor indicate the subtest's loading on *g*. *g* loadings were always matched to the age range of the groups involved in the comparison as close as possible. If the age range of the comparison groups comprised more than one age group of the IQ battery, we computed weighted average *g* loadings of all age groups of the IQ battery that fall within the age range of the comparison groups. Finally, Pearson correlations between d scores of the variables of interest and g loadings were computed.

1.10. General inclusion rules

For studies to be included in a meta-analysis three criteria had to be met: First, in order to obtain a reliable estimate of the true correlation between each of the variables and *g* loadings, the cognitive batteries had to be based on a minimum of six subtests. Second, the IQ test had to be well-validated. Finally, only studies published in English, Dutch, or German were used.

1.11. Choice of SD

When we compute the correlation $d \times g$ difference scores, (d) are computed by subtracting the score of the lower scoring group from the score of the higher scoring group and dividing the result by the best estimate of *SD* available. Our choice of *SDs*, in order of preference, is: First, the *SD* of a national standardization sample; second, the *SD* of a control group; third, a weighted average of the *SD* of the groups involved in the comparison. In cases in which we decide to deviate from this rule, we will state this explicitly.

1.12. Processing data

The data from the individual studies were processed by the second and third author. The third author was a graduate student working on his master thesis, and the second author was his supervisor and a senior intelligence researcher with extensive experience with meta-analyses. When differences of opinion arose they discussed until consensus was reached.

2. Study 1: iodine deficiency

The objective of this analysis is to explore the correlation $d \times g$ between children deficient in iodine and a control group that is not deficient in iodine.

2.1. Method

2.1.1. Searching and screening studies

An excellent and exhaustive meta-analysis of all studies on iodine and its relationship to cognitive development was carried out by Bleichrodt and Born (1994) comprising ten articles, book chapters, and reports; this is all published research on the subject in English-language research journals and books. Some of the studies in the meta-analysis are so specialist that they are extremely difficult to find but professor emeritus Nico Bleichrodt kindly supplied copies of three rare dissertations included in the meta-analysis. However, more than half a dozen requests in ten years to supply copies of the other seven studies did not lead to a response. This search yielded two studies.

2.1.2. Specific criteria for inclusion

Studies to be included in the meta-analysis had to consist of children deficient in iodine and a control group that is not deficient in iodine. The research design eligible was random assignment only. In the study of Bleichrodt et al. (1980) scores of cognitive ability and psychomotor tests were reported. We only included tests of cognitive ability and left psychomotor tests out of the comparison, because they are traditionally not included in IQ batteries.

2.1.3. Computation of score differences between an iodine deficient group and a comparison group

Score differences between an iodine deficient group and a control group (d) were computed by subtracting the mean score of the iodine deficient group of the particular test in question from the mean score of the control group, and then dividing the result by the *SD* of the control group.

2.2. Results

The results of the study on the correlation between g loadings and the score differences between iodine deficient groups and control groups are reported in Table 1. The Table gives data derived from one study, with participants numbering a total of 196. It also lists the reference for the study, the cognitive ability test used, the number of subtests, the correlation between g loadings and d, and the sample size. Correlations $d \times g$ range from substantially negative to substantially positive. Table 2 presents the results of the bare-bones meta-analysis of six data points. It shows the number of correlation coefficients (K), total sample size (N), the weighted mean correlation (mean r) and the standard deviation of the observed correlations (SD_r). The last column presents the percentage of variance explained by sampling errors (%VE). The analysis of all data points yields a weighted mean correlation of .01, with 51.09% of the variance in the observed correlations explained by sampling error.

3. Study 2: prenatal cocaine exposure

In the present study, we explore the correlation $d \times g$ between the magnitude of g loadings and difference scores on IQ battery subtest between children who were exposed to cocaine prenatally and a control/standardized group.

Table 1
Studies of correlations between g loadings and iodine deficiency.

Reference	Test	*N* subtests	*r*	*N*
Bleichrodt et al. (1980) (age range: 6–8 years; control group A)	Various tests of cognitive abilities	6	.37	21
Bleichrodt et al. (1980) (age range: 6–8 years; control group B)	Various tests of cognitive abilities	6	−.33	21
Bleichrodt et al. (1980) (age range: 9–12 years; control group A)	Various tests of cognitive abilities	11	−.03	41
Bleichrodt et al. (1980) (age range: 9–12 years; control group B)	Various tests of cognitive abilities	11	−.14	36
Bleichrodt et al. (1980) (age range: 13–20 years; control group A)	Various tests of cognitive abilities	11	.32	44
Bleichrodt et al. (1980) (age range: 13–20 years; control group B)	Various tests of cognitive abilities	11	−.23	33

Note. N = sample size computed by combining the *N*s of the two groups; r = correlation $d \times g$.

Table 2
Exploratory bare-bones meta-analytical results for correlations between g loadings and iodine deficient/iodine non-deficient score differences.

Variable	*K*	*N*	*rho*	SD_{rho}	%VE
Iodine deficiency	6	196	.01	.17	51.09

Note. Bare-bones meta-analytical results: score differences between an iodine deficient group a control group, and g loadings. K = number of correlations; N = total sample size; mean r = mean weighted correlation; SD_r = standard deviation of observed correlations; %VE = percentage of variance accounted for by sampling error.

3.1. Method

3.1.1. Searching and screening studies

We employed a threefold search strategy to identify studies containing IQ scores of children that were exposed to cocaine prenatally. First, an electronic search for published research using PsycINFO, ERIC, MEDLINE, PiCarta, Academic search premier, Web of science, Google Scholar, and PubMed was conducted. Keywords used were cocaine, prenatal(ly exposed), maternal, gestational, pregnancy, drug use, and crack, combined with the words: cognitive, mental ability, intelligence, IQ, WISC, Wechsler, and combinations of these concepts. Second, the reference lists of significant articles were analyzed in search of additional studies. Last, cited reference searches were conducted using Web of Science to identify the newest articles, citing already included key studies. This search yielded two studies. However, it should be emphasized that the search was quite superficial; most likely there are additional studies to be found with a thorough search.

3.1.2. Specific criteria for inclusion

For a study to be included in the meta-analysis it had to consist of a group of children who were exposed to cocaine prenatally; these were then compared to a control/standardized group. Also, the mean subtest scores had to be lower than the mean scores of a comparison group.

3.1.3. Computation of score differences between a prenatal cocaine exposure group and a control group

Score differences between a prenatal cocaine exposure group and a control group (d) were computed by subtracting the mean score of the prenatal cocaine exposure group from the mean score of the control group, and then dividing the result by the *SD* of the standardized group from the manual of the IQ battery.

3.2. Results

The results of the studies on the correlation between g loadings and the score differences between children exposed to cocaine prenatally and control groups (d) are shown in Table 3. The Table presents data derived from two studies, with participants numbering a total of 215. The correlations are opposite in sign with nearly the same mild magnitude. Table 4 presents the results of the bare-bones meta-analysis of the two data points. The analysis of both data points yields a mean weighted correlation of −.23, with 16.98% of the

Table 3
Studies of correlations between *g* loadings and prenatal cocaine exposure.

Reference	Test	*N* subtests	*r*	*N*
Singer et al. (2004)	WPPSI-R	6	−.31	190
Asanbe and Lockert (2006)	WISC-III	11	.39	25

Note. N = sample size computed by combining the *N*s of the two groups; r = correlation $d \times g$.

variance in the observed correlations explained by sampling errors.

4. Study 3: fetal alcohol syndrome

To explore the correlation $d \times g$ between the magnitude of *g* loadings and IQ subtest scores of individuals who suffered from fetal alcohol syndrome, an analysis was performed on the data from a study on subjects who suffered from FAS. Furthermore, we test the correlation between the magnitude of *g* loadings and IQ subtest scores of individuals with different degrees of severity of FAE/FAS.

4.1. Method

4.1.1. Searching and screening studies

We employed several searches. First, an electronic search for published research using PsycINFO, PiCarta, Academic search premier, Web of science, and PubMed was conducted. The following combinations were used to conduct the searches for studies concerning alcohol: the keywords alcohol, alcoholism, alcoholic, Korsakoff, and fetal alcohol syndrome in combination with the keywords IQ, intelligence, intellectual, cognitive, cognition, Wechsler, WAIS, and WISC. Also, the book by Wechsler (1958) was scanned for suitable studies. Second, the reference lists of significant articles were analyzed in search of additional studies. Last, cited reference searches were conducted using Web of Science to identify the newest articles, citing already included key studies. This search yielded one study, a German-language dissertation containing a lot of detailed information.

4.1.2. Specific criteria for inclusion

Studies to be included in the meta-analysis had to consist of individuals who suffered from fetal alcohol syndrome. Since FAE are lower in severity than FAS 1st° and 2nd° we expected that participants with FAE have a higher full scale IQ than participants with FAS 1st° and 2nd°. However, this is not the case. The FAE group has a Full Scale IQ range of 46 to 117 with a mean of 77, the FAS 1st° group has a Full scale IQ score range from 44 to 132 with a mean of 79, and the FAS 2nd° group has a Full Scale IQ range from 61 to 94 with a mean of 78. Therefore we left comparisons between FAE, FAS 1st°, and FAS 2nd° out of the study.

4.1.3. Computation of score differences between an FAS/FAE group and a control group

Score differences between an FAS/FAE group and a control group (d) were computed by subtracting the mean score of the FASFAE group from the mean score of the control group, and then dividing the result by the *SD* of the control group. To use the subtest scores and the *SD* of a standardized group is

Table 4
Exploratory bare-bones meta-analytical results for correlations between *g* loadings and prenatal cocaine exposed subjects/control subjects score differences.

Variable	*K*	*N*	*rho*	SD_{rho}	%VE
Prenatal cocaine exposure	2	215	−.23	.20	16.98%

Note. Bare-bones meta-analytical results: score differences between a group prenatally exposed to cocaine, control group, and *g* loadings. K = number of correlations; N = total sample size; mean r = mean weighted correlation; SD_r = standard deviation of observed correlations; %VE = percentage of variance accounted for by sampling errors.

also a theoretical option. However, the manual of the IQ battery was not available. *g* loadings of the HAWIK/E-R were not available, so *g* loadings of the HAWIK/E were used instead.

4.1.4. Computation of score differences between a FAS/FAE groups of different severity

Difference scores between FAS 1st°, 2nd°, 3rd°, and FAE are computed by subtracting the FAS/FAE group of higher severity (the ranking from lowest to highest is: FAE, FAS 1st°, 2nd°, and 3rd°) from the FAS/FAE group of lower severity. The difference is divided by the standard deviation of the control group of the FAE/FAS groups. Since there is only one control group for all FAS/FAE conditions, no further computation concerning the scores of the control group is needed. The Full Scale IQ of FAE, FAS 1st°, and 2nd° differed by one IQ point only. We assume that this difference is not large enough to make a meaningful comparison, so we left the comparison between these groups out of the analysis.

4.2. Results

The results of the study on the correlation between *g* loadings and score differences between FAE/FAS and a control group are presented in Table 5. The Table gives data derived from one study, with participants numbering a total of 110. The correlation is positive and small in magnitude. The results of the study on the correlation between *g* loadings and score differences between different degrees of FAE/FAS are presented in Table 6. The Table gives data derived from one study, with participants numbering a total of 125. The correlations are small and positive as well as small and negative in sign. Table 7 presents the results of the bare-bones meta-analysis of the three data points. The analysis of all data points yields a mean weighted correlation of .12, with 83.04% of the variance in the observed correlations explained by sampling errors.

5. Study 4: traumatic brain injury

To explore the correlation $d \times g$ between the magnitude of *g* loadings and IQ subtest scores of individuals who suffered from

Table 5
Studies of correlations between *g* loadings and fetal alcohol syndrome/fetal alcohol effects.

Reference	Test	*N* subtests	*r*	*N*
Juretko (2006)	HAWIE-R/HAWIK-R	11	.16	110

Note. N = sample size; r = correlation $d \times g$. The HAWIE-R is the German version of the WAIS-R. The HAWIK-R is the German version of the WISC-R.

Table 6
Studies of correlations between g loadings and different degrees of fetal alcohol syndrome/fetal alcohol effect.

Reference	Test	N subtests	r	N
Juretko (2006) FAE–FAS III°	HAWIE-R/HAWIK-R	11	−.07	51
Juretko (2006) FAS I°–FAS III°	HAWIE-R/HAWIK-R	11	.19	42
Juretko (2006) FAS II°–FAS III°	HAWIE-R/HAWIK-R	11	.34	32

Note. N = sample size computed by combining the Ns of the two groups; r = correlation $d \times g$. The HAWIE-R is the German version of the WAIS-R. The HAWIK-R is the German version of the WISC-R.

TBI, an exploratory psychometric meta-analysis was performed on a number of studies that reported IQ scores from TBI subjects.

5.1. Method

5.1.1. Searching and screening studies

To identify studies for inclusion in the meta-analysis, both electronic and manual searches were conducted for studies that contained cognitive ability data of TBI. Three methods were used to obtain scores of the traumatic brain injured from published studies for the present meta-analysis. First, an electronic search for published research using PsycINFO, Picarta, Academic search premier, Web of science, and PubMed was conducted. The following combinations were used to conduct the searches: any keyword that contains the word 'traumatic brain injury', or 'brain trauma' in combination with any keyword that contains one of the following words: IQ, g, general mental ability, GMA, cognitive ability, general cognitive ability, intelligence, Wechsler, cognitive ability test.

Second, we browsed the tables of content of several major research journals with a strong focus on the traumatic brain injured: *Brain Injury* 2000–2010, *Applied Neuropsychology* 1999–2010, *Journal of the International Neuropsychological Society* 1997–2010, and *Journal of Neurotrauma* 1993–2010. Third, 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. This procedure resulted in 40 articles and reports on the concurrent topics of traumatic brain injury and mental ability. Eight studies met all criteria for inclusion in the meta-analysis, yielding 14 data points.

5.1.2. Specific criteria for inclusion

For a study to be included in the meta-analysis it had to consist of individuals who suffered from TBI. Also, the mean subtest scores had to be lower than the mean scores of the comparison group.

Table 7
Exploratory bare-bones meta-analytical results for correlations between g loadings and different degrees of fetal alcohol syndrome/fetal alcohol effect.

Variable	K	N	rho	SD_{rho}	%VE
FAE/FAS	3	125	.12	.07	83.04%

Note. Bare-bones meta-analytical results: score differences between different degrees of FAE/FAS, and g loadings. K = number of correlations; N = total sample size; mean r = mean weighted correlation; SD_r = standard deviation of observed correlations; %VE = percentage of variance accounted for by sampling errors.

5.1.3. Computation of score differences between a TBI group and a standardized/control group

Score differences between a TBI group and a standardized group, or a control group (d) were computed by subtracting the mean score of the TBI group of the particular test in question from the mean score of the standardization group/control group, and then dividing the result by the *SD* of the standardization group. The standardization group scores were obtained from the manual of the IQ battery.

5.2. Results

The results of the studies on the correlation between g loadings and the score differences between a TBI group and control/standardized groups (d) are shown in Table 8. The Table reports data derived from nine studies, with participants numbering a total of 629. The correlations show no clear pattern with regard to magnitude or sign. Table 9 presents the results of the bare-bones meta-analysis of the 14 data points. The analysis of 14 data points yields an mean weighted correlation of .07, with 35.43% of the variance in the observed correlations explained by sampling error. Sample sizes were highly comparable, which most likely led to much lower %VE.

6. General conclusion

Table 10 gives an overview of all exploratory meta-analytical correlations between g loadings and differences on four variables we obtained in our studies. It is clear that all mean correlations are very close to zero. The unweighted average is exactly 0.00.

An important lesson we can learn from meta-analysis is that a collection of studies on the same topic will *not* have highly similar outcomes, but due to sampling error and other

Table 8
Studies of correlations between g loadings and traumatic brain injury.

Reference	Test	N_s	r	N
Tremont, Mittenberg, and Miller (1999)	WISC-III	12	−.11	30
Langeluddecke and Lucas (2005) (moderate TBI-Control group)	WMS-III	11	.15	44
Langeluddecke and Lucas (2005) (severe TBI-Control group)	WMS-III	11	.29	86
Langeluddecke and Lucas (2005) (extremely severe-Control group)	WMS-III	11	.24	50
Langeluddecke and Lucas (2003) (moderate TBI-Control group)	WAIS-III	13	−.29	35
Langeluddecke and Lucas (2003) (severe TBI-Control group)	WAIS-III	13	−.40	74
Langeluddecke and Lucas (2003) (extremely severe-Control group)	WAIS-III	13	−.16	41
Demakis et al. (2001)	WAIS-R	11	−.10	48
Cattelani, Lombarbi, Brianti, and Mazzucchi (1998)	WISC	10	−.33	20
Cattelani et al. (1998)	WAIS	11	−.07	20
Blake, Fichtenberg, and Abeare (2009)	WAIS-III	11	.15	57
Bittner and Crowe (2007a)	WAIS-III	11	−.09	23
Bittner and Crowe (2007b)	WAIS-III	11	−.06	40
Allen, Thaler, Donohue, and Mayfield (2010)	WISC-IV	10	−.43	61

Note. N = sample size computed by combining the Ns of the two groups; r = correlation $d \times g$; N_s = number of subtests in IQ battery.

Table 9

Exploratory Bare-bones meta-analytical results for correlations between *g* loadings and TBI groups and control/standardized groups score differences.

Predictor	*K*	*N*	*rho*	SD_{rho}	%VE
Traumatic brain injury	14	629	−.07	.20	35.43%

Note. Bare-bones meta-analytical results: Score differences between a TBI group, control group, and *g* loadings. K = number of correlations; N = total sample size; mean r = mean weighted correlation; SD_r = standard deviation of observed correlations; %VE = percentage of variance accounted for by sampling errors.

statistical artifacts there will always be variability (see Hunter & Schmidt, 2004). Correcting for statistical artifacts will strengthen the conclusions from our meta-analyses. In their meta-analysis of whether vectors of Flynn effect gains correlate with vectors of *g* loadings, te Nijenhuis and van der Flier (2013) discuss the influence of various statistical artifacts: sampling error, reliability of the *g* vector, reliability of the *d* vector, restriction of range in the vector of *g* loadings, and imperfectly measuring the construct of *g*.

To these could be added the unreliability of score differences that are used in the present meta-analyses. Jensen (1998, pp. 380–383) suggested that statistical artifacts strongly influence the outcomes of tests of Spearman's hypotheses and he makes plausible that correcting for these statistical artifacts will strongly increase the correlation between *g* loadings and effects (*d*). The two meta-analyses by te Nijenhuis and co-authors on, respectively, test-retest effects (te Nijenhuis, van Vianen, & van der Flier, 2007) and Flynn effect gains (te Nijenhuis & van der Flier, 2013) show that the weighted average correlation of all the studies in the meta-analyses differs strongly from the meta-analytical correlation (*rho*) that is corrected for several statistical artifacts. The meta-analytical *rho* may easily be 25% higher than the average observed correlation.

A fundamental point concerning the present paper is that correcting for psychometric artifacts will almost guarantee that our conclusions will be strengthened: correcting the means will have negligible effects, and the percentage variance explained between data points will increase strongly. When an observed correlation of .05 is corrected upwards by 25% it still remains only .06, a negligible effect; as the observed correlation is very small even a quite substantial correction will have almost no effect. Our outcomes suggest the true value of all the studies on different topics combined is zero or very close to zero, so correcting for statistical artifacts will not change the conclusions based on the mean one bit.

Three of our meta-analyses are based on only two to six studies, yet two of them already explain a substantial amount of variance. The fourth study is based on 14 data points, but explains only 35% of the variance, but this is most likely due to highly similar sample sizes, which can strongly reduce the amount of variance explained. Correcting for the statistical artifacts the way te Nijenhuis and co-authors do in all likelihood will strongly increase the amount of variance explained. This will mean that the reliability interval around the meta-analytical correlation will be quite small, an indication that there is little variability in the data points.

Table 10

Correlations of *g* loadings with biological–environmental variables.

Variable	*K*	*N*	Mean *r*
Iodine deficiency	6	196	.01
Prenatal cocaine exposure	2	215	−.23
Fetal alcohol syndrome/fetal alcohol effects	1	110	.16
Degree of fetal alcohol syndrome	3	125	.12
Traumatic brain injury	14	629	−.07

Note. K = number of correlations; N = total sample size or harmonic N; r = mean observed correlation (sample size weighted).

Moreover, the values of the mean weighted *r* of the various meta-analyses are clearly in line with each other. Our exploratory meta-analysis on the fetal alcohol syndrome is based on just three studies, but already has a highly reliable value of mean *r*, as reflected in the very high percentage of variance explained in the data points. Our exploratory meta-analysis on cocaine exposure is the weakest as it is based on only two data points, which is the minimum for a meta-analysis. But the value of mean *r* is highly similar to the values of mean *r* from the other analyses, two of which involve a number of studies. This increases the plausibility of the findings. We conclude that the outcomes of our four exploratory meta-analyses appear to be quite robust.

7. Discussion

All four meta-analyses of the score gaps between those with these afflictions and normal subjects subtest by subtest creates a hierarchy that has no clear correlation with the subtest *g*-loading hierarchy. However, their Full Scale IQ deficits are clear and substantial. Therefore, we are faced with a choice: how do we describe those IQ deficits? We believe that they would be accurately described as intelligence differences. However, let us set aside the word "intelligence" and focus on whether they are cognitively significant.

They are significant in negatively affecting the ability of the afflicted to solve cognitively demanding problems. The fact that we regard them as such is shown by the fact that we attempt to help these people improve their problem-solving skills. Our target, if possible, it to help them improve so that they will match normal subjects on all ten subtests. Because their deficits tend to very without regard to the *g* loading of the subtests, the gains that erase these deficits would not correlate with the *g* loadings of the subtests. Nonetheless, raising all their scores to the average level would be a therapeutic triumph.

Flynn (in press) suggests that we should add a third term to the "intelligence difference" or "hollow difference" dichotomy, namely, "cognitively significant difference". This would solve the two paradoxes. A trend that shows blacks gaining on whites over time would be cognitively significant and capable of diminishing their real-world intelligence gap (Flynn, 2008). IQ trends over time would be cognitively significant and constitute a cognitive advantage of this generation over the last. The rationale for the new term is that there is a radical difference between: comparing the cognitive environments of generations separated by time; and comparing the intelligence differences between individuals (and sometimes groups with similar environments) that IQ tests are designed to measure.

Take two people at a given place and time sharing the same cognitive environment (two brothers in the same home). If one accesses that environment better than the other, it makes sense to say he has the better mind. Moreover, he may outstrip his brother in accord with cognitive complexity. The less able brother may not fall far behind for simple cognitive skills but fall further behind for complex ones. After all, they live in a shared cognitive environment: both are subject to hot-house parenting, both will enjoy much the same amount of schooling, both have modern habits of mine, and so forth.

Contrast this with comparing two people, the first in 1900 when hot-house parenting did not exist, when the median year of schooling was closer to 6 years than 13 years (which much restricted their vocabularies and general information), when schools taught rote learning rather than logical analysis (Genovese, 2002), and, most of all, when people were focused on the concrete world rather than on classifying, taking the hypothetical seriously, and using logic on generalizations. Clearly the difference between these people is not what IQ tests were designed to measure. If one takes IQ scores literally as intelligence differences, the people of 1900 had a mean IQ of 70 and bordered on mental retardation. They were not mentally retarded, of course. There is a big difference between people who were not exposed to the modern cognitive environment and people who are exposed to it but cannot access it (Flynn, 2013a).

Therefore, we will not say that the last generation was less intelligent than we are, but we will not deny that there is a significant cognitive difference. Today we can simply solve a much wider range of cognitively complex problems than our ancestors could, whether we are schooling, working, or talking (the person with the larger vocabulary has absorbed the concepts that lie behind the meaning of words and can now convey them). Flynn (2009) has used the analogy of a marksmanship test designed to measure steadiness of hand, keenness of eye, and concentration between people all of whom were shooting a rifle. Then someone comes along whose environment has handed him a machine gun. The fact that he gets far more bulls eyes hardly shows that he is superior for the traits the test was designed to measure. However, it makes a significant difference in terms of solving the problem of how many people he can kill.

On one level, IQ scores over time reflect a large difference in the cognitive environments of two generations and thereby offer a measure in terms of what proportion of people could perform certain cognitive tasks. But strictly speaking, to confuse this with IQ scores as measuring intelligence traits is a perversion. When IQ tests reflect who has better accessed a relatively homogeneous cognitive environment, they signal an intelligence difference. When they measure who lived at a time that afforded a better cognitive environment, they are measuring something else — albeit something significant. People of previous generations were cognitively different but they were not dumb. As for people today “who have been rendered atypical by some peculiar affliction and cannot fully access the current environment”, why not just use *that phrase* to say what they are? Adding the label “dumb” adds no cognitive content.

So a division of labor is proposed. The concept of *g* and its attendant label of intelligence will be used to measure *individual differences* within a generation (with certain exceptions); and the concept of a shifting cognitive environment and its attendant concept of cognitive progress will be used to assess *generational differences* over time. Cognitive history will identify the new complex problems we have learned to solve over time but it will also be clear about what kind of cognitive progress they entail. This should preserve honor on both sides. The reason the two kinds of difference have been confused is that they are kissing kin. Both have to do with enhanced ability to solve cognitively complex problems. In one case, we identify individuals with enhanced skills that correlate with the hierarchy of cognitive complexity. In the other case, we identify the enhanced skills of a new generation that can better solve problems of cognitive complexity but not in rank order.

7.1. Limitations

It could be argued that due to the many small and quite small samples in our exploratory meta-analysis its power is quite modest, which would forbid drawing strong conclusions. Indeed, three of the four meta-analyses have total *N*s between 125 and 215. However, we are of the opinion that our data implies strong conclusions. First, the central outcome in our meta-analyses is a mean weighted correlation, and means are known to be reliable measures. Secondly, the mean *r*s from the four exploratory meta-analyses are all quite similar in that they are all quite close to zero. Thirdly, while the meta-analysis on iodine has only six studies, which is not large, it is still able to explain more than half of the variance in these six data points, which means that the value of the mean *r* is relatively stable. The meta-analysis on fetal alcohol syndrome is even smaller with only three studies, but a very large amount of variance is explained by sampling error, which means that the value of the mean *r* is highly stable. This is an impressive finding for meta-analyses, the small number of samples and generally small studies notwithstanding. On the other hand, while the meta-analysis on traumatic brain injury has a good 14 data points and a quite large total *N*, sampling error explains only a quite modest amount of variance in the data points, which would imply that the value of the mean *r* is unreliable. However, as we argued above, this is in all likelihood caused by the highly similar sample sizes. In effect, it could be argued that while three of these studies have modest power, the four of them combined are impressive in terms of similarity of the values of the mean *r* of the four meta-analyses and the amount of variance explained by sampling error only.

Our findings suggest that when combining data from various studies to test the method of correlated vectors it is not always necessary to carry out a full-fledged meta-analysis and collect all available studies, but rather a good exploratory meta-analysis can also yield quite stable findings. In sum, while power analyses might suggest a priori that the findings from exploratory meta-analyses with a total *N* between 125 and 215 will give unstable results, our empirical tests suggest the contrary.

References marked with an asterisk were included into the meta-analyses and exploratory studies.

References

*Allen, D. N., Thaler, N. S., Donohue, B., & Mayfield, J. (2010). WISC-4 profiles in children with traumatic brain injury: Similarities to and differences from the WISC-3. *Psychological Assessment*, *22*, 57–64.

Arendt, R. E., Short, E. J., Singer, L. T., Minnes, S., Hewitt, J., Flynn, S., et al. (2004). Children prenatally exposed to cocaine: Developmental outcomes and environmental risks at seven years of age. *Developmental and Behavioral Pediatrics*, *25*, 83–90.

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.

*Asanbe, C. B., & Lockert, E. (2006). Cognitive abilities of African-American children with prenatal cocaine/polydrug exposure. *Journal of Health Care for the Poor and Underserved*, *17*, 400–412.

Bandstra, E. S., Morrow, C. E., Anthony, J. C., Accornero, V. H., & Fried, P. A. (2001). Longitudinal investigation of task persistence and sustained attention in children with prenatal cocaine exposure. *Neurotoxicology and Teratology*, *23*, 545–559.

Batty, G. D., Deary, I. J., & Gottfredson, L. S. (2007). Premorbid (early life) IQ and later mortality risk: Systematic review. *Annals of Epidemiology*, *17*, 278–288.

*Bittner, R. M., & Crowe, S. F. (2007). The relationship between working memory, processing speed, verbal comprehension and FAS performance following traumatic brain injury. *Brain Injury*, *21*, 709–719.

*Blake, T. M., Fichtenberg, N. L., & Abeare, C. A. (2009). Clinical utility of demographically corrected WAIS-3 subtest scores after traumatic brain injury. *The Clinical Neuropsychologist*, *23*, 373–384.

Bleichrodt, N., & Born, M. Ph. (1994). A metaanalysis of research on iodine and its relationship to cognitive development. In J. B. Stanbury (Ed.), *The damaged brain of iodine deficiency* (pp. 195–200). New York: Cognizant Communication Cooperation.

*Bleichrodt, N., Drenth, P. J. D., & Querido, A. (1980). Effects of iodine deficiency on mental and psychomotor abilities. *American Journal of Physical Anthropology*, *53*, 55–67.

*Bleichrodt, N., Garcia, I., Rubio, C., de Escobar, Morreale G., & del Rey, Escobar F. (1987). Developmental disorders associated with severe iodine deficiency. In B. S. Hetzel, J. T. Dunn, & J. B. Stanbury (Eds.), *The prevention and control of iodine deficiency disorders* (pp. 65–84). : Elsevier Science Publishers.

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis.* UK, Chichester: Wiley.

*Cattelani, R., Lombarbi, F., Brianti, R., & Mazzucchi, A. (1998). Traumatic brain injury in childhood: Intellectual, behavioural and social outcome into adulthood. *Brain Injury*, *12*, 283–296.

Coyle, T. R., & Pillow, D. R. (2008). SAT and ACT predict college GPA after removing *g*. *Intelligence*, *36*, 719–729.

Dehaene, P. (1995). *La grossesse et l'alcool [Pregnancy and alcohol].* Paris: Presses Universitaires de France.

*Demakis, G. J., Sweet, J. J., Sawyer, T. P., Moulthrop, M., Nies, K., & Clingerman, S. (2001). Discrepancy between predicted and obtained WAIS-R IQ scores discriminates between traumatic brain injury and insufficient effort. *Psychological Assessment*, *13*, 240–248.

Dikmen, S. S., Machamer, J. E., Winn, H. R., & Temkin, N. R. (1995). Neuropsychological outcome at 1-year post head injury. *Neuropsychology*, *9*, 80–90.

Flynn, J. R. (2008). *Where have all the liberals gone? Race, class, and ideals in America.* Cambridge: Cambridge University Press.

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

Flynn, J. R. (2013). *Intelligence and human progress.* London: Elsevier.

Flynn, J. R. (2013). The "Flynn Effect" and the Flynn paradox. *Intelligence*, *41*, 851–857.

Flynn, J. R. (2014). The march of reason: What was hidden in our genes. In S. Goldstein, J. Naglieri, & D. Princiotta (Eds.), *Handbook of intelligence: Evolutionary theory, historical perspective, and current concepts*. New York: Springer (in press).

Genovese, J. E. (2002). Cognitive skills valued by educators: Historic content analysis of testing in Ohio. *Journal of Educational Research*, *96*, 101–114.

Gottfredson, L. S., & Deary, I. J. (2004). Intelligence predicts health and longevity, but why? *Current Directions in Psychological Science*, *13*, 1–4.

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

Hunter, J. E., & Schmidt, F. L. (2004). *Methods of meta-analysis* (2nd ed.). London: Sage.

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

*Juretko, K. (2006). *Das Muster kognitiver Funktionstörungen bei Patienten mit fetalem Alkoholsymdrom und fetalen Alkoholeffekten; Schwerpunkt: Die Intelligenz [The pattern of problems in cognitive functioning in patients with fetal alcohol syndrom and fetal alcohol effect with a focus on intelligence].* Germany, Münster: Westfälischen Wilhelms-Universität Münster.

*Langeluddecke, P. M., & Lucas, S. K. (2003). Wechsler Adult Intelligence Scale — Third edition findings in relation to severity of brain injury in litigants. *The Clinical Neuropsychologist*, *17*, 273–284.

*Langeluddecke, P. M., & Lucas, S. K. (2005). WMS-3 findings in litigants following moderate to extremely severe brain trauma. *Journal of Clinical and Experimental Neuropsychology*, *27*, 576–590.

Lewis, B. A., Singer, L. T., Short, E. J., Minnes, S., Arendt, R., Weishampel, P., et al. (2004). Four-year language outcomes of children exposed to cocaine in utero. *Neurotoxicology and Teratology*, *26*, 617–627.

Löser, H. (1995). *Alkoholembryopathie und Alkoholeffekte [Alcohol embryopathy and the effects of alcohol].* Germany, Stuttgart: Fischer-Verlag.

Morrow, C. E., Vogel, A. L., Anthony, J. C., Ofir, A. Y., Dausa, A. T., & Bandstra, E. S. (2004). Expressive and receptive language functioning in preschool children with prenatal cocaine exposure. *Journal of Pediatric Psychology*, *29*, 543–554.

Rindermann, H., & Thompson, J. (2013). Ability rise in NAEP and narrowing ethnic gaps? *Intelligence*, *41*, 821–831.

Ritchie, S. J., Bates, T. C., Der, G., Starr, J. M., & Deary, I. J. (2014). Education is associated with later-life IQ scores, but not with the speed of cognitive mechanisms. *Psychology & Aging* (in press).

Rothstein, H. R., Sutton, A. J., & Borenstein, M. (2005). *Publication bias in meta-analysis: Prevention, assessment, and adjustments.* UK, Chichester: Wiley.

Rushton, J. P., Čvorović, 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. (2006). The totality of available evidence shows race–IQ gap still remains. *Psychological Science*, *17*, 921–922.

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

*Singer, L. T., Minnes, S., Short, E., Arendt, R., Farkas, K., Lewis, B., et al. (2004). Cognitive outcomes of preschool children with prenatal cocaine exposure. *Journal of the American Medical Association*, *291*, 2448–2455.

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

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

*Tremont, G., Mittenberg, W., & Miller, L. J. (1999). Acute intellectual effects of pediatric head trauma. *Child Neuropsychology*, *5*, 104–114.

*van den Briel, T., West, C. E., Bleichrodt, N., van de Vijver, F. J. R., Hautvast, J. G. A. J., & Ategbo, E. A. (2000). Improved iodine status is associated with improved mental performance of schoolchildren in Benin. *American Journal of Clinical Nutrition*, *72*, 1179–1185.

Wechsler, D. (1958). *The measurement and appraisal of adult intelligence.* Baltimore, MD: Williams & Wilkins.

Woodley, M. A. (2012). A life history model of the Lynn–Flynn effect. *Personality and Individual Differences*, *53*, 152–156.

Woodley, M. A. (2012). The social and scientific temporal correlates of genotypic intelligence and the Flynn effect. *Intelligence*, *40*, 189–204.

Woodley, M. A., Figueredo, A. J., Ross, K. C., & Brown, S. D. (2013). Four successful tests of the cognitive differentiation–integration effort hypothesis. *Intelligence*, *41*, 832–842.