Two psychometric meta-analyses and three exploratory meta-analyses on *g* loadings and IQ scores: The relation of giftedness, mental retardation, alcohol and cocaine abuse, and depression with general intelligence

Rosina van Bloois & Lise-Lotte Geutjes

Master thesis
University of Amsterdam
Work and Organizational Psychology
Supervisors: Dr. Ing. I.E. de Pater and Dr. Jan te Nijenhuis
May 2009

Acknowledgement

We would like to thank Dr. Jan te Nijenhuis, Dr. I.E. de Pater, Dr. Mariët van der Molen, Dr. Albert Ponsioen, and Dr. Katherine Koenig for their enthusiastic support and helpful guidance on our project.

Table of contents

Abstract

Many studies have shown that IQ test scores can increase and decrease over time as a result of treatment or educational interventions. We focused on the question of the extent to which these scores were linked to the *g* factor, that is, to what extent were increases or decreases in test scores also increases or decreases in *g*. Based on a review of a large number of highly varied empirical findings we hypothesize that with regard to correlations between a vector of *g* loadings and a second vector a value of +1 implies that variation in scores on the second variable is caused by biological factors, a value of –1 implies that variation is caused by non-biological factors, and a value close to zero implies that variation is caused by both biological and non-biological factors having relatively comparable effects. Two psychometric meta-analyses and three exploratory studies were carried out using psychometric meta-analytical techniques. We predicted a strong positive correlation between gifted and mentally retarded persons on the one hand, and the vector of *g* loadings on the other hand; a strong positive correlation between alcohol abuse and cocaine abuse on the one hand, and the vector of *g* loadings on the other hand; a weak positive correlation between depression and the vector of *g* loadings.

Confirming our hypotheses, the two meta-analyses showed strong positive values, with true correlations for giftedness and mental retardation of, respectively, 1.01 (total $N = 4823$) and .74 (total $N = 2729$). The secular increase in IQ scores (Flynn effect) was tested as moderator among both the gifted and the mental retarded, but almost no evidence was found. Further, creating a homogeneous sample of educable mental retarded and the percentage of Blacks in the MR population did not moderate the value of the true correlation for mental retardates. Post-hoc analyses of the Wechsler subtest pattern for mental retardation showed a strong outlier, namely much lower scores on the Digit Span than predicted by its *g* loading, partially explaining why r (g x d) is smaller than +1. Among lower-IQ score groups, Digit Span is a good predictor of self-sufficiency.

Contrary to our hypotheses, the exploratory analyses of alcohol and cocaine abuse showed uncorrected correlations of, respectively, -.04 (total $N = 211$), and -.04 (total $N = 158$). However, the exploratory analyses of children prenatally exposed to alcohol did show a strong uncorrected correlation of .68, which is highly supportive of the hypothesis. Consistent with our hypothesis, exploratory analyses of depression showed uncorrected correlations of .12 (total $N = 268$).Taken together, the findings of giftedness and mental retardation increase the likelihood of the hypothesized link between *g* loadings and a dimension of biological causation. The negative findings for alcohol and cocaine seemed to be due to the stronger effects on nonverbal abilities

that resulted in negative correlations with *g*. Limitations of the theory and implications for interventions aimed at increasing *g* are discussed.

Introduction

IQ test scores is by far the best predictor of workplace performance and school achievement. As a result, IQ tests are effective instruments for selection and placement. High IQ scores define giftedness and in large part determine acceptance in the most selective schools and jobs. Low IQ scores define mental retardation and largely determine who should receive special education. So, having a low IQ score or a high IQ score can result in dramatically different opportunities and chances for individuals. If intelligence can be improved it would result in fewer people being diagnosed as mentally retarded and more people being considered gifted. Consequently, this paper focuses on the question whether intelligence can be improved.

Many studies have investigated whether IQ scores can increase or decrease as a result of treatment. An increase of IQ has been shown by van IJzendoorn, Juffer, and Poelhuis (2005). A meta-analysis of adopted children reported that they scored higher on IQ tests than their nonadopted siblings or than peers who remained in institutional care. Several meta-analyses of Headstart studies showed that children in the program outscored children in control groups (Caruso, Taylor, & Detterman, 1982; Ramey, Bryant, & Suarez, 1985; Nelson, Westhues, & MacLeod, 2003). Many studies have also demonstrated that IQ scores have been increasing over the last half century. This secular increase in test score, termed Flynn effect, has been reported in countries on all continents (Flynn, 2006).

However, many studies have also demonstrated that IQ test scores can decrease. Infectious diseases have been shown to lower IQ scores (Alcock & Bundy, 2001). Severe malnutrition leads to poorer intellectual development and school achievement (e.g., see Grantham-McGregor, Ani, & Fernald, 2001; Nwuga, 1977; Richardson, Birch, & Hertzig, 1973). High levels of hookworm, whipworm, and roundworm have been shown to impair mental performance (Watkins & Pollitt, 1997). However, lower IQ scores can also be the result of congenital or hereditary condition. Schull and Neel (1965) demonstrated that the offspring of consanguineous marriages have lower average IQ scores than non-inbred peers in the same community and age cohort.

The increase or decrease of IQ test scores implies that intelligence is a variable that can be manipulated or trained. An essential research question is to what extent such increases or decreases of IQ test scores represent a permanent increase or decrease in general mental ability (the *g* factor). In this paper we examined the correlations between a number of variables and *g* loadings by carrying out two psychometric meta-analyses on: giftedness and mental retardation. In addition, we collected studies on alcohol and cocaine abuse and on depression, and conducted

exploratory psychometric meta-analysis. The collective results of these methods allowed us to assess which effects are associated with a true increase in general mental ability and which effects produced only "hollow" gains in test scores.

But why does g matter?

A difficulty in understanding the value of IQ tests is the fact that people have trouble understanding the relationship between the content of several IQ tests and the demands of everyday life. For example, it is not obvious in what way solving arithmetic analogies is related to the cognitive capacities we use in daily life. Gottfredson (1997) shed light on this question by pointing out that *g* is the ability to deal with cognitive complexity. All life tasks vary strongly in their complexity, and subsequently, the advantage of having a high general mental ability is greater in some situations in life, but smaller in other situations. Therefore, complexity can be seen as the ‘active ingredient’ of the *g* factor. Any kind of item content ─ words, numbers, figures, pictures, symbols, blocks, mazes, and so on ─ can be used to create *g*-loaded tests and test items varying in levels of complexity. Differences in *g* loadings of tests reflect the amount of mental manipulation they require.

IQ scores in the Hierarchical Intelligence Model

Jensen (1998) hypothesized that scores on IQ batteries are best described by hierarchical intelligence models, such as Carroll’s (1993) three-stratum hierarchical factor model of cognitive abilities. At the highest level of the hierarchy (stratum III) is general intelligence or *g*. One level lower (stratum II) is occupied by the broad abilities of Fluid Intelligence, Crystallized Intelligence, General Memory and Learning, Broad Visual Perception, Broad Auditory Perception, Broad Retrieval Ability, and Broad Cognitive Speediness or General Psychomotor Speed. One level lower still (stratum I) comprises the narrow abilities, including Sequential Reasoning, Quantitative Reasoning, Verbal Abilities, Memory Span, Visualization, and Perceptual Speed. The lowest level of the hierarchy consists of large numbers of specific tests and subtests. Some tests, despite seemingly very different formats, have been demonstrated empirically to cluster into one narrow ability (Carroll, 1993).

A well-established empirical finding—the manifold of positive correlations among measures of various mental abilities—is putative evidence of a general factor in all of the measured abilities. The method of factor analysis makes it possible to determine the degree to which each of the variables is correlated (or loaded) with the factor that is common to all the variables in the analysis. Spearman termed this *g* to represent a general factor that is manifested in individual differences on all mental tests, regardless of content (Jensen, 1998, p. 18). Spearman’s *g* is best understood as a measure of cognitive complexity (Gottfredson, 1997), and is

usually defines operationally as the loading on the first unrotated factor in a principal-axis factor analysis of a varied set of IQ tests (Jensen & Weng, 1994). Thus, tests demanding higher cognitive complexity are high on *g* (have high *g* loadings), and tests demanding lower cognitive complexity are low on *g* (have low *g* loadings).

As noted, an interesting question is to what extent general mental ability can be increased or decreased and which effects will produce only "hollow" gains in test score. Jensen (1998) argued that training effects are most clearly manifested at the lowest level of Carroll's hierarchy, particularly on specific tests that most resemble the trained skills. One hierarchical level higher, the training effect is still evident for certain narrow abilities, depending on the nature of the training. However, the gain virtually disappears at the level of broad abilities and is altogether undetectable at the highest level, *g*. This implies that the transfer of training effects is strongly limited to tests or tasks all of which are dominated by one particular narrow skill or ability. Hence, there is virtually no transfer to tasks dominated by different narrow abilities, and none at the level of *g* itself. Thus any increase in narrow abilities or test-specific ability is independent of *g*.

Test-specific ability is defined as that part of a given test's true-score variance that is not common to any other test; i.e., it lacks the power to predict performance on any other tasks except those that are highly similar. Gains on test specificities are therefore not generalizable, but rather are 'empty 'or 'hollow'. Only the *g* component is highly generalizable. Jensen (1998, ch. 10) gives various examples of empty score gains, including a detailed analysis of the Milwaukee Project, which claimed very large increases in IQ score. However, Jensen's analysis indicates that there was no increase in *g*. Another example of empty score gains is given by Christian, Bachnan, and Morrison (2001) who state that increases due to schooling show very little transfer across domains.

The Method of Correlated Vectors

The question still remains as to which variables are associated with an increase or decrease in general intelligence. The method of correlated vectors is a means of identifying variables that are associated with Spearman's *g*, the general factor of mental ability. This method involves calculating the correlation between: (a) the column vector of the *g* factor loadings of the subtests of an intelligence test or similar battery, and (b) the column vector of the relation of each of those same subtests with the variable in question. When the latter variable is dichotomous, the relations are usually calculated in terms of an effect size statistic. When the latter variable is continuous (or nearly so), the relations are usually calculated in terms of a correlation coefficient (Ashton & Lee, 2005).

Although little has been written about the distribution of the values of the correlation between a *g* vector and a second vector, a clear picture has emerged from the individual studies and meta-analyses (see Table 1). First of all, applying the method of correlated vectors to biological variables such as head size, brain volume, brain's gray matter, brain's evoked potential, brain glucose metabolic rate, peripheral nerve conduction velocity, brain pH, body symmetry, inbreeding, and hybrid vigor results in high positive correlations. After applying the statistical corrections typically carried out in psychometric meta-analysis (Hunter & Schmidt, 1990) it is not unlikely that the true correlation between *g* loadings and most of these biological variables approaches +1.00.

Second, a psychometric meta-analysis of correlations between vectors of *g* loadings and vectors of test-retest score gains based on a very large sample yielded a true correlation of -1.00 (te Nijenhuis, van Vianen, & van der Flier, 2007): the non-biological variable showing a perfect negative correlation with *g*. An exploratory study on learning potential in South-Africa (te Nijenhuis et al., 2007) reported a correlation of -.39 between score gains and the magnitude of *g* loadings of the items of Raven's Progressive Matrices. Correction for unreliability would probably yield a correlation of about -.80. Braden's (1989) study of the IQ scores of the non-genetic deaf found strong negative correlations between *g* loadings and the score difference between hearing and deaf groups. These negative correlations argue that the score gains or score differences are "hollow", that is, they are non-biological and do not represent true gains (or differences) in *g*.

Third, Jensen (1998, p. 320-321) was the first to ask the question whether the secular increase in test scores (the Flynn effect) is also correlated with *g* loadings. He reported data on four test batteries and found that these test's *g* loadings were not highly correlated with the amount of change. Subsequently, seventeen studies have examined whether secular trends are related to *g* (e.g., see Colom, Juan-Espinosa, & García, 2001; Flynn 1999ab, 2000; Must, Must, & Raudik, 2003; te Nijenhuis, & van der Flier, 2007; Wicherts et al., 2004) and have produced conflicting results. A psychometric meta-analysis based on a very large sample which reported correlations between *g* loadings and standardized score gains on all studies having seven or more subtests yielded a true correlation of -.33; correction for statistical artifacts explained all the variance in the data points (te Nijenhuis & van der Flier, submitted, see Table 1). This correlation implies that less than half of the Flynn effect constitutes a real gain on *g*: the gain on non-*g* abilities being stronger than the gain on *g* suggesting that biological causes are less of an influence than non-biological causes. Therefore it is likely that the secular IQ gains only partially reflect a functional increase of real-life problem solving (general mental) ability.

These findings suggest the following theory: If the correlation between the *g* vector and a second vector is close to +1.00, variation in scores on the variable is caused by biological factors; If the correlation is close to -1.00, the variation in scores is caused by non-biological factors; If the correlation is close to 0.00, the variation is caused by both biological and non-biological factors. The last possibility also suggests that *g* and non-*g* skills play equal roles in effects.

Technically, a correlation of +1 or -1 implies that the variation in scores on specific variables can be perfectly predicted using *g* loadings, while a correlation of 0 implies that *g* loadings are useless for predicting the variation in the scores on the variables. However, we hypothesize that when these correlations are obtained by using the method of correlated factors additional information can be gleaned, namely the degree to which variation in scores on the specific variables is caused by biological and non-biological factors. Not only do very high positive or negative correlations yield important information, but correlations close to zero do as well.

Table 1

Various Studies on the Correlation Between a g Vector and a Second Vector

study	variable	*r*	*N*
Biological variables			
Jensen (1994)	head size	.64	286
Wickett, Vernon, & Lee (1994)	brain volume	.65	80
Schoenemann (1997)	brain volume	.51	72
	brain's cortical gray matter	.66	72
Schafer (1985)	brain's evoked potential habituation index	.77	52
Eysenck & Barrett (1985)	brain's averaged evoked potential	.95	219
Haier, Siegel, Tang, Abel, & Buchsbaum (1992)	brain's glucose metabolic rate	.79	8
Vernon (1992, 1993)	peripheral nerve conduction velocity	.44	85
Rae et al. (1996)	intercellular brain pH	.63	42
Colom, Jung, & Haier (2006)	brain gray matter	.82	23
	brain gray matter	.36	25
Lee et al. (2006)	brain activity	.61	36
Prokosch, Yeo, & Miller (2005)	body symmetry	.98	78
Schull & Neel (1965)	inbreeding	.79	865
Badarudozza & Afzal (1993)	inbreeding	.83	50
Nagoshi & Johnson (1986)	hybrid vigor	.52	2,096
te Nijenhuis & Jongeneel-Grimen (2007)	heritability coefficients	**1.01**[1]	2,590
Mix biological/non-biological variables			
te Nijenhuis & van der Flier (submitted)	Flynn effect gains	**-.33**[1]	12,732
Non-biological variables			
te Nijenhuis et al. (2007)	test-retest gains	**-1.00**[1]	26,990
	learning potential training gains	-.39	95
te Nijenhuis & Jongeneel-Grimen (2007)	Headstart gains	**-.80**[1]	602
te Nijenhuis & Jongeneel-Grimen (2007)	adoption gains	**-1.06**[1]	664
Braden (1989)	IQ scores of non-genetic deaf	-.76	325

Note. n.r. = not reported or could not be obtained. Many of the correlations were taken from Jensen (1998), but the authors of the original studies are listed in the Table. Schoenemann (1997) is cited in Jensen (1998, p. 147); sample sizes are not reported by Jensen and were taken from Schoenemann's dissertation.

Haier et al. (1992) show that there is an inverse relationship between brain glucose metabolic rate and psychometric measures of intelligence. A negative correlation is reported and we reversed the sign. Colom et al. (2006) report a collection of 28 correlations (Table 3) and 26 correlations (Table 5) on brain gray matter yielding the average correlation presented in the present Table. Lee et al. (2006) report in their Table 2 data on the activity in several brain regions. The average value of the sixteen correlations is reported in the present Table. Prokosch et al. (2004) report data on IQ scores and body symmetry. They also report the association between the rank-order of *g* loadings of five cognitive tests and its body symmetry association. We used their data to compute the rank-order correlation between rank-ordered *g* loadings and body symmetry association, which is r_s = .98. Schull & Neel (1965) tested 865 children from consanguineous marriages and 989 children from non-consanguineous marriages. Jensen

(1983) uses the same data. Badaruddozza & Afzal (1993) tested 50 inbred and 50 non-inbred control children. Braden (1989) reports the correlation of the differences in IQ scores between normal and hearing-impaired individuals and *g* loadings. Braden reports a median r = -.76 for six studies, but the three largest studies are criticized by Isham & Kamin (1993). We take the r = -.76 as an estimate of the mean correlation for the remaining three studies (combined N = 325). [1] These correlations are based on meta-analyses and are corrected for artifacts.

Research Questions

The question of whether certain factors are associated with an increase or decrease in IQ scores led to a theory of a relation between biological and non-biological factors, and *g*. To investigate this theory, two full psychometric meta-analyses were carried out to examine: (1) the true correlation between score differences between a gifted group and an average group and the magnitude of *g* loadings of the scores, and (2) the true correlation between score differences between a mentally retarded group and an average group and the magnitude of *g* loadings of the scores.

Additionally, three exploratory studies, using bare-bones meta-analytical techniques were carried out to test: (3) the mean weighted correlation between an alcohol abusing group and a comparison group and the magnitude of the *g* loadings, (4) the mean, weighted correlation between score differences between a cocaine abusing group and an comparison group and the magnitude of the *g* loadings, and (5) the mean, weighted correlation between score differences between a depressed group and an average group and the magnitude of the *g* loadings.

Giftedness and mental retardation

Galton (Galton, 1883) was among the first to assess the relative roles of nature and nurture in the development of intelligence and concluded that nature prevails strongly over nurture. In contrast, the environmental research program assumed that the influence of genetics on measured intelligence would diminish with age and the influence of environmental influences would increase. In fact, the opposite is true. As reported by Gottfredson (2003), the heritability of intelligence increases with age, from 20% in infancy, to 60% by adolescence, to 80% by adulthood. Further, studies of adoptive children consistently find that over time adopted siblings become less like their non-adopted siblings and parents and more like the biological ones they have never met. Further, the correlation between heritability coefficients and *g* loadings of subtests of a battery shows a meta-analytical correlation of 1 (te Nijenhuis & Jongeneel-Grimen, 2007). These well established results indicate that biological and genetic factors have a higher contribution to the development of IQ than cultural-environmental factors. Therefore, we hypothesize that mental retardation and giftedness are most likely caused by genetic and

biological factors, which, in turn, leads to the following hypotheses: (1) the true correlation between score differences between a gifted group and an average group and the magnitude of the *g* loading of the scores is strongly positive in sign, and (2) the true correlation between score differences between a mentally retarded group and an average group and the magnitude of the *g* loading of the scores is also strongly positive in sign.

Flynn Effect Moderator. Yet, ever since the introduction of standardized IQ tests there has been a systematic and pervasive rise in IQ scores all over the world. Over time, as IQ test norms become dated, people improve their performances on the test, raising the mean IQ by several points within a matter of years (Kanaya, Scullin, & Ceci, 2003). However, a negative *r* between the Flynn effect and *g* of -.33 was found (te Nijenhuis & van der Flier, submitted). Therefore, the hypothesized positive correlation for gifted and retarded samples could be masked by a Flynn effect. We test this moderator both for giftedness and mental retardation since IQ score gains over generations have been shown for both these groups. First, we expect the Flynn effect to cause a decrease in the correlation of *r* (giftedness x *g*) and of *r* (mental retardation and *g*). Second, we expect the Flynn effect to increase the percentage of variance accounted for by artifactual errors.

Educable Mental Retardation (EMR) Moderator. The population of mentally retarded is highly heterogeneous regarding the degree and nature of impairment and therefore some criteria are necessary to differentiate among this population. Jensen (1972) showed that two broad categories of mental retardation can be distinguished. The first category consists of a variety of severe mental defects resulting in IQs mainly below 50 which are accompanied by physical abnormalities or clear signs of neurological damage. This category of severe defects is usually caused by 'major gene' defects, chromosomal defects, or physical brain damage. The second category is comprised of retardates, those in the IQ range from 50 to 70, who represent the low end of the normal population distribution of intelligence. Within this category, at least 75 percent of the individuals appear clinically normal, and show no signs of neurological damage, sensory defects, or physical stigmata (Jensen, 1972). Also, Jensen argues that in pre-literatate and pre-industrial societies most persons in the IQ range from 70 to 85 would not be perceived as retarded or occupationally disadvantaged. Therefore, it is possible that all the individuals in the second category will not be detected and diagnosed as mentally retarded.

To investigate whether the meta-analytical results of mental retardation are influenced by the heterogeneity of the dataset, we tested for educable mental retardation as a moderator. We expected that splitting the sample into two homogeneous datasets would increase the size of the correlation as well as the percentage of variance explained.

Blacks-Whites Moderator. Black children have been shown to perform - on average -less well on cognitive tests such as intelligence tests, ability tests, and preschool readiness tests than do White children (Loehlin, Lindzey, & Spuhler, 1975; Scarr, 1981; Scarr-Salapatek, 1971; Weinberg, 1989). Therefore, low IQ groups usually have a large percentage of Blacks. Furthermore, Jensen (1998) has shown that Black/White score differences correlate about .70 with *g* loadings. Consequently, differences in test scores between Blacks and Whites could have a substantial effect on the correlation between mental retardation and *g*. Therefore we tested the moderating effect of percentage of Blacks and Whites on *r* (*d* x *g*). We expect a homogeneous Black sample to have a correlation close to .70.

Exploratory studies

Alcohol Abuse. Long-term excessive use of alcohol has been associated with damage to cortical and subcortical areas of the brain resulting in intellectual impairments (Smith & Smith, 1977). A distinction has to be made between adult alcohol abusers and children prenatally exposed to alcohol. In this thesis, however, when using the term alcohol abusing group we refer to both of these types of alcohol exposure. Children with Fetal Alcohol Spectrum Disorder (FASD) have deficits in intelligence as well as behavioral and emotional problems as a result of prenatal alcohol exposure (Aragón et al., 2008). The effects of alcohol on the human body, especially the functioning of the brain, are severe. Many exposure victims also have deficits of vision, hearing, speech, and locomotor function, as well as structural and functional brain damage and many secondary disabilities including mental health problems, trouble with law, confinement, alcohol and drug abuse, and dropping out of school (Streissguth, 1997). As the effects of alcohol are highly damaging to the brain, we hypothesize that the *g* factor is strongly affected. Therefore, we expect the mean weighted correlation between an alcohol abusing group and a comparison group and the magnitude of the *g* loadings to be strongly positive in sign.

Cocaine Abuse. Many investigators have demonstrated the deleterious effects of cocaine abuse on different areas of the human brain such as motor system dysfunctioning (O'Mally & Gawin, 1990), memory, visuospatial abilities, and concentration (Berry et al., 1993). Deleterious effects are also seen on Arithmetic and Symbol Digiting Modalities tests (O'Mally, Adamse, Heaton, & Gawin, 1992). A distinction has to be made between cocaine abusing adults and children prenatally exposed to cocaine. In this thesis, when using the term cocaine abusing group, we refer to both types of cocaine exposure. Research reveals that prenatal cocaine exposure is related to fewer positive emotions, less arousal, and less instrumental responding (Alessandri, Sullivan, Imaizumi, & Lewis, 1993). Also, heavy exposure early in pregnancy was related to poorer recognition memory and information processing (Jacobson et al., 1996). Cocaine abuse

diminishes several cognitive capabilities and has a large influence on brain functioning. As the effects of cocaine are highly damaging to the brain, we hypothesize that the *g* factor is strongly affected. Therefore, we expect the mean weighted correlation between a cocaine abusing group and a comparison group and the magnitude of the *g* loadings to be strongly positive in sign.

Depression. Depressive symptoms have a large impact on brain functioning. Memory deficits have been linked to dysfunctions of the hippocampus during depression and impairment in executive tasks has been linked to working memory, motor speed, and attentional set-shifting (Mandelli et al., 2006). These impairments are also manifested on cognitive test performance of mood disorder patients. The most common finding for depressed individuals on the WAIS-Revised (WAIS-R) is the discrepancy between verbal and nonverbal abilities, with lower Performance IQ (PIQ) relative to Verbal IQ (VIQ) (Groth-Marnat, 1997; Kluger & Goldberg, 1990; Pernicano, 1986; Sackeim et al., 1992; Zillmer et al., 1991). After remission, some improvement of cognitive functioning has been observed, but also stable deficits have been reported both during depression and remission. These reports have a focus on improvement of specific cognitive functions after antidepressant treatment, such as memory function or attention. In mood disorders, performance on the WAIS has been mainly investigated in acute-depressed patients, cognitively-impaired patients, or elderly subjects (Mandelli et al., 2006). However, acute depression might have different effects on brain functioning than severe disorders. Evidence from genetic and biochemical studies suggest that, for instance, bipolar disorder has substantial heritability. Family pedigree studies and twin studies reveal that some people inherit a predisposition to develop bipolar disorder (Craddock & Jones, 1999; Gershon & Nurnberger, 1995). Low activity of the neurotransmitters norepinephrine and serotonin have been strongly linked to the depressive episode of bipolar disorder (Sobczak, Honig, van Duinen, & Ridel, 2002). Thus, the inconsistencies in results regarding the improvement of cognitive functioning after remission might be related to the severity of the depressive disorder indicating that depression might have a permanent effect on general intelligence. These effects tend to appear most strongly on the Performance Scale of a Wechsler battery. Compared to the VIQ scale of the Wechsler battery, the PIQ scale consists of subtests containing - on average - lower *g* loadings indicative of lower cognitive complexity and less *g* saturation. Thus, there seems to be an effect of depression on general intelligence, though it is small. Therefore, we expect the mean weighted correlation between a depressed group and an average group and the magnitude of the *g* loadings to be weakly positive in sign.

General Method

Psychometric meta-analysis (Hunter & Schmidt, 1990) estimates what the results of studies would have been if all studies had been conducted without methodological limitations or flaws. The results of perfectly conducted studies would allow a clearer view of underlying construct-level relationships (Schmidt & Hunter, 1999). The goal of the present psychometric meta-analyses is to provide reliable estimates of the true correlation between a number of variables and the magnitude of *g* loadings. These variables are: giftedness and mental retardation. Also three exploratory studies were carried out on: the effects of alcohol, the effects of cocaine, and the effects of depression. These exploratory studies yielded the mean, weighted correlations between these variables and *g*.

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. Finally, Pearson correlations between each of the five variables (score differences between a gifted group and an average group, a mentally retarded group and an average group, an alcohol abusing group and an average group, a cocaine abusing group and an average group, a depressed group and an average group, and the *g* loadings) were computed.

There has been a discussion whether one should use Pearson *r* or Spearman's rho when applying the method of correlated vectors. The answer depends on whether one assumes an interval or an ordinal measurement level for IQ scores. Ranking of IQ scores can be seen as a way of categorizing intelligence levels on an ordinal scale. For instance, an IQ score of 150 indicates a higher level of intelligence compared to an IQ score of 75. However, the inference that an IQ score of 150 indicates a doubling in level of intelligence compared to an IQ score of 75 cannot be drawn.

In order to obtain our results, mean IQ scores were used to calculate the score differences between groups (*d*). Score differences have the characteristics of an interval scale: arithmetical operations can be conducted, and the effects (*d*) have values ranging from negative to positive. Thus, the choice for Pearson *r* or Spearman's rho depends on whether the underlying construct on which calculations are carried out are more important or the calculations themselves. Colom, Juan-Espinosa, Abad, & Garcia (2000) consider both Pearson *r* and Spearman's rho as suitable measures of the degree of relationship between two vectors. We decided to use Pearson *r* following earlier conducted meta-analyses using Pearson *r* in the method of correlated vectors (te

Nijenhuis, van Vianen, & van der Flier, 2007; te Nijenhuis & Jongeneel-Grimen, 2007). This has the advantage that the results of the present studies can be compared directly against those of the earlier studies.

General Inclusion Rules

For a study to be included in the meta-analyses and in the exploratory studies, three criteria had to be met: First, in order to obtain a reliable estimate of the true correlation between each of the two variables (giftedness and mental retardation) and the *g* loadings and a reliable estimate of the mean, weighted correlation between each of the three variables (effects of alcohol, effects of cocaine, and effects of depression) and the *g* loadings, the cognitive batteries had to have a minimum of seven subtests. Second, the test had to be well-validated. Third, since studies with a test-retest effect would influence the 'true' correlation between *d* and *g* (see discussion below) – they were excluded. That is, studies using a counterbalanced design and the scores of the re-administration of an IQ battery within a test-retest design were not included. In a counterbalanced design, participants are administered two IQ batteries, X and Y, in different orders. Half of the participants take test X first, then test Y and vice versa.

Test-retest Effects on 'true' g

First, when people take the same test battery a second time test-retest effects are, by definition, at 100% of their strength. For instance, they can take the WISC on one occasion, and then again on a second occasion. The effect of taking a test twice is modest – only a few IQ points (see Jensen, 1980; see meta-analyses by Kulik and colleagues). Additional training can increase the size of the score gains. Kulik, Bangert-Drowns, and Kulik's (1984) meta-analysis on test preparation studies resulted in effect sizes on intelligence tests for practice and additional coaching of 0.25 *SD* and 0.51 *SD*, respectively. Second, the true correlation between test-retest score gains and *g* loadings has been shown to be -1, based on a psychometric meta-analysis with a very large total sample (te Nijenhuis, van Vianen, & van der Flier, 2007). Te Nijenhuis et al. argue that the gains are linked to test-specific variance only, and not at all to the variance associated with general, broad, or narrow abilities.

Third, there are also occasions when people take two comparable test batteries, such as the WISC and the WISC-R, or the WISC and the WAIS, for instance when they are approximately sixteen years of age. In some studies people also take test batteries that are non-comparable, i.e. constructed according to different psychometric principles, such as, the WISC-R and the Kaufman-ABC. The size of the test-retest effect is most strongly influenced by the degree of similarity of the two test batteries taken. With subtests of the same name, the same cognitive principles have to be used to solve the items. These principles, having been discovered during the

taking of the first test, can be applied when the similar subtest from the second test battery based on the same principles is taken. Because, for example, the WISC and the WISC-R are not identical, the test-retest effect will most likely not be one hundred percent, so, not as strong as when either test is taken twice. When two non-comparable tests are taken, such as first the WISC-R and then the Kaufman-ABC, these tests both measure the *g* factor, but with different flavors (see Carroll, 1993). So, a number of the principles in the subtests of the WISC-R cannot be applied to the subtests of the K-ABC. Therefore, the test-retest effect will be weaker.

Overall, test-retest effects mask the theoretically expected true correlation of +1 for the gifted and the retarded. Therefore, we excluded all studies with test-retest designs.

Corrections for Artifacts

Psychometric meta-analytical techniques (Hunter & Schmidt, 1990, 2004) were applied using the software package developed by Schmidt and Le (2004). Psychometric meta-analysis is based on the principle that there are artifacts in every dataset and that most of these artifacts can be corrected. In the present meta-analyses we corrected for five artifacts identified by Hunter and Schmidt (1990) that alter the value of outcome measures. These are: (1) sampling error, (2) reliability of the vector of *g* loadings, (3) reliability of the vector of a specific variable of theoretical interest (4) restriction of range of *g* loadings, and (5) deviation from perfect construct validity. In the present exploratory studies, using bare-bones meta-analytical techniques, we corrected for only one artifact, namely sampling error.

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.

Correction for Reliability of the Vector of g Loadings

The values of *r* (*g* × giftedness) and *r* (*g* × mental retardation) 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 that differed little on background variables were compared. For the comparisons using children we chose samples that were highly comparable with regard to age. Samples of children in the age of 3 to 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 to 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 to 95 years.

We collected correlation matrices from test manuals, books, articles, and technical

reports. The large majority came from North America, with a large number from European countries, and also a substantial number from Korea, China, Hong Kong, and Australia. This resulted in about 700 data points, which yielded 385 comparisons of *g* loadings of comparable groups from which to estimate the reliability for that group. To give an illustration of the procedure, van Haasen et al. (1986) report correlation matrices of the Dutch and the Flemish WISC-R for 22 samples in the age of 6-16 years. We compared samples of children in the age of 6 to 17 years with other samples of children who do not differ by more than 1.5 years. Because the samples of children reported in van Haasen et al. (1986) were between 6 and 17 years we only compared children who did not differ more than 1.5 years. The *N*s in these samples were comparable. The resulting average correlation was .78 (combined *N* = 3,018; average *N* = 137).

A scatter plot of reliabilities against *N*s should show that the larger *N* becomes, the higher the value of the reliability coefficients, with an asymptotic function between *r* (*g* x *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* x *g*) and *N* is not reported. For the same reason we divided Figure 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*.

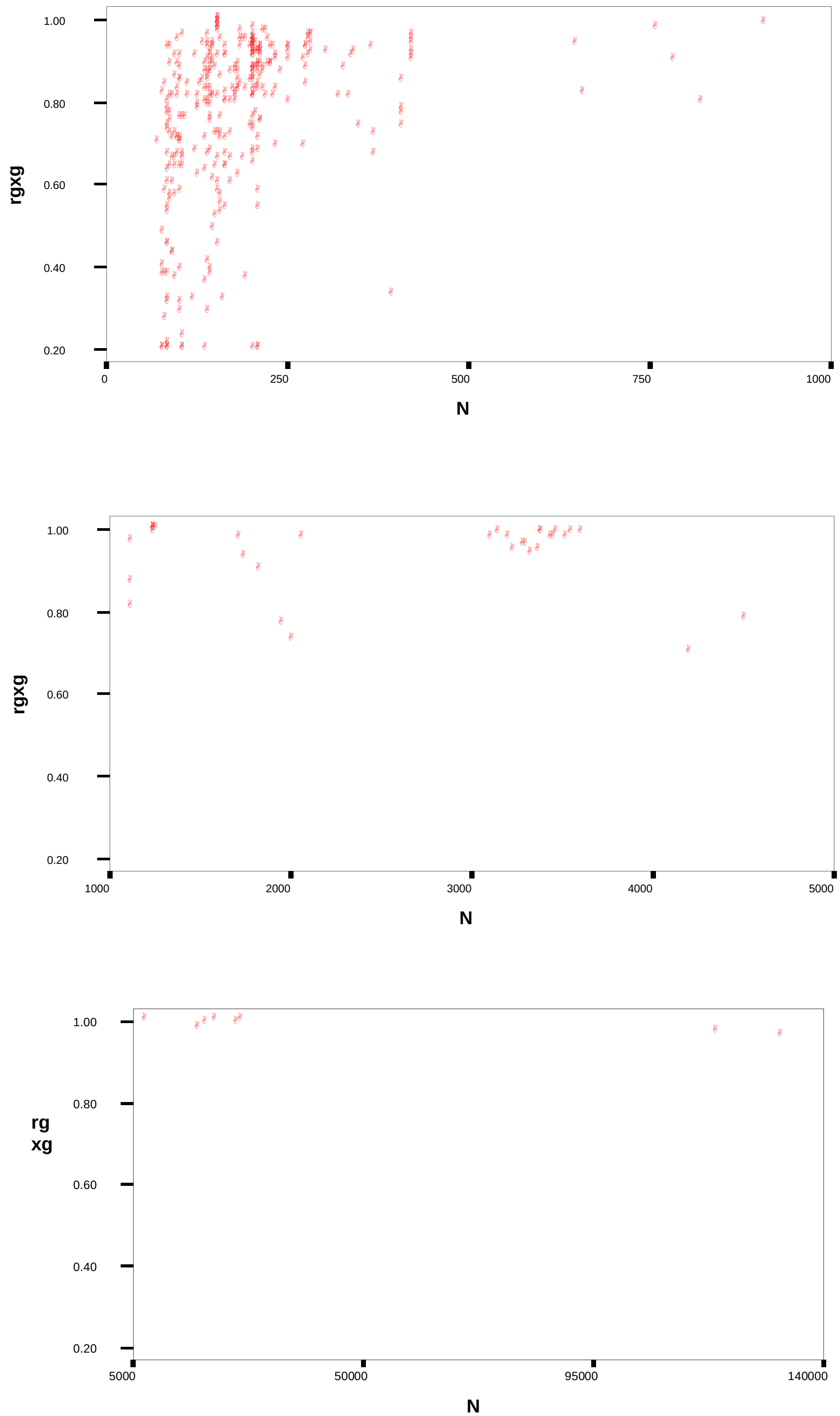

Figure 1

Three Scatter Plots of Reliability of the Vector of g Loadings and Sample Size Each for a Different Range of N

Correction for Reliability of the Vector of the Second Variable. The values of *r* (*g* × giftedness) and *r* (*g* × mental retardation) are attenuated by the reliability of the vector of the second variable, respectively giftedness and mental retardation, 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. The reliability of the vector of giftedness and the vector of mental retardation were each estimated using the present datasets, comparing samples that took the same test, and that differed little on background variables. For the comparisons using children we chose samples that were highly comparable with regard to age, and for the comparisons of adults we chose samples that were roughly comparable with regard to age.

Correction for Restriction of Range of g Loadings. The values of *r* (*g* × giftedness) and *r* (*g* × mental retardation) 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 (1998, 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 variation in range is to define a reference population and express all correlations in terms of it. 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 $u = SD_{\text{study}}/SD_{\text{ref}}$. As references we used tests that are broadly regarded as exemplary for the measurement of intelligence, namely the 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 children and adults, respectively.

The Hunter and Schmidt meta-analytical program computes only the aforementioned four corrections. We will refer to the observed correlation corrected for sampling error, unreliability of the vector of *g* loadings and the second vector, and range restriction as rho-4.

Correction for Deviation from Perfect Construct Validity. The deviation from perfect construct validity in *g* attenuates the values of *r* (*g* × giftedness) and *r* (*g* × mental retardation). In making up any collection of cognitive tests, we do not have a perfectly representative sample of the entire universe of all possible cognitive tests. Therefore any one limited sample of tests will not yield exactly the same *g* as another such sample. The sample values of *g* are affected by psychometric sampling error, 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*. The values of *r* (*g* × giftedness) and *r* (*g* × mental retardation) are attenuated by psychometric sampling error in each of the batteries from which a *g* factor has been extracted.

The more tests and the higher their *g* loadings, the higher the *g* saturation is of the composite score. The Wechsler tests have a large number of subtests with quite high *g* loadings, yielding a highly *g*-saturated composite score. Jensen (1998, p. 90–91) states that the *g* score of the Wechsler tests correlates more than .95 with the tests' IQ score. However, shorter batteries with a substantial number of tests with lower *g* loadings will lead to a composite with somewhat lower *g* saturation. Jensen (1998. ch. 10) states that the average *g* loading of an IQ score as measured by various standard IQ tests lies in the +.80s. When we take this value as an indication of the degree to which an IQ score is a reflection of "true" g, we can estimate that a tests' *g* score correlates about .85 with "true" *g*. As *g* loadings represent the correlations of tests with the *g* score, it is most likely that most empirical *g* loadings will underestimate "true" *g* loadings; therefore, empirical *g* loadings correlate about .85 with "true" *g* loadings. As the Schmidt and Le (2004) computer program only includes corrections for the first four artifacts, the correction for deviation from perfect construct validity was carried out on the values of *r* (*g* × giftedness) and *r* (*g* × mental retardation) after correction for the first four artifacts. To limit the risk of overcorrection, we conservatively chose the value of .90 for the correction. We will refer to the observed correlation corrected for sampling error, unreliability, range restriction, and imperfect construct validity as rho-5.

Study 1: Giftedness

To test whether there is a strong positive correlation between the magnitude of the *g* loadings of tests and giftedness, we carried out a psychometric meta-analysis of all studies of giftedness that reported scores for at least seven subtests.

Method

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 the gifted. Four methods were used to obtain scores of the gifted 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 “gifted”, or “exceptional” in combination with any keyword that contains one of the following words; IQ, *g*, general mental ability, GMA, cognitive ability, general cognitive ability, intelligence, intelligence test, Wechsler, Stanford Binet, cognitive ability test. Second, we browsed the tables of content of several major research journals with a strong focus on the gifted: *Psychology in the Schools* 1964-2008, *Gifted Child Quarterly* 1977-2008, *Roeper Review* 1990-2008, *Journal for Advanced Academics* 1996-2008, and *Exceptional Children* 1934-2008. 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. Finally, several well-known researchers who have conducted cognitive ability research of the gifted were contacted in order to obtain any additional articles or supplementary information.

Specific criteria for inclusion. Only studies reporting a mean Full scale IQ score of 125 or higher were included in the meta-analysis. Application of these inclusion rules yielded twenty-two papers resulting in twenty-two correlations between *g* and score differences between a gifted group and an average group.

Computation of score differences between a gifted group and an average group. Score differences between a gifted group and an average group (*d*) were computed by subtracting the mean of the standardization group of the particular test in question from the mean of the gifted group, and then dividing the result by the *SD* of the standardization group. *g* Loadings matching the average age of the gifted group were used to compute *r* (*d* x *g*). However, the WISC manual, for instance, only contains three age groups. The mean age of the gifted group is not always comparable to the mean age of the standardization group. Therefore, we averaged several standardization groups to get as close to the mean age of the gifted group as the data allowed. Subsequently, *g* loadings of the subtests were also averaged for several standardization groups

when using the WISC.

Psychometric meta-analytical techniques (Hunter & Schmidt, 1990, 2004) were applied to the resulting twenty-two *r* (*g* × giftedness)s using the software package developed by Schmidt and Le (2004). In the present study we corrected for the five artifacts (mentioned above) that alter the value of outcome measures listed by Hunter and Schmidt (1990).

Correction for reliability of the vector of giftedness. The value of *r* (*g* × giftedness) is attenuated by the reliability of the vector of giftedness for a given battery. It was estimated using the present datasets, comparing samples that took the same test and that were comparable in regard to age and sample size. As an illustration of the procedure, the following rules were set in order to analyze studies that were highly comparable. First of all, only studies using the same test and the same version of this test were taken together. Second, studies containing less than a hundred participants were considered to be highly comparable as long as the difference in *N* between two studies was lesser than or equal to sixty. Third, studies containing more than a hundred participants were considered to be highly comparable as long as the difference in *N* between two studies was lesser than or equal than hundred-fifty. Fourth, the difference in average age of participants in separate studies was three years or less. Finally, the date of publication between two studies did not differ more than ten years.

A scatter plot of reliabilities against *N*s should reveal that the larger *N* becomes, the higher the value of the reliability coefficients, with an asymptotic function between *r* (*d* x *d*) and *N* 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 giftedness and sample size, and the logarithmic curve that fitted optimally.

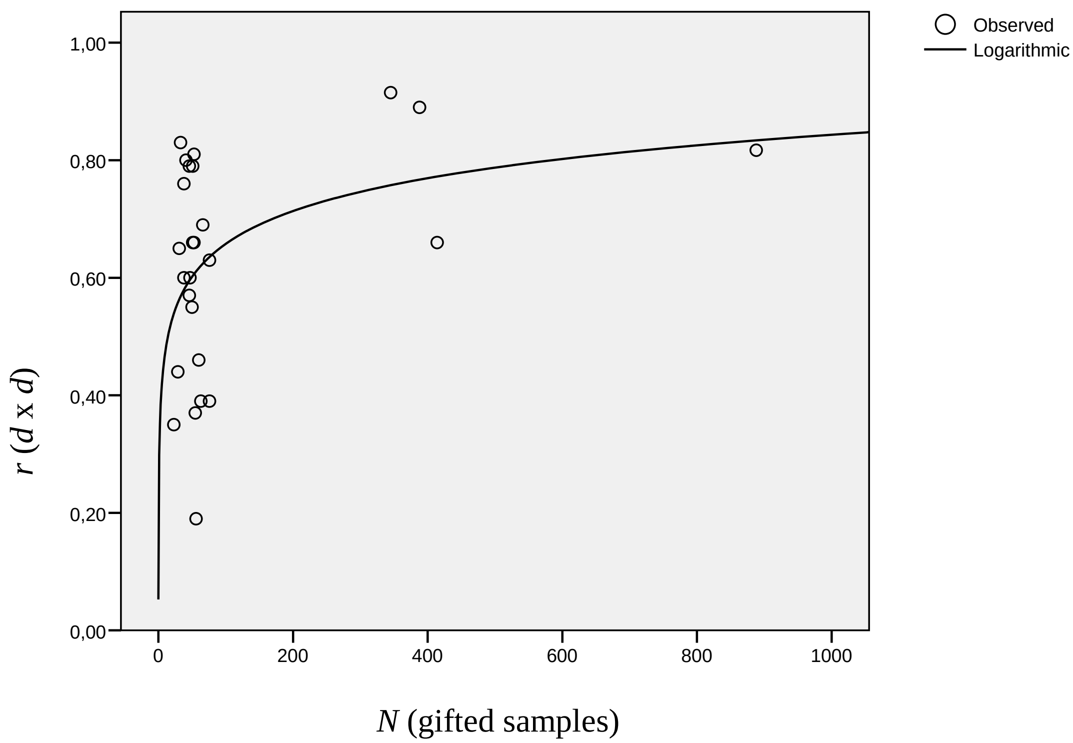

Figure 2

Scatter Plot of Reliability of the Vector of Giftedness and Sample Size and Regression Line

Results

The results of the studies on the correlation between *g* loadings and the score differences between gifted groups and average groups (*d*) are shown in Table 2. The Table gives data derived from twenty-two studies, with participants numbering a total of 5,129. It also lists the reference for the study, the cognitive ability test used, the correlation between *g* loadings and *d*, the sample size, and the mean age (and range of age). It is clear that the large majority of the correlations are strongly positive.

Table 3 presents the results of the psychometric meta-analysis of the twenty-two data points. It shows (from left to right): the number of correlation coefficients (*K*), total sample size (*N*), the mean observed correlations (*r*) and their standard deviation (SD_r), the correlations one can expect once artifactual error from unreliability in the *g* vector, the giftedness vector, and range restriction in the *g* vector have been removed (rho-4), and their standard deviation ($SD_{\text{rho-4}}$), and the true correlation one can expect when corrections for all five artifacts have been carried out (rho-5). The next two columns present the percentage of variance explained by artifactual

errors (%VE), and the 80% confidence interval (80% CI). This interval denotes the values one can expect for rho in sixteen out of twenty cases.

The analysis of all 22 data points yields an estimated correlation (rho-4) of .84, with only 19% 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 analysis, 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. Figure 3 shows the scatter plot of correlations *r* (*d* x *g*) against sample size. We chose to leave out one extreme outlier, with a value of *r* more than 11 *SD* beneath the average *r* of the final sample of twenty-one data points. This resulted in a small change in the value of the correlation (rho-4), a large decrease in the *SD* of rho-4 with 41%, and a tripling of the amount of variance explained in rho-4 by artifacts: 57% of the variance is now explained. Finally, a correction for deviation from perfect construct validity in *g* took place, using the conservative value of .90. This resulted in a value of 1.01 for the final estimated true correlation between *g* loadings and giftedness.

Flynn Effect Moderator. We hypothesized that *r* (*d* x *g*) for giftedness would be strongly positive in sign. However, the final estimated true correlation of 1.01 could be moderated by the Flynn effect. To test whether IQ score gains over generations moderate the relation between giftedness and *g*, the dataset has been split up into two groups based upon the estimated differences in years between the publication of a test battery and administration of the test battery (see Table 3). This resulted in a group with an estimated difference of 0-6 years, and a second group with an estimated difference of 9-27 years. The results from the meta-analyses of the Flynn effect are compared to the results of the final data set of the gifted. That is, the values of rho-4 and the percentage of variance from the moderator analyses are compared to the dataset of the gifted where outliers have been left out of the analysis. We hypothesized that a Flynn effect will moderate the data resulting in an increase in percentage of variance accounted for by artifactual errors. Also, a decrease of the correlation (rho-4) is expected. In the first dataset rho-4 increases from .91 to .98, disconfirming the Flynn effect as a moderator. However, the percentage of variance increased from 57% to 120%. Regarding the results of the second dataset, rho-4 decreases from .91 to .88, indicating that here the Flynn effect does act as a moderator. However, the percentage of variance decreased from 57% to 53%. So, there appears to be some support for the Flynn effect to act as a minor moderator though.

Table 3 shows a percentage of variance explained of 120%. This phenomenon is called "second-order sampling error", and results from the sampling of studies in a meta-analysis. Percentages of variance explained greater than 100% are not uncommon when only a limited number of studies are included in an 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).

Post-hoc analysis: Subtest pattern of the vector giftedness. A true correlation of 1.00 results in a figure pattern where the data points ($d_{weighted}$ x g) are all on the regression line. In Figure 4, only data on the subtests of the WISC-R were used, since this test battery was administered in the majority of the studies. $d_{weighted}$ is computed by multiplying the score differences between a gifted group and an average group (*d*) by the number of participants (*N*) for a certain subtest of the WISC-R. These scores (*d* x *N*) are aggregated for all studies using the same subtest and divided by the sum of the total number of participants. *g* Loadings of the subtests of the WISC-R of all ages were used. We used the general method to compute the *g* loadings.

The effects shown are uncorrected for artifacts, but even this figure shows that the data points are virtually on the regression line or very close to the regression line. The true correlation (*d* x *g*) was 1.01 which means that the amount of deviation from the regression line is miniscule.

Table 2

Studies of Correlations Between g Loadings and Giftedness

references	test battery	*r*	*N*	mean age (range)
Brown & Yakimowksi (1987)	WISC-R	.94	320	9.95 (5.3-16.9)
Brown, Hwang, Baron, & Yakimowski (1991)	WISC-R	.80	158	9.6
Chae, Kim, & Noh (2003)	Korean WISC-R	.64	106	7.7 (6-9)
Fishkin & Kampsnider (1996)	WISC-III	.67	42	10.8 (9-14)[1]
Henry & Wittman (1981)	WISC-R	.70	40	9[2] (6-12)[1]
Karnes & Brown (1980)	WISC-R	.85	946	9.9 (6-16)
Linville & Rust (1999)	WISC-III	.53	58	10.3 (6-14.3)
Lucito & Gallagher (1960)	WISC	.55	50	9[2] (7-11)[1]
Macmann, Plasket, Barnett, & Siler (1991)	WISC-R	.74	829	9.2 (6-14.9)
Namay (1967)	WISC	.25	32	9.5[2] (9-10)[1]
Reams, Chamrad, & Robinson (1990)	WISC-R	.83	66	8.6 (6-13.1)
Sapp, Chissom, & Graham (1985)	WISC-R	.81	371	9.5 (7.5-11.5)
Sattler & Covin (1986)	WISC-R	.30	85	8.25 (6-11.5)
Sevier & Bain (1994)	WISC-R	.75	35	9.1 (7-12.7)
Speer, Hawthorne, & Bucellato (1986)	WPPSI	-.10	306	4.9 (4.2-6.4)
Spangler & Sabatino (1995)	WISC-R	.29	66	8.3
Swanson, Brandenburg-Ayres, & Wallace (1989)	K-ABC	.67	169	8 (6-10.6)
Thompson & Finley (1962)	WISC	.66	400	10.23
Watkins, Greenawalt, & Marcell (2002)	WISC-III	.58	505	8
Wechsler (2003)	WISC-IV	.91	63	11[2] (6-16)
Wheaton & Vandergriff (1978)[3]	WISC-R	.71	26	10.8 (9.5-11.5)
Wilkinson (1993)	WISC-R	.77	456	8.8 (7.4-9.8)

Note: [1]These studies report the range in grades instead of age. We converted these into age. [2] Estimated mean age based on the age range. [3]Wheaton et al. (1978) used a test-retest design. Each subject was administered the WISC-R and then the WISC. We only used the data on the WISC-R to rule out any test-retest effects.

Table 3

Meta-analytical Results for Correlation Between Giftedness and g Loadings After Corrections for Reliability, Restriction of Range, and Imperfect Construct Validity

predictor	*K*	*N*	*r*	SD_r	rho-4	SD_{rho-4}	rho-5	%VE	80% CI
Giftedness[1]	22	5129	.69	.24	.84	.22	.93	19%	.56 - 1.13
Giftedness minus one outlier[2]	21	4823	.74	.14	.91	.09	1.01	57%	.79 - 1.02
moderator									
Flynn effect (0-6 years)[3]	7	1344	.80	.09	.98	0	1.09	120%	.98 - .98
Flynn effect (9-27 years)[3]	14	3479	.71	.14	.88	.10	.98	53%	.75 - 1.01

Note. [1] Meta-analytical results for correlations between *g* loadings and *d* (giftedness). [2] The study by Speer, Hawthorne, & Bucellato (1986) is considered an extreme outlier and therefore left out of the analysis . [3] Meta-analytical results for a Flynn effect within the gifted population. The dataset has been split into two groups based upon the estimated difference between the publication of a test battery and administration of the test battery. *K* = number of correlations; *N* = total sample size; *r* = mean observed correlation (sample-size weighted); SD_r = standard deviation of observed correlation; rho-4 = observed correlation corrected for sampling error, unreliability, and range restriction); SD_{rho-4} = standard deviation of correlation; rho-5= true correlation (observed correlation corrected for sampling error, unreliability, range restriction, and imperfect construct validity); %VE = percentage of variance accounted for by artifactual errors; 80% CI = 80% credibility interval.

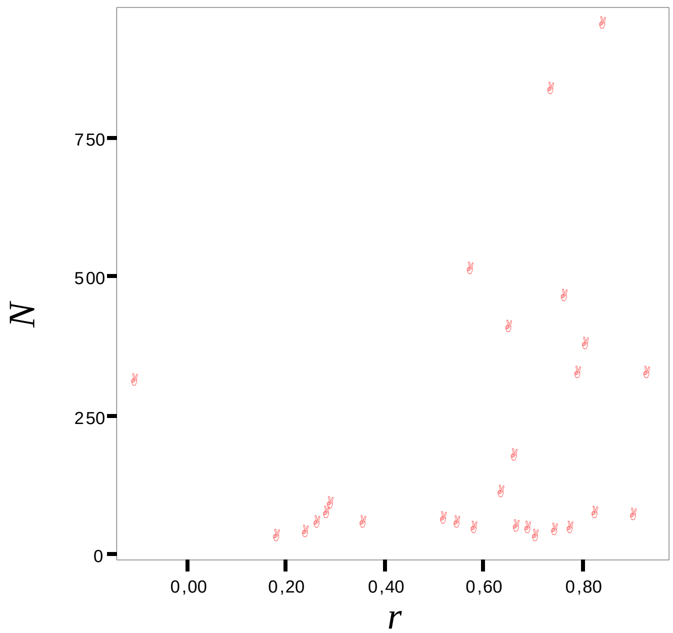

Figure 3

Scatter Plot of Correlations (*d* x *g*) *and Sample Size of the Variable Giftedness*

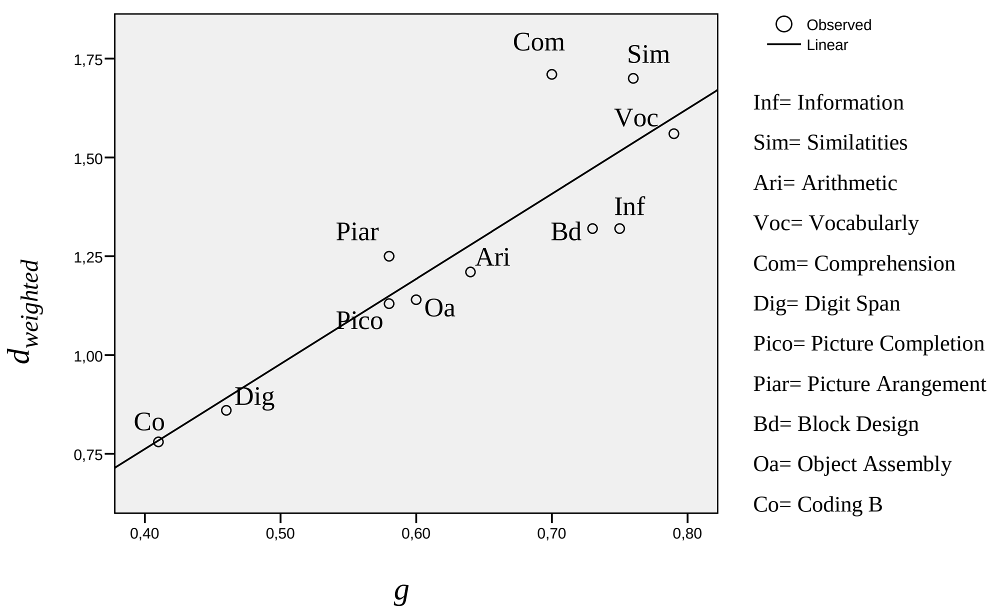

Figure 4

Scatter Plot of Subtest Pattern ($d_{weighted}$ x *g*) *of the Variable Giftedness*

Study 2: Mental retardation

To test whether there is a strong positive correlation between the magnitude of the *g* loadings of tests and mental retardation, we carried out a psychometric meta-analysis of all studies concerning mental retardation which included at least seven subtests.

Method

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 the mentally retarded. Three methods were used to obtain data about mental retardation from both published and unpublished 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 keyword combinations were used to conduct the searches: the keywords mental retardation and cognitive disabilities in combination with the keywords IQ, intelligence, intellectual development, *g*, *g* factor, GMA, general mental ability, Wechsler, WISC, cognitive ability, and general cognitive ability. Also, three additional search terms were used for topics that are related to mental retardation, namely Autism, Down syndrome, and educable mentally retarded.

Second, we browsed the tables of content of several major research journals with a strong focus on the mentally retarded: *Psychology in the Schools* 1964-2008*, Exceptional Children* 1934-2009*, Journal of Applied Research in Intellectual Disabilities* 2000-2009*, Journal of Applied Research in Cognitive Disabilities 2000-2008, American Journal on Mental Retardation* 1997-2008, and *Journal of Intellectual Disability Research* 1992-2008. 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. Finally, several well-known researchers who have conducted cognitive ability research of mentally retarded were contacted in order to obtain any additional articles or supplementary information.

Specific criteria for inclusion. As a general rule, we only included studies with a mean sum score of 75 or lower on the IQ batteries. However, the population of low-IQ individuals consists of groups with highly divergent scores on the subscales of IQ batteries. For instance, some learning disabled children score far below 75 on the Verbal scale of a Wechsler battery, but higher than 75 on the Performance scale of a Wechsler battery. For the majority of the learning disabled, this profile results in a Full Scale IQ exceeding a score of 75. However, for a minority of the learning disabled, this profile results in a Full Scale IQ of 75 or lower with a high variability between the subtest scales. We decided to keep the population of the mentally retarded as homogeneous as possible. We therefore excluded mentally retarded children with a high

variability between the subtest scales from the analysis. Consequently, only studies were included that reported an IQ score of 75 or less on both the Full Scale IQ and on at least one of the subtest scales of an IQ battery. Application of these inclusion rules yielded twenty-four papers resulting in thirty-four correlations between *g* and score differences between a mentally retarded and an average group.

Computation of score differences between a mentally retarded group and an average group. Score differences between a mentally retarded group and an average group (*d*) were computed by subtracting the mean of the mental retardation group on a particular test from the mean of a standardized group on a particular test. These differences were then divided by the *SD* of the standardized group. *g* Loadings matching the average age of participants in a particular study were used for computing *r* (*g* x *d*). However, specific IQ tests like the WISC, WAIS, WAIS-III, and WAIS-R consists of only a few norm samples of age groups. The mean age of the mentally retarded group is not always comparable to the mean age of the norm sample. Therefore, we averaged several standardization groups to get as close to the mean age of the mentally retarded group as the data allowed. Subsequently, *g* loadings of the subtests were also averaged for several standardization groups when using the WISC, WAIS, WAIS-III, and WAIS-R.

Psychometric meta-analytical techniques (Hunter & Schmidt, 1990, 2004) were applied to the resulting thirty-four *r* (*g* x *d*)s using the software package developed by Schmidt and Le (2004). In the present study we corrected for the five artifacts (mentioned above) that alter the value of outcome measures listed by Hunter and Schmidt (1990).

Correction for reliability of the vector of mental retardation. The value of *r* (*d* x *d*) is attenuated by the reliability of the vector of mental retardation for a given battery. The reliability of the vector of mental retardation was estimated using the present datasets, comparing samples that took the same test and that were comparable in age and sample size. As an illustration of the procedure, the following rules were set in order to analyze studies that were highly comparable. First, only studies using the same test and the same version of this test were taken together. Second, studies containing less than a hundred participants were considered to be highly comparable as long as the difference in *N* between two studies did not outnumber sixty. Third, studies containing more than a hundred participants were considered to be highly comparable as long as the difference in *N* between two studies was hundred-fifty or less. Fourth, differences in average age between participants of two separate studies of three years or less were allowed. For specific versions of the Wechsler Adult Intelligence Scale the acceptable age differences were less strict, given the fact that the WAIS is meant for all adults, ranging from eighteen to eighty

years old. Finally, the date of publication between two studies did not differ more than ten years.

A scatter plot of reliabilities against *N*s should reveal that the larger *N* becomes, the higher the value of the reliability coefficients, with an asymptotic function between *r* (*d* x *d*) and *N* expected. We checked which curve gave the best fit to the expected asymptotic function. Figure 5 shows the scatter plot of reliability of the vector of mental retardation (*d* x *d*) and sample size, and the logarithmic curve that fitted optimally.

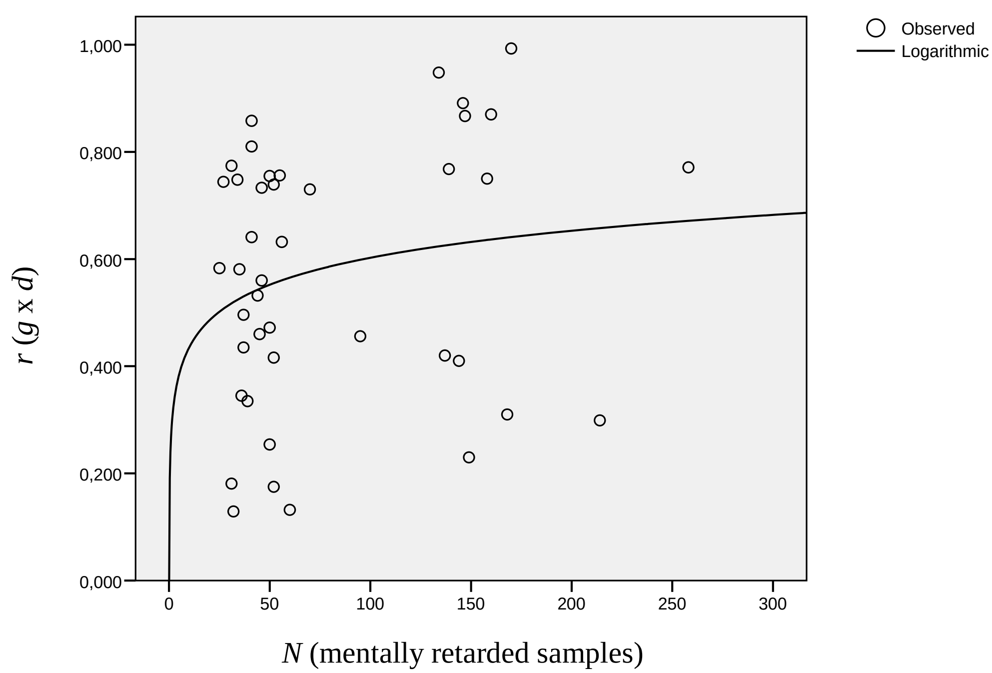

Figure 5

Scatter Plot of Reliability of Mental Retardation and Sample Size and Regression Line

Results

The results of the studies on the correlation between *g* loadings and mental retardation are shown in Table 4. The Table contains data derived from 24 studies with a total sample size of 3,057 participants. It also lists the reference for the study, the cognitive ability test used, the correlation between *g* loadings and mental retardation, the sample size, and the mean age as well as the age ranges. It is clear that the correlations of studies concerning mental retardation vary strongly.

Table 5 presents the results of the psychometric meta-analysis of the 34 data points. It shows (from left to right): the number of correlation coefficients (*K*), total sample size (*N*), the

mean observed correlations (r) and their standard deviation (SD_r), the correlations one can expect once artifactual error from unreliability has been removed in the *g* vector and the mental retardation vector, in range restriction in the *g* vector (rho-4), their standard deviation ($SD_{\text{rho-4}}$), and the true correlations one can expect when correction for all five artifacts have been done (rho-5). The next two columns present the percentage of variance explained by artifactual errors (%VE), and the 80% confidence interval (80% CI). This interval denotes the values one can expect for rho in sixteen out of twenty cases.

The estimated true correlation has a value of .70, with only 42% 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. The outliers can be clearly spotted in the scatter plot of *r* (*g* x *d*) (see Figure 6). We chose to leave out two strong outliers with values of *r* more than 5.0 *SD* above the average *r* of the sample of 32 final data points. This resulted in a small change in the value of the true correlation, no decrease in the *SD* of rho-4, and a small increase of 15% in the amount of variance explained in the observed correlations by artifacts. After the two outliers are excluded artifacts explain 57% 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 .74 for the final estimated true correlation between *g* loadings and mental retardation.

Moderators

We hypothesized that the true correlation of mental retardation would result in a high positive correlation, and it indeed reaches a value of .74. However, only 57 percentage of variance is accounted for by artifactual errors. Therefore, several moderators were tested in an attempt to reduce the heterogeneity in the data points.

Flynn Effect Moderator. To test whether IQ score gains over generations moderate the relation between mental retardation and *g*, the dataset was split up into two groups based upon the estimated difference in years between the publication of a test battery and administration of the test battery (see Table 5). According to the hypothesis, a moderating Flynn effect should increase the percentage of variance accounted for by artifactual errors in both datasets. In addition, the Flynn effect should decrease the value of rho-4 compared to the value of rho-4 in the final dataset without outliers. First of all, Table 5 clearly shows that the Flynn effect has a large effect on the percentage of variance explained in the first dataset (93%), but the percentage of variance explained in the second dataset (31%) actually decreases strongly. Second, the

decrease in rho-4 is miniscule. It is concluded that there is hardly any evidence for the Flynn effect acting as a moderator.

Educable Mental Retardation (EMR) Moderator. In order to investigate whether the meta-analytical results of mental retardation were influenced by the heterogeneity of the dataset, a homogeneous dataset was created consisting of fourteen studies that contained only educable mentally retarded participants. Results revealed that the percentage of variance accounted for by artifactual errors did not increase in comparison to the percentage of variance in the heterogeneous dataset (see Table 5). Thus, separating a homogeneous sample like EMR does not moderate the outcomes.

Blacks-Whites Moderator. Since it was hypothesized that a large percentage of Blacks could decrease the true correlation, differences in percentages of Blacks and Whites on *g* were studied as a moderator. However, our database only contained a few studies reporting separate test scores for Blacks and for Whites. The small number of correlations from studies with predominantly Black participants and from studies with predominantly White participants yielded substantially lower correlations for Whites than for Blacks, respectively $r = -.55$ and $r = .55$; this is the opposite of what was predicted. Considering the fact that only 8 studies reported separate information about scoring profiles of Blacks and Whites, further meta-analytical analysis would provide no additional information. Thus, a moderating effect of differences in percentages between Blacks and Whites could not be confirmed in this study.

Post-hoc analysis: Subtest pattern of the vector of mental retardation

Since the investigated moderators could not reduce the heterogeneity of the data points, we decided to examine an alternate explanation. A true correlation of .74 should result in a subtest pattern where the effects ($d_{weighted}$ x g) vary substantially around the regression line (see Figure 7). $d_{weighted}$ is computed by multiplying the score differences between a mentally retarded group and an average group (d) by the number of participants N for a certain subtest of the WISC-R. These scores (d x N) are aggregated for all studies using the same subtest and divided by the sum of the total number of participants N. *g* Loadings of the subtests of the WISC-R of all ages were used. For computation of *g* loadings, we refer to the general method. For the subtest patterns, only data on the subtests of the WISC-R were used, since this test battery was administered in the majority of the studies. The effects shown are uncorrected for artifacts. The pattern will become more pronounced when all corrections have been made. Figure 7 clearly shows the scatter of the data points around the regression line. The data points with the largest distance from the regression line showing score gaps that are larger than predicted from *g* loadings are Digit Span, Information, and Vocabulary. Ellis (1963) hypothesized that mental

retardates can essentially be characterized by a deficit in short term memory. The large distance of the subtest Digit span from the regression line (Figure 7) clearly supports the hypothesis. Further, Jensen (1972) proposed a link between memory span and the ability to become self-sufficient. He argued that Digit Span is the essential subtest for defining one's associative learning ability. Low-IQ persons showing sufficient associative learning ability may seem much brighter and more self-sufficient than their IQs would lead one to expect. Consequently, these children will never be tested and their low IQs will never be noticed. There is considerable support for this theory, but most of the studies measuring a deficit in short term memory were conducted on institutionalized retardates (Jensen, 1972). Therefore, we decided to investigate the differences in subtest pattern between institutionalized retardates and non-institutionalized retardates. This resulted in 11 studies containing exclusively institutionalized mental retardates and 11 studies containing exclusively non-institutionalized mental retardates.

For the institutionalized mental retardates the Digit Span, Block Design, and Information subtests account for the strongest decrease in IQ (see Figure 8). In Figure 9, the results of the non-institutionalized show that the subtests Digit Span, Information and Similarities have the strongest decrease in IQ. First, it can be concluded that the effects for institutionalized mental retardates on the Digit Span subtest are somewhat smaller than the effects for non-institutionalized retardates. However, these differences are miniscule. Overall, the subtest patterns do not show large differences. So, it would seem that the people tested in our studies are a subset of the low-IQ population, namely those that are not self-sufficient and score low on Digit Span. The strongly outlying data point for Digit Span has a substantial effect on the correlation between *g* loadings and effects, leading to a correlation lower than +1.

Table 4

Studies of Correlations Between g Loadings and Mental Retardation

reference	test battery	*r*	*N*	mean age (range)
Box, Law, Moracco (1980)	WISC-R	.463	30	9.3[3] (7.1-11.5)
Bernard & Clarizio (1981)	WISC-R	.485	141	11.5 (4.0-19.0)
Goldman, Rosenfeld, & Rubin (1985)	WISC-R	-.016	41	13.5 (12.0-15.0)
	WAIS-R	.392	20	18.0 (15.0-21.0)
Beebe, Koepke, & Schooler (1978)	WISC-R	.272	127	12.1[3] (6.0-16.0
Covin & Sattler (1985)[3]	WISC-R	.632	147	13.5 (9.0-16.0)
Sexton & Street (1985)	WISC-R	.356	278	9.5 (6.0-15.9)
Engin, Hankins, Vance, & Wallbrown (1978)	WISC-R	.361	238	10.9 (5.7-16.7)
Kaufman & Van Hagen (1977)	WISC-R	.199	80	11.5 (6.3-16.9)
Lazarus & Nagle (1979)	WISC-R	.564	30	16.6 (16.0-16.11)
	WAIS	.627	30	16.1 (16.0-16.1)
Naglieri (1985)	WISC-R	.139	33	10.3 (8.1-12.5)
	K-ABC	.507	33	9.3 (8.1-12.5)
Catron & Catron (1977)	WISC	.566	62	13.6 (12.6-14.6)
	WISC-R	.674	62	13.6 (12.6-14.6)
Gironda (1977)	WISC	.343	20	11.0 (9-12.5)
	WISC-R	.499	20	14 (12.0-15.5)
Moore, Petersen, & Teeter (1982)	WISC-R	.495	189	11.1 (6.0-16.10)
	WISC-R	.451	189	11.1 (6.0-16.10)
Webster (1988)	WISC-R	.154	72	13.9 (13.6-14.11)
Bolen (1998)	WISC-III	.761	70	10.7 (6.4-13.8)
Henry & Wittman (1981)[1]	WISC-R	.188	40	9.0[4] (6.0-12.0)
Aldenkamp, Alpherts, Dekker, Forceville, & Schelvis (1992)	WAIS	-.252	56	31.5 (16.7-50.2)
Calhoun & Mayes (2003)	WISC-III	.135	53	4.8 (3-7)
Douglas, Houghton, Powell (1997)	K-ABC	.067	8	10.3 (4.9-40.9)
	K-ABC	-.242	32	22.1 (4.9-40.9)
Chang, Lau, & Tang (1996)	WISC-R (China)	.440	130	28.5 (17-48)
Davis & Reschley (1977)	WISC	.633	48	11.1 (7.9-16.1)
	WISC-R	.382	48	11.1 (7.9-16.1)
Thompson & Finley (1962)[2]	WISC	.737	309	10.9 (6.0-16.0)
Koenig (2008)[2]	WAIS-III	-.025	42	28.8 (17.5-39.6)
	WAIS-III	-.013	19	26.6 (15.0-38.1)
Wechsler (2003)	WISC-IV	.357	63	11.0 (6-16)
	WISC-IV	.709	57	11.0 (6-16)

Note. [1] Estimated age mean and age range. Based on grade levels (1 to 7). [2] Estimated age range. Based on studies conducted on the same amount of participants and the same test. [3] In this study, the data has been split up according to the sex of the participants. For the analysis, the group was taken in its entirety.

Table 5

Meta-analytical Results for Correlation Between Mental Retardation and g Loadings After Corrections for Reliability, Restriction of Range, and Imperfect Construct Validity

predictor	*K*	*N*	*r*	SD_r	rho-4	SD_{rho-4}	rho-5	%VE	80% CI
Mental Retardation[1]	34	2817	.43	.23	.63	.18	.70	42%	.32-.94
Mental Retardation[2]	32	2729	.45	.20	.67	.18	.74	57%	.44-.90
moderator									
Flynn effect 0-6 years[3]	17	1437	.45	.15	.61	.06	.68	93%	.53-.68
Flynn effect 7-27 years[3]	15	1459	.42	.25	.63	.28	.70	31%	.34-1.00
Educable Mentally Retarded (EMR)[4]	14	765	.43	.21	.61	.19	.68	56%	.36-.86

Note. [1] Meta-analytical results for correlations between *g* loadings and *d* (mental retardation). [2] Meta-analytical results for correlations between *g* loadings and *d* (mental retardation) without outliers. Two outliers were left out for further analysis: Aldenkamp, Alpherts, Dekker, Forceville, & Schelvis (1992) and Douglas, Houghton, & Powell (1997). [3] The dataset has been split into two groups based upon the estimated difference between the publication of a test battery and administration of the test battery [4] Meta-analytical results for correlations between *g* loadings and *d* (educable mental retardation). *K* = number of correlations; *N* = total sample size; *r* = mean observed correlation (sample size weighted); SD_r = standard deviation of observed correlation; rho-4 = observed correlation corrected for sampling error, unreliability, and range restriction); SD_{rho-4} = standard deviation of correlation; rho-5= true correlation (observed correlation corrected for sampling error, unreliability, range restriction, and imperfect construct validity); %VE = percentage of variance accounted for by artifactual errors; 80% CI = 80% credibility interval.

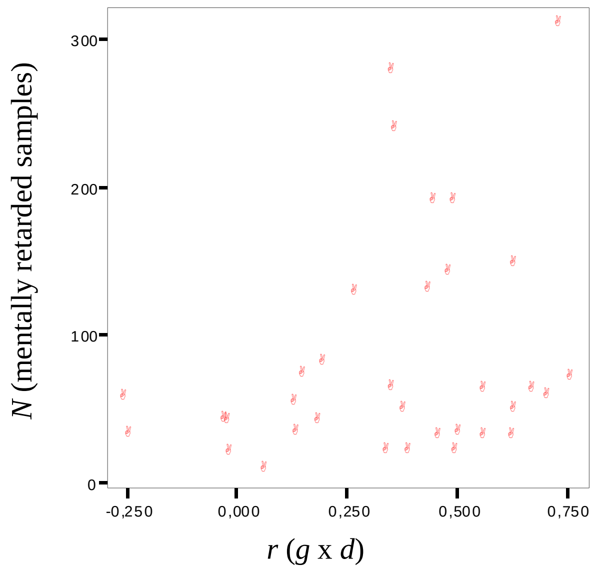

Figure 6

Scatter Plot of Reliability of Mental Retardation and Sample Size

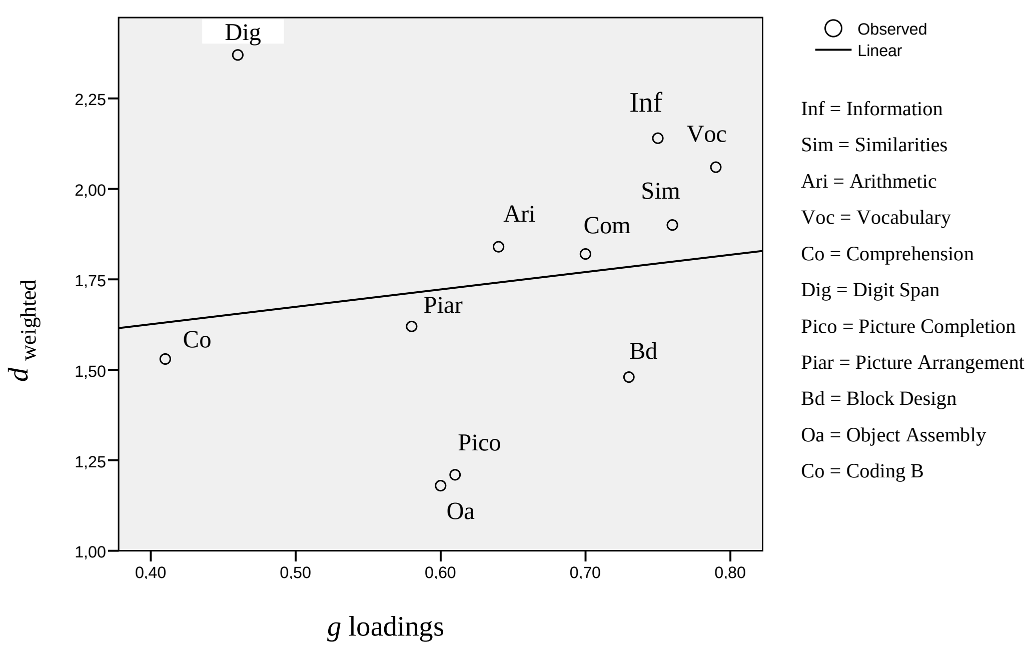

Figure 7

Scatter Plot of Subtest Pattern ($d_{weighted}$ x g) of the Mentally Retarded

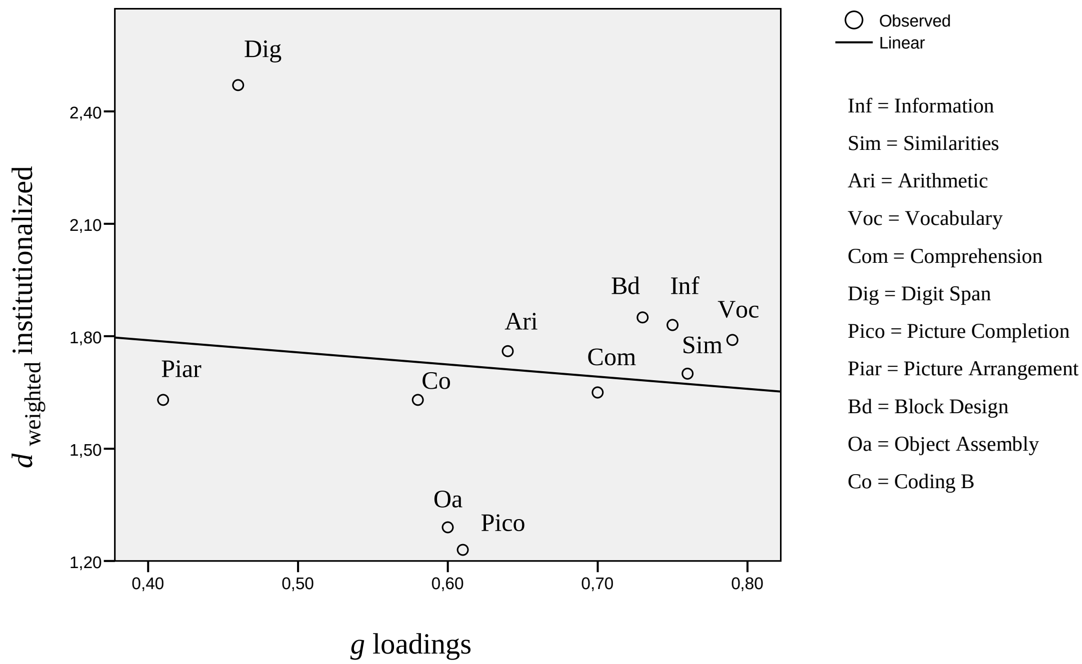

Figure 8

Scatter plot of Subtest Pattern ($d_{weighted}$ *x g) of the Institutionalized Mentally Retarded*

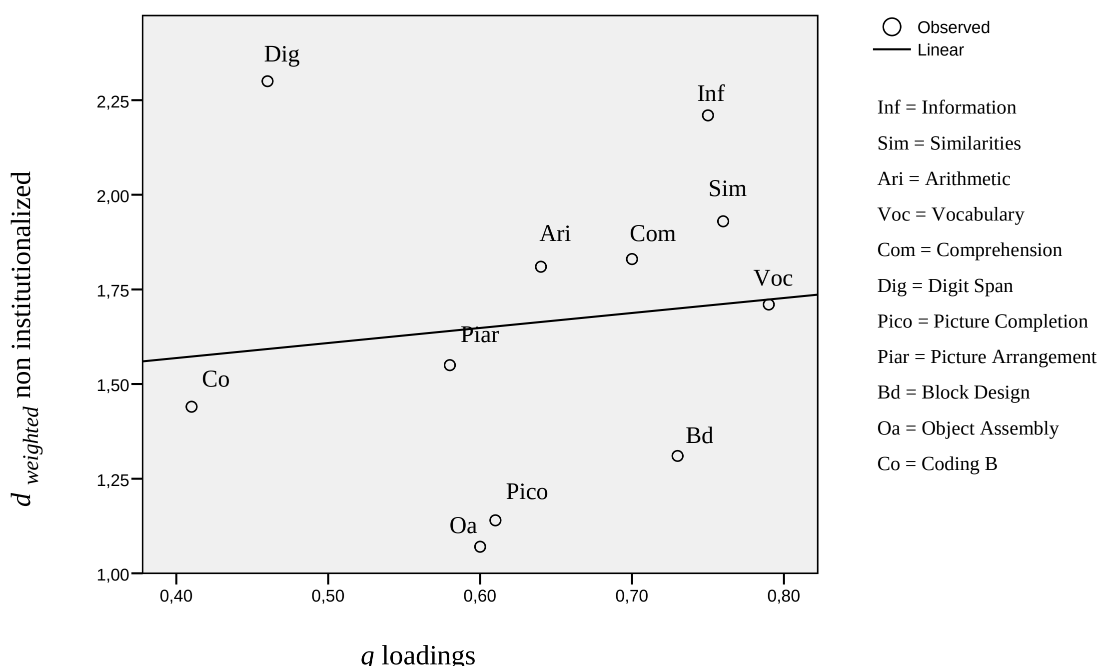

Figure 9

Scatter Plot of Subtest Pattern ($d_{weighted}$ x *g*) *of the Non-institutionalized Mentally Retarded*

Studies 3, 4, and 5: Three Exploratory Studies

To test whether there is a strong positive correlation between *g* loadings of tests and alcohol (*d*), whether there is a strong positive correlation between *g* loadings of tests and cocaine (*d*), and whether there is a weak positive correlation between *g* loadings of tests and depression *(d)*, we carried out three exploratory studies, using bare-bones meta-analytical techniques. A bare-bones psychometric meta-analysis estimates how much of the observed variance in findings across studies is due to sample size alone.

Method

Searching and screening studies. For the three exploratory studies, convenience samples using various versions of the Wechsler batteries were used. 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.

For the studies on cocaine we searched for the keyword cocaine in combination with the keywords: IQ, intelligence, cognitive impairment, cognitive disabilities, Wechsler, WISC, *g* factor, WAIS, and GMA. Also, the *American Journal of Drug and Alcohol Abuse* (1974-2009) was scanned for suitable studies.

Studies on depression were identified using the keyword depression in combination with the keywords: subtest, intelligence, Wechsler, WAIS, and WISC. Second, the reference lists of the collected studies were scanned for other suitable studies. Third, the tables of content of the *Journal of Clinical and Experimental Neuropsychology* 1998-2008 were scanned.

Application of the general inclusion rules produced nine papers yielding thirteen correlations between *g* and *d* (score differences between an alcohol abusing group and a comparison group), four papers yielding eight correlations between *g* and *d* (score differences between a cocaine abusing group and a comparison group), and five papers yielding five correlations between *g* and *d* (score differences between a depressed group and a standardization group).

Computation of score differences between an alcohol abusing group and a comparison group, a cocaine abusing group and a comparison group, and a depressed group and a standardization group. The goal of the present exploratory studies is to obtain the mean and the weighted correlations between a number of variables and *g*. In order to do so, score differences between the variable of interest and a comparison group have to be computed. We compared the

gifted and mentally retarded to the standardization group of the test battery in question. However, some studies about alcohol or cocaine abuse also compared their participants to a control group. Comparisons made with a control group are expected to be more accurate, especially when the group is large. Therefore, score differences between an alcohol/cocaine abusing group and the control group were also computed. Thus, score differences between an alcohol abusing group and a standardization group, a cocaine abusing group and a standardization group, and a depressed group and a standardization group were computed. Further, score differences between an alcohol abusing group and a control group, and a cocaine abusing group and a control group were computed.

Score differences were computed by subtracting the mean of the variable of interest, i.e. alcohol abusing group, cocaine abusing group, or depressed group, respectively, from the mean of the comparison group, i.e. the standardization group or control group, of the particular test in question (*d*), and then dividing the result by the *SD* of the standardization group. In order to obtain comparable groups, the *SD* of the standardization group was chosen.

Bare-bones meta-analytical techniques (Hunter & Schmidt, 1990, 2004) were applied to the resulting twelve r ($g \times d$)s (score differences between an alcohol abusing group and a comparison group), eight r ($g \times d$)s (score differences between a cocaine abusing group and a comparison group), and five r ($g \times d$)s (score differences between a depressed group and a standardization group), using the software package developed by Schmidt and Le (2004).

Results

Results of the exploratory study of score differences between an alcohol abusing group and a comparison group. The results of the studies on the correlation between *g* loadings and score differences between an alcohol abusing group and a comparison group are shown in Table 9. The Table presents data derived from nine studies with participants numbering a total of 434. It lists the reference for the study, the cognitive ability test used, the correlation between *g* loadings and score differences between an alcoholic group and a comparison group, the sample size, and the mean age together with the age range. Table 9 shows that correlations between the alcohol abusing group and the control group vary between modestly negative and modestly positive. The same pattern of correlations is found between the alcohol abusing group and the standardization group, but correlations vary even more strongly.

Results of the exploratory study of score differences between a cocaine abusing group and a comparison group. The results of the studies on the correlation between *g* loadings and score differences between a cocaine abusing group and a comparison group are shown in Table 10. The Table presents data derived from four studies, with participants numbering a total of 158.

It lists the reference for the study, the correlation between *g* loadings and score differences between a cocaine abusing group and a comparison group, and the sample size. It clearly shows that the correlations between a cocaine abusing group and a control group vary between modestly positive and modestly negative. However, the correlations are all modestly positive when the standardization group is used as comparison group.

Results of the exploratory study of score differences between a depressed group and a standardization group. The results of the studies on the correlation between *g* loadings and score differences between a depressed group and a standardization group are shown in Table 11. The Table presents data derived from five studies, with participants numbering a total of 268. It lists the reference for the study, the correlation between *g* loadings and score differences between a depressed group and a standardization group, and the sample size. It shows a large range of negative and positive correlations between the depressed group and the standardization group.

Bare-bones meta-analytical results of the exploratory studies. Table 12 shows the results of the three exploratory studies, using bare-bones meta-analytical techniques. A strong positive value was hypothesized for the relation between alcohol and *g*, and for the relation between cocaine and *g*. Promising results were found for the comparisons between the cocaine abusing group, the standardization group, and *g*: A rho= .61was found, using only one simple correction. Also, the percentage of variance accounted for by artifactual errors is 171%. As mentioned earlier, this is most plausibly a result of second order sampling so it can be interpreted as meaning that all of the variance has been explained. However, the results regarding the relation between alcohol and *g* are less optimistic. A rho=.08 was found with only 6% of variance accounted for by sampling error. A weak positive correlation was hypothesized for the relation between depression and *g*. Results show a rho=.12, which is in support of the hypothesis. The percentage of variance accounted for by sampling error is only 11%, but will probably increase when corrections for additional statistical artifacts are made.

However, the results for alcohol and cocaine are based upon the use of the standardization group as comparison group. As stated earlier, more informative results are expected using a control group as comparison group. Table 12 shows a rho=.-04 for the comparison between the cocaine abusing group and a control group as well as for the alcohol abusing group and a control group. Further, whereas the percentage of variance explained increases enormously from 6% to 79% for the alcohol abusing group, a huge decrease can be seen in the percentage of variance explained for the cocaine abusing group: from 171% to 34%. However, correcting for additional statistical artifacts, will, in all likelihood, substantially increase the amount of variance explained. Nonetheless, these results do not confirm the strong positive value hypothesized for the relations

between alcohol, and cocaine and *g*.

Prenatally exposed versus alcoholics. The strong positive value hypothesized between alcohol and *g* was not confirmed. In order to find an alternate explanation, we decided to look at the subtest pattern of the alcohol abusing group. First, the term alcohol abusing group refers to two different groups, namely, children exposed to alcohol prenatally, and people who personally abuse alcohol. A remarkable difference can be seen between the relation of alcoholics and *g*, and children prenatally exposed and *g* using standardization groups as comparison (see Table 9): While the correlations regarding alcoholics are all negative in sign, the correlations regarding children prenatally exposed to alcohol are all positive. Since only a few studies contain data using control groups, these correlations are acquired by comparisons made with the standardization groups, which are less informative than comparisons made with the control groups.

Post-hoc analysis: Subtest patterns of children prenatally exposed to alcohol and alcoholics. Following the differences in sign of the correlations for children prenatally exposed to alcohol as opposed to alcoholics, different subtest patterns are expected for these groups. $d_{weighted}$ is computed by multiplying the score differences between alcoholics/children prenatally exposed to alcohol and a standardization group (d) by the number of participants N for a certain subtest of the Wechsler battery. These scores (d x N) are aggregated for all studies using the same subtest and divided by the sum of the total number of participants N. *g* Loadings of the subtests of the WISC-III of all ages were used for children prenatally exposed to alcohol. For alcoholics, *g* loadings of the subtest of the WAIS of all ages were used. For computation of *g* loadings, we refer to general method.

Figure 10 shows the subtest pattern for alcoholics. This pattern reveals that alcohol has the largest effect on Performance IQ, specifically the Digit Symbol (Coding), Object Assembly, and Block Design subtests. The subtest pattern for children prenatally exposed to alcohol is shown in Figure 11. Results show that in this group alcohol has the largest effect on Verbal IQ, specifically on the Vocabulary, Comprehension, and Arithmetic subtests. When comparing children prenatally exposed to alcohol to alcoholics, the following observations can be made: Overall, the effect sizes for prenatally exposed children are larger on all subtests than the effect sizes for the alcoholics. It can be concluded that these children have a lower overall intellectual ability. Further, whereas alcohol has a more pronounced effect on Performance IQ for the alcoholics, it seems to show a larger effect on Verbal IQ for prenatally exposed children.

A similar distinction between prenatal exposure and drug abuse can be drawn for the cocaine abusing group. However, since only four studies were collected for this comparison, the amount of data was too small to obtain informative results separately for children prenatally

exposed to cocaine and adults abusing cocaine. However, we expect similar subtest patterns here since cocaine is hypothesized to have a more pronounced effect on brain development for prenatally exposed children.

Table 9

Studies of Correlations Between g Loadings and Score Differences Between an Alcohol Abusing Group and Comparison Groups

reference	tests	comparison group	*r*	*N*	mean age (range)
Aragón, Coriale, Fiorentino, Buckley, Gossage, Ceccanti, Michell, & May (2008)[1]	Italian WISC-R	Standardization group	.36[5]	23	6.13 (6-7)
		Control group	.13		
Blusewicz, Schenkenberg, Dustman, & Beck (1977)	WAIS	Standardization group	-.30[4]	20	33
		Control group	.20		
Hollan, Levi & Watson (1979)	WAIS	Standardization group	-.36[4]	79	41.95
Howell, Lynch, Platzman, Smith, & Coles (2006)[3]	WISC-III	Standardization group	.83[5]	128	15.1 (13-17)
		Control group	-.02		
O'Mahony & Doherty (1993)	WAIS-R	Standardization group	-.29[4]	43	55 (37-76)
Rasmussen, Horne, & Witol (2006)	WISC-III	Standardization group	.43[5]	50	9.5 (6-15.7)
Smith & Smith (1977)	WAIS	Standardization group	-.73[4]	40	51.1 (35-64)
		Control group	-.33		
Wechsler (1958)	WAIS	Standardization group	-.53[4]	31	(35-55)
Wechsler (1958)[2]	W-B I	Standardization group	-.38[4]	29	(36-55)

Note. [1] Data from the Italian WISC-R for the comparison group were not available. Therefore, data from the WISC-R were used to calculate the effect. [2] Data from the W-B I were not available for the standardization group. Therefore, we took a mean of 10 and a standard deviation of 3 to calculate the effect. [3] In this study, the alcohol abusing group has been split into two groups. The group was taken in its entirety for the analysis. Therefore, the weighted mean of the scaled scores were computed. [4] Studies regarding alcoholism with adults as research participants. [5] Studies regarding children prenatally exposed to alcohol.

Table 10

Studies of Correlations Between g Loadings and Score Differences Between a Cocaine Abusing Group and Comparison Groups

reference	tests	comparison group	*r*	*N*	mean age (range)
Easton & Bauer (1996)	WAIS-R[3]	Standardization group	.58	12	38 (15-61)
		Control group	.45		
Frank, Rose-Jacobs, Beeghly, Wilbur, Bellinger, & Cabral (2005)	WPPSI-R[1,2]	Standardization group	.65	91	4
		Control group	-.22		
Robinson, Heaton, & O'Malley (1999)	WAIS-R[3]	Standardization group	.45	30	28.7 (25-34)
		Control group	-.11		
Asanbe & Lockert (2006)	WISC-III[2]	Standardization group	.65	25	7.2 (6-8)
		Control group	.45		

Note. [1]Data from the WPPSI-R for the standardization group were not available. Therefore, data from the WPPSI were used to calculate the effect. [2] Studies regarding children prenatally exposed to cocaine. 3 Studies regarding cocaine abuse with adults as research participants.

Table 11

Studies of Correlations Between g Loadings and Score Differences Between a Depressed Group and a Standardization group

reference	tests	*r*	*N*	mean age (range)
Gorlyn, Keilp, O'Quendo, Burke, Sackeim, & Mann (2006)	WAIS-III	.28	121	38.40
Iverson, Turner, & Green (1999)	WAIS-R	-.51	70	34.3
Mandelli, Serretti, Colombo, Florita, Santoro, Rossini, Zanardi, & Smeraldi (2006)[1]	Italian WAIS-R	.32	51	52.3
Pernicano (1986)	WAIS-R	-.66	12	(18+)
Piedmont, Sokolove, & Fleming (1989)	WAIS-R	-.58	48	30.8 (16-85)

Note. [1] Data from the Italian WAIS-R for the standardization group were not available. Therefore, data from the WAIS-R were used to calculate the effect.

Table 12

Exploratory Bare-bones Meta-analytical Results for Correlations Between g Loadings and Alcohol Abuse-Standardization Score Differences, Alcohol Abuse-Control Score Differences, Cocaine Abuse-Standardization group Score Differences, Cocaine Abuse- Control Score Differences and Depression-Standardization Score Differences

predictor	*K*	*N*	rho	SD_{rho}	%VE
Alcohol Abuse-Standardization[1]	9	434	.08	.58	6%
Alcoholics-Standardization group[3]	6	242	-.43	.15	74%
Prenatally exposed-Standardized group[4]	3	201	.68	.20	11%
Alcohol Abuse-Control[2]	4	211	-.04	.16	79%
Cocaine Abuse-Standardization[1]	4	158	.61	.08	171%
Cocaine Abuse-Control[2]	4	158	-.04	.28	34%
Depression-Standardization[1]	5	268	.12	.42	11%

Note. [1] Bare-bones meta-analytical results: Score differences between the variable of interest, a standardization group and *g* loadings. [2] Bare-bones meta-analytical results: Score differences between the variable of interest, a matched control group, and *g* loadings. [3] Bare-bones meta-analytical results for score differences between alcoholics, a standardization group, and *g* loadings. [4] Bare-bones meta-analytical results for score differences between children prenatally exposed to alcohol, a standardization group, and *g* loadings. *K* = number of correlations; *N* = total sample size; rho = true correlation (observed correlation corrected for sample size); SD_{rho} = standard deviation of true correlation; %VE = percentage of variance accounted for by artifactual errors.

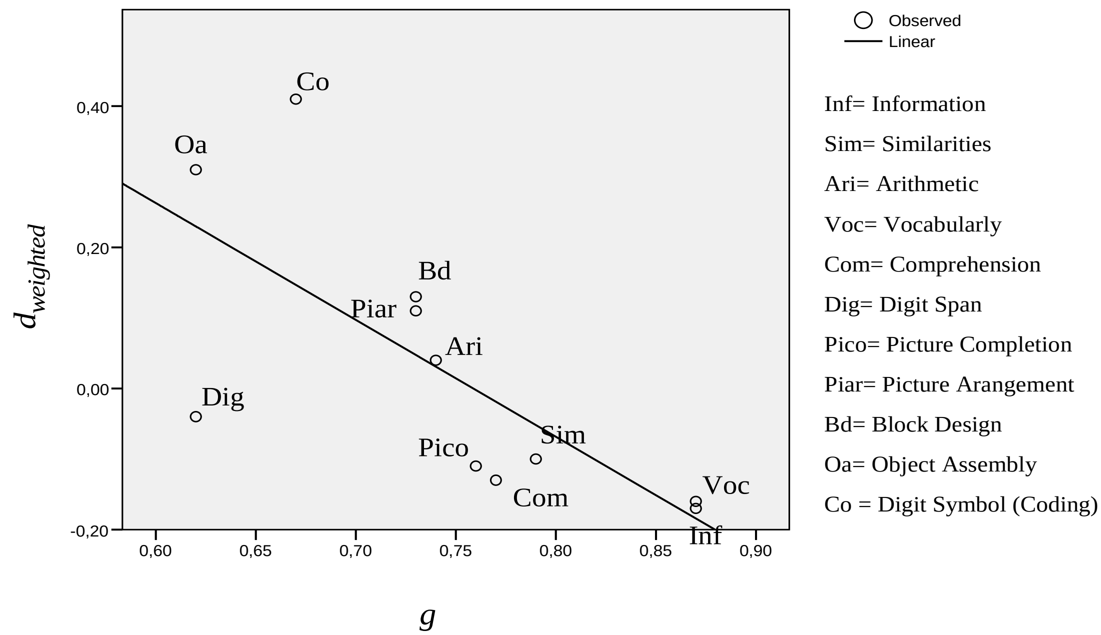

Figure 10

Pattern of Alcoholics Compared to a Standardized Group ($d_{weighted}$ x g)

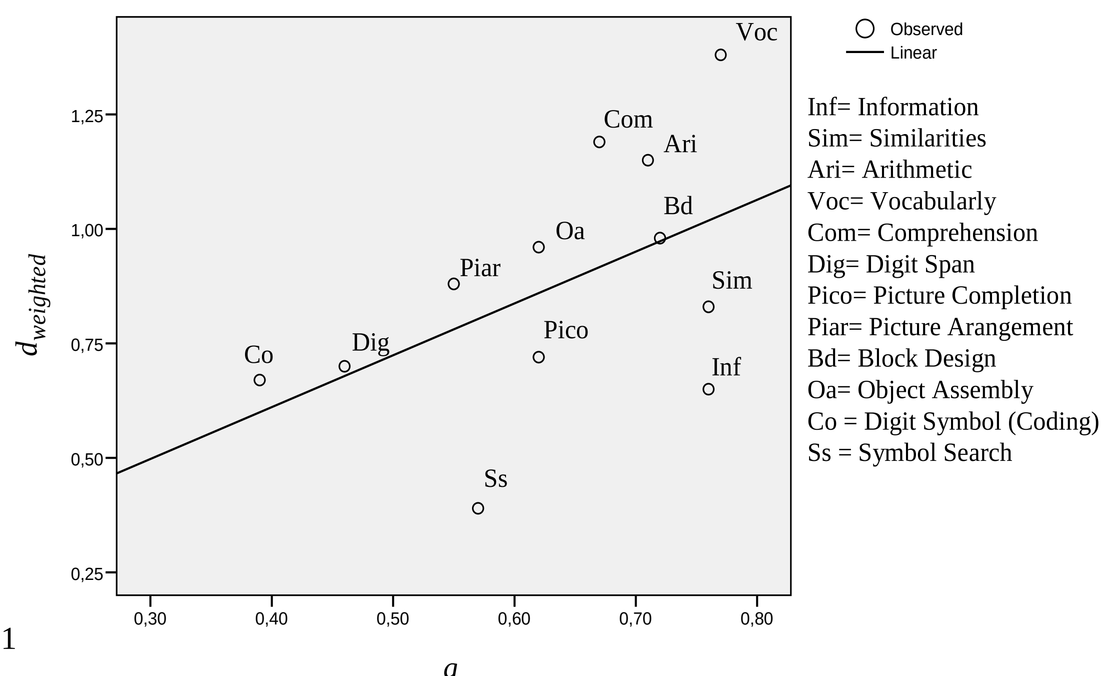

Figure 11

Subtest Pattern of Children Prenatally Exposed to Alcohol Compared to a Standardized Group ($d_{weighted}$ x g)

Discussion

Scores on IQ tests predict work achievement, school achievement, and thousands of other important real life variables. Numerous studies have shown that IQ test scores can increase over time as a result of treatment or educational interventions. The question is to what extent are these changes true changes in general mental ability? Various researchers have used the method of correlated vectors on a wide variety of research topics. When all the findings are taken into account, the following theory emerges: When there is a correlation of +1 between the *g* vector and a second vector, variation in scores on the variable is caused by biological factors. When the correlation is –1 the variation in scores on the variable is caused by non-biological factors. When the correlation is close to zero the variation in scores on the variable is caused by roughly comparable biological and non-biological factors. Therefore, a link was hypothesized between *g* loadings and a dimension of biological causation versus non-biological causation.

To test for biological causation, the correlations between scores on two variables and their *g* loadings were tested by carrying out two psychometric meta-analyses on variables with strong biological components: giftedness and mental retardation. In line with the expectations, the correlations with the vector of *g* loadings were perfectly positive for the IQ score differences between average individuals and gifted individuals. A possible moderating Flynn effect could be ruled out, meaning that the perfect positive correlation was not masked by an IQ score gain over generations. This provides even stronger support for the biological basis of giftedness.

As predicted, the correlations with the vector of *g* loadings were highly positive for the IQ score differences between average individuals and mentally retarded individuals. However, a perfectly positive correlation of +1 was not obtained. Several moderators were investigated in order to identify effects that could increase the value of the true correlation. However, the Flynn effect, Educable Mental Retardation, and the percentage of Blacks in the mentally retarded population were shown to have only negligible moderating effects.

In order to find an alternate explanation for the true correlation, we conducted post-hoc analyses to examine the subtest patterns between gifted people and average people and between mentally retarded people and average people. The perfect correlation of the gifted population resulted in a subtest pattern in which all data points are virtually on the regression line.

The data points for memory span of the mentally retarded population showed the largest distances to the regression line of *d* on *g*. Both institutionalized mental retardates and non-institutionalized mental retardates show low self-sufficiency. Most likely the non-institutionalized had already been detected by their teachers due to their low performance and

low self-sufficiency at school. Consequently, these children were tested in order to gain insight into their IQ scores. That is, their self-sufficiency at school was questioned and on the basis of the test results it was strongly recommended that these children should continue to be educated at special education institutes. This data point is a strong outlier and substantially lowers the correlation between *g* and *d*.

In line with the expectations, the two resulting correlations were essentially +1 for the gifted and very close to +1 for the mentally retarded. Taken together, these findings increase the plausibility of the hypothesized link between *g* loadings and a dimension of biological causation. Yet, further meta-analytical studies are recommended to increase the plausibility of the theory and to investigate the dimension of non-biological causation more thoroughly.

In addition, collections of studies were analyzed using exploratory psychometric meta-analytic techniques. The strong positive value hypothesized for the relation between alcohol abuse and general intelligence was not confirmed: A weak negative correlation was found. This might be due to the heterogeneity of the alcohol abusing group. When a distinction is made between children prenatally exposed to alcohol and alcoholics, different effects of alcohol on cognitive functioning are found. First, alcohol has a more pronounced effect on general intelligence for prenatally exposed children as opposed to adult alcohol abuse. This might be due to the disruption of the brain when it is most vulnerable to external influences, i.e. at the early stages of development. These impairments are measured using cognitive tests resulting in a lower overall IQ and lower scores on all subtests of an IQ battery for children prenatally exposed to alcohol when compared to normal children. Further, whereas alcohol affects mainly the verbal abilities for prenatally exposed children, it mainly affects the non-verbal abilities of adult abusers. Future research investing the relation between alcohol and general intelligence should separate these groups.

The positive value hypothesized for the relation between cocaine and general intelligence was not confirmed: A weak negative correlation was found. Like the alcohol abusing group, the cocaine abusing group can be separated into prenatally exposed children and abusing adults. Due to the small amount of studies, results for these groups could not be obtained. Future research should reveal whether the relation between cocaine and general intelligence is similar to that for alcohol.

The weak positive value hypothesized for the relation between depression and general intelligence was confirmed. However, research has shown inconsistencies in results regarding the improvement of cognitive functioning after remission. Future research should examine whether this might be related to the severity of the depressive disorder.

Practical implications

The present study makes a strong empirical contribution to the important discussion as to which interventions raise *g* and which do not. IQ tests are important instruments for selection and placement. Consequently, in today's society low IQ scores may lead to placement in special education whereas high IQ scores may lead to the placement in advanced training programs. The solution to lower the IQ gap between disadvantaged children and advantaged children referring to intelligence was assumed to be compensatory education. Success of compensatory education was measured by IQ gains and improvement in scholastic achievement. However, increases due to schooling show very little or no transfer to general intelligence, suggesting that the massive sums spent on such programs have little chance of success. These resources would be better spent on interventions based on biological variables that have been shown to be linked to the *g* factor. Drug abuse such as alcoholism and cocaine addiction have proven to severely damage the brain resulting in cognitive impairments, especially for prenatally exposed children, and there may be a link to *g*. However, it is clear that the emotional, social, and economic costs of prenatal exposure to drugs are lifelong for affected individuals, their families, and the community. Therefore, government funding of mandatory rehabilitation programs for pregnant women who are severely addicted to alcohol and/or cocaine merit serious consideration.

Limitations of the studies

In this study we conducted two psychometric meta-analyses and three exploratory meta-analyses in order to test the link between a dimension of biological causation and *g*. It was hypothesized that true correlations between non-biological factors and *g* would lead to a perfectly negative value of -1. In our study we did find negative correlations, namely between alcohol abuse by adults, cocaine abuse by adults, and *g*. However, the cause of the negative correlation did not seem to be the fact that alcohol is non-biological, since drug abuse has highly damaging biological effects. The negative correlation seems to stem from high score differences at the Performance Scale combined with the Performance Scale being less *g* loaded than the Verbal Scale. Stronger effects correlated with lower *g* loadings lead to a negative value of *r*. Thus, a negative value does not necessarily imply a non-biological link to *g*. This is a limitation of the use of the method of correlated vectors, which is brought into clear focus by the present study.

The Flynn effect was tested as a moderator for giftedness as well as for mental retardation. We decided to split the dataset into two groups based upon the estimated difference in years between the publication of a test battery and administration of the test battery. This

resulted in a group with an estimated difference of 0-6 years, and a second group with an estimated difference of 9-27 years for giftedness and 7-27 years for mental retardation. However, the estimated difference of the second group (respectively 18 years and 20 years) is much larger than the estimated difference of the first group (6 years). In order to obtain more accurate results it could be argued that the second group needs to be split into two separate groups. This would result in three groups with more comparable estimated differences in years.

In this study, we decided to apply the method of correlated vectors using Pearson *r* which has the advantage of comparing the outcomes to earlier conducted meta-analyses also using Pearson *r*. However, we did not investigate whether the use of Spearman's rho would alter the robustness of the method. Future research should conduct meta-analyses using Spearman's rho and Pearson *r* separately in order to compare the results.

Conclusion

Based on a large number of empirical studies and employment of the method of correlated vectors we developed a data-driven theory of a link between *g* loadings and a dimension of biological causation versus non-biological causation. Testing this theory by two meta-analytical studies and three exploratory studies yielded a much stronger empirical basis for the theory and thereby increased its plausibility. Given the strong positive correlations of giftedness and mental retardation with *g*, a link between general intelligence and biological causation is supported. Further meta-analytical studies are recommended to increase the plausibility of the theory still further.

References

References marked with an asterisk indicate studies included in the meta-analyses and in the exploratory studies.

Alcock, K. J., & Bundy, D. A. P. (2001). The impact of infectious disease on cognitive development. In R. J. Sternberg & E. L. Grigorenko (Eds.), *Environmental effects on cognitive abilities.* Mahwah, NJ: Erlbaum.

Alessandri, S. M., Sullivan, M. W., Imaizumi, S., & Lewis, M. (1993). Learning and emotional responsitivity in cocaine-exposed infants. *Developmental Psychology, 29,* 989-997.

*Aragón, A. S., Giovanna, C., Fiorentino, D., Kalberg, W. O., Buckley, D., Gossage, J. O., Ceccanti, M., Mitchell, E. R., & May, P. A. (2008). Neuropsychological characteristics of Italian children with Fetal Alcohol Spectrum Disorders. *Clinical and Experimental Research, 32,* 1909-1919.

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

Ashton, M. C., & Lee, K. (2005). Problems with the method of correlated vectors. *Intelligence, 33*, 431-444.

Badaruddoza, & Afzal, M. (1993). Inbreeding depression and intelligence quotient among North Indian children. *Behavior Genetics, 23,* 343-347.

Berry, J., Van Gorp, W. G., Herzberg, D. S., Hinkin, C., Boone, K., Steinman, L., & Wilkins, J. N. (1993). Neuropsychological deficits in abstinent cocaine abusers: Preliminary findings after two weeks of abstinence. *Drug and Alcohol Dependence, 32,* 231-237.

*Blusewicz, M. J., Schenkenberg, T., Dustman, R. E., & Beck, E. (1977). WAIS performance in young normal, young alcoholic, and elderly normal groups: An evaluation of organicity and mental aging indices. *Journal of Clinical Psychology, 33,* 1141-1153.

*Bolen, L. M. (1998). WISC-III score changes for EMH students. *Psychology in the Schools, 35,* 327-332.

Braden, J. P. (1989). Fact or artifact? An empirical test of Spearman’s hypothesis. *Intelligence, 13*, 149-155.

*Brown, S. W., & Yakimowski, M. E. (1987). Intelligence scores of gifted students on the WISC-R. *Gifted Child Quarterly, 31,* 130-134.

*Brown, S. W., Hwang, M. T., Baron, M., & Yakimowski, M. E. (1991). Factor analysis of responses to the WISC-R for gifted children. *Psychological Reports, 69,* 99-107.

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. Cambridge University Press.

Caruso, D. R., Taylor, J. J., & Detterman, D. K. (1982). Intelligence research and intelligent policy. In D.K. Detterman & R.J. Sternberg (Eds.), *How and how much can intelligence be increased* (pp. 45-65). Norwood, NJ: Ablex Publishing Corporation.

*Catron, D. W., & Catron, S. S. (1977). WISC-R vs. WISC: A comparison with educable mentally retarded children. *Journal of School Psychology, 15,* 264-266.

*Chae, P. K., Kim J. H., & Noh, K. S. (2003). Diagnosis of ADHD among gifted children in relation to KEDI-WISC and T.O.V.A. performance. *Gifted Child Quarterly, 47,* 192-201.

Christian, K., Bachnan, H. J, & Morrison, F. J. (2001). Schooling and cognitive development. In R. J. Sternberg & E. L. Grigorenko (Eds.), *Environmental effects on cognitive abilities* (pp. 287-335). Mahwah, NJ: Erlbaum.

*Clarizio, H., & Bernard, R. (1981). Recategorized WISC-R scores of learning disabled children and differential diagnosis. *Psychology in the Schools, 18,* 5-12.

Colom, R., Juan-Espinosa, M., & García, L. F. (2001). Secular increase in test scores is a "Jensen effect". *Personality and Individual Differences, 30,* 553-559.

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

Colom, R., Jung, R. E., & Haier, R. J. (2006). Distributed brain sites for the *g*-factor of intelligence. *NeuroImage, 31,* 1359-1365.

*Covin, T. M., & Sattler, J. M. (1985). A longitudinal study of the Stanford-Binet and WISC-R with special education students. *Psychology in the Schools, 22,* 274-276.

Craddock, N., & Jones, I. (1999). Genetics of bipolar disorder. *Journal of Medical Genetics, 36,* 585-594.

*Easton, C., & Bauer, L.O. (1996). Neuropsychological correlates of urine toxicology results. *Progess in Neuro-Psychopharmacology and Biological Psychiatry, 20,* 969-982.

Ellis, N. R. (1963). *The stimulus trace and behavioral inadequacy* . In N. R. Ellis (Eds.), Handbook of mental deficiency (pp 134-158). New York: McGraw-Hill.

Eysenck, H. J., & Barrett, P. (1985). Psychophysiology and the measurement of intelligence. In C. R. Reynolds & P. C. Wilson (Eds.), *Methodological and statistical advances in the study of individual differences* (pp. 1-49). New York: Plenum.

*Fishkin, A. S., & Kampsnider, J. J. (1996). Exploring the WISC-III as a measure of giftedness. *Roeper Review, 18,* 226-231.

Flynn, J. R. (1999a). Evidence against Rushton: The genetic loading of WISC-R subtests and the causes of between-group IQ differences. *Personality and Individual Differences, 26,* 373-379.

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

Flynn, J. R. (2000). IQ gains, WISC subtests, and fluid *g*: *g* theory and the relevance of Spearman's hypothesis to race. In G. B. J. Goode (Ed.), *The nature of intelligence* (pp. 202-227). New York: Wiley.

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

Forceville, E. J. M., Dekker, M. J. A., Aldenkamp, A. P., Alpherts, J. C. E., & Schelvis, A. J. (1992). Subtest profiles of the WISC-R and WAIS in mentally retarded patients with epilepsy. *Journal of Intellectual Disability Research, 36,* 45-59.

*Frank, D. A., Rose-Jacobs, R., Beeghly, M., Wilbur, M., Bellinger, D., & Cabral, H. (2005). Level of prenatal cocaine exposure and 48-month IQ: Importance of preschool enrichment. *Neurotoxicology and Teratology, 27,* 15-28.

Galton. F. (1883). *Inquiries into human faculty and its development*. London: Macmillan.

Gershon, E. S., & Nurnberger, J. I. (1995). Bipolar illnes. In J. M. Oldman & M. B. Riba (Eds.). *American Psychiatric Press review of psychiatry* (Vol.14). Washington, DC: American Psychiatric.

*Gironda, R. J. (1977). A comparison of WISC and WISC-R results of urban educable mentally retarded students. *Psychology in the Schools, 14,* 271-275.

*Gorlyn, M., Keilp, J. G., Oquendo, M. A., Burke, A. K., Sackeim, H. A., & Mann, J. J. (2006). The WAIS-III and major depression: Absence of VIQ/PIQ differences. *Journal of Clinical and Experimental Neuropsychology, 28*, 1145-1157.

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

Gottfredson, L. S. (2003). *The science and politics of intelligence in gifted education.* In N. Colangelo and G. A. Davis (Eds). The handbook of gifted education (3rd ed.; 24-40).

Boston: Allyn & Bacon.

Grantham-McGregor, S., Ani, C., & Fernald, L. (2001). The role of nutrition in intellectual development. In R. J. Sternberg & E. L. Grigorenko (Eds.), *Environmental effects on cognitive abilities.* Mahwah, NJ: Erlbaum.

Groth-Marnat, G. (1997). The Wechsler intelligence scales. In G. Groth-Marnat (Ed.), *Handbook of psychological assessment, third edition* (pp. 132–204). New York: John Wiley & Sons.

Haier, R. J., Siegel, B., Tang, C., Abel, L., & Buchsbaum, M. S. (1992). Intelligence and changes in regional cerebral glucose metabolic rate following learning. *Intelligence, 16*, 415-426.

*Henry, S. A., & Wittman, R. D. (1981). Diagnostic implications of Bannatyne's recategorized WISC-R scores for identifying learning disabled children. *Journal of Learning Disabilities, 14*, 517-520.

*Holland, T. R., & Levi, M. (1979). Multivariate structure of associations between verbal and nonverbal intelligence among brain-damaged, schizophrenic, neurotic, and alcoholic patients. *Journal of Abnormal Psychology, 88,* 354-360.

*Howell, K. K., Lynch, M. E., Platzman, K. A., Smith, G. H., & Coles, C. D. (2006). Prenatal alcohol exposure and ability, academic achievement, and school functioning in adolesence: A longitudinal follow-up. *Journal of Pediatric Psychology, 31,* 116-126.

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.

*Iverson, G. L., Turner, R. A., & Green, P. (1999). Predictive validity of WAIS-R WIQ-PIQ splits in persons with major depression. *Journal of Clinical Psychology, 55,* 519-524.

Jacobson, S. W., Jacobson, J. L., Sokol, R. J., Martier, S. S., & Chiodo, L. M. (1996). New evidence for neurobehavioral effects of in utero cocaine exposure. *Journal of Pediatrics, 129,* 581–590.

Jensen, A. R. (1972). Genetics and education. New York: Harper and Row.

Jensen, A. R. (1980) Bias in Mental Testing. London: Methuen.

Jensen, A. R. (1994). Psychometric *g* related to differences in head size. *Personality and Individual Differences, 17*, 597-606.

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

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

Kanaya, T., Scullin, M. H., & Ceci, S. J. (2003). The Flynn effect and US policies: The impact of rising IQ scores on American Society via mental retardation diagnoses. *American Psychologist, 58*, 778-790.

*Karnes, F. A., & Brown, K. E. (1980). Factor analysis of the WISC-R for the gifted. *Journal of Educational Psychology, 72,* 197-199.

*Kaufman, A. S., & Van Hagen, J. (1977). Investigation of the WISC-R for use with retarded children: Correlation with the 1972 Stanford-Binet and comparison of WISC and WISC-R profiles. *Psychology in the Schools, 14,* 10-14.

Kluger, A., & Goldberg, E. (1990). IQ patterns in affective disorder, lateralized and diffuse brain damage. *Journal of Clinical and Experimental Neuropsychology, 12,* 182–194.

*Koenig, K. A. (2007). *Performance of elementary cognitive tasks in Down Syndrome and Fragile X Syndrome*. Unpublished doctoral dissertation, Case Western Reserve University, Cleveland, OH.

Kulik, J. A., Bangert-Drowns, R. L., & Kulik, C. C. (1984) Effectiveness of coaching for aptitude tests. *Psychological Bulletin, 95,* 179-188.

*Law, J. G., Box, D., & Moracco, J. C. (1980). A validation study of recategorized Wisc-R scores of learning disabled children. *Education, 101,* 195-199.

Lee, K. H., Choi, Y. Y., Gray, J. R., Cho, S. H., Chae, J.-H., Lee, S., & Kim, K. (2006). Neural correlates of superior intelligence: Stronger recruitment of posterior parietal cortex. *NeuroImage*, *29,* 578-586.

*Linville, J. N., & Rust, J. O. (1999). The information and picture completion dyad of the WISC-III as a screening test for gifted. *Journal of Instructional Psychology, 26,* 98-104.

Loehlin, J. C, Lindzey, G., & Spuhler, J. N. (1975). *Race differences in intelligence.* San Francisco: W. H. Freeman.

*Lucito, L., & Gallagher, J. (1960). Intellectual patterns of highly gifted children on the WISC. *Peabody Journal of Education, 38,* 131-136.

*Macmann, G. M., Plasket, C. M., Barnett, D. W., & Siler, F. S. (1991). Factor structure of the WISC-R of superior intelligence. *Journal of School Psychology, 29,* 19-36.

*Mandelli, L., Serretti, A., Colombo, C., Florita, M., Santoro, A., Rossini, D., Zanardi, R., & Smeraldi, E. (2006). Improvement of cognitive functioning in mood disorder patients with depressive symptomatic recovery during treatment: An exploratory analysis. *Psychiatry and Clinical Neurosciences, 60,* 568-604.

*Mayes, S. D., & Calhoun, S. L. (2003). Analysis of WISC-III, Stanford-Binet: 4, and academic achievement test scores in children with autism. *Journal of Autism and Developmental Disorders, 33(3),* 329-341.

Must, O., Must, A., & Raudik, V. (2003). The secular rise in IQs: In Estonia the Flynn effect is not a Jensen effect. *Intelligence*, *167,* 1-11.

*Nagle, R. J., & Lazarus, S. C. (1979). The comparability of the WISC-R and WAIS among 16-year-old EMR children. *Journal of School Psychology, 17,* 362-367.

*Naglieri, J. A. (1985). Use of the WISC-R and K-ABC with learning disabled, borderline mentally retarded, and normal children. *Psychology in the Schools, 22,* 133-141.

Nagoshi, G. T., & Johnson, R. C. (1986). The ubiquity of *g*. *Personality and Individual Differences, 7,* 201-207.

*Namy, E. (1967). Intellectual and academic characteristics of fourth grade gifted and pseudogifted students. *Exceptional Children, 34,* 15-18.

Nwuga, V. C. B. (1977). Effect of severe kwashiorkor on intellectual development among Nigerian children. *American Journal of Clinical Nutrition*, *30*, 1423-1430.

O'Malley, S., Adamse, M., Heaton, R. K., & Gawin, F. H. (1992). Neuropsychological impairment in chronic cocaine abusers. *Drug Alcohol Abuse*, *18*, 131-144.

O'Malley, S., & Gawin, F. H. (1990). Abstinence symptomatology and neuropsychological impairment in chronic cocaine abusers. In J. W. Spencer and J. J. Boren (Eds.), *Residual Effects of Abused Drugs on Behavior* (pp 179-191). U.S. Government Printing Office: Washington D.C.

*Pernicano, K. M. (1986). Score differences in WAIS-R scatter for schizophrenics, depressives, and personality disorders: A preliminary analysis. *Psychological Reports, 59,* 539-543.

*Piedmont, R. L., Sokolove, R. L., & Fleming, M. Z. (1989). Discriminating psychotic and affective disorders using the WAIS-R. *Journal of Personality Assessment, 53,* 739-748.

*Powell, L., Houghton, S., & Douglas, G. (1997). Comparison of etiology-specific cognitive functioning profiles for individuals with Fragile X and individuals with Down Syndrome. *The Journal of Special Education*, *31*, 362-376.

Prokosch, M. D., Yeo, R. A., & Miller, G. F. (2005). Intelligence tests with higher *g*-loadings show higher correlations with body symmetry: Evidence for a general fitness factor mediated by developmental stability. *Intelligence*, *33*, 203-213.

Rae, C., Scott, R. B., Thompson, C. H., Kemp, G. J., Dumughn, I., Styles, P., Tracey, I., & Radda, G. K. (1996). Is pH a biochemical marker of IQ? *Proceedings of the Royal Society* (London), *263*, 1061-1064.

*Rasmussen, C., Horne, K., & Witol, A. (2006). Neurobehavioral functioning in children with Fetal Alcohol Spectrum Disorder. *Child Neuropsychology, 12,* 453-468.

*Reams, R., Chamrad, D., & Robinson, N. (1990). The race is not necessarily to the swift: Validity of WISC-R bonus points for speed. *Gifted Child Quarterly, 34,* 108-110.

*Reschly, D.J., & Davis, R.A.(1977). Comparability of WISC and WISC-R scores among borderline and mildly retarded children. *Journal of Clinical Psychology, 33,* 1045-1048.

Richardson, S. A., Birch, H. G., & Hertzig, M. (1973). School performance of children who were severely malnourished in infancy. *American Journal of Mental Deficiency, 77,* 623-632.

*Robinson, J. E., Heaton, R. K., & O'Malley, S. S. (1999). Neuropsychological functioning in cocaine abusers with and without alcohol dependence. *Journal of the International Neuropsychological Society, 5,* 10-19.

*Rubin, H. H., Goldman, J. J., & Rosenfeld, J. G. (1985). A comparison of WISC-R and WAIS-R IQS in a mentally retarded residential population. *Psychology in the Schools, 22,* 392-397.

Sabatino, D. A., Spangler, R. S., & Vance, H. B. (1995). The relationship between the Wechsler Intelligence Scale for Children-Revised and the Wechsler Intelligence Scale for Children-III scales and subtests with gifted children. *Psychology in the Schools, 32,* 18-23.

Sackeim, H.A., Freeman, J., McElhiney, M., Coleman, E., Prudic, J., & Devanand, D.P. (1992). Effects of major depression on estimates of intelligence. *Journal of Clinical and Experimental Neuropsychology, 14*, 268–288.

*Sapp, G. L., Chissom, B., & Graham, E. (1985). Factor analysis of the WISC-R for gifted students: A replication and comparison. *Psychological Reports, 57,* 947-951.

*Sattler, J. M., & Covin, T. M. (1986). Comparison of the Slosson Intelligence Test, revised norms, and WISC-R for children with learning problems and for gifted children. *Psychology in the Schools, 23,* 259-264.

Scarr, S. (1981). *Race, social class, and individual differences*. In I.Q. Hillsdale, NJ: Erlbaum.

Scarr-Salapatek, S. (1971). Race, social class, and IQ. *Science, 74,* 1285-1295.

Schafer, E. W. P. (1985). Neural adaptability: A biological determinant of *g* factor intelligence. *Behavioral and Brain Sciences, 8,* 240-241.

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. University of Iowa, Department of Management & Organization, Iowa City, Iowa, 42242.

Schoenemann, P. T. (1997). The evolution of the human brain and its relationship to behavior. Unpublished doctoral dissertation. Berkeley: University of California.

*Schooler, D. L., Beebe, M. C., & Koepke, T. (1978). Factor analysis of WISC-R scores for children identified as learning disabled, educable mentally impaired, and emotionally impaired. *Psychology in the Schools, 15,* 478-485.

Schull, W. J., & Neel, J. V. (1965). *The effects of inbreeding on Japanese children*. New York: Harper & Row.

*Sevier, R. C., & Bain, S. K. (1994). Comparison of WISC-R and WISC-III for gifted students. *Roeper Review, 17,* 39-42.

*Sexton, L. C., & Street, S. (1985). A comparison of WISC-R score profiles of students placed in special education and those referred, but not placed. *Psychology in the Schools, 22*, 142-145.

*Smith, H. H., & Smith, L. S. (1977). WAIS functioning of cirrhotic and non-cirrhotic alcoholics. *Journal of Clinical Psychology, 33,* 309-313.

Sobczak, S., Honig, A., van Duinen, M. A., & Riedel, W. J. (2002). Serotonergic dysregulation in bipolar disorders: A literature review of serotonergic challenge studies. *Bipolar Disorders, 4*, 347-356.

*Spangler, R. S. & Sabatino, D. A. (1995). Temporal stability of gifted children's intelligence. *Roeper Review, 17,* 207-210.

Spearman, C. (1923). *The nature of 'intelligence' and the principles of cognition***.** London: Macmillan.

Speer, S. K., Hawthorne, L. W., & Buccellato, L. (1986). Intellectual patterns of young gifted children on the WPPSI. *Journal for the Education of the Gifted*, *10,* 57-63.

Streissguth, A. P. (1997). *Fetal alcohol syndrome: A guide for families and communities*. Baltimore, MD: Paul H. Brookes Pub.

*Swanson, H. L., Brandenburg-Ayres, S., & Wallace, S. (1989). Construct validity of the K-ABC with gifted children. *Journal of Special Education, 23,* 342-352.

*Tang, C. S., Lau, B. H., & Chang, S. S. (1996). Factor structure of the Chinese version of the WAIS-R for Chinese adults in the lowest percentiles of IQ. *Journal of Clinical Psychology, 52,* 345-355.

*Teeter, A., Moore, C. L., & Petersen, J. D. (1982). WISC-R verbal and performance abilities of native american students referred for school learning problems. *Psychology in the Schools, 19,* 39-44.

te Nijenhuis, J., & Jongeneel-Grimen, B. (2007). *Can people get smarter: Three psychometric meta-analyses and three exploratory meta-analyses on g loadings and IQ scores*. Unpublished manuscript, University of Amsterdam, Amsterdam, NL.

te Nijenhuis, J., & van der Flier, H. (2007). The secular rise in IQs in the Netherlands: Is the Flynn effect on g? *Personality and Individual Differences, 43,* 1259-1265.

te Nijenhuis, J., & van der Flier, H. *Is the Flynn effect on g?: A meta-analysis*. Manuscript submitted for publication.

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

*Thompson, J. M., & Finley, C. J. (1962). A further comparison of the intellectual patterns of gifted and mentally retarded children. *Exceptional children, 28,* 379-381.

Streissguth, A. P. (1997). *Fetal alcohol syndrome: A guide for families and communities*. Baltimore, MD: Paul H. Brookes Pub.

van der Molen, M. J. (2009). *Working memory in children with mild intellectual disabilities: Abilities and training potential.* Unpublished doctoral dissertation, University of Amsterdam, The Netherlands.

van Haasen, P. P., de Bruyn, E. E. J., Pijl, Y. J., Poortinga, Y. H., Lutje Spelberg, H. C., Vander Steene, G., Coetsier, P., Spoelders, R., & Stinissen, J. (1986). *WISC-R: Wechsler Intelligence Scale for Children-Revised; Nederlandstalige uitgave* [WISC-R: Wechsler Intelligence Scale for Children-Revised; Dutch edition]. Lisse, the Netherlands: Swets.

van IJzendoorn, M. H., Juffer, F., & Klein Poelhuis, C. W. (2005). Adoption and cognitive development: A meta-analytic comparison of adopted and nonadopted children's IQ and school performance. *Psychological Bulletin, 131*, 301-316.

*Vance, H., Hankins N., Wallbrown, F., & Engin, A. (1978). Analysis of cognitive abilities for mentally retarded children on the WISC-R. *The psychological record, 28,* 391-397.

Vernon, P. A. (1992). Intelligence, reaction times, and peripheral nerve conduction velocity. *Intelligence, 16*, 273-288.

Vernon, P. A. (1993). Intelligence and neural efficiency. In D. K. Detterman (Ed.), *Current topics in human intelligence,* Vol. 3 (pp. 171-187). Norwood, NJ: Ablex.

Wallace, A. K., Kennedy, M. S., Van De Riet, H., Van De Riet, V., Smith, H., Rapp, D., & Paine, W. (1960). A multidimensional study of mathematically gifted adolescents. *Child Development, 31,* 655-666.

*Watkins, M. W., Greenawalt, C. G., & Marcell, C. M. (2002). Factor structure of the Wechsler Intelligence Scale for Children-Third Edition among gifted students. *Educational and Psychological Measurement, 62,* 164-172.

Watkins, W. E., & Pollitt, E. (1997). "Stupidity or worms": Do intestinal worms impair mental performance? *Psychological Bulletin, 121*, 171-191.

*Webster, R. E. (1988). Statistical and individual temporal stability of the WISC-R for cognitively disabled adolescents. *Psychology in the Schools, 25,* 365-372.

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

*Wechsler, D. (2003). *Wechsler Intelligence Scale for Children-Fourth Edition.* San Antonio, TX: The Psychological Corporation.

Weinberg, R. A. (1989). Intelligence and IQ: Landmark issues and great debate. *American Psychologist, 44,* 98-104.

*Wheaton, P. J., & Vandergriff, A. F. (1978). Comparison of WISC and WISC-R scores of highly gifted students in public school. *Psychological Reports, 43*, 627-630.

Wicherts, J. W., Dolan, C. V., Oosterveld, P., van Baal, G. C. V., Boomsma, D. I., & Span, M. M. (2004). Are intelligence tests measurement invariant over time? Investigating the nature of the Flynn effect. *Intelligence, 32,* 509-537.

Wickett, J. C., Vernon, P. A., & Lee, D. H. (1994). In *vivo* brain size, head perimeter, and intelligence in a sample of healthy adult females. *Personality and Individual Differences, 16*, 831-838.

*Wilkinson, S. C. (1993). WISC-R profiles of children with superior intellectual ability. *Gifted Child Quarterly, 37,* 84-91.

Zillmer, E.A., Ball, J.D., Fowler, P.C., Newman, A.C., & Stutts, M.L. (1991). Wechsler Verbal-Performance IQ discrepancies among psychiatric inpatients: Implications for subtle neuropsychological dysfunctioning. *Archives of Clinical Neuropsychology, 6*, 61–71.