The Causes of Group Differences in Intelligence Studied Using the Method of Correlated Vectors and Psychometric Meta-Analysis

Daniel Metzen

Master thesis

Daniel Metzen

5693403

University of Amsterdam

Department of Industrial and Organizational Psychology

Supervisor: Dr. Jan te Nijenhuis

2012

Version 05

Acknowledgements

First, I would like to thank the students that worked with the method of correlated vectors before me. Furthermore I would like to thank Dr. Jan te Nijenhuis for supervising this thesis and making datasets available. I would also like to thank Prof. Dr. Heiner Rindermann, Jakob Pietschnig MA, Prof. dr. Marise Born, Prof. dr. Richard Lynn, and dr. Reagan Murphy for making datasets available.

Special note. This master thesis is part of an ongoing meta-analytical project where a combination of the Method of Correlated Vectors and psychometric meta-analysis is applied to various topics of practical and theoretical interest. Previous undergraduate students in this ongoing research project were Birthe Jongeneel-Grimen, Rosina van Bloois, Lise-Lotte Geutjes, Jan Smit, Joep Dragt, Dennys Franssen, Jasper Repko, Denise Willigers, and Evelien van Meerveld. In this project, I used material from previous reports and this is indicated in a system of footnotes and references. This material entails paragraphs (adapted and unadapted), verbatim use of the formulation of hypotheses, verbatim use of the formulation of method, verbatim use of the formulation of reporting of results; I also used the format of graphs and tables, including verbatim use of the formulation of graph and table descriptions. As one of the purposes of this master thesis is to contribute to an ongoing research project, which builds on the work of my predecessors, this approach ensures maximum comparability with regard to earlier findings of this study project. If a complete paragraph is taken from an earlier study of this research project and used without adaptation this is indicated with a footnote at the heading of the paragraph in form of a number and the text “Paragraph taken from (source, for example, te Nijenhuis & Franssen (2010, p. 7))”. The paragraph ends with a reference to the study in question. If a paragraph is taken from an earlier study of this research project, but has been adapted to the focal topic, this is indicated with a footnote at the heading of the paragraph in form of a number and the text “Paragraph taken and adapted from (source, for example, te Nijenhuis & Franssen (2010, p. 7))”. The paragraph ends with a reference to the study in question. If a taken, or taken and adapted, paragraph is used in-text, the footnote will be placed at the end of the in-text paragraph. If a table is taken and adapted this is indicated with a footnote at the end of the description of the table in form of a number and the text “Table taken and adapted from (source, for example te Nijenhuis & Franssen (2010))” If I make verbatim use of the formulation of hypotheses of earlier studies of this research project, this is indicated with a footnote at the end of the hypothesis in form of the letter a, and the text “Verbatim use of formulation of hypothesis from (source, for example, te Nijenhuis & Franssen (2010))”. If I make verbatim use of the formulation of method of earlier studies of this research project, this is indicated with a footnote at the heading of the paragraph in form of the letter b, and the text “Verbatim use of formulation of method from (source, for example, te Nijenhuis & Franssen (2010))”. If I make verbatim use of the formulation of reporting of results from earlier studies of this research project, this is indicated with a footnote at the heading of the paragraph in form of the letter c, and the text “Verbatim use of formulation of reporting of results from (source, for example, te Nijenhuis & Franssen (2010))”. If I adapt the format of a graph or a table and make verbatim use of the formulation of graph/ table descriptions of earlier studies of this research project, this is indicated with a footnote at the end of the table/ graph descriptions in form of the letter d, and the text “Format of table/ graph adapted from (source, for example, te Nijenhuis & Franssen (2010)). Verbatim use of formulation of table/ graph descriptions from (source, for example, te Nijenhuis & Franssen (2010))”.

Table of Contents

Abstract

The huge IQ gap between non-Western immigrants and ethnic Dutch has emerged as one of the primary explanations for the large differences in school and work achievement between these groups. Is there a genetic component in the IQ gap between immigrants and ethnic Dutch? Meta-analyses have shown that the group differences on IQ subtests correlate almost perfectly with the cognitive complexity of these subtests; moreover, the cognitive complexity correlates perfectly with heritability and strongly with physical characteristics of the brain. If no other causes for IQ differences show a strong correlation with *g* loadings, this would point to a strong genetic component in IQ differences between immigrants and ethnic Dutch. In the present study, we first seek support for the hypothesis that only variables under genetic influence show a strong positive relationship with general intelligence. These are group differences, heritability, and physical characteristics of the brain. Second, we test whether differences in IQ due to variables not under genetic influence, namely biological-environmental factors, aging, and autism show a negligible to weak correlation with general intelligence. Support for both hypotheses would suggest that group differences are primarily driven by genetic factors and only to a minor extent by non-genetic factors. Therefore, group differences between non-Western immigrants and ethnic Dutch should be regarded as stable over time.

Concerning the first analysis, we first conducted a full-fledged meta-analysis on reaction time differences between Whites and higher-IQ groups, and Whites and lower-IQ groups, and we conducted several bare-bones meta-analyses and analyses of individual studies on differences in IQ profile between groups of different ethnicity. Second, we explored subgroups on school type, and religion. Third, we carried out a meta-analysis on the question whether *g*-loadedness of reaction time measures predicted the heritability of these measures. Fourth, we carried out a meta-analysis on the link between *g* loadings and brain volume. Concerning the second analysis, we first conducted several bare-bones meta-analyses and analyses of individual studies on biological-environmental variables. Second, we conducted bare-bones meta-analyses on the psychological phenomena autism and aging.

The hypothesis was strongly supported: heritabilities and most group differences showed moderate to strong positive correlations with *g*, but the correlation of brain volume with *g* was quite modest. All other phenomena showed no strong positive correlation with *g*.

It is concluded that these findings are strongly in line with a substantial genetic component in group differences in intelligence. This suggest that the large group differences in school achievement and work achievement are stable and that I/O psychologists should find ways to deal with them instead of ways of trying to change them.

Introduction

The integration of non-Western immigrants into the Dutch labor market has been tried and it has failed (see Scheffer, 2011). Problems of integration have been linked to differences in culture between the country of origin and the host country (see Erickson, 1987), and to differences between the socio-economic status (SES) of the first immigrant generations and the average SES of the host country population (White, 1982). It has often been assumed that when the cultural and the socio-economic gap will disappear, the integration of non-Western immigrants into the Dutch labor market should not be a problem anymore. However, this assumption only holds if non-Western immigrants are not structurally different on any other factor that drives the success of the integration into the Dutch labor market. In the present study, we address such a factor, namely the huge IQ gap between the Western population and non-Western immigrants. Previous research in applied psychology has shown that both school performance and job performance are excellently predicted by IQ scores (Salgado, Anderson, Moscoso, Bertua, de Fruyt, & Rolland, 2003; Schmidt & Hunter, 1998, 2004) . In addition, it is a firmly established scientific fact that non-Western immigrants have a mean IQ of only 85, which leads to large group differences in school and job performance (te Nijenhuis, de Jong, Evers, & van der Flier, 2004). These findings in combination suggest that the IQ gap needs to be reduced to give non-Western immigrants the same job opportunities as the Western population.

Te Nijenhuis et al. (2004) show that the IQ scores of the second generation of immigrants strongly improve in comparison with the IQ scores of the first generation. This most likely means that the Dutch environment was better than the non-Western environment causing an increase in IQ scores. However, the gap between the second generation and ethnic Dutch is still large; what are the causes of the gap between second-generation immigrants and ethnic Dutch? How strong is the genetic component in these group differences and how strong is the environmental component? The fundamental question is what part of the gap will disappear for the third and fourth generation of non-Western immigrants. A related question is what part of the integration-related gap in school achievement and work achievement will disappear.

Previous research indicated that score differences on IQ test batteries can be distinguished into two broad categories (te Nijenhuis, van Vianen, & van der Flier, 2007). Differences in general intelligence, or *g*, and hollow, test specific differences. The former refers to score differences on subtests with high loadings on *g* and the latter refers to score differences on subtests with low loadings on *g*. *g* loadings are the loadings on the first unrotated factor in a principal component analysis of a varied set of IQ tests (Jensen & Weng, 1994), and can be best understood as a measure of cognitive complexity (Gottfredson, 1997). 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). In broader terms, differences on tests high in cognitive complexity represent differences in general intelligence, and differences on tests low in cognitive complexity represent differences in abilities specific to the subtest in question. Whether score differences between two groups are more pronounced on test with high or low *g* loadings can be tested with the method of correlated vectors. This method is based on the reasoning that when a vector of score differences on subtests is correlated with the *g* loadings of these subtests a high correlation would indicate that higher score differences correspond to higher *g* loadings, and lower score differences correspond to lower *g* loadings. Hence, the higher a positive correlation of *g* loadings with differences on subtests is, the more strongly are differences related to the cognitive complexity of these subtest. A negligible negative or negligible positive correlation indicates that differences on subtests do not depend on the cognitive complexity of subtests, and the stronger a negative correlation of *g* loadings with differences on subtests is, the less strong are differences related to cognitive complexity of these subtests. In this sense, the method of correlated vectors provides a test to determine whether group differences are related to general intelligence.

The essential question is whether the IQ gap between second-generation immigrants and ethnic Dutch represents differences in general intelligence, thus whether subtests differences show a strong positive correlation with *g* loadings. If yes, should differences in general intelligence be regarded as stable, or can they change over time? Concerning the first question, two previous meta-analyses showed that the pattern in immigrant-Dutch score differences on subtests of an IQ battery is very similar to the pattern in cognitive complexity: large group differences on subtests of high cognitive complexity and small group differences on subtests of low cognitive complexity (te Nijenhuis & Dragt, 2010; te Nijenhuis & Willigers, 2011; see also te Nijenhuis & Repko, 2011). Therefore, we can conclude that the IQ gap between second-generation immigrants and ethnic Dutch represent differences in general intelligence. Concerning the second question, of the many hypothesized causes of group differences in intelligence the only two that have been shown to have a strong positive correlation with the cognitive complexity of subtests are heritability (te Nijenhuis & Jongeneel-Grimen, 2007; te Nijenhuis & Franssen, 2010) and various physical characteristics of the brain (see te Nijenhuis & Jongeneel-Grimen, 2007, for a review). Other hypothesized environmental causes of group differences in intelligence show negligible correlations with cognitive complexity, for example, cocaine and alcohol abuse (te Nijenhuis, van Bloois, & Geutjes, 2009), or even strong to very strong negative correlations with cognitive complexity, for example, headstart programs and adoption (te Nijenhuis & Grimen, 2007). These findings in combination are suggestive of a substantial genetic and a weak environmental component in group differences. In this sense, differences in general intelligence between second generation immigrants and ethnic Dutch should be regarded as stable over time.

To further explore the causes of group differences in general intelligence between second generation immigrants and ethnic Dutch, in the present study, we test the hypothesis that only factors under genetic influence show strong positive correlations with *g* loadings and that factors that are not under genetic influence, predominantly biological-environmental factors, do not. If we find support for this hypothesis, this would imply that, although changes in IQ can occur through various non-genetic factors, differences in general intelligence sustain. Hence, the IQ gap between second-generation immigrants and ethnic Dutch could be reduced, but differences in general intelligence could not[1].

We will test this hypothesis in the following way: To test whether variables under genetic influence show a strong and positive correlation with *g* loadings, we first conduct analyses on reaction time differences between Whites and higher-IQ groups and Whites and lower-IQ groups, and IQ subtest score differences between Jewish and non-Jewish-Whites, Jews and Arabs, and German and migrant children. Second, we explore within one specific group how subgroup differences can be explained by cognitive complexity; we focus on differences between schools and differences between religious groups. Can the differences between groups be generalized to differences between subgroups within groups? Third, it will be tested whether the heritabilities of

---

[1] It stands to reason how strong a positive correlation needs to be to give a clear indication that differences in a variable, for instance, subtest differences between groups, are predominantly related to general intelligence. We take a conservative approach, and assume that only correlations above .7 indicate a strong relationship between a variable and general intelligence. Correlations between .4 and .7 indicate a modest relationship. Correlations in the range of .4 to -.4 indicate either a negligible to weak positive or negligible to weak negative relationship, respectively, correlations between -.4 and -.7 indicate a modest negative, and -.7 indicate a strongly negative relationship. It needs to be emphasized that the interpretation of correlation coefficients derived from the method of correlated vectors is problematic if they do not exceed a certain value in magnitude. A strong, or at least modest, positive correlation indicates a positive relationship between a variable and general intelligence. Conversely, a strong, or at least modest, negative relationship indicates a negative relationship with general intelligence. Any correlation in between, however, does not give a clear indication on the relationship between a variable and general intelligence. The only statement we can make is that the explanatory value of general intelligence with regard to observed differences in a variable is rather limited. In sum, to determine whether a variable is related to general intelligence, we only accept a strong, or at least modest, positive correlation, as a clear indication. All correlations between .4 and -.4 are interpreted as a not pronounced relationship with general intelligence, and modest to strong negative correlations are interpreted as a clear negative relationship with general intelligence. This approach boils down to the question whether a variable is moderately to strongly positively related to general intelligence, in which case we assume that there is a positive relationship, or whether a variable is not moderately to strongly positively related to general intelligence, in which case we assume that there is no positive relationship.

reaction time measures are strongly predicted by the cognitive complexity of these measures. Fourth, we test the hypothesis that a physical characteristic of the brain, namely brain volume, correlates highly and positively with cognitive complexity. To find support for the hypothesis that factors that are not under genetic influence do not show strong and positive correlations with cognitive complexity we will test a vast range of variables that have previously been shown to affect IQ scores. The compilation of variables might seem arbitrary. Still, it constitutes a near exhaustive collection of phenomena that are a) not under genetic influence, b) testable with the method of correlated vectors, and c) not tested with a combination of the method of correlated vectors and psychometric meta-analysis, yet. First, it will be tested whether various biological-environmental variables show an absence of a strong positive correlation with cognitive complexity. These are iodine supplementation/ deficiency, prenatal cocaine exposure, fetal alcohol syndrome, air pollution, traumatic brain injury, and malnutrition. Second, some other factors have not been hypothesized to explain group differences in intelligence, such as aging and autism. However, we study these factors to test the hypothesis that there is no strong positive correlation with cognitive complexity.

In sum, if general intelligence is only related to heritabilities, group differences, and physical characteristics of the brain, but not to biological-environmental variables, aging, and autism, it could be interpreted as meaning that genetic factors are mainly responsible for differences in general intelligence. Hence, although a reduction in the IQ gap between second-generation immigrants and ethnic Dutch is possible, the gap in general intelligence is unlikely to close. Therefore, if the large group differences in school achievement and work achievement represent differences in general intelligence they should be seen as stable over time, and I/O psychologists should find ways to deal with them instead of ways of trying to change them.

Cognitive Tests[2]

Scores on cognitive tests provide the best general predictor of accomplishments in education, job training, and work. IQ tests predict performance in primary school with a correlation of about .70-.80; in secondary and tertiary education, these correlations become lower, but IQ tests are still excellent predictors of school performance (Jensen, 1980, 1998). IQ tests predict work performance with a correlation of about .50 (Salgado, Anderson, Moscoso, Bertua, de Fruyt, & Rolland, 2003; Schmidt & Hunter, 1998, 2004). As a result, cognitive tests are widely used for selection and placement in organizations, and increasingly also in educational settings. (te Nijenhuis & Franssen, 2010)

Group Differences in Tests of Cognitive Abilities and Their Causes

[2] Paragraph taken from te Nijenhuis & Franssen (2010, p. 7)

Group differences in average IQ between Blacks, Whites, and Asians have been the subject of extensive study (Jensen, 1998). The finding that there are differences with Blacks scoring in the range of one *SD* lower than Whites, and Asians scoring slightly better than Whites has been widely confirmed by research (Lynn & Vanhanen, 2002; te Nijenhuis, de Jong, Evers, & van der Flier, 2004). The causes, to which these differences can be attributed, however, are not clear. In general, there are two schools of thought concerning group differences in IQ: The first primarily tries to explain differences in cognitive profile between Blacks and Whites by differences in socioeconomic status between both groups. According to this school, a non-stimulating environment and a culture that does not particularly value high societal and career achievement restrict the development of cognitive abilities. It is argued that when differences in socioeconomic status between groups diminish, differences in IQ will diminish subsequently (Capron & Duyme, 1989). The second school explains differences in IQ between ethnic groups predominantly with genetic factors.

Spearman's hypothesis states that group differences in IQ are strongly linked to general intelligence, or *g*, and several meta-analyses (te Nijenhuis & Dragt, 2010; te Nijenhuis & Repko, 2011; te Nijenhuis & Willigers, 2011) have shown that the hypothesis is strongly confirmed. Te Nijenhuis and Jongeneel-Grimen (2007) showed that *g* loadings have a perfect true correlation with heritability coefficients. These findings combined are suggestive of a substantial or even a strong genetic component in group differences.

The fundamental hypothesis that we test is that only group differences, heritabilities, and physical characteristic of the brain and no other factors show highly positive correlations with general intelligence. Some of these other factors have been hypothesized to explain group differences in intelligence, such as a wealth of environmental factors, both biological-environmental and cultural-environmental. Support for their causality should come in the form of strong positive correlations with *g* loadings. Some other factors have not been hypothesized to explain group differences in intelligence, such as physical characteristics, and psychological phenomena such as aging and autism; indeed, they are to be regarded as effects and not as causes. However, we study these factors to test how the strong positive correlation of cognitive complexity is limited to group differences, heritability, and physical characteristics of the brain. Two other lines of research focus on the link between group differences and *g*. First, we test the hypotheses that all other group differences can also be explained with *g*; we focus on, for instance, Jewish-non-Jewish-White differences, and Jewish-Arab differences. Secondly, we explore within one specific group whether subgroup differences can also be explained using *g* loadings, for instance looking at differences between schools. In sum, in the present study we systematically test a large number of variables' correlation with *g*.

General Intelligence (g)[3]

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 defined 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). (te Nijenhuis & Franssen, 2010)

Hierarchical Intelligence Model[4]

Jensen (1998) hypothesized that hierarchical intelligence models, such as Carroll's (1993) three-stratum hierarchical factor model of cognitive abilities, best describe scores on IQ batteries. At the highest level of the hierarchy (stratum III) is general intelligence or *g*. The level below (stratum II) incorporates 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. The level further below (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 empirically demonstrated to cluster into one narrow ability (Carroll, 1993). (te Nijenhuis & Franssen, 2010)

Research With the Method of Correlated Vectors (MCV)[5]

The MCV 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. When the latter variable is continuous (or nearly so), the relations are usually calculated in terms of a correlation coefficient (Ashton & Lee, 2005). (te Nijenhuis & Franssen, 2010)

---

[3] Paragraph taken from te Nijenhuis & Franssen (2010, p. 7)
[4] Paragraph taken from te Nijenhuis & Franssen (2010, p. 8)
[5] Paragraph taken from te Nijenhuis & Franssen (2010, p. 9)

In a large number of studies using the MCV it was shown that a large number of variables are associated with Spearman's *g*. We review this literature and make a distinction into four groups of studies. We then indicate which new variables will be studied.

Stuy1: Group Differences

Study 1a: Difference Between Whites and Higher-IQ and Lower-IQ Groups in Reaction Time

Many studies have shown that White samples have higher IQ scores than, for instance, Blacks and Hispanics in the US, non-Western immigrants in Europe, Roma in Serbia, and Blacks in Africa. Three meta-analyses have shown that these group differences are almost perfectly linked to *g* loadings (te Nijenhuis & Dragt, 2010; te Nijenhuis & Repko, 2011; te Nijenhuis & Willigers, 2011). As reaction time measures show a moderate to substantial correlation with scores on Raven's standard progressive matrices (Jensen, 1983), which is one of the best predictors of general intelligence, or *g* (Jensen, 1993), in this study, we will explore whether group differences in reaction time measures are also linked to *g* in a similar way as group differences on IQ batteries are. Reaction time can be understood as a measure of an individual's speed of cognitive processing. "Typical reaction time measures assess the speed of scanning alphanumeric symbols in short-term memory, the speed of retrieval of verbal information from long-term memory, or the efficiency with which individuals can simultaneously store and process verbal and/ or numerical information in short short-term memory" (Vernon, 1988). As reaction time measures show a modest to substantial correlation with Raven's progressive matrices (Jensen, 1983), we expect that

> (1) the true correlation between score differences on reaction time measures between Whites and lower-IQ groups and the magnitude of *g* loadings is modestly to strongly positive in sign.[a]

Conversely, various studies have shown that South Koreans, Japanese, and Chinese have somewhat higher mean IQ scores than Whites (Lynn & Vanhanen, 2002). In the present study, we will test whether differences in reaction time between Whites and higher-IQ groups can be explained by *g*. In line with Hypothesis 1, we expect that

> (2) the true correlation between score differences on reaction time measures between Whites

and higher-IQ groups and the magnitude of *g* loadings is modestly to strongly positive in sign.[a]

Study 1b: Differences Between Germans and Immigrants

Germany has a high number of migrants from Turkey, South Italy, Russia, and Poland. The literature on comparisons between the IQ scores of Germans and non-Western – predominantly Turkish – immigrants and Italian immigrants from the middle of the 20th century is quite modest in amount, but suggest lower to much lower scores for both immigrant groups. The PISA studies showed that a higher percentage of migrant children attend lower school types. In the school year 2003/2004 the percentage of migrant children in *Hauptschule*, the lowest regular school type pupils can attend to, was 18.6%. The percentage in the higher school types *Realschule*, a school type in which pupils of average capability are educated, and *Gymnasium*, a school type in which pupils of high capability are educated, had a percentage of migrant children of only 7% and 4%, respectively (Daseking, Lipsius, Petermann, & Waldmann, 2008). The Pisa studies also showed that migrant children that speak another language than German at home show worse scores on tests of scholastic aptitude than migrant children that report to speak German at home (Prenzel et al., 2007). In the present study, we will explore whether differences in IQ profile between migrant children and German children are due to differences in general intelligence. Since migrant children score lower on tests of scholastic aptitude even if the tests are not depended on enhanced language comprehension (Ramm, Prenzel, Heidemeier, & Walter, 2004), we hypothesize that

(3) the true correlation between score differences of German children and migrant children in Germany and the magnitude of *g* loadings is modestly to strongly positive in sign.[a]

Study 1c: Differences Between European Jews and Oriental Jews, and European Jews and Non-Jewish Whites

Lynn and Vanhanen (2002) have shown that group differences in IQ are more the rule than the exception and that the average IQ in the world is about 85. Whites have an IQ of about 100, which is about one *SD* higher than the world mean. However, average IQ scores are still higher for South Koreans, Japanese, and Chinese, with a value of about 105. Among the Jews a distinction has to be made between Ashkenazi Jews (European Jews) and Sephardic Jews (Oriental Jews), with the former having a mean IQ of 110-115 (Lynn, 2006; Lynn & Vanhanen, 2002) and the latter having

---

[a] Verbatim use of formulation of hypothesis from te Nijenhuis & Franssen (2010)

an about one *SD* lower IQ (David & Lynn, 2007). European Jews also differ from Oriental Jews with regard to their intelligence profile. Differences on subtest scores of verbal ability are larger than difference scores on subtests that measure other dimensions of intelligence (Zeidner, 1987; Rim, 1983). Similar differences in IQ profile were found between European Jews and American Whites (Lynn, 2004). Cochran, Hardy, and Harpending (2006) argued that this difference is due to the history of European Jews in medieval times. In contrast to the business enterprises of Christian Europeans, whose religion forbid them to lend money to other people, Jewish Europeans' business enterprises had a higher chance to prosper. Thereby European Jews acquired higher verbal and mathematical skills, than their Christian counterparts (Cochran. et al., 2006). However, as Lamarckian inheritance is not an option, obviously, it stands to reason how the acquisition of verbal and mathematical skills in medieval times could have had an effect on the verbal and mathematical skills of subsequent generations. From a Darwinist perspective, it is possible to argue that the initially similar distributions of cognitive abilities of European Jews and non-Jewish Europeans moved apart from each other because a criterion concerning the likelihood of reproduction for both groups changed. Given the assumption that cognitive abilities were normally distributed for Jews, there were Jews with higher verbal and mathematical abilities who could easily capitalize on enterprises involving the lending of money, and there were Jews who had low verbal and mathematical abilities and could not capitalize on them. Those who could capitalize became more prosperous and therefore could provide better nurturing for their offspring, which resulted in higher chance of survival compared to the offspring of Jews with low verbal and mathematical skills. In this sense, natural selection reduces the variance with regard to the mathematical and verbal abilities at the lower end of the gene pool, and the distribution of mathematical and verbal abilities shifts to the higher end, resulting in higher mean verbal and mathematical abilities. The distribution of mathematical and verbal abilities of non-Jewish Europeans, however, stayed stable, because they have not been exposed to a similar selection criterion.

Since verbal subtests have on average higher *g* loadings than performance subtests this specific profile – large *d*s on verbal and smaller *d*s on performance – will most likely result in finding high correlations *d* x *g* when comparing European Jews and Oriental Jews, and European Jews and non-Jewish Whites. Therefore, we expect that

(4a) the true correlation between score differences of European Jews and Oriental Jews and the magnitude of *g* loadings is strongly positive in sign.[a]

[a] Verbatim use of formulation of hypothesis from te Nijenhuis & Franssen (2010)

(4b) the true correlation between score differences of European Jews and non-Jewish Whites and the magnitude of *g* loadings is strongly positive in sign.[a]

Study 1d: Differences Between European Jews and Arabs

The data provided by Professor Lynn on the IQ scores of Jews (see Lynn, 2010) contained comparison with non-Jewish Whites in Western countries and comparisons of Jews and Arabs in Israel. These studies all show higher IQ scores for Jews compared to Arabs (Kugelmass, Lieblich, & Bossik, 1974; Lieblich, Kugelmass, & Ehrlich, 1975; Lieblich, & Kugelmass, 1981). In the present study, we will test whether score differences on IQ tests have a strongly positive relationship with general intelligence. We expect that

(5) the true correlation between score differences of Jews and Arabs and the magnitude of *g* loadings is strongly positive in sign.[a]

Study 2: Subgroup Differences

The method of correlated vectors is a means to explore whether differences on IQ test batteries between different groups are related to general intelligence. Since there is an extensive body of IQ data on a large number of different groups, for instance in the book by Lynn and Vanhanen (2002), we want to explore whether differences in IQ between groups, of which we can find usable data, are related to general intelligence. We will conduct an exploratory analysis of IQ differences between groups that attend different school types, and groups of different religious belief. The question is whether IQ differences between those groups are modestly to strongly related to general intelligence. The purpose of this study is to test whether the cluster of variables that previously have been shown to have highly positive correlations with general intelligence, that is heritabilities, (ethnic) group differences, and physical characteristics of the brain, can be extended to subgroup differences. As there is no evidence as to whether subgroup differences are under genetic influence or not, the study of subgroup differences is entirely explorative in nature. Table 1 shows the results of previous studies on group differences.

---

[a] Verbatim use of formulation of hypothesis from te Nijenhuis & Franssen (2010)

Table 1

Correlations between g Loadings and Group Differences

study	*variable*	*K*	*N*	*r*	rho
Te Nijenhuis, de Pater, van Bloois, & Geutjes (2009)	gifted	22	4,823		1.01 (.91)
	mentally retarded	34	2,729		.74
te Nijenhuis & Dragt (2010)	group differences on IQ batteries	32	19,267	.72	.91
te Nijenhuis & Dragt (2010)	Dutch/immigrant differences on educational and training criteria	9	909	.79	
te Nijenhuis & Repko (2011)	Group differences on Raven PM item scores	28	2,830	.43	.83
te Nijenhuis & Willigers (2011)	Dutch/immigrant differences on IQ tests	28	4,048	.61	

Note. *K* = Number of correlations; *N* = Total sample size or harmonic *N*; *r* = mean observed correlation (sample size weighted); rho = true correlation (observed correlation corrected for unreliability, range restriction, and imperfectly measuring the construct of *g*). Results for the value of the meta-analytical rho are always reported for the final analysis, so the one without outliers. A rho value higher than 1is due to the correction of deviation from perfect construct validity (see Hunter & Schmidt, 1990). As this value is overcorrected, obviously, we also show the rho value without this correction in brackets.[d]

Study 2a: Explorative Comparison of School Types

Te Nijenhuis, de Pater, van Bloois, and Geutjes (2009) compared scores of groups of gifted and mentally retarded children with scores of nationally representative samples. These researchers found a correlation *d* x *g* of +1 for gifted children and a correlation *d* x *g* of .74 for children with mental retardation. For the mentally retarded children the score on subtests of short-term memory acted as outlier; leaving out the scores on this outlier will bring the value of the correlation quite close to the value found for gifted children. The comparison of IQ scores between pupils of different school types yield less substantial differences between IQ scores than comparisons

[d] Format of table adapted from te Nijenhuis & Franssen (2010). Verbatim use of formulation of table descriptions from te Nijenhuis & Franssen (2010)

between gifted children with average children and mentally retarded children with average children. Still, there is a difference of approximately 10 IQ points between pupils of different school types in the Netherlands and Germany (van Dijk & Tellegen, 1994; Daseking, Lipsius, Petermann, Waldmann, 2008). In the present study we will explore whether IQ differences between school types have a modest to strong positive relationship with general intelligence.

In the Netherlands and Germany, pupils are allocated to different school types of secondary education based on the level of achievement they showed during elementary school. In the Dutch education system, pupils with low levels of achievement follow the tracks IVBO (*Individueel Voorbereidend BeroepsOnderwijs,* [*preparatory individual job* education]) and VBO (*Voorbereidend BeroepsOnderwijs,* [*preparatory job* education]). Pupils with average levels of achievement follow the track MAVO (*Middelbaar Algemeen Voortgezet Onderwijs,* [*continued general* education]). The tracks for pupils with high to very high levels of achievement are HAVO (*Hoger Algemeen Voortgezet Onderwijs*, [*continued higher general education*]) and VWO (*Voorbereidend Wetenschappelijk Onderwijs,* [*preparatory program for scientific* education]), respectively. The German education system differentiates between school types of secondary education for students with low levels of achievement (Hauptschule), average levels of achievement (Realschule), and high levels of achievement (Gymnasium). With respect to the age of students, IVBO, VBO, MAVO, HAVO, VWO, Hauptschule, Realschule, and Gymnasium are comparable to middle and high school in the US. Since there is no theoretical foundation on which we can base a clear expectation about the magnitude and sign of the correlation d x g, we will explore whether IQ differences between school types are related to general intelligence. We will test whether

(6) the true correlation between score differences of different school types and the magnitude of g loadings is modestly to strongly positive in sign.[a]

Study 2b: Explorative Comparisons of Religious Groups

Religious belief, the extent to which an individual beliefs in the existence of one or more higher entities with theistic, deistic, or pantheistic properties, has been found to correlate negatively with IQ (Lynn, Harvey, & Nyborg, 2008). However, it is not clear whether people that are more religious do not want to enlarge their knowledge about the world and their cognitive abilities and thereby do not enhance their intelligence, or whether people of lower intelligence are more prone to belief in a supernatural being of some sort. In the Netherlands of the 1960s, the Dutch population consisted of mainly Catholics and Protestants, and to a lesser extent of Atheists and groups of non-

[a] Verbatim use of formulation of hypothesis from te Nijenhuis & Franssen (2010)

Christian belief. Verhage (1964) found a small difference of four IQ points in the favor of Protestants compared to Catholics. As religious beliefs, or at least formal membership of religion or confession, are passed on from one generation to the next, the gene pool within different groups of religious belief should be stable over time. However, it stands to reason which mechanism should have lead the distributions of general intelligence of different religious groups to move apart from each other. In this sense, differences in IQ between religious groups might just be hollow, test-specific differences rather than differences in general intelligence. Still, we want to explore whether the pattern of high positive correlations between heritabilities, (ethnic) group differences, and physical characteristics of the brain, and general intelligence can be expanded to subgroup differences. As there is no theoretical foundation on which we can base a clear expectation about the magnitude and sign of the correlation *d* x *g*, we will explore whether IQ differences between groups of different religious belief are related to general intelligence. We will test whether

(7) the true correlation between score differences of groups with different religious beliefs and the magnitude of *g* loadings is modestly to strongly positive in sign.[a]

Study 3: Heritability

The method of correlated vectors has been used in studies of heritability. Te Nijenhuis and Grimen (2007) meta-analyzed the literature on IQ-test batteries and report a perfect true correlation between *g* loadings and heritability coefficients (see Table 2). Te Nijenhuis and Franssen (2009) looked at inbreeding depression, which is assumed to be very strongly or even perfectly under genetic influence, and report a meta-analytic correlation of rho = .84 between subtest score differences of inbred children and comparison groups and *g* loadings. Differences in subtests score due to hybrid vigor showed a correlation of .52 with *g* loadings (Nagoshi & Johnson, 1986; cited in Jensen, 1998). Hybrid Vigor refers to the enhancement of any biological quality or function in the offspring that is caused by the intermingling of the different genetic contributions of the parents, which is assumed to be very strongly or even perfectly under genetic influence. The conclusion of these studies is that there is a) a strong positive correlation between *g* loadings and heritability coefficients of IQ batteries and b) a modest to strong positive correlation between *g* loadings and subtests score differences due to variables under genetic influence. In the present study, we will test the finding under a) in a new context, namely reaction time measures.

h^2 Reaction Times

It is an established finding that cognitive abilities are heritable to a large extent. The general factor of intelligence (*g*) has a heritability of about 0.50 (Bouchard & McGue, 1981). Research has shown that the heritabilities for specific cognitive abilities can range between 0.20-0.70 (Plomin & DeFries, 1979). It has also been shown that cognitive abilities show a somewhat substantial correlation with speed of information processing (Vernon, 1989). We test whether measures of speed of information processing, usually some sort of reaction time measure (RT), have a heritability pattern similar to the cognitive abilities described above. The heritability coefficient of reaction time measures is estimated by comparing the correlations of reaction time subtests between monozygotic twins and the correlations of reaction time subtests between dizygotic twins. The *g* loadedness of reaction time measures is determined by the correlation of reaction time measures with Full scale IQ. As reaction time measures show a somewhat substantial positive correlation with cognitive abilities, and cognitive abilities show a strong correlation with heritability coefficients we expect that general intelligence, as it is reflected in reaction time measures, shows a moderate to strong positive correlation with heritability coefficients. Therefore, we expect that

(8) the true correlation of h^2 of reaction time measures and the magnitude of *g* loadings is moderately to strongly positive in sign.[a]

Table 2

Correlations of g Loadings with Various Genetic Variables

study	*topic*	*K*	*N*	*r*	rho
te Nijenhuis & Grimen (2007)	h^2 IQ batteries	7	2,590	.69	1.01
te Nijenhuis & Franssen (2009)	Inbreeding depression	3	1,783	.56	.84
Nagoshi & Johnson (1986)	Hybrid vigor	1	2,096	.52	

Note. *K* = Number of correlations; *N* = Total sample size or harmonic *N*; *r* = mean observed correlation (sample size weighted); rho = true correlation (observed correlation corrected for unreliability, range restriction, and imperfectly measuring the construct of *g*). Results for the value of the meta-analytical rho are always reported for the final analysis, so the one without outliers.[d]

[a] Verbatim use of formulation of hypothesis from te Nijenhuis & Franssen (2010)

[d] Format of table adapted from te Nijenhuis & Franssen (2010). Verbatim use of formulation of table descriptions from te Nijenhuis & Franssen (2010)

Study 4: Physical Characteristics of the Brain

Jensen (1998) is a rich source of studies using the MCV, and many of the studies discussed in this paragraph are taken from it. The MCV has been applied to physical characteristics of the brain such as head size, brain volume, brain's gray matter, brain's evoked potential, brain glucose metabolic rate, peripheral nerve conduction velocity, and brain pH (te Nijenhuis & Franssen, 2010) (see Table 3), and without exception result in substantial to high positive correlations. These correlations might even approach a value of +1 when the statistical corrections associated with psychometric meta-analysis (Hunter & Schmidt, 1990) are administered (te Nijenhuis & Franssen, 2010). The fourth research question is whether the correlations of brain volume with subtests scores have a strongly positive correlation with the *g* loadings of subtests.

Brain Volume

The prediction of intelligence by physical properties of the brain has been an issue since the early days of the previous century. Jensen (1994) showed a positive correlation between head size might seem a rather low correlation at first, we should keep in mind that head circumference does not necessarily show a high correlation with actual brain size or weight. Since measurement of brain volume is now much easier and precise, the relationship between physical properties of the brain and intelligence can be explored in detail. McDaniel (2005) conducted a meta-analysis on the relationship of in vivo brain volume and *g* and found a correlation of .33. The hypothesis in the present study is that

> (9) the true correlation between mean scores on IQ subtests and brain volume and the magnitude of *g* loadings is moderately to strongly positive in sign.[a]

Study 5: Biological-Environmental Factors

Spitz (1987) speculates that the pattern in biological-environmental variables mimics the pattern in genetic variables, so that both variables would have strong positive correlations with *g* loadings. In previous studies where the MCV was combined with psychometric meta-analysis, this was also hypothesized (te Nijenhuis & Grimen, 2007; te Nijenhuis & Franssen, 2009; te Nijenhuis & Smit, 2010). However, Table 4 shows there is definitely no clear-cut support for this hypothesis. The weighted *r*s from three exploratory meta-analyses show one negative and two positive correlations of weak to moderate strength. The findings for these three topics suggest that the

[a] Verbatim use of formulation of hypothesis from te Nijenhuis & Franssen (2010)

Table 3

Correlation of g Loadings with Physical Characteristics of the Brain

study	*variable*	*r*	*N*
Jensen (1994)	head size	.64	286
Wickett, Vernon, & Lee (1994)	brain volume	.65	80
	brain volume	.51	72
Schoenemann (1997)	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

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. Adapted from te Nijenhuis & Franssen (2010).[6] (te Nijenhuis & Franssen, 2010).

[6] Table taken and adapted from te Nijenhuis & Franssen (2010)

hypothesis that biological-environmental variables correlate strongly with *g* is wrong. A more plausible explanation might be that the effects of biological-environmental variables are specific to certain broad and narrow abilities. However, it may still be that the hypothesis of a strong positive correlation will be confirmed for other biological-environmental variables. Both of these hypotheses will be tested on the biological-environmental variables iodine deficiency/ supplementation, prenatal cocaine exposure, fetal alcohol syndrome, air pollution, traumatic brain injury, and malnutrition. If we do not observe a strong correlation *d* x *g*, we will also check whether differences in IQ profile lie on lower levels of the intelligence hierarchy. These levels are the broad and narrow abilities in Carroll's three-stratum model of cognitive abilities (Carroll, 1993).

Study 5a: Iodine Deficiency/ Supplementation

According to the WHO, iodine deficiency is the single most preventable cause for mental retardation in the world. A diet that has a strong deficiency in the trace element iodine can lead to endemic goiter and endemic cretinism. The former refers to a swelling of the thyroid glance that

Table 4

Correlations of g Loadings with Biological-Environmental Variables

study	*Topic*	*K*	*N*	*r*	rho
te Nijenhuis, de Pater et al. (2009)	Alcohol abuse	6	242		-.43
te Nijenhuis & Smit (2010)	Lead exposure	2	101	.18	
te Nijenhuis & Smit (2010)	Prenatal cocaine exposure	1	25	.38	

Note. *K* = Number of correlations; *N* = Total sample size or harmonic *N*; *r* = mean observed correlation (sample size weighted); rho = true correlation (observed correlation corrected for unreliability, range restriction, and imperfectly measuring the construct of *g*). Results for the value of the meta-analytical rho are always reported for the final analysis, so the one without outliers. Te Nijenhuis & Smit (2010) also analyzed data using only Wechsler Full Scale, Verbal, and Performance scores, but in this Table only their data on subscales of test batteries with at least seven subtests are reported.[d]

leads to a swollen neck. The latter refers to a restriction in physical and mental growth due to a lack of thyroid hormones (Bleichrodt, Drenth, & Querido, 1980). Additional abnormalities in endemic cretinism include bilateral hearing loss or deaf-mutism and neurological abnormalities such as paralysis (Bleichrodt, Garcia, Rubio, Morreale de Escobar, & Escobar del Rey, 1987). In a meta-analysis on IQ differences between iodine deficient and control groups, an IQ difference of 14 IQ points has been reported (Bleichrodt & Born, 1994). Iodine deficiency has also been shown to have

[d] Format of table adapted from te Nijenhuis & Franssen (2010). Verbatim use of formulation of table descriptions from te Nijenhuis & Franssen (2010)

a detrimental effect on psycho-motor development (Bleichrodt, Garcia, Rubio, Morreale de Escobar, & Escobar del Rey, 1987). The supplementation with iodine of individuals with iodine deficiency has been shown to lead to an increase in IQ (van den Briel, West, Bleichrodt, van de Vijver, Hautvast, & Ategbo, 2000). Since there is no clear support for an either strongly positive or strongly negative correlation between biological-environmental variables, such as iodine, and general intelligence, we expect that

> (10a) the true correlation between score differences of iodine deficient groups and control groups and the magnitude of *g* loadings will not be moderately to strongly positive.[a]
>
> (10b) we hypothesize that the true correlation of score differences between iodine deficient groups that were supplemented with iodine and iodine deficient control groups that received a placebo and the magnitude of *g* loadings will not be moderately to strongly positive.[a]

Study 5 b: Prenatal Cocaine Exposure

Cocaine is a stimulant drug that has powerful effects on the central nervous system. If consumed during pregnancy, cocaine can have adverse effects on the developing brain through alterations of the central monoamine systems as well as through maternal vascular disruptions (Singer, et al., 2004). Prenatal cocaine exposure has been found to have detrimental effects on a wide range of cognitive abilities, such as general knowledge, arithmetic skills, visual-spatial skills (Singer et al., 2004), attention span (Bandstra, Morrow, Anthony, Accornero, & Fried, 2001), and verbal comprehension (Lewis, et al., 2004, Morrow, Vogel, Anthony, Ofir, Dausa, & Bandstra, 2004). Since a wide range of cognitive abilities is affected by prenatal cocaine exposure, in the present meta-analysis we explore whether lower IQ scores of children exposed to cocaine prenatally are related to general intelligence, or to broad and narrow cognitive abilities. In a previous analysis by te Nijenhuis & Smit (2010) on the relationship between differences scores of children prenatally exposed to cocaine and comparison groups and *g* loadings of subtests of an IQ battery a correlation *d* x *g* of .38 was obtained. This result does not point to a strong relationship between prenatal cocaine exposure and general intelligence. Since research results on the relationship between biological-environmental variables and general intelligence did not point to an either strongly positive or strongly negative relationship, we expect that

[a] Verbatim use of formulation of hypothesis from te Nijenhuis & Franssen (2010)

(11) the true correlation of score differences between groups prenatally exposed to cocaine and control groups and the magnitude of *g* loadings will not be strongly positive.[a]

Study 5c: Fetal Alcohol Syndrome

The term fetal alcohol syndrome (FAS) comprises all neurological, intellectual, and behavioral abnormalities in an individual that can be attributed to the exposure of the toxin alcohol through the use or abuse of alcohol by the individual's mother during pregnancy (Juretko, 2006). Since even relative small amounts of alcohol can cause devastating effects in the children of mothers with a low neurological resistance to the toxin alcohol, mothers of children with FAS are not necessarily heavy drinkers (Löser, 1995). However, if the neurological or organic tolerance of mother and/ or embryo is high, sustained alcohol intake of mothers does not necessarily lead to FAS in their children (Dehaene, 1995).There are three different degrees of FAS (FAS 1st°, FAS 2nd°, and FAS 3rd°) whereby a higher degree corresponds to more severe damage to the brain. A less severe form of FAS is known as fetal alcohol effects (FAE). Although individuals are diagnosed with FAE and FAS 1st°, FAS 2nd°, or FAS 3rd° based on intellectual and behavioral functioning, there is no clear neurological definition of what constitutes an FAS or FAE. Neuropsychological damage can have different causes and intellectual and behavioral abnormalities can also be caused by environmental influence and heritage. Gottfredson and Deary (2004) suggest that individuals with a lower IQ are more likely to fail to manage the challenges of maintaining good health and are more prone to expose themselves to unhealthy environments and lifestyles. Therefore it is reasonable to assume that heavy drinking mothers, that are obviously either indifferent of or uninformed about the hurtful effects of chronic or excessive alcohol consume during pregnancy, have on average a lower IQ, than not alcohol consuming mothers. Since IQ has been found to be highly heritable (te Nijenhuis & Grimen, 2007) deficits in intellectual performance of individuals with FAS or FAE might to some extent be due to genetic predisposition rather than toxic damage caused by alcohol. Previous research has shown that recreational drugs like cocaine or nicotine have detrimental effects on the cognitive development of a child when received prenatally (te Nijenhuis & Smit, 2010). However, research failed to show a pronounced relationship between these biological-environmental variables and general intelligence, or *g*. Therefore

(12) we hypothesize that the true correlation of score differences between groups that were diagnosed with fetal alcohol syndrome and control groups and the magnitude of *g* loadings will not be strongly positive.[a]

Study 5d: Air Pollution

In cities with a large population, air pollution has become a significant environmental problem. Air pollution has been found to cause stroke-related sickness and death (Hong, Lee, Kim, & Kwon, 2002; Maheswaran, et al., 2006). Air pollution has also been found to have a detrimental effect on cognitive abilities. Calderon-Garciduenas, et al. (2008) found that children from a city with low air pollution score better on an IQ test battery than children from a high-polluted city. However, one should not forget that it is theoretically also possible that families with higher IQs earn more money, so they can choose to live in less polluted surroundings. In other words, the causality question is still open. However, the experimental and the control group were matched with regard to parental occupation and income; therefore, a priori differences in IQ that are unrelated to the exposure of polluted air seem rather implausible. Since previous research did not show an either strongly positive or strongly negative relationship between IQ differences caused by biological-environmental variables and general intelligence, we expect that

(13) the true correlation of score differences between groups exposed to air pollution and control groups and the magnitude of *g* loadings will not be strongly positive.[a]

Study 5e: Traumatic Brain Injury

Traumatic brain injury (TBI) refers to severe damage to the brain resulting from external force. Since severity of damage can vary extremely from case to case, we classify severity of damage following the criteria of the Glasgow Coma Scale (GCS). The Glasgow Coma Scale is a widely used instrument to classify TBI. Classification is based on a patient’s verbal, motor, and eye movement reactions to various stimuli. Scores of this classification range from 3 to 15, whereby a score higher than or equal to 13 leads to a diagnosis of mild TBI. A score lower than 13 and higher than or equal to 8 leads to a diagnosis of moderate TBI, and a score lower than 8 leads to a diagnosis of severe TBI. Since TBI also often leads to memory impairment (Dikmen, Machamer, Winn, & Temkin, 1995), a second classification is based on the time a patient suffers from post-traumatic amnesia (PTA). PTA refers to a state of confusion immediately after suffering a TBI. PTA

[a] Verbatim use of formulation of hypothesis from te Nijenhuis & Franssen (2010)

can include retrograde amnesia as well as anterograde amnesia. The former refers to the loss of memories before the accident. For example, a patient might not be able to recall his name, address, or other autobiographic memories. The latter refers to the inability to store memories of events that happened after the injury in memory. Patients who suffer from PTA less than a day are diagnosed with mild TBI. A PTA of one to seven days indicates a moderate TBI. If a patient suffers more than seven days from PTA a diagnosis of severe TBI is made. A third classification is based on the time a patient suffered from loss of consciousness (LOC). If a patient had an LOC of less than 30 minutes, this indicates a mild TBI. LOC of 30 minutes to 24 hours points to a moderate TBI. A LOC longer than 24 hours indicates a severe TBI. Since scores on the Glasgow Coma scale, the duration of post traumatic amnesia and the duration of loss of consciousness do not necessarily lead to the same classification with regard to severity of TBI, the classification of severity depends to some extent on the judgment of the medical staff in question.

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

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

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

---

[a] Verbatim use of formulation of hypothesis from te Nijenhuis & Franssen (2010)

(14) the true correlation between score differences between groups that have had a traumatic brain injury and control groups and the magnitude of *g* loadings is not strongly positive in sign.[a]

Study 5f: Malnutrition

Malnutrition refers to a lack of or a strong imbalance in food intake that leads to physical and psychological dysfunction. During critical growth periods of the brain malnutrition can have adverse effects on the cognitive development of children (Kaplan, 1972). Malnutrition during the first year of life has been found to lead to lower IQ scores in school-aged children (Ellis & Hill, 1975). Malnutrition can occur for several reasons. Most obvious, if food is not available on a regular basis, as is the case in poor African countries, a healthy food intake is not possible. However, even if food is easily available individuals can limit themselves to a low food intake for psychological reasons. Individuals that suffer from anorexia nervosa, for instance, have a disturbed body perception, that causes an extremely low food intake, which eventually can even lead to self-starvation. Minimal food intake due to anorexia nervosa has been found to lead to lower IQ scores (Gillberg, Rastam, Wentz, & Gillberg, 2007). In the present study, we will explore whether lower IQ scores due to anorexia nervosa have a pronounced relationship with general intelligence. Since previous research on the relationship between biological-environmental variables and general intelligence did not lead to either a pronounced positive or pronounced negative correlation *d* x *g*, we expect that

(15) the true correlation between score differences of malnourished groups and control groups and the magnitude of *g* loadings will not be strongly positive.[a]

Study 6: Aging and Autism

In an exploratory fashion, a limited number of psychological phenomena have been studied concerning their correlation with *g* loadings. Absent a thorough knowledge of the literature, one might simply speculate that the lower IQ found in cases of schizophrenia and epilepsy is caused by damage to the information processing systems in the brain; general damage might lead to a lowering on the *g* factor. The studies up to now do not lend clear-cut support for this speculation

[a] Verbatim use of formulation of hypothesis from te Nijenhuis & Franssen (2010)

(see Table 5). It might also be that schizophrenia or epilepsy damages specific parts of the brain. In an attempt to widen the nomological net of correlates of *g* loadings, the psychological phenomena aging and autism will be tested with the MCV.

Table 5

Correlations of g Loadings with various Psychological Phenomena

Study	*topic*	*K*	*N*	*r*	rho
te Nijenhuis & Franssen (2010)	schizophrenia	5	315	-.50	-.87
te Nijenhuis & Franssen (2010)	epilepsy	5	445	.44	.64
te Nijenhuis, de Pater, van Bloois, & Geutjes (2009)	Clinical depression	5	268	.12	.42

Note. *K* = Number of correlations; *N* = Total sample size or harmonic *N*; *r* = mean observed correlation (sample size weighted); rho = true correlation (observed correlation corrected for unreliability, range restriction, and imperfectly measuring the construct of *g*). Results for the value of the meta-analytical rho are always reported for the final analysis, so the one without outliers.[d]

Study 6a: Aging

An extensive body of literature shows that cognitive abilities are affected differently by aging (Ardila, 2007). Cattell (1963) established a taxonomy of cognitive abilities that differentiates between tests that measure Crystallized Intelligence and tests that measure Fluid Intelligence. Crystallized Intelligence refers to knowledge about the world that is gathered over one’s whole lifetime. It comprises explicit and implicit knowledge and depends largely on long-term memory. Crystallized Intelligence is measured by tasks as, for example, V*ocabulary*, *General Information*, *Verbal Comprehension*, *Ideational Fluency*, and *Arithmetic Reasoning* (Horn & Cattell, 1967). Fluid Intelligence refers to the speed of problem solving, information processing and the understanding of novel information. The measurement of Fluid Intelligence comprises tasks such as *Letter-Grouping*, *Matrix Reasoning*, *Figure Classification*, and *Inductive Reasoning* (Horn & Cattell, 1967).

With concern to age decline, it has been shown that scores on crystallized tests remain relatively stable into old age, but scores on all other tests decline fast in old age (Juan-Espinosa et al., 2002). A longitudinal study on decline of cognitive abilities in aging individuals by Schaie and Willis (1993) suggests that around age 68 cognitive abilities start to decrease extremely. In the present study, we will test the hypothesis that differences in cognitive abilities between groups younger than 68 and groups older than 68 do not have a strong positive relationship with general

[d] Format of table adapted from te Nijenhuis & Franssen (2010). Verbatim use of formulation of table descriptions from te Nijenhuis & Franssen (2010)

intelligence. Since Crystallized Intelligence is measured by subtests of *Vocabulary*, *Verbal Reasoning*, and *Verbal Comprehension*, which usually have high *g* loadings there is one more reason not to expect a high correlation *d* x *g*. We expect that

(16) the true correlation between score differences of different age groups and the magnitude of *g* loadings is not strongly positive[a]

Study 6b: Autism

Autism is a pervasive developmental disorder with onset in infancy that has been linked to a genetic predisposition. The disorder is characterized by anomalous speech development and language use, an overall lack of interest in social interaction, and a preference for repetitive stereotype behaviors (Rutter, 1978; Szatmari, 1998). About three-fourths of autistic children are also diagnosed with mental retardation (Lincoln, et al., 1988). However, the score deviation of subtests, when they are compared to a standardized group, can be very different. Autistic children tend to show less deviation on tests of perceptual motor organization skills and rote memory skills, but more deviation on tests of verbal abstraction and verbal comprehension (Lincoln, et al., 1988). We will systematically test the difference in the intelligence profile of autistic children and the general population in two steps. First, we explore whether difference in intelligence of autistic children are related to general intelligence, or *g*. Second, we will test whether differences are smaller on tests of perceptual motor organization skills and rote memory skills and larger on tests of verbal abstraction and verbal comprehension. We expect that

(17a) the true correlation between score differences of autistic groups and control groups and the magnitude of *g* loadings is not strongly positive[a]
(17b) score differences are larger on tests of verbal abstraction and verbal comprehension and smaller on tests of perceptual motor organization and rote memory skills[a]

General Method[7]

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

[a] Verbatim use of formulation of hypothesis from te Nijenhuis & Franssen (2010)

results of perfectly conducted studies would allow a clearer view of underlying construct-level relationships (Schmidt & Hunter, 1999). The goal of the present thesis is to explore the nomological net of *g*. The fundamental hypothesis is that the differences in cognitive profile due to ethnicity, heritability, and physical characteristics are strongly predicted by the *g*-loadedness of the subtests in an IQ battery, and that differences due to biological environmental variables, aging, and autism are not strongly related to general intelligence.

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

There has been a discussion whether one should use Pearson's *r* or Spearman's rho when applying the method of correlated vectors (Colom, Juan-Espinosa, Abad, & Garcia, 2000). 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, and 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; te Nijenhuis, de Pater, van Bloois, & Geutjes, 2009; te Nijenhuis & Franssen, 2010; te Nijenhuis & Dragt, 2010; te Nijenhuis & Repko, 2011; te Nijenhuis & Willigers, 2011). This has the advantage that the results of the present studies can be compared directly against those of the earlier studies. (te Nijenhuis & Franssen, 2010)

---

[7] Paragraph taken from te Nijenhuis & Franssen (2010, p. 15)

General Inclusion Rules[8]

For studies to be included in a meta-analysis three criteria had to be met: First, in order to obtain a reliable estimate of the true correlation between each of the variables and *g* loadings, the cognitive batteries had to be based on a minimum of seven subtests. Second, the IQ test had to be well-validated. Finally, only studies published in English, Dutch, or German were used. (te Nijenhuis & Franssen, 2010)

Corrections for Artifacts[9]

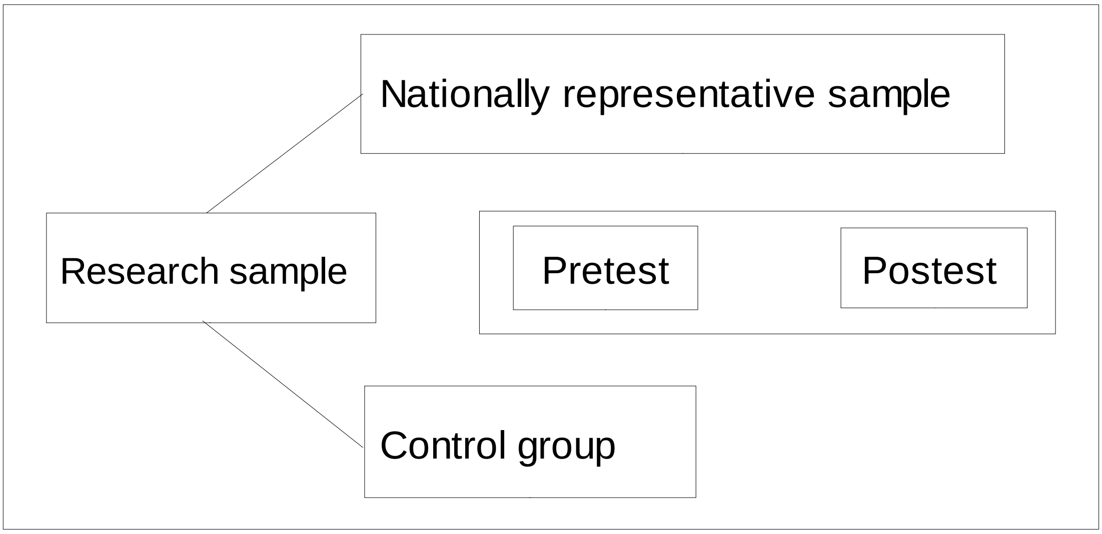
Psychometric meta-analytical techniques (Hunter & Schmidt, 1990, 2004) were applied using the software package developed by Schmidt and Le (2004). Psychometric meta-analysis is based on the principle that there are artifacts in every dataset and that most of these artifacts can be corrected. A full psychometric meta-analysis was carried out only for Spearman's hypothesis tested on reaction time measures. In the full-fledged present meta-analyses carried out in this master thesis 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 most cases, however, we carried out exploratory studies, using bare-bones meta-analytical techniques, where we corrected for only one artifact, namely sampling error. (te Nijenhuis & Franssen, 2010)

Designs for Group Comparisons

When two groups are being compared using the method of correlated vectors various designs can be used. In the following we will describe four designs that are used in this master thesis (see Figure 1).

The first design is a comparison of a research sample and a nationally representative sample of the general population. This design is used when we want to find the difference in the intelligence profile between a research sample with a certain attribute and the general population. For instance, we can answer the question: Do gifted children differ from average children with regard to cognitive abilities? In this case, the research sample should consist of gifted children, which are then compared to average children in form of a nationally representative sample of children. The scores of representative samples are usually reported in the manual of IQ batteries. An advantage of nationally representative samples is that they are usually much larger than control groups, making them more reliable. However, for specific comparisons a small, but carefully matched control group is strongly preferred above a large nationally representative sample as will be described in the following section.

[8] Paragraph taken and adapted from te Nijenhuis & Franssen (2010, p. 16)
[9] Paragraph taken from te Nijenhuis & Franssen (2010, p. 17)



The second design is a comparison of a research sample and a control group. It is used when

Figure 1. Designs for group comparisons

we assume that the research sample differs from the general population on characteristics that are not the subject of the comparison. For instance, when studying the IQs of children of crack-addicted mothers most likely these mothers will have below-average IQs. So, a matched group should also consist of mothers with below-average IQs. A control group increases the plausibility that differences in IQ are actually caused by a certain attribute and not by differences on possible other dimensions on which the research sample might differ compared to a non-matched sample.

The third design is a comparison of pre and posttest outcomes of a research sample that undergoes any kind of change between pre and posttest. For instance, change can occur by external sources as training on an IQ test and a medical treatment, or internal sources such as aging. In principle, any change between pre and posttest that has a potential effect on IQ can be the subject of such a comparison. We use this design when we want to explore whether a treatment or any kind of event that causes change in an individual has an effect on the intelligence profile of an individual. An advantage of such a test-retest design is that variance is reduced to within-subjects variance. Therefore effects caused by differences between groups as can be found in a between subjects design are eliminated. In consequence, any observed difference between pre and posttest is more likely to be caused by the treatment or any kind of change the subjects underwent between pre and posttest. On a duller note: When testing cognitive abilities, subjects tend to increase their score on the second test irrespective of treatment of any kind. Therefore, test-retest effects can obscure the actual change due to treatment between pre and posttests.

The fourth design is a comparison of research samples that have differ degrees of an attribute. For example, if two research samples received a diagnosis of moderate and severe traumatic brain injury (TBI), respectively, we can test whether differences in IQ between different degrees of severity of TBI are related to *g*. Such a comparison is problematic for several reasons. First, the categorization by diagnosis might not be as clear-cut as it seems. Therefore, individuals that received a diagnosis of moderate TBI might have received a diagnosis of severe TBI under slightly other conditions, and vice versa. Second, although there might be an overall effect of TBI on general intelligence, there might be no difference in IQ impairment between degrees of severity. Therefore, suffering from TBI in general could lead to an IQ impairment strongly related to *g.* The severity with which one suffers from TBI might however not lead to a further impairment in IQ that is strongly related to *g.* Since outcomes are hard to predict, such a comparison is usually of exploratory nature.

Choice of *SD*

When we compute the correlation *d* x *g* difference scores (*d*) are computed by subtracting the score of the lower scoring group from the score of the higher scoring group and dividing the result by the best estimate of *SD* available (te Nijenhuis & Dragt, 2010). Our choice of *SDs*, in order of preference, is: First, the *SD* of a national standardization sample; second, the *SD* of a control group; third, a weighted average of the *SD* of the groups involved in the comparison. In cases in which we decide to deviate from this rule, we will state this explicitly.

The *SD* of a national standardization sample. The *SD* of a national standardization sample is the first choice, because it is usually based on a large sample size and is therefore very reliable. However, if no scores from a standardization sample can be obtained, or both groups involved in the comparison are part of a standardization sample, for example, when we compare scores of different school types that are reported in the manual of an IQ battery there are different possibilities to compute the *SD*.

The *SD* of a control group. We choose to use the *SD* of the control group as the best approximation of the true *SD* of the general population, when the *SD* of the standardization sample is not available. The *SD* of a control group is usually based on a sample size that is very small compared to the sample size of a standardized group. Therefore, the *SD* score is less reliable. However, the control group will have a stronger resemblance to a standardized group than an experimental group. Therefore, we assume that the *SD* score of the control group yields a better estimate of the true *SD* of the general population than the *SD* score of the experimental group. Theoretically, there is also the possibility to compute a weighted average of the *SD* of the experimental and the control group. On the one hand, this procedure enhances the overall sample size on which the *SD* is based. Therefore, the *SD* should be more reliable. On the other hand, we compute the correlation *d* x *g*, because the experimental group has different subtests scores than a comparison group. If mean subtests scores are different, also score distributions on subtests might be different and in consequence the *SD*. Therefore, a weighted average of the *SD* of experimental and control group does not necessarily yield a better estimate of the true *SD* of the general population (as the *SD* of the standardization sample would do).

A weighted average of the *SD* of the groups involved in the comparison. If there is no *SD* score from either a standardized group or a control group available, or if comparisons do not involve experimental and control conditions, but groups with different characteristics, we will use a weighted average as the best approximation of the true value of the *SD*. For instance, if all groups involved in the comparison are subsamples of a standardization sample, or the comparison takes place between groups with different manifestations of the same characteristic (for example comparisons between groups of different age, to explore the relationship of *g* and aging) we compute a weighted average of all groups included in the comparisons. This means, if we make a comparison between subsamples A and B of a standardization sample (or age 29 and age 39 when comparing different age groups), and subsequently we also compare subsamples B and C, C and D, A and D, etc., we use a weighted average of the *SD* of all subsamples (A, B ,C, and D). The weighted average *SD* is always computed as the pooled *SD* with the formula $pooledSD = \sqrt{(((N_1-1)*SD_1^2+(N_2-1)*SD_2^2)/(N_1 + N_2 - 2))}$ for two samples. For three samples the pooled *SD* formula changes to $pooled\ SD = \sqrt{(((N_1-1)*SD_1^2+(N_2-1)*SD_2^2+(N_3-1)*SD_3^2)/(N_1 + N_2 + N_3 - 3))}$. For four or more samples the formula changes in the same manner. When the samples have the same *N* the pooled *SD* is equal to the ordinary average of the *SDs* (Agresti & Franklin, 2007).

Analysis of Broad Cognitive Abilities

If differences in IQ profile between groups cannot be explained by differences in general intelligence it is possible to test whether differences between groups are more pronounced on one or more broad cognitive abilities on stratum II of Carroll’s three-stratum model of cognitive abilities. The most widespread cognitive abilities tests used in our analyses are the various versions of the Wechsler tests. All subtests of Wechsler tests fall within one of four different broad cognitive abilities. These abilities are: Fluid Intelligence, Crystallized Intelligence, Broad Visual Perception, and General Memory and Learning. To test on which of the four broad cognitive abilities differences are more pronounced between groups of interest, we first average all *d* scores of subtests obtained in a study that fall within a broad cognitive ability. Thereafter we conduct a bare-bones meta-analysis of the average effect sizes for each broad cognitive ability that we obtained in a study with the meta-analysis program of Schmidt and Lee (2004). The resulting meta-analytic effect size can be interpreted as the true effect size for differences on tests of the broad cognitive ability in question. In cases in which we could only obtain a limited amount of studies, that did not report subtest scores for all broad cognitive abilities a meta-analysis could not be conducted. In these cases we decided to compute an average of all subtest that fall within a broad cognitive ability for all studies.

Measures of Broad and Narrow Abilities in Wechsler Batteries[10]

The various versions of the Wechsler IQ battery are by far the most used instrument in the studies in the present thesis, with the WISC-R (Wechsler, 1974) the most commonly used. Figure 2 depictures the various broad and narrow abilities measured by all the subtests of the various Wechsler batteries, namely the WISC (Wechsler, 1949), WISC-R (Wechsler, 1974), WISC–III (Wechsler, 1991), WISC-IV (Wechsler, 2003), WBIS (Wechsler, 1939), WAIS (Wechsler, 1955), WAIS-R (Wechsler, 1981), WAIS–III (Wechsler, 1997), WAIS-IV (Wechsler, 2008), WPPSI (Wechsler, 1967), WPPSI-R (Wechsler, 1989), and WPPSI–III (Wechsler, 2002).

Wechsler batteries are tools to assess the cognitive abilities of populations of different ages. The WPPSI (Wechsler, 1967) is used to assess preschool children (age four to age six and a half), with the WISC (Wechsler, 1949) children between age six to age sixteen can be tested, and the WAIS (Wechsler, 1955), previously WBIS (Wechsler, 1939), is used for adult assessment. The upper bound of the age range of the WAIS (Wechsler, 1955) differs between versions. First, we will describe the first version of the WISC (Wechsler, 1949). After that we will report differences in subtests between the different version of the WISC (Wechsler, 1949), differences in subtests between the WISC (Wechsler, 1949), the WPPSI (Wechsler, 1967), and the WAIS (Wechsler, 1955), and differences in subtests in the different versions of the WAIS (Wechsler, 1955) and the WPPSI (Wechsler, 1967). Subtests with the same name in the WISC (Wechsler, 1949), WAIS (Wechsler, 1955), and WPPSI (Wechsler, 1967) differ with regard to administration and difficulty of content, but measure essentially the same broad and narrow abilities.

Subtests of the WISC.[11]

1. *Information*; the examinee has to answer verbally all kinds of general questions, some of which have several possible correct answers. This subtest measures general information, which is a measure of Crystallized Intelligence at stratum II.
2. *Picture Completion*; the examinee has to find out which essential part of a picture is missing, within a given time. This subtest measures Closure Speed at stratum I, which makes it a measure of Broad Visual Perception at stratum II.
3. *Similarities*; the examinee has to find a similarity between two objects or concepts. There are several correct answers. This subtest is a measure of Induction at stratum I, which makes it a measure of Fluid Intelligence at stratum II.
4. *Picture Arrangement*; the examinee has to order a series of pictures in such a way that the pictures form a comprehensive story, within a given time. This subtest is a measure of

[10] The formulation of descriptions of subtests in this paragraph is adapted from te Nijenhuis & Dragt (2010)
[11] Paragraph taken and adapted from te Nijenhuis & Dragt (2010, p. 28)

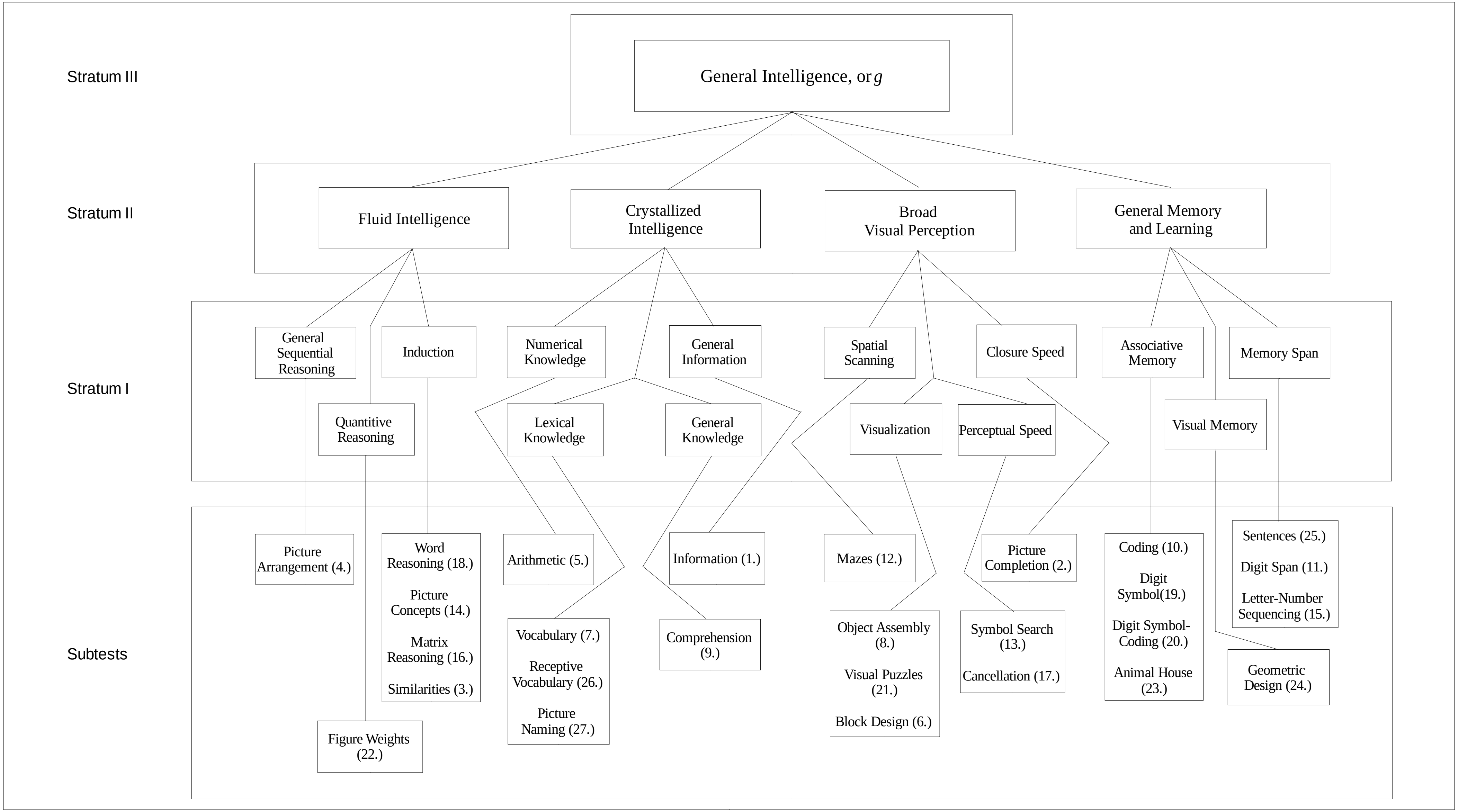


Figure 2. All subtests used in Wechsler batteries in the three-stratum model of cognitive abilities (Carroll,1993). Numbers in brackets refer to the descriptions of subtests.

General Sequential Reasoning at stratum I, which makes it a measure of Fluid Intelligence at stratum II.

5. *Arithmetic*; the examinee has to solve arithmetic problems. Since arithmetic problems are of rather low difficulty the subtest can be considered a measure of Numerical Knowledge at stratum I, which falls under Crystallized Intelligence at stratum II.
6. *Block Design*; the examinee has to replicate pattern presented on a card with blocks. This subtest is a measure of Visualization at stratum I, which measures Broad Visual Perception at stratum II.
7. *Vocabulary*; the examinee has to give the meaning of a presented word. This subtest measures Lexical Knowledge at stratum I, which makes it a measure of Crystallized Intelligence at stratum II.
8. *Object Assembly*; the examinee has to put different pieces of cardboard together to copy a given figure within a given time. This subtest is a measure of Visualization at stratum I, which measures Broad Visual Perception at stratum II.
9. *Comprehension*; the examinee has to answer different questions in which they have to give their insight and judgment about everyday life issues. This subtests measures general knowledge which is a measure of Crystallized Intelligence at stratum II and is a measure of General Sequential Reasoning at stratum I which is a measure of Fluid Intelligence at stratum II.
10. *Coding*; the examinee has to put a sign in a series of figures (code A) or under a series of numbers (code B). The sign belonging to the figure of number was presented to the examinee earlier. This subtest is a measure of Associative Memory at stratum I, which falls under General Memory and Learning at stratum II.
11. *Digit Span*; the examinee has to repeat a series of numbers exactly in the sequence presented to them auditorily (Forward Digit Span) or in reverse order starting with the last number they heard back to the first number (Backward Digit Span). This subtest is a measure of Memory Span at stratum I, which makes it a measure of General Memory and Learning at stratum II.
12. *Mazes*; the examinee has to trace the way out of a maze (presented on paper) with a pencil within a given time. The examinee is not allowed to enter a dead end. This subtest is a measure of Spatial Scanning at stratum I, which falls under Broad Visual Perception at stratum II.

The WISC (Wechsler, 1949) and the WISC-R (Wechsler, 1974) have the same subtest, therefore a description of WISC-R (Wechsler, 1974) subtests is not necessary. (te Nijenhuis & Dragt, 2010)

Changes in subtests between the WISC and the WISC–III. The WISC–III (Wechsler, 1991) has an additional subtest Symbol search:

13. *Symbol Search*; examinees are presented with a group of target symbols and a group of search

symbols. The child scans the two groups and indicates whether or not a target symbol appears in the search group. Symbol search is a measure of Perceptual Speed at stratum I, which falls under Broad Visual Perception at stratum II.

Changes in subtests between the WISC–III and the WISC-IV. The WISC-IV (Wechsler, 2003) does not include the tests *Picture Arrangement*, *Object Assembly,* and *Mazes*, anymore, and now includes five new tests:

14. *Picture Concepts*; examinees are asked to look at rows of pictures of objects and indicate the single picture from each row that shares a characteristic in common with the single picture(s) from the other row(s). This test is a measure of Induction at stratum I, which falls under Fluid Intelligence at stratum II.
15. *Letter-Number Sequencing*; examinees are asked to listen to and remembering a string of digits and letters read aloud at a speed of one per second. Subsequently, children try to repeat the numbers in chronological order, followed by the letters in alphabetical order. This subtest is a measure of Memory Span at stratum I, which falls under General memory and learning at stratum II.
16. *Matrix Reasoning*; examinees are presented with incomplete matrices and have to choose the missing part from a set of possibilities. This subtest is a measure of Induction at stratum I, which falls under Fluid Intelligence at stratum II.
17. *Cancellation*; examinees are asked to scan both a random and structured arrangement of pictures and mark target pictures within a specified time limit. This subtest is a measure of Perceptual Speed at stratum I, which falls under Broad Visual Perception at stratum II.
18. *Word Reasoning*; examinees are asked to identify the common concept being described in a series of clues. This test is a measure of Induction at stratum I, what makes it a measure of Fluid Intelligence at stratum II. (Descriptions are taken from Wechsler, 2003)

Subtests of the WAIS. Although administration and difficulty of content of subtests differ between WAIS (Wechsler, 1955) and WISC (Wechsler, 1949), the broad and narrow abilities measured by subtests are similar. There are two exceptions: The subtest *Mazes* is left out and the subtest *Digit Symbol* replaces the subtest *Coding*. *Digit Symbol* can be a considered a more difficult version of the subtest Coding (B) from the WISC (Wechsler, 1949).

19. *Digit Symbol*; the examinee receives a list of pairs consisting of digits and corresponding symbols. Subsequently, the adult receives a list with the same digits and has to fill in the corresponding symbols. The number of correct symbols after a time period of 90 or 120 seconds is measured. This subtest is a measure of Associative Memory at stratum I, which falls under General Memory and Learning at stratum II.

The WBIS or Wechsler Bellevue Intelligence Scale (Wechsler, 1939) is an earlier version of the

WAIS (Wechsler, 1955), which did not have the subtest *Vocabulary*, but was similar to the WAIS (Wechsler, 1955) in all other subtests. The WAIS (Wechsler, 1955) and the WAIS-R (Wechsler, 1981) have the same subtests, therefore a description of WAIS-R (Wechsler, 1981) subtests is not necessary.

Changes in subtests between the WAIS and the WAIS–III. The WAIS–III (Wechsler, 1997) has three additional subtests (*Letter-Number Sequencing*, *Symbol Search*, *Matrix Reasoning*) and an extended version of the test *Digit Symbol*: *Digit-Symbol Coding*.

Letter-Number Sequencing; (see Changes in subtests between the WISC–III and the WISC-IV ;15. *Letter-Number Sequencing*)

Matrix Reasoning; (see Changes in subtests between the WISC–III and the WISC-IV;16. *Matrix Reasoning*)

Symbol search; (see Changes in subtests between the WISC and the WISC–III; 13. *Symbol search)*

20. *Digit symbol-coding*; this test is the same as *Digit-symbol* in previous versions of the WAIS except that after the test paired and free recall of symbols is assessed. This subtest is a measure of Associative Memory at stratum I, which falls under General Memory and Learning at stratum II.

Changes in subtests between the WAIS-III and the WAIS-IV. The WAIS-IV(Wechsler, 2008) does not include the subtests *Picture Arrangement* and *Object Assembly*, and the subtest *Digit Symbol-Coding* is changed to the subtest *Coding*. Three subtests are added:

21. *Visual Puzzles*; examinees are presented with a picture and are subsequently asked to choose from a set of puzzle pieces which can be combined to reproduce the picture. This subtest is a measure of Visualization at stratum I, which falls under Broad Visual Perception at stratum II.

22. *Figure Weights*; examinees are presented with pictures of three scales in equilibrium. Two of these scales have geometrical objects in different colors on each side. A third scale has geometrical objects on one side only. The examinee has to infer quantitative relations between objects from the first two scales, for example one blue ball equals two yellow stars, and has to choose the correct set of objects from several possibilities, so that the third scale remains in equilibrium too. This subtest is a measure of Quantitative Reasoning at stratum I, which makes it a measure of Fluid Intelligence at stratum II.

Cancellation; (see Changes in subtests between the WISC–III and the WISC-IV; 17. *Cancellation*)

Coding; examinees are presented with number-sign combinations. Subsequently they are presented numbers only, and have to fill in the appropriate sign (also see Subtests of the WISC; 10. Coding, Code B). This subtest is a measure of Associative Memory at stratum I, which falls under General Memory and Learning at stratum II.

Subtests of the WPSSI. In comparison to the WISC (Wechsler, 1949), the WPPSI (Wechsler, 1967) has easier items on most subtests. On four subtests (*Arithmetic*, *Similarities*, *Mazes*, and *Block Design*) content and mode of administration has changed. Still, these subtests measure the same broad and narrow abilities as the subtests of the WISC (Wechsler, 1949). The subtests *Digit Span* and *Coding* are replaced by the subtests *Sentences* and *Animal House,* respectively. The subtests *Picture Arrangement* and *Object Assembly* are left out. The three additional or changed subtests of the WPPSI (Wechsler, 1967) are:

23. *Animal House*; this subtest is similar to the subtest Coding in the WISC (Wechsler, 1949), except that the child has to associate symbols with animal pictures instead of signs and numbers. This subtest is a measure of Associative Memory at stratum I, which falls under General Memory and Learning at stratum II.
24. *Geometric Design*; examinees are asked to reproduce a geometric figure. This subtest is a measure of Visual Memory as stratum I, which falls under General Memory and Learning at stratum II.
25. *Sentences*; examinees are asked to reproduce a sentence. This subtest is a measure of Memory Span at stratum I, which falls under General Memory and Learning at Stratum II.

Changes in subtests between the WIPPSI and the WIPPSI-R. The WPPSI-R (Wechsler, 1989) retains all subtests of the WPPSI (Wechsler, 1967) and has an additional subtest *Object Assembly*. The subtest *Animal House* is changed to *Animal Pegs*, but remains basically the same.

*Object Assembly***;** (see Subtests of the WISC; 8. *Object Assembly)*

Changes in subtests between the WIPPSI-R and the WIPPSI-III. In the WPPSI-III (Wechsler, 2002) the subtests *Arithmetic*, *Animal Pegs*, *Geometric Design*, *Mazes*, and *Sentences* are left out. The subtests *Receptive Vocabulary*, *Picture Naming*, *Word Reasoning*, *Matrix Reasoning*, *Picture Concepts*, *Coding*, and *Symbol Search* are added:

26. *Receptive Vocabulary*; the examinee is presented with several pictures and is asked to point to the picture that is described by the examiner. This subtest is a measure of Lexical Knowledge at stratum I, which falls under Crystallized Intelligence at stratum II.
27. *Picture Naming*; examinees name pictures that are presented to them. This subtest is a measure of Lexical Knowledge at stratum I, which falls under Crystallized Intelligence at stratum II.

Word Reasoning; (see Changes in subtests between the WISC–III and the WISC-IV; 18. *Word Reasoning)*

Matrix Reasoning; *(*see Changes in subtests between the WISC–III and the WISC-IV; 16. *Matrix Reasoning)*

Picture Concepts; (see Changes in subtests between the WISC–III and the WISC-IV; 14. *Picture concepts)*

Coding (see Subtests of the WISC; 10. *Coding*, Code A*)*

Symbol Search; (see Changes in subtests between the WISC and the WISC–III; 13. *Symbol search)*

Study 1: Group Differences

Study 1a: Difference Between Whites and Higher-IQ and Lower-IQ Groups in Reaction Time

In almost all studies that were included in this meta-analysis the reaction time measure of Jensen (1993) were used. The reaction time measure of Jensen (1983) consists of an apparatus with a *home button* in the middle and several *reaction buttons* located in a half circle around the *home button.* This apparatus is connected to a computer screen on which stimuli is presented. The participant has to rest his hand on the *home button* until a stimulus is presented. When the stimulus is presented the participant removes his hand from the *home button* and presses the correct *reaction button.* There are three different tasks: A) *Simple reaction time* (SRT): In this task, the participant has to press a designated *reaction button*. There is no uncertainty which button will light up, but there is uncertainty with regard to when the button will light up. This task measures pure reaction time. B) *Choice reaction time* (CRT): In this task the participant has to press a not designated *reaction button*. Hence, there is uncertainty with regard to which button will light up and when the button will light up. This task is slightly more complex than the SRT. The time needed to choose the correct *reaction button* is an additional indication of the speed of cognitive processing. C) *Discrimination reaction time* (Oddman RT): In this task three different *reaction buttons* light up. The button that is most far away from the other two buttons is the correct *reaction button*. In addition to the uncertainty concerning which *reaction buttons* light up and when they light up, the participant has also to identify the correct *reaction button*. Since the Oddman RT involves the most cognitive processing of the three tasks, Jensen (1983) suggests that the Oddman RT should show the highest correlation with IQ from the three tasks.

Four different reaction time measures are recorded. The RT (reaction time) consists of the time interval between the onset of the stimulus and the removal of the hand from the *home button*. The MT (movement time) consists of the time interval between the removal of the hand from the *home button* and the activation of the *reaction button*. SDRT (Standard deviation of reaction time) measures the intraindividual variability in RTs. Every participant has a mean reaction time with a standard deviation. The SDRT is the average of this *SD* of all participants; SDMT (Standard deviation of movement time) is the analogue measure for movement time. So, for all measures a low score is a good score and a high score is a bad score.

In the study of Vernon and Jensen (1984) different reaction time measures were used. In this study the measures DIGIT (Speed of processing), SD2 & SA2 (Speed of retrieval of information from long-term memory), DT2 and DT3 (Efficiency of short term memory storage and processing), and RT (Simple and choice reaction time or decision making) were used. For more information, we refer to the original article of Vernon and Jensen (1984).

The purpose of this study is to determine whether the correlation between the magnitude of *g* loadings and difference scores on a set of reaction time measures between Whites and lower-IQ groups, and Whites and higher-IQ groups is strongly positive in sign. We will test this by conducting two full-fledged psychometric meta-analyses on studies that reported reaction time measures for Whites, Blacks, and Asians.

Method

Searching and screening studies. All articles were taken from the archive of dr. Jan te Nijenhuis, who published extensively on Spearman's hypothesis and corresponded with the small group of persons who published empirical papers on the topic. Dr. te Nijenhuis assured that the studies assembled here constitute all published papers on the topic.

Specific criteria for inclusion. First, only studies that reported several reaction time measures were included. Second, we could only include sets of reaction time measures for which *g* loadings were available or could be estimated. In sum, we yielded four studies on differences in reaction between Whites and Blacks and five studies on differences between Whites and Asians.

Computation of score differences between groups of different ethnic background. Studies on Spearman's hypothesis are carried out with all kinds of measures and comparing all kinds of groups. It is therefore important that the comparisons are made in similar ways, so the outcomes of all studies and meta-analyses of studies can be easily compared. For instance, on IQ batteries a confirmation of Spearman's hypothesis entails a positive score, meaning that on tasks with high *g* loadings there are large differences (high-IQ group much better score) and on tasks with low *g* loadings there are small differences (high-IQ group slightly better score).

Computation of *g* scores. Unlike the computation of *g* loadings for IQ batteries, we could not use the correlation matrix of subtest scores of reaction time measures to compute *g* loadings. Of course, we could have conducted a principal component analysis on the correlation matrix of subtest scores of reaction time measures. However, it is hard to argue from a theoretical perspective that the resulting loadings on the first factor of this analysis indeed are loadings on *g*. This is because, unlike IQ tests, reaction time is not a primary measure of intelligence. An alternative way to obtain *g* loadings for reaction time measures was proposed by Jensen (1993). Here the correlation between a reaction time subtest (for example the SDRT) and the score on Raven's standard progressive matrices (SPM) is computed. Raven's SPM is known to be one of the best measures of Spearman's *g* (Jensen, 1983). So, the correlation of a reaction time subtest with the Raven's score yields a good estimate of the *g* loading of the reaction time subtest. In the case of reaction time all measures should correlate negatively with the SPM; however, there is no such thing as a negative *g* loading, so we decided to multiply a negative value of the correlation coefficient with -1.

An additional consideration is that due to (1) the many non-large samples and (2) the quite modest *g*-loadedness of some measures, the correlations in some cases had the opposite sign of what one would theoretically expect. That is, some reaction time measures showed an initial positive correlation with SPM, which means that the slower the reaction time the higher is the score on the SPM. Obviously, from a theoretical perspective, this makes no sense. However, this is exactly what one would expect in a meta-analysis (see Schmidt, 1992). This effect becomes clear when all correlations from different studies were put beside each other in a Table. A visual inspection showed that some correlations departed from the pattern one would theoretically expect. Still, the *g* loadings need to have positive signs, so that results derived from correlation between *g* loadings and reaction time differences are meaningful and can be compared to previous studies on this topic. There are four solutions to this problem: First, we could score reaction time subtests that showed an initially negative correlation with the SPM contra-indicatively. Therefore, a slower reaction time becomes the better score. However, a) from a theoretical perspective, a slower reaction time as a better indicator of intelligence does not make a lot of sense, and b) from a practical perspective, the computation of the reliability of the *d* vector becomes virtually impossible. The second solution would be, to assign initially negative correlations a *g* loading value of zero. However, this would lower the *g*-loadedness in the analyses where we combined *g* loadings from various studies. The third solution would be to exclude all measures with negative *g* loadings from the analysis; however, this would have led to an exclusion of too many subtests. We chose the fourth solution, that is, we reversed the sign on the value of the correlation to arrive at an estimate of the *g* loading for that specific measure. After this adjustment, all *g* loadings are positive in in sign. We admit there is an element of arbitrariness in these choices, and will evaluate our choices in Discussion.

Computation of *d* scores. Following the practice in testing Spearman's hypothesis first, the scores of low-*g* groups were subtracted from the scores of the high-*g* group. Scores of Blacks were subtracted from scores of Whites; and scores of Whites were subtracted from scores of North-East Asians, as the North-East Asians had higher mean scores on Raven's Progressive Matrices. As a low score is a good score and a high score is a bad score, a negative difference score means that the high-*g* group does better. However, in all other tests of Spearman's hypothesis a positive difference score means that the high-*g* people do better, so we multiplied the difference scores by -1. The resulting product is divided by the *pooled SD* of both groups involved in the comparison. The formula is $pooledSD = \sqrt{(((N_1-1)*SD_1^2+(N_2-1)*SD_2^2)/(N_1 + N_2 - 2))}$. This allows a classical interpretation of the value of *d*: positive values of *d* mean the high-*g* group has a better score, and negative values of *d* mean that the lower-g group has a better score.

Corrections for artifacts.[12] 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. A full psychometric meta-analysis was carried out for Spearman's hypothesis tested on reaction time measures. In the full-fledged present meta-analyses carried out in this master thesis, 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. (te Nijenhuis & Franssen, 2010)

Correction for sampling error.[13] 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. (te Nijenhuis & Franssen, 2010)

Correction for Reliability of the vector of g loadings.[14] The value of r ($g \times d$) for Spearman's hypothesis tested on reaction time is attenuated by the reliability of the vector of *g* loadings for a given battery of RT measures. 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 *g* loadings was estimated using the present datasets, comparing samples that took the same test and that were comparable with regard to age and sample size. Age of participants from all studies fell within a range of nine to eleven years. Sample sizes were considered comparable if the *N* of the larger sample in the comparison was less than 200% of the smaller sample. Although the sample sizes were highly comparable, we accounted for differences in sample size by using the Harmonic mean as the sample size for a correlation between two vectors of *g* loadings.
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 (g x g) and *N* expected. We checked to see which curve gave the best fit to the expected asymptotic function. Figure 3a shows the scatter plot of reliability of the *g* vector of RT measures and sample size, and the logarithmic

[12] Paragraph taken and adapted from te Nijenhuis & Franssen (2010, p. 17)
[13] Paragraph taken from te Nijenhuis & Franssen (2010, p. 18)
[14] Paragraph taken and adapted from te Nijenhuis & Franssen (2010, p. 18)

curve that fitted optimally. Of course, there is no such thing as a negative reliability, and the occurrence of such a large number of negative correlations between various vectors of *g* loadings indicates that these studies require a hard look at the reliability of the *g* loadings. From the initial sample of 51 correlations, we excluded 17 correlations that had a value equal to or smaller than 0.0 (see Figure 3b). The final sample consisted of 34 correlations larger than 0.0. We estimated the reliability of a specific sample by taking the value of the reliability as the point on the regression line for the *N* of that study. (te Nijenhuis & Franssen, 2010)

Correction for reliability of the vector of the second variable.[15] The values of r ($g \times d$) for Spearman's hypothesis tested on reaction time measures are attenuated by the reliability of the *d* vector 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 group differences in reaction time measures were each estimated using the present datasets, comparing samples that took the same test, and that differed little on background variables.

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 4a shows the scatter plot of reliability of the *d* vector RT measures and sample size, and the logarithmic curve that fitted optimally. Of course, there is no such thing as a negative reliability, and the occurrence of such a large number of negative correlations between various vectors of *d* loadings indicates that these studies require a hard look at the reliability of *d*. From the initial sample of 23 correlations, we excluded 7 correlations that had a value equal to or smaller than 0.0 (see Figure 4b). The final sample consisted of 16 correlations larger than 0.0. We estimated the reliability of a specific sample by taking the value of the reliability as the point on the regression line for the *N* of that study. (te Nijenhuis & Franssen, 2010)

Correction for restriction of range of g loadings.[16] The values of r ($g \times d$) for Spearman's hypothesis tested on reaction time measures are attenuated by the restriction of range of *g* loadings in the collections of reaction time measures, just as there is attenuation in many of the standard test batteries when testing Spearman's hypothesis on 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

[15] Paragraph taken and adapted from te Nijenhuis & Franssen (2010, p.20)
[16] Paragraph taken and adapted from te Nijenhuis & Franssen (2010, p. 20)

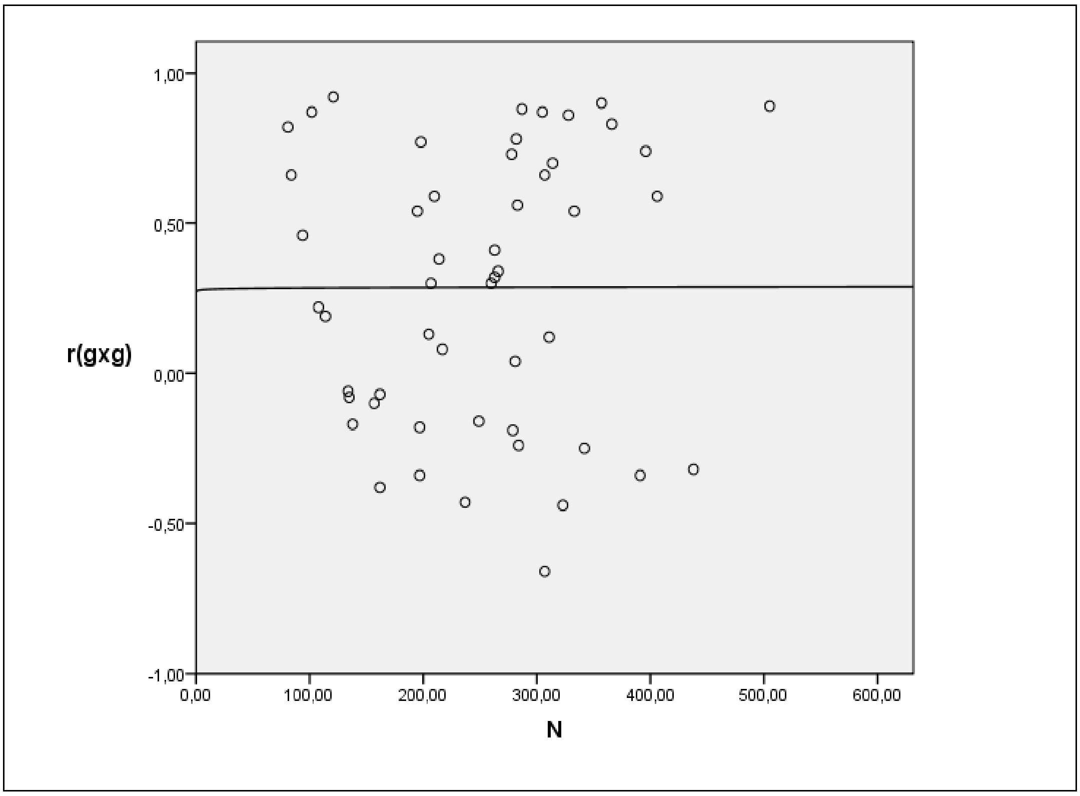


Figure 3a. Scatter plot with logarithmic regression line of reliability of the vector of *g* loadings taken from White, Black, and Asian samples and sample size.

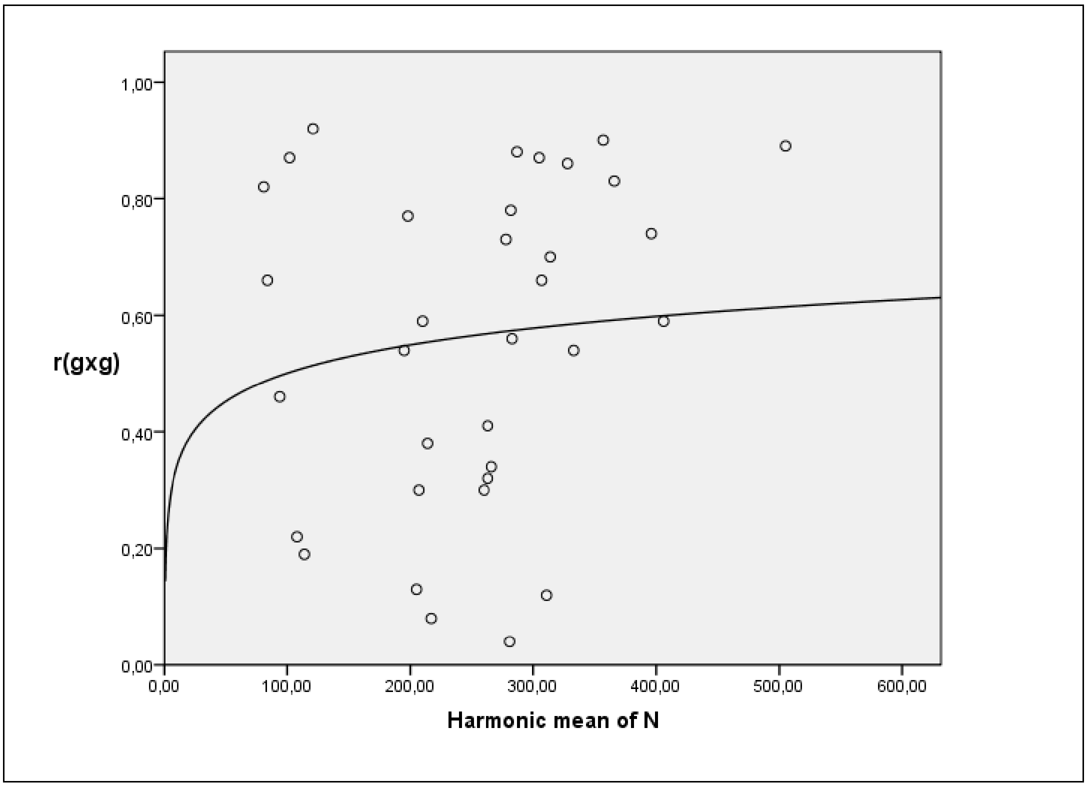


Figure 3b. Scatter plot excluding negative correlations with logarithmic regression line of reliability of the vector of *g* loadings taken from White, Black, and Asian samples and sample size.

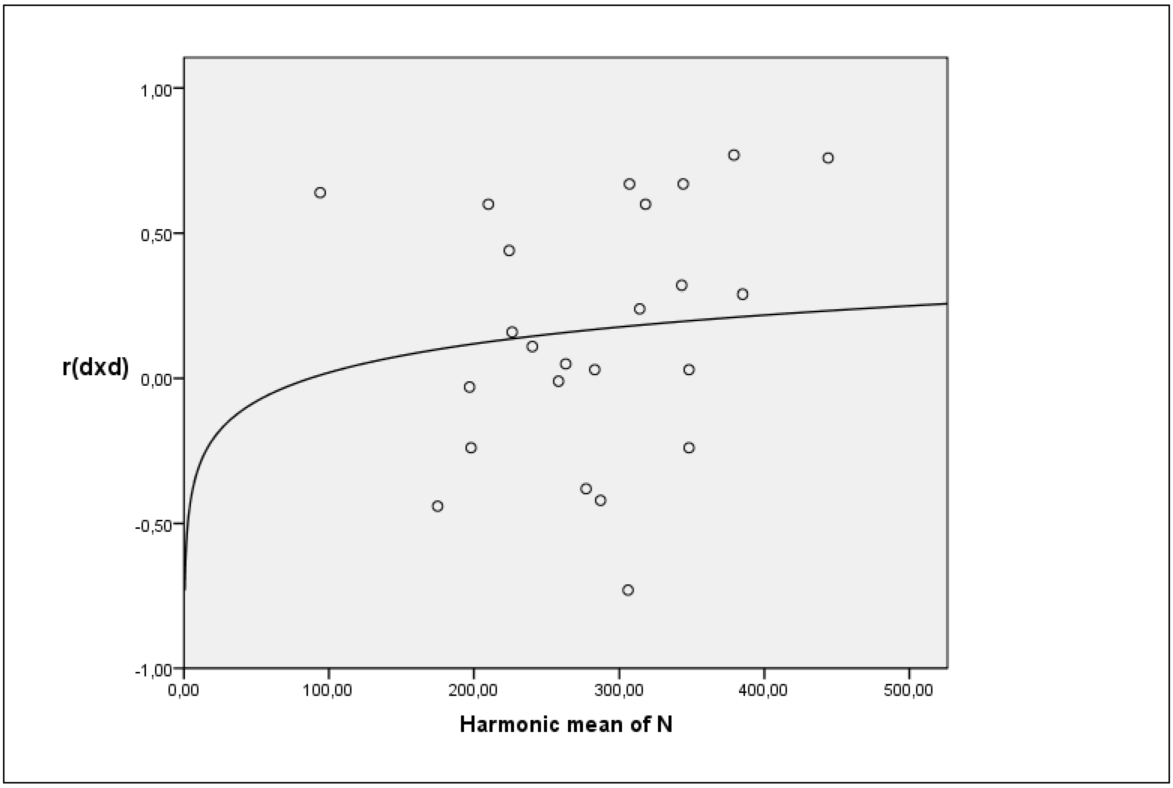


Figure 4a. Scatter plot with logarithmic regression line of reliability of the vector of reaction time differences between Asians, Whites, and Blacks and sample size.

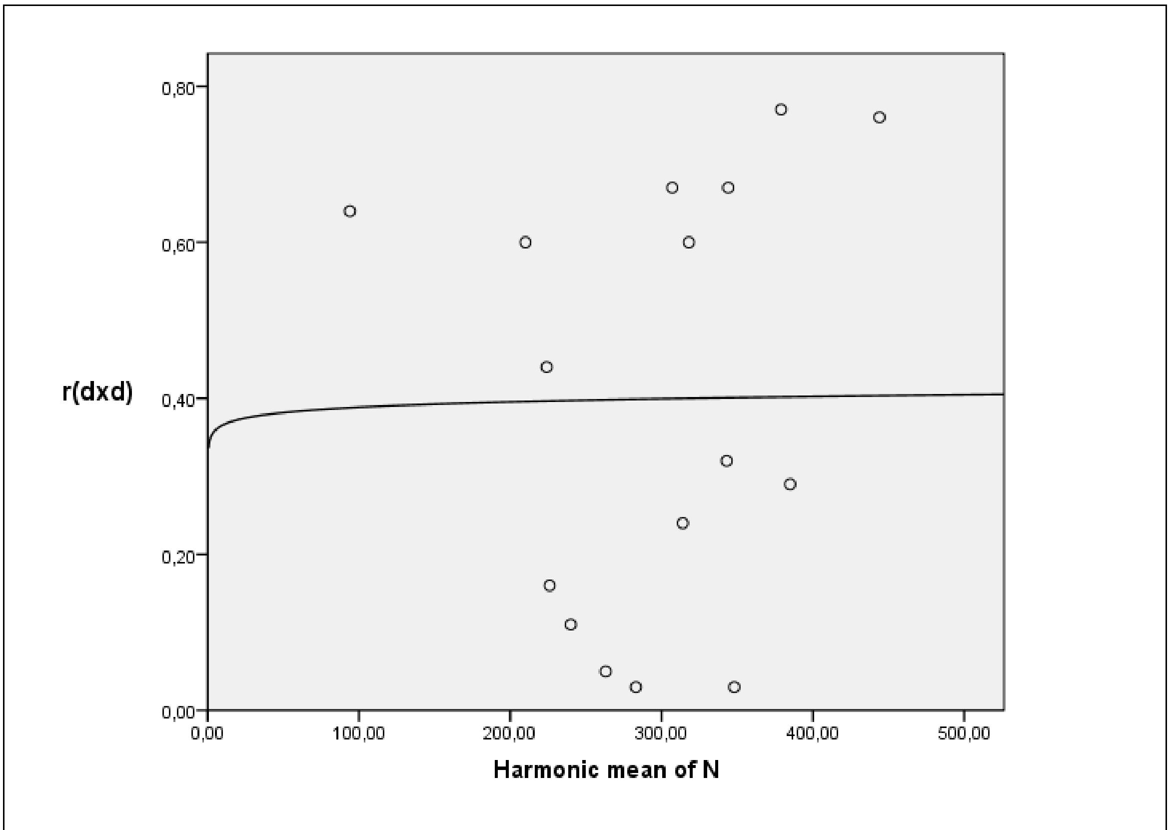


Figure 4b. Scatter plot excluding negative correlations with logarithmic regression line of

population would be if the standard deviation were the same as in the reference population. The reliability of the vector of reaction time differences between Asians, Whites, and Blacks and sample size.

standard deviations can be compared by dividing the standard deviation of the study population by the standard deviation of the reference group, that is $u = SD_{\text{study}}/SD_{\text{ref}}$.

There are several ways to define a reference population – and use its *SD* of *g* loadings – in specific tests of Spearman's hypothesis. Frank Schmidt (personal communication with Jan te Nijenhuis, 2010) suggested to use the maximal range of $g = 0$ to $g = 1$ for every meta-analytical test of Spearman's hypothesis. However, the ranges for different measures are quite different, as shown in Table 6. Moreover, this might possibly lead to values of rho-5, which are much larger than 1.00, suggesting there is massive overcorrection. Finally, a practical problem is how to compute the *SD* of *g* loadings within this theoretical range from 0 to 1. This suggests that this could be a less sound approach, for both practical and theoretical reasons.

Jensen (1998, pp. 381-382) could be interpreted as suggesting that every meta-analytical test of Spearman's hypothesis should have its own reference population, resulting in every meta-analysis having its own *SD* of *g* loadings. Indeed, Table 6 gives rough estimates of the range of *g* loadings for different measures and shows highly different ranges; for instance, the range for IQ batteries is much higher than the range for reaction time measures. This most likely will also lead to substantial differences in the value of SD_g, although this is not necessarily so.

In the present study we choose to elaborate Jensen's approach and we constructed a standard range of *g* loadings, yielding a standard value of SD_g, where the *SD*s of *g* loadings in all the individual studies can be compared to, yielding a value of *u* for every data point. First, the *g* loadings of all collections of reaction time measures in the individual studies were collected and *SD*s of this collection of *g*s were computed for every individual study. Second, all the *g* loadings from all individual studies in the meta-analysis were entered into an SPSS file, and a histogram was computed (Figure 5). Third, the *SD* of the complete collection of *g* loadings was computed. This SD_g is then the standard to which the *SDg* of the individual datasets are being compared.

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. (te Nijenhuis & Franssen, 2010)

Correction for deviation from perfect construct validity.[17] The deviation from perfect construct validity in *g* attenuates the values of r ($g \times d$). In making up any collection of cognitive

[17] Paragraph taken from te Nijenhuis & Franssen (2010, p. 21)

Table 6

Range of g Loadings When Testing Spearman's Hypothesis for Various Measures

	g loadings	
measure	Lowest	highest
IQ battery	.20	.80
Educational criteria	.50	.80
Reaction times	.01	.35

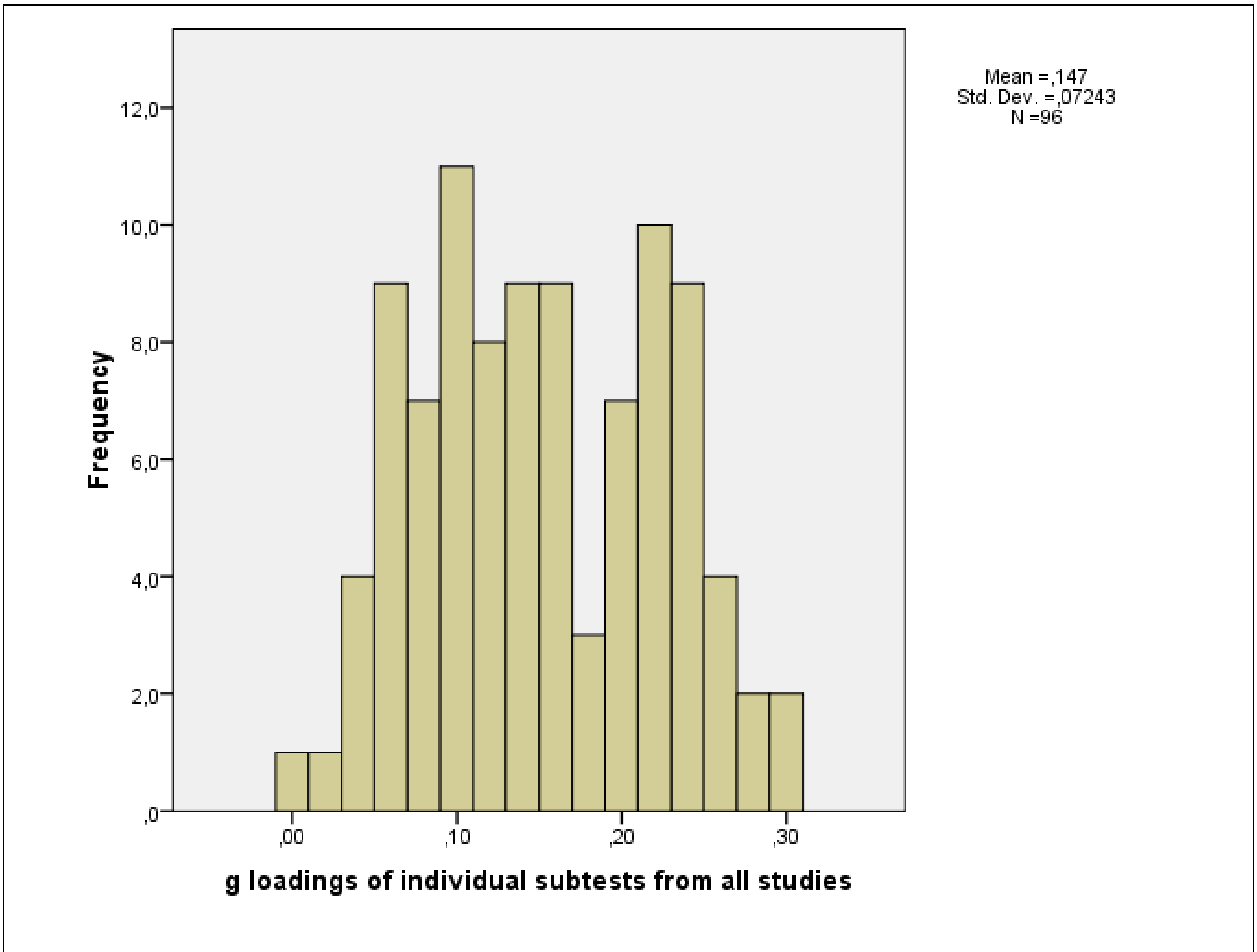


Figure 5. Histogram of *g* loadings of individual subtests from all studies.

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* × *d*) are attenuated by psychometric sampling error in each of the batteries from which a *g* factor has been extracted. We carried out a separate study to estimate the values for this correction.
The more tests and the higher their *g* loadings, the higher the *g* saturation of the composite score is. 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 this value is taken as an indication of the degree to which an IQ score is a reflection of “true” *g*, it can be estimated that a test’s *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 has to be carried out on the values of *r* (*g* × *d*) after correction for the first four artifacts. To limit the risk of overcorrection, in previous studies a conservative choice of the value of .90 for the correction was made (te Nijenhuis, van Vianen, & van der Flier, 2007; te Nijenhuis, & Grimen, 2007; te Nijenhuis, de Pater, van Bloois, & Geutjes, 2009; te Nijenhuis & Franssen, 2009; te Nijenhuis & van der Flier, 2009). Te Nijenhuis & Dragt (2010) showed that this correction value for imperfectly measuring *g* is too strong. These researchers estimated the value of the correlation between *g* scores and “true *g*” scores to be .925. The observed correlation corrected for sampling error, unreliability, range restriction, and imperfect construct validity is referred to as rho-5, as it is corrected for five statistical artifacts. (te Nijenhuis & Franssen, 2010)

Results[c]

To test whether differences in reaction time between higher-IQ groups and lower-IQ groups are related to *g* we conducted meta-analyses on the correlations *d* x *g* between Asians and Whites, and Whites and Blacks. The meta-analysis on the correlation *d* x *g* between Asians and Whites included eight data points. Figure 6 depicts a scatterplot for the obtained correlations *d* x *g* and the harmonic mean. The outcomes of the meta-analysis are reported in Table 7. It shows the number of studies (*K*), the number of participants (*N*), the bare-bones meta-analytic correlation *d* x *g* (*r*), the standard deviation of this correlation (SD_r), the meta-analytic correlation *d* x *g* corrected for four statistical artifacts (rho-4), the SD of this correlation ($SD_{\text{rho-4}}$), correlation *d* x *g* corrected for all five

[c] Verbatim use of formulation of reporting of results from te Nijenhuis & Franssen (2010)

statistical artifacts (rho-5), the percentage of variance explained by all artefactual errors (%VE), and the 80% confidence interval (80%CI). The meta-analysis yields a correlation *d* x *g* with a value of -.64 with 1% of explained variance only. This percentage of variance explained is so low, we tested the origin of comparisons as a moderator. Most of the values *d* x *g* derived from reaction time measures reported in the same study. However, we did also compute vectors of reaction time measures derived from different studies. When we exclude the latter comparisons, the rho-5 becomes .48. The percentage of explained variance is still low (3%).

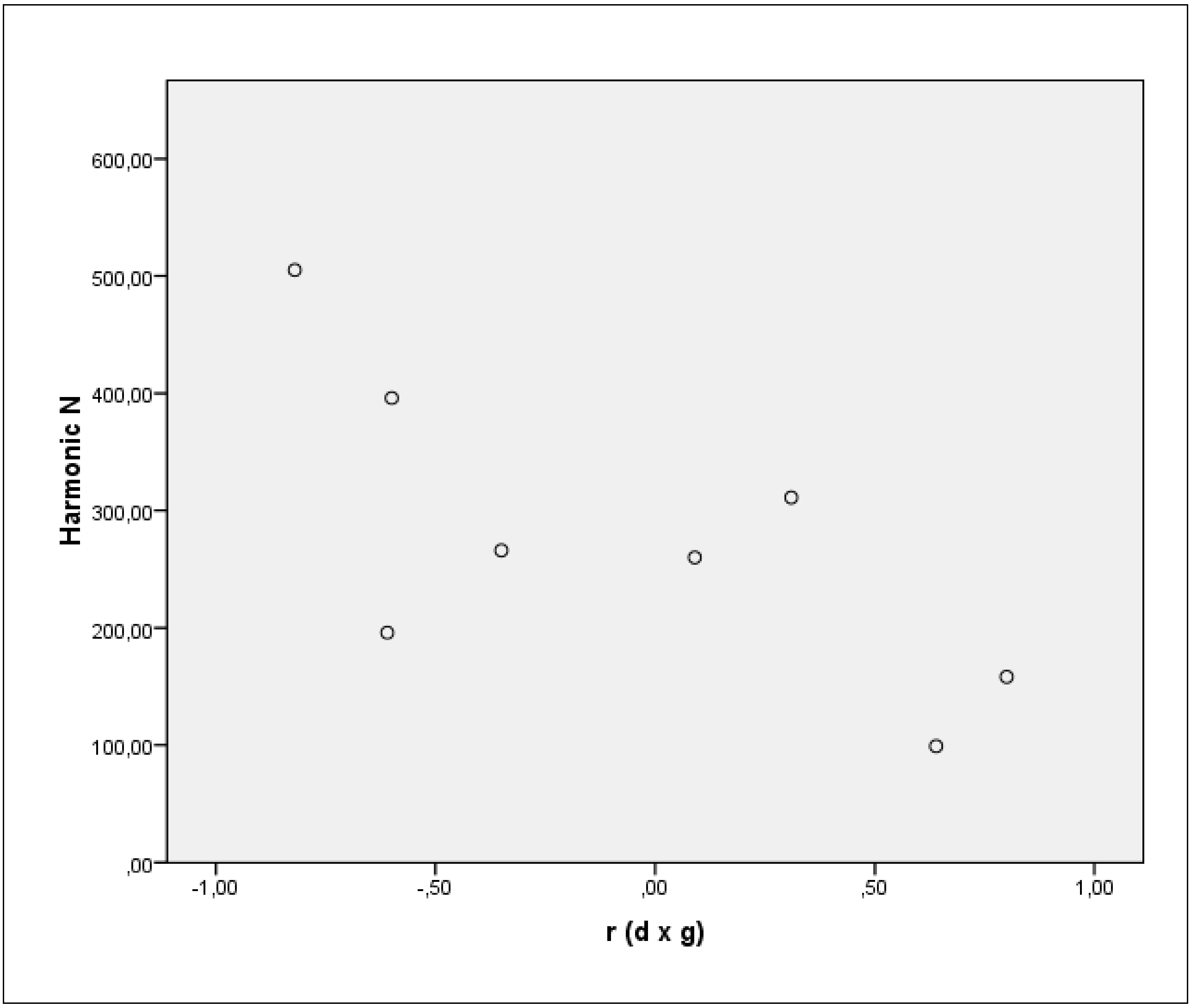


Figure 6. Scatterplot with logarithmic regression line for correlations *d* x *g* between Asians and Whites and harmonic mean.

Table 7

Meta-Analytical Output of r(dxg) for Asian-White Differences in Reaction Time

K	*N*	*r*	SD_r	rho-4	$SD_{\text{rho-4}}$	rho-5	%VE	80% CI
8	2191	-.25	.53	-.59	1.1	-.64	1.13	-2 - .81
Moderator								
Original study								
5	1094	.20	.38	.44	.86	.48	2.73	-.66 – 1.54

Note. *K* = number of correlations; *N* = total sample size; *r* = meta-analytic correlation *d* x *g* corrected for sample size; rho-4 = correlation *d* x *g* corrected for four statistical artefacts; $SD_{\text{rho-4}}$ = standard deviation of rho-4; rho-5 = correlation *d* x *g* corrected for five statistical artefacts; %VE = percentage of variance accounted for by artefactual error.[d]

The meta-analysis on the correlation *d* x *g* for Black-White differences included six data points. Figure 7 depicts a scatterplot between the obtained correlations *d* x *g* and the harmonic means. The outcomes of the meta-analysis are reported in Table 8. It shows the number of studies (*K*), the number of participants (*N*), the bare-bones meta-analytic correlation *d* x *g* (*r*), the standard deviation of this correlation (SD_r), the meta-analytic correlation *d* x *g* corrected for four statistical artifacts (rho-4), the SD of this correlation ($SD_{\text{rho-4}}$), correlation *d* x *g* corrected for all five statistical artifacts (rho-5), the percentage of variance explained by all artefectual errors (%VE), and the 80% confidence interval (80%CI). The meta-analysis yields a correlation *d* x *g* with a value of .15 with 1% of explained variance only. This percentage of variance explained is so low, we tested the origin of comparisons as a moderator. Most of the values *d* x *g* derived from reaction time measures reported in the same study. However, we did also compute vectors of reaction time measures derived from different studies. When we exclude the latter comparisons, the rho-5 becomes .37. However, there is virtually no increase in explained variance.

Conclusion

The present study was designed to test whether differences in reaction time measures between Whites and lower-IQ groups, and Whites and higher-IQ groups are related to general intelligence. We intended to broaden the datasets by combining reaction time scores derived from different studies. However, since reaction time scores presumably are vunerable to even slight changes in the measurement apparatus, we decided to base our conclusions on the results of original studies. The results of the analysis on data derived from original studies offer moderate support for

[d] Format of table adapted from te Nijenhuis & Franssen (2010). Verbatim use of formulation of table descriptions from te Nijenhuis & Franssen (2010)

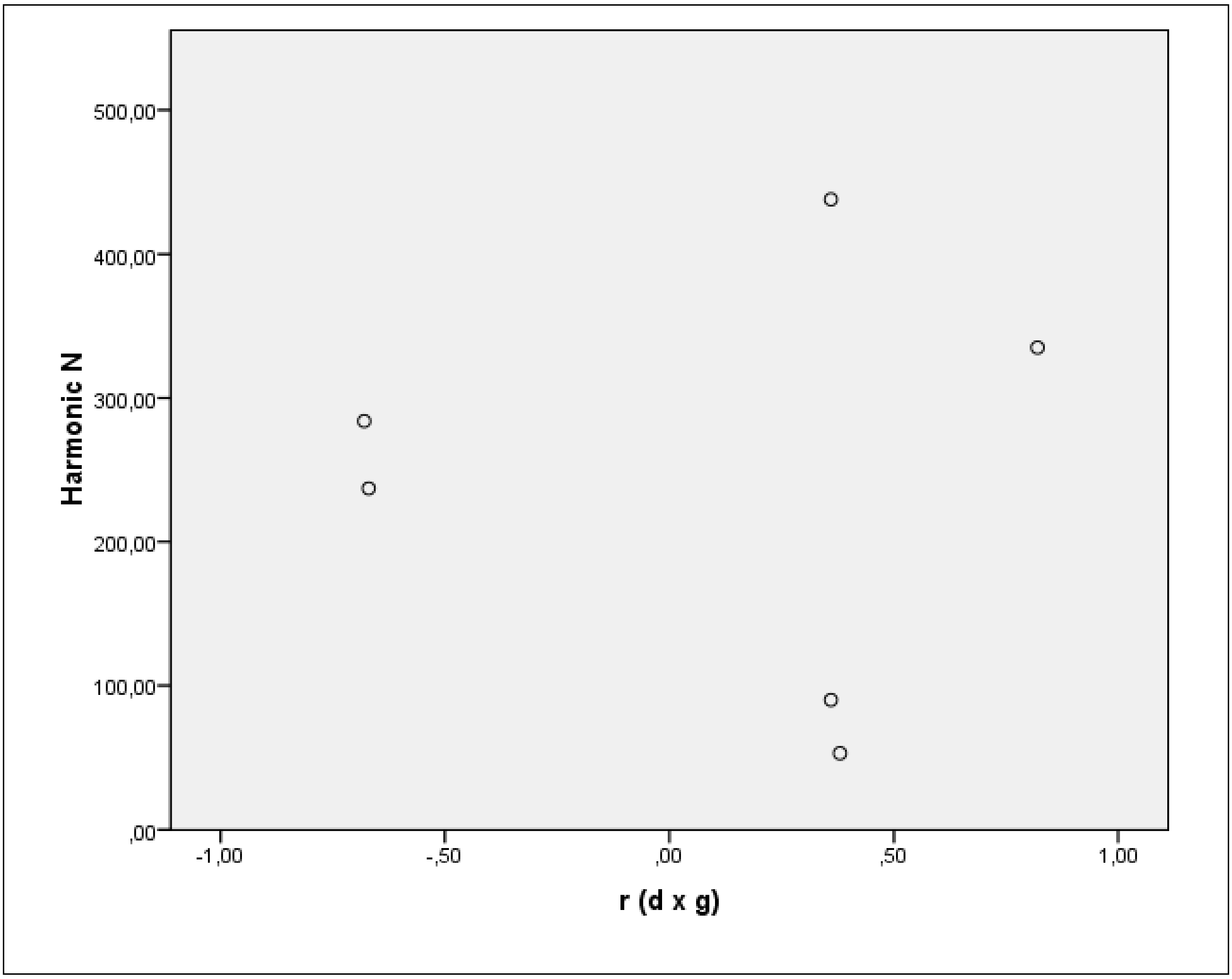


Figure 7. Scatterplot with logarithmic regression line for correlations *d* x *g* between Blacks and Whites and harmonic mean.

Table 8

Meta-Analytical Output of r(dxg) for Black-White Differences in Reaction Time

K	N_H	r	SD_r	rho-4	$SD_{\text{rho-4}}$	rho-5	%VE	80% CI
6	1436	.09	.60	.14	1.33	.15	.92	-1.56 – 1.84
Moderator								
Original study								
4	762	.18	.68	.34	1.39	.37	1.07	-1.43 – 2.12

Note. *K* = number of correlations; *N* = total sample size; *r* = meta-analytic correlation *d* x *g* corrected for sample size; rho-4 = correlation *d* x *g* corrected for four statistical artefacts; $SD_{\text{rho-4}}$ = standard deviation of rho-4; rho-5 = correlation *d* x *g* corrected for five statistical[d]

[d] Format of table adapted from te Nijenhuis & Franssen (2010). Verbatim use of formulation of table descriptions from te Nijenhuis & Franssen (2010)

artefacts; %VE = percentage of variance accounted for by artefactual error.[d]

our hypothesis. The meta-analytical correlation *d* x *g* for reaction time differences between Asians and Whites had a moderate value of value of rho-5 = .48.

Study 1b: Differences Between Germans and Immigrants

To test whether there is a strong positive correlation between the magnitude of *g* loadings and the difference on IQ subtests scores between migrant and German children in Germany, an analysis was performed on the data from one study that reported IQ scores of subtests from migrant children in Germany.

Method[b]

Searching and screening studies. A short exploratory search was conducted in the library at the University of Cologne. Professor Heiner Rindermann from the Technical University in Chemnitz in Germany has published extensively on national and international intelligence questions and was so friendly to supply us with some German articles on group differences in IQ from his private archive. This yielded one usable study.

Computation of score differences between a group consisting of German migrant children and a group consisting of German children. Score differences between a migrant group and a German group (*d*) were computed by subtracting the mean score of the migrant group from the mean score of the German group, and then dividing the result by the *SD* of the standardized group from the manual of the IQ battery. *g* loadings were obtained by matching them to the age range of the study group in question as close as possible.

Results[c]

The results of the study on the correlation between *g* loadings and the score differences between German and migrant children (*d*) are shown in Table 9. The Table gives data derived from one study, with participants numbering 218. It also lists the reference for the study, the cognitive ability test used, the correlation between *g* loadings and *d*, and the sample size. The correlation is substantially positive.

Conclusion

The goal of this study was to test whether differences in IQ profile between migrant and German children in Germany are strongly related to general intelligence, or *g*. We obtained a

---

[b] Verbatim use of formulation of method from te Nijenhuis & Franssen (2010)

[c] Verbatim use of formulation of reporting of results from te Nijenhuis & Franssen (2010)

substantially high correlation *d* x *g* of .68. Findings from this single study are in line with previous

Table 9

Studies of Correlations Between g Loadings and Differences Between German and Migrant Children

reference	test	*r*	*N*
Daseking, Lipsius, Petermann, & Waldmann (2008)	HAWIK-IV	.68	218

Note. *N* = sample size; *r* = correlation *d* x *g*. The HAWIK-IV is the German version of the WISC-IV.[d]

meta-analyses on group differences. We therefore conclude that our analysis, although limited by a modest sample size, offers further support for the Hypothesis that differences in IQ between the Western population and non-Western immigrants are related to *g*, and, in consequence, are stable over time. However, additional studies to confirm this result are required.

Study 1c: Differences between European Jews and Oriental Jews, and European Jews and non-Jewish Whites

The purpose of this study is to determine whether the correlation between the magnitude of *g* loadings and difference scores on IQ subtest scores between European Jews and Oriental Jews, and between European Jews and non-Jewish Whites is strongly positive in sign. We will test this by performing an exploratory psychometric meta-analysis on studies that reported IQ-subtests scores from Jews.

Method[b]

Searching and screening studies. Richard Lynn (2010) wrote a book called *The chosen people: A study of Jewish intelligence*, where he reviews the empirical literature on the intelligence of Jews. Professor Lynn is one of the world experts on group differences in intelligence, and most likely has made a thorough search, and is known to contact as many experts from all over the world as possible. Professor Lynn kindly gave us copies of the empirical studies he used for his book. So, most likely the collection of four empirical studies is exhaustive or near exhaustive.

Computation of score differences between a European Jewish group and a group of Oriental Jews. To compute the score differences between a European Jewish group and a group of Oriental Jews (*d*), the mean score of the Oriental Jewish group was subtracted from the mean score of the European Jewish group. The difference was divided by the *SD* of the standardization sample.

[d] Format of table adapted from te Nijenhuis & Franssen (2010). Verbatim use of formulation of table descriptions from te Nijenhuis & Franssen (2010)

Computation of score differences between a European Jewish group and a group of non-Jewish Whites. To compute the score differences between a European Jewish group and a group of non-Jewish Whites (*d*), the mean score of the US standardization sample was subtracted from the mean score of the European Jewish group. The difference was divided by the *SD* of the US standardization sample.

Results[c]

The results of the studies on the correlation between *g* loadings and the score differences between non-Jewish Whites and European Jews (*d*) are shown in Table 10. The Table shows data derived from four studies, with participants numbering a total of 302. It also lists the reference for the study, the cognitive ability test used, the correlation between *g* loadings and *d*, and the sample size. The correlations are positive in sign and substantial in magnitude. Table 11 presents the results of the bare-bones meta-analysis of the four data points. The preschool children tested by Levinson (1959) and Levinson (1960) had a mean age of about 5.5 years. The comparison group, that is the lowest age range from the WISC, however, was two years older. Therefore, we conducted an additional analysis without these two data points. Table 11 shows the number of correlation coefficients (*K*), total sample size (*N*), the true correlation (rho), and its standard deviation (Sd_r). The last column presents the percentage of variance explained by sampling errors (%VE). The analysis of all four data points yields an estimated correlation (rho) of .52, with 11.50% of the variance in the observed correlations explained by sampling errors. When excluding the aforementioned studies from the analysis, the estimated correlation (rho) rises to .80 with 82% of variance explained by sampling errors.

The results of the studies on the correlation between *g* loadings and the score differences between European Jews and Oriental Jews (*d*) are shown in Table 12. The Table reports data derived from one study, with participants numbering a total of 870. It also lists the reference for the study, the cognitive ability test used, the correlation between *g* loadings and *d*, and the sample size. The correlations are positive in sign and substantial in magnitude. Table 13 presents the results of the bare-bones meta-analysis of the four data points. It shows the number of correlation coefficients (*K*), total sample size (*N*), the true correlation (rho) and their standard deviation (SD_r). The last column presents the percentage of variance explained by sampling errors (%VE). The analysis of four data points yields an estimated correlation (rho) of .70, with 40.81% of the variance in the observed correlations explained by sampling errors.

Analysis of Verbal and Performance component of the Wechsler scale. We obtained substantially high correlations *d* x *g* and might therefore conclude that differences between

---

[b] Verbatim use of formulation of method from te Nijenhuis & Franssen (2010)
[c] Verbatim use of formulation of reporting of results from te Nijenhuis & Franssen (2010)

European Jews and non-Jewish Whites, and differences between European Jews and Oriental Jews are indeed related to differences in general intelligence, or *g*. Since European Jews were also found to have better scores on verbal tests than on performance tests we also want to include an analysis with regard to scores on verbal and performance IQ tests included in the test battery. Since all studies report scores of Wechsler tests, we decided to compare mean *d* scores on tests of the Verbal IQ scale with mean *d* scores of tests of the Performance IQ scale. Table 14 presents the results of

Table 10

Studies of Correlations Between g Loadings and non-Jewish Whites - European Jews Differences

reference	test	*r*	*N*
Levinson (1959) preschool children	WISC	.45	57
Levinson (1959) elementary schoolchildren	WISC	.75	64
Levinson (1959) university students	WAIS	.85	64
Levinson (1960)	WISC	.26	117

Note. *N* = sample size; *r* = correlation *d* x *g*[d]

Table 11

Exploratory Bare-Bones Meta-Analytical Results for Correlations Between g Loadings and non-Jewish Whites - European Jews Differences

variable	*K*	*N*	rho	SD_r	%VE
Non-Jewish Whites-/ European Jews	4	302	.52	.23	11.50
moderator					
Two groups same age	2	128	.80	.02	82.29

Note. Bare-bones meta-analytical results: Score differences between non-Jewish Whites, European Jews, and *g* loadings. *K* = number of correlations; *N* = total sample size; rho = true correlation (observed correlation corrected for sample size); SD_r = standard deviation of true correlation; %VE = percentage of variance accounted for by sampling errors. Two groups of European Jews were younger than the comparison sample. The comparability of samples with regard to age is therefore treaded as a moderator.[d]

[d] Format of table adapted from te Nijenhuis & Franssen (2010). Verbatim use of formulation of table descriptions from te Nijenhuis & Franssen (2010)

Table 12

Studies of Correlations Between g Loadings and European Jews/ Oriental Jews Differences

reference	Test	*r*	*N*
Lieblich, Ninio, & Kugelmass (1972)	WPPSI	.71	259
Lieblich, Ninio, & Kugelmass (1972)	WPPSI	.74	211
Lieblich, Ninio, & Kugelmass (1972)	WPPSI	.74	217
Lieblich, Ninio, & Kugelmass (1972)	WPPSI	.60	183

Note. *N* = sample size; *r* = correlation *d* x *g*

Table 13

Exploratory Bare-Bones Meta-Analytical Results for Correlations Between g Loadings and European Jews/ Oriental Jews Score Differences

Variable	*K*	*N*	rho	SD_r	%VE
European Jews/ Oriental Jews	4	870	.70	.04	40.81

Note. [1]Bare-bones meta-analytical results: Score differences between European Jews, Oriental Jews, and *g* loadings. *K* = number of correlations; *N* = total sample size; rho = true correlation (observed correlation corrected for sample size); SD_r = standard deviation of true correlation; %VE = percentage of variance accounted for by sampling errors.[d]

the bare-bones meta-analysis of two data points for differences between European Jews and non-Jewish Whites on the Verbal scale and the Performance scale. It shows the number of *d* scores (*K*), total sample size (*N*), the true effect size (d_t) and their standard deviation (SD_d). The last column presents the percentage of variance explained by sampling errors (%VE). The analysis of two data points yields an estimated effect size (d_t) on the Verbal scale of 1.42 with 100% of the variance in the observed effect sizes explained by sampling errors. On the Performance scale, we found an estimated effect size of .26 with 100% of the variance in the observed effect size explained by sampling errors. These results are visualized in Diagram 1.

Table 15 presents the results of the bare-bones meta-analysis of two data points for differences between European Jews and Oriental Jews on the Verbal scale and the Performance scale. It shows the number of *d* scores (*K*), total sample size (*N*), the true effect size (d_t) and its standard deviation (SD_d). The last column presents the percentage of variance explained by sampling errors (%VE). The analysis of two data points yields an estimated effect size (d_t) on the

[d] Format of table adapted from te Nijenhuis & Franssen (2010). Verbatim use of formulation of table descriptions from te Nijenhuis & Franssen (2010)

Verbal scale of .71 with 100% of the variance in the observed effect sizes explained by sampling errors. On the Performance scale, we found an estimated effect size of .55 with 100 % of the variance in the observed effect size explained by sampling errors. The results are visualized in Diagram 2.

Table 14

Exploratory Bare-Bones Meta-Analytical Results for d Scores on Subtests of Wechsler's Verbal and Performance Scale for European Jews/ non-Jewish Whites

variable	*K*	*N*	d_t	SD_d	%VE
Verbal scale differences	2	128	1.42	0	590
Performance scale differences	2	128	.26	0	13,448

Note. Bare-bones meta-analytical results: Score differences between European Jews and non-Jewish Whites on subtests of Wechsler's verbal and performance scale. *K* = number of *d* scores; *N* = total sample size; d_t = true effect size corrected for sample size; SD_d = standard deviation of true effect size; %VE = percentage of variance accounted for by sampling errors.[d]

Table 15

Exploratory Bare-Bones Meta-Analytical Results for d Scores on Subtests of Wechsler's Verbal and Performance Scale for European Jews/ Oriental Jews

variable	*K*	*N*	*d*	SD_d	%VE
Verbal scale differences	4	870	.71	0	748
Performance scale differences	4	870	.55	0	1,271

Note. Bare-bones meta-analytical results: Score differences between European Jews and Oriental Jews on subtests of Wechsler's verbal and performance scale. *K* = number of *d* scores; *N* = total sample size; *d* = true effect size corrected for sample size; SD_d = standard deviation of true effect size; %VE = percentage of variance accounted for by sampling errors.[d]

Conclusion

The goal of this study was to explore whether differences in IQ profile between European Jews and non-Jewish Whites, and European Jews and Oriental Jews have a strong correlation with general intelligence, or *g*. We obtained a meta-analytic correlation of .80 for differences between European Jews and non-Jewish Whites, and a meta-analytic correlation of .70 for differences between European Jews and Oriental Jews. As these findings are based on only a limited amount of

[d] Format of table adapted from te Nijenhuis & Franssen (2010). Verbatim use of formulation of table descriptions from te Nijenhuis & Franssen (2010)

studies with a rather small total *N* it is not possible to draw strong conclusions. Still, this study offers support for the Hypothesis that IQ group differences between European Jews and non-Jewish Whites and European Jews and Oriental Jews are related to general intelligence, and, in consequence, should be stable over time. A further analysis of effect sizes on differences between Verbal scale subtests and Performance scale subtests, revealed a meta-analytic effect size of 1.42 for differences between European Jews and non-Jewish Whites on the Verbal scale and a meta-analytic effect size of .27 for differences between European Jews and non-Jewish Whites on subtests of the Performance scale. Since the effect size for the Verbal scale is more than five times larger than the

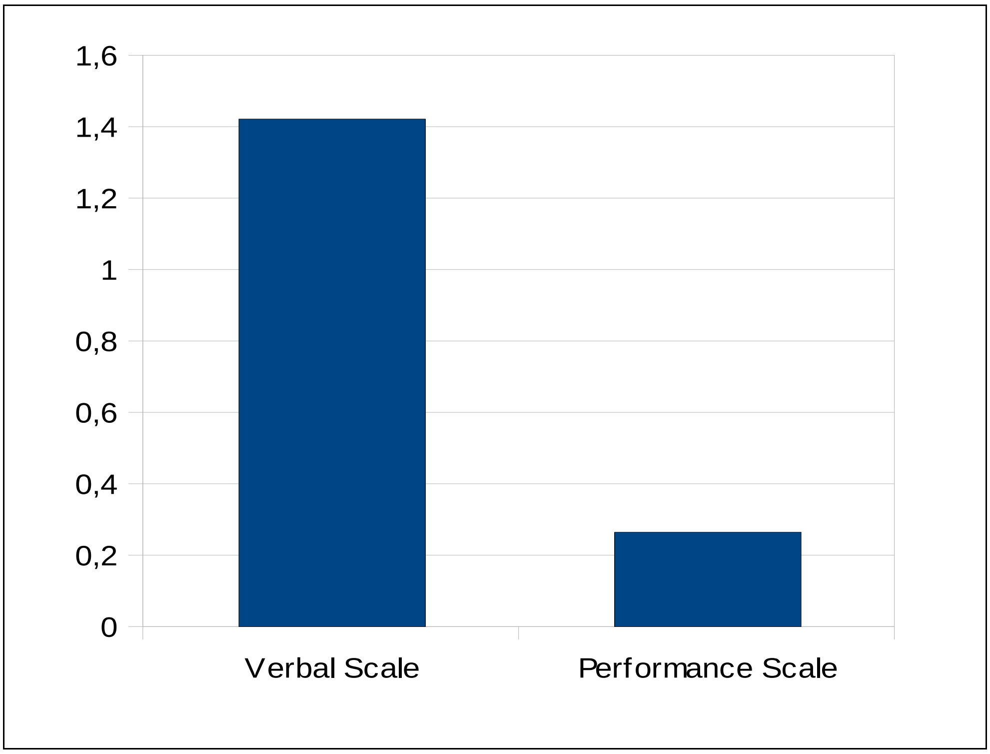


Diagram 1. Effect sizes for differences between European Jews and non-Jewish Whites on the verbal and the performance scale of Wechsler tests.

effect size of the Performance scale, we can conclude that differences in IQ profile between European Jews and non-Jewish Whites are substantially stronger for Verbal tests than for Performance tests. Differences in effect size between the Verbal and the Performance scale between European Jews and Oriental Jews were negligible.

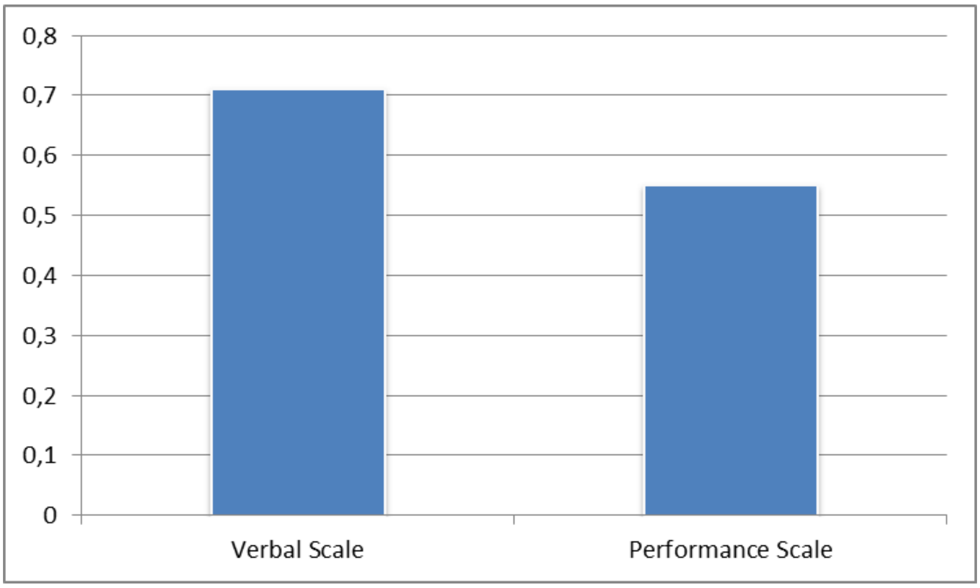


Diagram 2. Effect sizes for differences between European Jews and Oriental Jews on the verbal and the performance scale of Wechsler tests

Study 1d: Differences Between Jews and Arabs

The purpose of this study is to determine whether the correlation between the magnitude of *g* loadings and difference scores on IQ subtest scores between Jewish and Arab groups in Israel is strongly positive in sign. We will test this by performing an exploratory psychometric meta-analysis on studies that reported IQ scores of at least seven subtests from Jews and Arabs residing in Israel.

Method[b]

Searching and screening studies. Professor Richard Lynn kindly gave us copies of the empirical studies comparing the IQ scores of Jews and Arabs in Israel he used for his book *The chosen people: A study of Jewish intelligence* (2010). Most likely the collection of three empirical studies is exhaustive or near exhaustive.

Computation of score differences between a Jewish and an Arab group. Score differences between a Jewish and an Arab group (*d*) were computed by subtracting the mean score of the Arab group from the mean score of the Jewish group, and then dividing the result by the pooled *SD* of both groups. *g* loadings were obtained by matching them to the age range of the Jewish and Arab groups as close as possible.

Results[c]

The results of the studies on the correlation between *g* loadings and the score differences between Jews and Arabs (*d*) are shown in Table 16. The Table gives data derived from three studies,

[b] Verbatim use of formulation of method from te Nijenhuis & Franssen (2010)
[c] Verbatim use of formulation of reporting of results from te Nijenhuis & Franssen (2010)

with participants numbering a total of 1443. It also lists the reference for the study, the cognitive ability test used, the correlation between *g* loadings and *d*, and the sample size. The correlations are small and negative. Table 17 presents the results of the bare-bones meta-analysis of the five data points. It shows the number of correlation coefficients (*K*), total sample size (*N*), the true correlation (rho) and their standard deviation (SD_r). The last column presents the percentage of variance explained by sampling errors (%VE). The analysis of five data points yields an estimated correlation (rho) of -.25, with 216.96% of the variance in the observed correlations explained by sampling errors.

The analysis of data points yielded estimated effect sizes, with a percentage of variance explained by sampling errors larger than 100. 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).[18] (taken from te Nijenhuis & Franssen, 2010)

Table 16

Studies of Correlations Between g Loadings and Differences Between Jews and Arabs

reference	*test*	*r*	*N*
Zeidner (1987) Jews - Arabs	Various tests of cognitive ability	-.25	1294
Kugelmass, Lieblich, and Bossik (1974) Kindergarten	WPPSI	-.36	48
Kugelmass, Lieblich, and Bossik (1974) First Grade	WPPSI	-.12	34
Lieblich, Kugelmass, & Ehrlich (1975) Kindergarten	WPPSI	-.04	25
Lieblich, Kugelmass, & Ehrlich (1975) First Grade	WPPSI	-.35	20
Lieblich, Kugelmass, & Ehrlich (1975) Third Grade	WPPSI	-.30	22

Note. *N* = sample size; *r* = correlation *d* x *g*

[18] Paragraph taken and adapted from te Nijenhuis & Franssen (2010, p. 35)

Table 17

Exploratory Bare-Bones Meta-Analytical Results for Correlations Between g Loadings and Jewish/ Arab Score Differences

predictor	*K*	*N*	rho	SD_{rho}	%VE
Jews/ Arabs	5	1443	-.25	0	216.96

Note. Bare-bones meta-analytical results: Score differences between Jews, Arabs, 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 sampling errors.[d]

Analysis of verbal and performance component of the Wechsler scale. The bare-bones meta-analysis of the correlation *d* x *g* between Jews and Arabs yielded a correlation that points to a small negative relationship of IQ differences between Jews and Arabs, and general intelligence. We also conducted an analysis of difference scores on the Verbal and the Performance scale. We averaged *d* scores of subtests for the Verbal and the Performance scale in every study that uses a Wechsler test and conducted a bare-bones meta-analysis of effect sizes with the average *d* score of each study as data points. Table 18 presents the results of the bare-bones meta-analysis of 11 data points for the Verbal and the Performance scale. It shows the number of *d* scores (*K*), total sample size (*N*), the true effect size (d_t) and their standard deviation (SD_d). The last column presents the percentage of variance explained by sampling errors (%VE). The analysis of five data points yields an estimated effect size (*d*) of .42 for differences on the Verbal scale, with 195.45% of the variance in the observed effect sizes explained by sampling errors. For differences on the Performance scale we received an estimated effect size of .61, with 96.49 % of the variance in the observed effect sizes explained by sampling errors.

Conclusion

In the study on differences between the IQ profile of Jews and Arabs, we expected a strong correlation *d* x *g*. The results of this analysis clearly show no support for this hypothesis. The correlation *d* x *g* is small and negative in sign, which points to a rather minor role of general intelligence as an explaining factor for differences in IQ between Jews and Arabs. A further analysis on subtest scores on the Verbal and the Performance scale of Wechsler tests yielded comparable effect sizes for both scales. In conclusion, this study does not support the Hypothesis that group differences in IQ between Jews and Arabs residing in Israel can be explained by differences in general intelligence.

[d] Format of table adapted from te Nijenhuis & Franssen (2010). Verbatim use of formulation of table descriptions from te Nijenhuis & Franssen (2010)

Table 18

Exploratory Bare-Bones Meta-Analytical Results for d Scores on the Verbal and the Performance Scale for Jews/ Arabs Differences

variable	*K*	*N*	d_t	SD_d	%VE
Jewish-Arab differences on the Verbal scale	5	149	.42	0	195.45
Jewish-Arab differences on the Performance scale	5	149	.61	0.07	96.49

Note. Bare-bones meta-analytical results: Score differences between Arabs and Jews on subtests of the Verbal scale and the performance scale. *K* = number of *d* scores; *N* = total sample size; d_t = true effect size corrected for sample size; SD_d = standard deviation of true effect size; %VE = percentage of variance accounted for by sampling errors.[d]

Study 2: Subgroup Differences

Study 2a: Explorative Comparison of School Types

To test whether there is a strong positive correlation or a negligible correlation between the magnitude of *g* loadings and the differences on IQ subtest scores of children of different school types, an exploratory psychometric meta-analysis was performed on a number of studies that reported IQ scores of at least seven subtests from children of different school types.

Method[b]

Searching and screening studies. No structural search was conducted to identify articles on this topic. The subtest scores for children of different school types were found in manuals of four IQ batteries and one article was obtained when searching for articles on other topics that concern the influence of environmental variables on *g*.

Specific criteria for inclusion. In the present study, we made a comparison between school types and there had to be a substantial difference between the mean IQs of the different schools in the primary studies. When the mean IQs are similar, the children might have been as well in the same class room in the same school type. As the difference between Dutch school types is 10 IQ points, we took that value as the minimum difference between groups in primary studies. We state explicitly that this is a practical choice, and not a choice based on substantive theory. Still, at least

[d] Format of table adapted from te Nijenhuis & Franssen (2010). Verbatim use of formulation of table descriptions from te Nijenhuis & Franssen (2010)

[b] Verbatim use of formulation of method from te Nijenhuis & Franssen (2010)

in the Netherlands, a difference of some ten IQ points results in a homogeneous group, so the overwhelming majority of the children in a school class can ‘keep up with the teaching speed’. In the study by Fokkema and Dirkzwager (1968) girls in the lower school type outscored girls in the higher school type. As children in a higher school type should have higher mean scores than children in a lower school type we decided to exclude this data point from our meta-analysis. It should be taken into consideration that the present study is about explaining the pattern of the higher mean scores on the majority of subtests of children of a higher school type in comparison with children of a lower school type. However, if the children of the higher school type actually have lower scores the computation of difference scores and their explanation makes no sense.

In many of the studies we collected IQ scores are not reported, so a straightforward comparison of the total score between groups is not possible. However, we choose to estimate differences between groups on a test battery by comparing the average *d* values of the subtests with the three highest *g* loadings. The group difference on an IQ score will be highly similar to the group difference on a *g* score. The *g*-loadedness of an IQ score will be close to a value of 1.00. So, to estimate the group difference on a total score we took the values of *d* of the subtest with the three highest *g* loadings and multiplied it with a value of 1 / *g*. We presume that multiplying the *g* loading with a value of 1 / *g* results in a *g* loading of 1.00. So, $d * 1 / g$ is the value of *d* for a test with a perfect *g* loading, and highly similar to the difference on an IQ score. We average the transformed *d* for the subtests with the three highest *g* loadings to receive an even more reliable estimate of the difference in IQ between groups. A value of $d \geq .67$ means that difference between schools is at least 10 IQ points.

Computation of score differences between children of different school types. Score differences between children of different school types (*d*) were computed by subtracting the mean score of the lower school type from the mean score of the higher school type, and then dividing the result by the pooled *SD* of all school types reported in the manual.

Computation of score differences between children of different school types and a standardized group. Score differences between children of different school types and a standardized group were computed by subtracting the score of the standardized group from a school type that has a higher full scale IQ than the standardized group, and dividing the result by the *SD* of the standardized group. If the school type has a lower Full scale IQ than the standardized group, the score of the school type was subtracted from the score of the standardized group, and the result was divided by the *SD* of the standardized group. In several manuals, no score of a standardized group with regard to age was reported. In this case the scores of all available school types with the same age range were used to estimate the score of an average population of this age. Since the expected average IQ score for this population is approximately 100, we included school types in this computation of which the Full scale IQ scores lie one *SD* above the average score and one *SD* below the average IQ score. When no school types were reported of which the IQ score was 2 *SDs* below the average score, but a school type of which the Full scale IQ score was 2 *SDs* above the average Full scale score (or vice versa) the latter was not included in the computation of the standardized score. If the scores of the school types were reported in an article, the score of the standardized group was obtained in the manual of the IQ battery. Standardized scores were matched with regard to age as close as possible.

Results[c]

The results of the studies on the correlation between *g* loadings and the score differences between school types (*d*) are shown in Table 19. The Table reports data derived from three IQ battery manuals and one study, with participants numbering a total of 9849. It also lists the reference for the study, the cognitive ability test used, the correlation between *g* loadings and *d*, and the sample size. The correlations range from substantially negative to extremely positive. Table 20 presents the results of the bare-bones meta-analysis of 32 data points. It shows the number of correlation coefficients (*K*), total sample size (*N*), the true correlation (rho) and their standard deviation (SD_r). The last column presents the percentage of variance explained by sampling errors (%VE). The analysis of all data points yields an estimated correlation (rho) of .20, with 3.01% of the variance in the observed correlations explained by sampling errors. However, it is clear that the comparison between Mavo 2 and Lbo 2 in the study by Evers and Lucassen (1983) is an extreme outlier. Taking the reduced sample of 31 studies, the value of r = -.42 is more than three *SD* below the average sample-sized weighted correlation of .29. Taking out this one extreme outlier increased the percentage of variance to a value of 5.85.

Conclusion

An earlier meta-analysis by te Nijenhuis et al. (2007) showed a correlation *d* x *g* of +1 for differences in IQ profile between gifted individuals and standardized groups and a correlation *d* x *g* of .74 for differences in IQ profile between mentally retarded individuals and standardized groups. We expected that differences between school types would result in smaller IQ differences, which presumably are not substantially related to general intelligence. The results of our analysis show a positive meta-analytic correlation *d* x *g* of rho = .29, but only 5.85% variance explained by sampling errors. Based on these results we can conclude that the relationship between differences in IQ profile of different school types and general intelligence is modest, but not as strong as in the meta-analyses on gifted and mentally retarded; differences might be more pronounced on lower

[c] Verbatim use of formulation of reporting of results from te Nijenhuis & Franssen (2010)

Table 19

Studies of Correlations Between g Loadings and School Differences

references	test	*r*	*N*
Evers & Lucassen (1983) Mavo3/ Havo3	DAT'83	.38	1100
Evers & Lucassen (1983) Vwo3/ Mavo3	DAT'83	.25	1100
Evers & Lucassen (1983) Vwo3/ Havo3	DAT'83	.11	1100
Evers & Lucassen (1983) Mavo4/ Havo4	DAT'83	0.4	600
Evers & Lucassen (1983) Vwo4/ Mavo4	DAT'83	0.07	464
Evers & Lucassen (1983) Vwo4/ Havo4	DAT'83	-.15	464
Evers & Lucassen (1983) Mavo2/ Lbo2	DAT'83	-.42	1232
Fokkema & Dirkzwager (1968) Havo3(m)/ Mavo3(m)	DAT(1968)	.29	81
Fokkema & Dirkzwager (1968) Gym3(m)/ Mavo3(m)	DAT(1968)	.11	160
Fokkema & Dirkzwager (1968) Havo3(f)/ Mavo3(f)	DAT(1968)	.25	80
Fokkema & Dirkzwager (1968) Gym3(f)/ Mavo3(f)	DAT(1968)	.53	112
Fokkema & Dirkzwager (1968) Gym4(m)/ Mavo4(m)	DAT(1968)	-.02	77
Fokkema & Dirkzwager (1968) Gym4(f)/ Mavo4(f)	DAT(1968)	.11	46
Van Dijk & Tellegen (1994) Vbo2+3/ Ivbo2+3	GIVO(1994)	.47	220
Van Dijk & Tellegen (1994) Mavo2+3/ Vbo2+3	GIVO(1994)	.34	385
Van Dijk & Tellegen (1994) Mavo2+3/ Ivbo2+3	GIVO(1994)	.46	183
Van Dijk & Tellegen (1994) Havo2+3/ Mavo2+3	GIVO(1994)	-.13	241
Van Dijk & Tellegen (1994) Havo2+3/ Vbo2+3	GIVO(1994)	.34	310
Van Dijk & Tellegen (1994) Havo2+3/ Ivbo2+3	GIVO(1994)	.49	164
Van Dijk & Tellegen (1994) Vwo2+3/ Mavo2+3	GIVO(1994)	.1	310
Van Dijk & Tellegen (1994) Vwo2+3/ Vbo2+3	GIVO(1994)	.34	435
Van Dijk & Tellegen (1994) Vwo2+3/ Ivbo2+3	GIVO(1994)	.46	194
Van Dijk & Tellegen (1994) Vbo1/ Ivbo1	GIVO(1994)	.85	100
Van Dijk & Tellegen (1994) Mavo1/ Vbo1	GIVO(1994)	.78	74
Van Dijk & Tellegen (1994) Mavo1/ Ivbo1	GIVO(1994)	.86	51
Van Dijk & Tellegen (1994) (Havo1, Vwo1)/ Mavo1	GIVO(1994)	.55	68
Van Dijk & Tellegen (1994) (Havo1, Vwo1)/ Vbo1	GIVO(1994)	.81	196
Van Dijk & Tellegen (1994) (Havo1, Vwo1)/ Ivbo1	GIVO(1994)	.84	89
Daseking, Lipsius, Petermann, & Waldmann (2008) Hauptschule MG- Realschule MG	HAWIK-4	.64	51
Daseking, Lipsius, Petermann, & Waldmann (2008) Hauptschule MG- Gymnasium MG	HAWIK-4	.74	57
Daseking, Lipsius, Petermann, & Waldmann (2008) Hauptschule KG- Gymnasium KG	HAWIK-4	.70	57
Daseking, Lipsius, Petermann, & Waldmann (2008) Realschule KG- Gymnasium KG	HAWIK-4	.59	48

Note. *N* = sample size; *r* = correlation *d* x *g*[d]

[d] Format of table adapted from te Nijenhuis & Franssen (2010). Verbatim use of formulation of table descriptions from te Nijenhuis & Franssen (2010)

levels of the intelligence hierarchy.

Table 20

Exploratory Bare-Bones Meta-Analytical Results for Correlations Between g Loadings School Difference Scores

variable	*K*	*N*	rho	SD_{rho}	%VE
School Differences	32	9849	.20	.31	3.01
School Differences (without outliers)	31	8617	.29	.22	5.85

Note. Bare-bones meta-analytical results: Score differences between school types, 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 sampling errors.[d]

Study 2b: Explorative Comparison of Religious Groups

The analysis was performed on the data from a study that reported IQ scores of seven subtests from different religious groups.

Method[b]

Searching and screening studies.[19] Three methods were used to identify studies that contained correlations between religion and IQ subtest scores. First, an electronic search for published research using PsycINFO, ERIC, MEDLINE, PiCarta, Academic search premier, Web of science, Google Scholar, and PubMed was conducted. Keywords used were religion, religious groups, catholics, protestants, cognitive, mental*, intelligence, IQ, WISC, Wechsler, and combinations of these concepts (* is a truncation symbol to represent multiple spellings or endings; AND is a Boolean operator that combines search terms so that the search result contains all of the terms). Second, the reference lists of significant articles were analyzed in search of additional studies. Finally, cited reference searches were conducted using Web of Science, to search for articles citing significant articles. This procedure resulted in one article. (te Nijenhuis & Franssen, 2010)

[d] Format of table adapted from te Nijenhuis & Franssen (2010). Verbatim use of formulation of table descriptions from te Nijenhuis & Franssen (2010)

[b] Verbatim use of formulation of method from te Nijenhuis & Franssen (2010)
[19] Paragraph taken and adapted from te Nijenhuis & Franssen (2010, p. 22)

Computation of score differences between a groups of different religious background. Score differences between different religious groups (*d*) were computed by subtracting the mean score of the religious group with a lower Full scale IQ score from the mean score of the religious group with a higher Full scale IQ score, and then dividing the result by the pooled *SD* of all religious groups included in the study. *g* loadings were obtained by matching them as close as possible to the age range of the religious groups in question.

Results[c]

The results of the studies on the correlation between *g* loadings and the score differences between different religious groups in the Netherlands (*d*) are shown in Table 21. The Table reports data derived from one study, with participants numbering a total of 1913. It also lists the reference for the study, the cognitive ability test used, the correlation between *g* loadings and *d*, and the sample size. The correlations show no clear pattern with regard to magnitude and sign. Table 22 presents the results of the bare-bones meta-analysis of the six data points. It shows the number of correlation coefficients (*K*), total sample size (*N*), the true correlation (rho) and their standard deviation (SD_r). The last column presents the percentage of variance explained by sampling errors (%VE). The analysis of both data points yields an estimated correlation (rho) of -.21, with 1.31% of the variance in the observed correlations explained by sampling errors.

Table 21

Studies of Correlations Between g Loadings and Different Religious Groups

reference	test	*r*	*N*
Verhage (1964) Roman Catholic - Dutch Reformed Church	WISC-R	-.57	544
Verhage (1964) Roman Catholic - Reformed Church Service	WISC-R	.28	242
Verhage (1964) Roman Catholic - Nonmembers of Church	WISC-R	-.49	378
Verhage (1964) Dutch Reformed Church - Reformed Church Service	WISC-R	.80	224
Verhage (1964) Dutch Reformed Church - Nonmembers of Church	WISC-R	-.09	335
Verhage (1964) Reformed Church Service - Nonmembers of Church	WISC-R	-.61	190

Note. *N* = sample size; *r* = correlation *d* x *g*

[c] Verbatim use of formulation of reporting of results from te Nijenhuis & Franssen (2010)

Table 22

Exploratory Bare-Bones Meta-Analytical Results for Correlations Between g Loadings and Religious Groups Score Differences

variable	*K*	*N*	rho	SD_{rho}	%VE
Religious groups	6	1913	-.21	.47	1.31

Note. Bare-bones meta-analytical results: Score differences between different religious groups, 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 sampling errors.[d]

Conclusion

The study on the relationship between IQ profiles of different religious groups and general intelligence was of exploratory nature only. Based on a sample size of 1913 we obtained a meta-analytic correlation *d* x *g* of rho = -.21 with 1.31% variance explained by sampling errors. We therefore conclude that differences between religious groups in the Netherlands are not related to general intelligence.

Study 3: Heritability

h^2 Reaction Time Measures

To test whether the herditability coefficients of reaction time measures are related to general intelligence, or *g*, we conducted a bare-bones meta-analysis on two studies. To compute the correlation h^2 x g we had to obtain *g* loadings of RT measures. Since the correlation matrix of RT scores was not an option, we used the correlation of RT measures with IQ batteries like the ASVAB and the WISC-R as an estimate for the *g* loadedness of the RT measures.

Method[b]

Searching and screening studies.[20] Starting point of the search is the meta-analysis by Beaujean (2005). The reference list of this article was checked for studies that contained IQ scores combined with reaction times. Next, the following databases were searched: PsycINFO, ERIC, MEDLINE, PiCarta, Academic search premier, Web of science, Google Scholar, and PubMed, using

[d] Format of table adapted from te Nijenhuis & Franssen (2010). Verbatim use of formulation of table descriptions from te Nijenhuis & Franssen (2010)

[b] Verbatim use of formulation of method descriptions from te Nijenhuis & Franssen (2010)

the following keywords: heritability, heritability coefficient, twin*, twin stud*, cognitive task*, processing speed, reaction time, speed-of-information processing (SIP), information processing, perceptual speed, inspection time, IQ, intellectual development, intelligence, *g*, general mental ability, mental ability, cognitive development, cognitive ability, and combinations of these keywords (* is a truncation symbol to represent multiple spellings or endings; AND is a Boolean operator that combines search terms so that the search result contains all of the terms). Third, we browsed the tables of contents of several major research journals of genetics, development, and of intelligence, such as *Behavior Genetics* 1970-2007, *Acta Geneticae Medicae et Gemellologiae* 1953-1992, *Twin Research: The Official Journal of the International Society for Twin Studies* 1998-2007, *Intelligence* 1977-2007, *Psychological Science* 1990-2007, *Child Development* 1930-2007, *and Developmental Psychology* 1969-2007. Fourth, cited reference searches were conducted using Web of Science to identify the newest articles, citing key studies already in our possession. Finally, we checked the reference list of all currently included empirical studies to identify any potential articles that may have been missed by earlier search methods. This procedure yielded two data points from two studies. (te Nijenhuis & Franssen, 2010)

Specific criteria for inclusion. For a study to be included in the meta-analysis three criteria had to be met: First, only empirical studies reporting heritability coefficients (h^2) of several RT submeasures were included. Second, *g* loadings had to be available for the RT measures, otherwise the correlation h^2 x *g* could not be computed. Third, there were many studies that reported heritability coefficients of inspection time. Inspection time, however, consists of only one measure, so the method of correlated vectors could not be applied. In the study of Petrill, Thompson, and Detterman (1995) we left out all measures for which no h^2 was reported.

Results[c]

The results of the studies on the correlation between *g* loadings and h^2 of RT measures are shown in Table 23. The Table gives data derived from two studies, with participants numbering a total of 389. It also lists the reference for the study, the reaction time measures used, the correlation between h^2 and *g* loadings of reaction time measures, and the sample size. Both correlations are substantially positive. Table 24 presents the results of the bare-bones meta-analysis of the two data points. It shows the number of correlation coefficients (*K*), total sample size (*N*), the true correlation (rho) and their standard deviation (SD_r). The last column presents the percentage of variance explained by sampling error (%VE). The analysis of data points yields an estimated correlation (rho) of .51, with 65.22% of the variance in the observed correlations explained by sampling error.

[20] Paragraph taken and adapted from te Nijenhuis & Franssen (2010, p. 22)

Conclusion

Since scores of measures of cognitive processing were found to be substantially heritable and show a positive correlation with *g*, we explored whether the heritability pattern in reaction time measures is similar to the heritability pattern found in measures of IQ test batteries. Te Nijenhuis and Jongeneel-Grimen (2007) carried out a meta-analysis based on a sample size of 2,590 and found a correlation of +1 between the vector of *g* loadings and the vector of heritability coefficients

Table 23

Studies of Correlations Between g Loadings and h^2 *of RT Measures*

reference	test	*r*	*N*
Vernon (1989)	Various RT measures	.62	102
Petrill, Thompson, & Detterman (1995)	RT measures of the CAT	.47	287

Note. *N* = sample size; *r* = correlation *d* x *g*. Vernon (1989) used mean RT and intraindividual variability.

Table 24

Exploratory Bare-Bones Meta-Analytical Results for Correlations Between g Loadings and h^2 *of RT Measures*

variable	*K*	*N*	rho	SD_{rho}	%VE
h^2 reaction times	2	389	.51	.04	65.22

Note. Bare-bones meta-analytical results: h^2 reaction times 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 sampling error.[d]

of IQ batteries. Based on the assumption A) that reaction time measures have a small to substantial correlation with general intelligence and B) that reaction time measures are substantially heritable, in the present study we expected a substantial positive correlation between the vector of *g* loadings and the vector of heritability coefficients of reaction time measures. Based on a bare-bones meta-analysis of two data points we found a rho of .51 with 65.22% of the observed variance explained by sampling error. Although the correlation we found is not as positive as the correlation between heritability coefficients and *g* loadings of IQ batteries in te Nijenhuis and Jongeneel-Grimen (2007), we can conclude that the present, relatively small, dataset confirms our hypothesis that general intelligence, as it is reflected in measures of reaction time, is substantially heritable.

[c] Verbatim use of formulation of reporting of results from te Nijenhuis & Franssen (2010)

[d] Format of table adapted from te Nijenhuis & Franssen (2010). Verbatim use of formulation of table descriptions from te Nijenhuis & Franssen (2010)

Study 4: Physical characteristics of the brain

Brain Volume

The purpose of this study is to determine whether the correlation between the magnitude of *g* loadings and the correlation between brain volume and IQ subtest scores is strongly positive in sign. We will test this by performing an exploratory psychometric meta-analysis on all studies that report the correlation of brain volume with IQ subtest scores.

Method[b]

Searching and screening studies.[21] Three methods were used to identify studies that contained correlations between brain volume and IQ subtest scores. First, an electronic search for published research using PsycINFO, ERIC, MEDLINE, PiCarta, Academic search premier, Web of science, Google Scholar, and PubMed was conducted. Keywords used were brain volume, brain weight, brain size, cognitive, mental*, intelligence, IQ, WISC, Wechsler, and combinations of these concepts (* is a truncation symbol to represent multiple spellings or endings; AND is a Boolean operator that combines search terms so that the search result contains all of the terms). Second, the reference lists of significant articles were analyzed in search of additional studies. Finally, cited reference searches were conducted using Web of Science, to search for articles citing significant articles. Additionally, Pietschnig (2011) presented a paper on his meta-analysis in progress, sampling more broadly than McDaniel (2005). Mr. Pietschnig MA was so friendly to supply us with the studies he used in his meta-analysis. This procedure resulted in nine articles and reports on the concurrent topics of the relationship between brain volume and mental ability. Four studies met all criteria for inclusion in the meta-analysis. (te Nijenhuis & Franssen, 2010)

Computation of the correlation r x g between subtest g loadings and the correlation of these subtests with brain volume. The studies included in this meta-analysis report correlations between scores on subtests and the brain volume of participants. These correlations were correlated with the *g* loadings of the respective subtests. This comparison follows the logic that if brain volume correlates with general intelligence, we should observe a strong positive correlation between the correlation of brain volume and subtest scores and *g* loadings of these subtests.

---

[b] Verbatim use of formulation of method from te Nijenhuis & Franssen (2010)

[21] Paragraph taken and adapted from te Nijenhuis & Franssen (2010, p. 22)

Results[c]

The results of the studies on the correlation between r (*brain volume* x *subtest scores*) (r_1) and g loadings are shown in Table 25. The Table gives data derived from two studies, with participants numbering a total of 246. It also lists the reference for the study, the cognitive ability test used, the correlation r_1, and the sample size. Table 26 presents the results of the bare-bones meta-analysis of the four data points. It shows the number of correlation coefficients (K), total sample size (N), the true correlation (rho) and their standard deviation (SD_r). The last column presents the percentage of variance explained by sampling errors (%VE). The analysis of both data points yields an estimated correlation (rho) of .07, with 10.15% of the variance in the observed correlations explained by sampling errors. However, it is clear that the study by Flashman, Andreasen, Flaum, and Swayze (1998) is an extreme outlier: taking the reduced sample of three studies, the value of $r_1 = -.41$ is more than four *SD* below the average sample-sized weighted correlation of .35. Taking out this one extreme outlier increased the percentage of variance to a value of 34.21.

Table 25

Studies of Correlations Between g Loadings and r_1

Reference	test	r_1	*N*
Wickett, Vernon, & Lee (2000)	MAB	.11	68
Egan, Chiswick, Santosh, Naidu, Rimmington, & Best (1994)	WAIS-R	.51	48
Flashman, Andreasen, Flaum, & Swayze (1998)	WAIS-R	-.41	90
Wickett, Vernon, & Lee (1994)	MAB	.56	40

Note. N = sample size; r = correlation d x g.

Conclusion

Similar to other physical characteristics of the brain, we expected that brain volume shows a strongly positive correlation with g. After excluding one outlier from the analysis we derived at a true correlation between brain volume and g of .35. Compared to other characteristics of the brain this correlation is rather small. However, one needs to consider that the sample size of the present study was relatively small.

[c] Verbatim use of formulation of reporting of results from te Nijenhuis & Franssen (2010)

Table 26

Exploratory Bare-Bones Meta-Analytical Results for Correlations Between g Loadings and the Correlation r_1

predictor	*K*	*N*	rho	SD_{rho}	%VE
r (*brain Volume* x *subtest scores*)	4	246	.07	.38	10.15
r (*brain Volume* x *subtest scores*) (without outlier)	3	156	.35	.17	34.21

Note. Bare-bones meta-analytical results: Correlation between brain volume, subtests scores, 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 sampling errors.[d]

Study 5: Biological-environmental factors

Although Spitz (1987) suggested that biological-environmental variables should mimic the pattern found in genetic variables, previous research did not find a strong positive correlation *d* x *g* for biological-environmental variables. Therefore, we want to explore whether changes in IQ profile due to biological-environmental variables are strongly related or unrelated to differences in general intelligence. We expect increased or decreased IQ subtest scores due to biological-environmental influences to have a correlation close to zero with the subtests' *g* loadings. If there is a correlation close to zero, we will also explore whether differences can be found on broad or narrow cognitive abilities measured by Wechsler tests.

Study 5a: Iodine supplementation/ deficiency

The objectives of this analysis are twofold. First, we explore the correlation *d* x *g* between the magnitude of *g* loadings and difference scores on IQ tests of children with iodine deficiency that were supplemented with iodine and children with iodine deficiency that received a placebo. Second, we explore the correlation *d* x *g* between children deficient in iodine and a control group that is not deficient in iodine.

[d] Format of table adapted from te Nijenhuis & Franssen (2010). Verbatim use of formulation of table descriptions from te Nijenhuis & Franssen (2010)

Method[b]

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

Specific criteria for inclusion. Only empirical studies that report at least seven IQ subtest scores were included. In the study of Bleichrodt, Drenth, and Querido (1980) scores of cognitive ability and psychomotor tests were reported. We only included tests of cognitive ability and left psychomotor tests out of the comparison, because it is not clear to what extent such tests measure IQ.

Computation of score differences between an iodine deficient group that receives supplementation with iodine and an iodine deficient group that receives a placebo. Score differences between a treatment group and a placebo group were computed by subtracting the score of the placebo group from the score of the treatment group. The result is divided by *SD* of the placebo group.

Computation of score differences between an iodine deficient group and a comparison group. Score differences between an iodine deficient group and a control group (*d*) were computed by subtracting the mean score of the iodine deficient group of the particular test in question from the mean score of the control group, and then dividing the result by the *SD* of the control group. Bare-bones meta-analytical techniques (Hunter & Schmidt, 1990, 2004) were applied to the resulting six r ($g \times d$)s for iodine deficiency using the software package developed by Schmidt and Le (2004).

Results[c]

The results of the study on the correlation between *g* loadings and the score differences between an iodine deficient group that was supplemented with iodine and an iodine deficient group that received a placebo (*d*) are shown in Table 27. The Table gives data derived from one study, with participants numbering a total of 72. It also lists the reference for the study, the cognitive ability test used, the correlation between *g* loadings and *d*, and the sample size. The correlation *d* x *g* is substantially negative in sign. The results of the study on the correlation between *g* loadings and

[b] Verbatim use of formulation of method from te Nijenhuis & Franssen (2010)
[c] Verbatim use of formulation of reporting of results from te Nijenhuis & Franssen (2010)

the score differences between iodine deficient groups and control groups are reported in Table 28. The Table gives data derived from one study, with participants numbering a total of 196. It also lists the reference for the study, the cognitive ability test used, the correlation between *g* loadings and *d*, and the sample size. Correlations *d* x *g* range from substantially negative to substantially positive. Table 29 presents the results of the bare-bones meta-analysis of six data points. It shows the number of correlation coefficients (*K*), total sample size (*N*), the true correlation (rho) and their standard deviation (SD_r). The last column presents the percentage of variance explained by sampling errors (%VE). The analysis of all data points yields an estimated correlation (rho) of .01, with 51.09% of the variance in the observed correlations explained by sampling error.

Table 27

Studies of Correlations Between g Loadings and Iodine Supplementation

reference	test	*r*	*N*
Shrestha, West, Bleichrodt, van der Vijver, & Hautvast (1994)	Various tests of cognitive abilities	-.54	72

Note. *N* = sample size; *r* = correlation *d* x *g*.[d]

Conclusion

The goal of this study was to tests whether differences in IQ profile due to iodine deficiency are related to general intelligence. One study on the effects of iodine supplementation of iodine deficient children on change in IQ profile yielded a correlation *d* x *g* of -.54. Furthermore, a bare-bones meta-analysis of a study on differences in IQ profile between iodine deficient groups and non-iodine deficient control groups yielded a rho of .01 with 51 % variance explained by sampling error. Taken together, the results of this study confirm our expectations as they indicate that the effect of Iodine supplementation/ deficiency is not strongly related to *g*.

Study 5 b: Prenatal Cocaine Exposure

In the present study, we explore the correlation *d* x *g* between the magnitude of *g* loadings and difference scores on IQ battery subtest between children who were exposed to cocaine prenatally and a control/ standardized group. We conducted an exploratory psychometric meta-analysis on a number of studies that reported IQ scores of at least seven subtests from subjects

[d] Format of table adapted from te Nijenhuis & Franssen (2010). Verbatim use of formulation of table descriptions from te Nijenhuis & Franssen (2010)

exposed to cocaine prenatally. If we observe a negligible correlation, we will explore whether observed differences between IQ are related to broad dimensions of cognitive abilities.

Table 28

Studies of Correlations Between g Loadings and Iodine Deficiency

reference	test	*r*	*N*
Bleichrodt, Drenth, & Querido (1980) (age range: 6-8 years; Control group A)	Various tests of cognitive abilities	.37	21
Bleichrodt, Drenth, & Querido (1980) (age range: 6-8 years; Control group B)	Various tests of cognitive abilities	-.33	21
Bleichrodt, Drenth, & Querido (1980) (age range: 9-12 years; Control group A)	Various tests of cognitive abilities	-.03	41
Bleichrodt, Drenth, & Querido (1980) (age range: 9-12 years; Control group B)	Various tests of cognitive abilities	-.14	36
Bleichrodt, Drenth, & Querido (1980) (age range: 13-20 years; Control group A)	Various tests of cognitive abilities	.32	44
Bleichrodt, Drenth, & Querido (1980) (age range: 13-20 years; Control group B)	Various tests of cognitive abilities	-.23	33

Note. *N* = sample size; *r* = correlation *d* x *g*.[d]

Table 29

Exploratory Bare-Bones Meta-Analytical Results for Correlations Between g Loadings and Iodine Deficient/ Iodine Non-Deficient Score Differences

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

Note. Bare-bones meta-analytical results: Score differences between an iodine deficient group a control group, and *g* loadings. *K* = number of correlations; *N* = total sample size; rho = true correlation (observed correlation corrected for sample size); SD_{rho} = standard deviation of true correlation; %VE = percentage of variance accounted for by sampling error.[d]

---

[d] Format of table adapted from te Nijenhuis & Franssen (2010). Verbatim use of formulation of table descriptions from te Nijenhuis & Franssen (2010)

Method[b]

Searching and screening studies.[22] Te Nijenhuis and Smit (2010) employed a threefold search strategy to identify studies containing IQ scores of children that were exposed to cocaine prenatally. First, an electronic search for published research using PsycINFO, ERIC, MEDLINE, PiCarta, Academic search premier, Web of science, Google Scholar, and PubMed was conducted. Keywords used were cocaine, prenatal(ly exposed), maternal, gestational, pregnancy, drug use, and crack, combined with the words: cognitive, mental ability, intelligence, IQ, WISC, Wechsler, and combinations of these concepts. Second, the reference lists of significant articles were analyzed in search of additional studies. Last, cited reference searches were conducted using Web of Science to identify the newest articles, citing already included key studies. This search yielded nine studies. However, it should be emphasized that the search was quite superficial; there most likely are additional studies to be found with a thorough search. (te Nijenhuis & Franssen, 2010)

Specific criteria for inclusion. For a study to be included in the meta-analysis, two criteria had to be met: First, only empirical studies reporting at least seven IQ-subtest scores of a prenatal cocaine exposure group were included. Second, the mean subtest scores had to be lower than the mean scores of the control group.

Computation of score differences between a prenatal cocaine exposure group and a control group. Score differences between a prenatal cocaine exposure group and a control group (*d*) were computed by subtracting the mean score of the prenatal cocaine exposure group from the mean score of the control group, and then dividing the result by the *SD* of the standardized group from the manual of the IQ battery. Bare-bones meta-analytical techniques (Hunter & Schmidt, 1990, 2004) were applied to the resulting two r ($g \times d$)s using the software package developed by Schmidt and Le (2004).

Results[c]

The results of the studies on the correlation between *g* loadings and the score differences between children exposed to cocaine prenatally and control groups (*d*) are shown in Table 30. The Table presents data derived from two studies, with participants numbering a total of 215. It also lists the reference for the study, the cognitive ability test used, the correlation between *g* loadings and *d*, and the sample size. The correlations are opposite in sign with nearly the same mild magnitude. Table 31 presents the results of the bare-bones meta-analysis of the two data points. It shows the number of correlation coefficients (*K*), total sample size (*N*), the true correlation (rho) and its standard deviation (SD_r). The last column presents the percentage of variance explained by

[b] Verbatim use of formulation of method from te Nijenhuis & Franssen (2010)
[22] Paragraph taken and adapted from te Nijenhuis & Franssen (2010, p. 22)
[c] Verbatim use of formulation of reporting of results from te Nijenhuis & Franssen (2010)

sampling errors (%VE). The analysis of both data points yields an estimated correlation (rho) of -.23, with 16.98% of the variance in the observed correlations explained by sampling errors.

Table 30

Studies of Correlations Between g Loadings and Prenatal Cocaine Exposure

Reference	test	*r*	*N*
Singer, Minnes, Short, Arendt, Farkas, Lewis, Klein, Russ, Min, & Kirchner (2004)	WPPSI-R	-.31	190
Asanbe and Lockert (2006)	WISC-III	.39	25

Note. *N* = sample size; *r* = correlation *d* x *g*.[d]

Table 31

Exploratory Bare-Bones Meta-Analytical Results for Correlations Between g Loadings and Prenatal Cocaine Exposed Subjects/ Control Subjects Score Differences

variable	*K*	*N*	rho	SD_{rho}	%VE
prenatal Cocaine exposure	2	215	-.23	.20	16.98%

Note. Bare-bones meta-analytical results: Score differences between a group prenatally exposed to cocaine, control group, and *g* loadings. *K* = number of correlations; *N* = total sample size; rho = true correlation (observed correlation corrected for sample size); SD_{rho} = standard deviation of true correlation; %VE = percentage of variance accounted for by sampling errors.[d]

Analysis of broad cognitive abilities. Since we found a rho of -.23 in a bare-bones meta-analysis based on two data points that are nearly completely the opposite in magnitude and sign, we also explored whether differences in IQ profile between children exposed to cocaine prenatally and control groups not exposed to cocaine prenatally are due to differences on broad or narrow abilities. The study of Singer et al. (2004) reports only subtests of Crystallized Intelligence and Broad Visual Perception. Therefore, a meta-analysis of effect sizes could not be conducted. Instead, we computed a weighted average of all *d* scores of broad abilities of both studies. This analysis yielded mean *d* scores shown in Table 32. These results are also visualized in Diagram 3.

[d] Format of table adapted from te Nijenhuis & Franssen (2010). Verbatim use of formulation of table descriptions from te Nijenhuis & Franssen (2010)

Table 32

Mean d Scores in Studies of Differences Between Prenatal Cocaine Exposure Groups/ Control Groups

Broad ability	mean *d* score
Fluid Intelligence	.54
Crystallized Intelligence	.16
Broad Visual Perception	.13
General Memory and Learning	.19

Note. *d* scores of Crystallized Intelligence and Broad Visual Perception are based on the weighted average of two studies. Scores of Fluid Intelligence, and General Memory and Learning were reported in only one study.

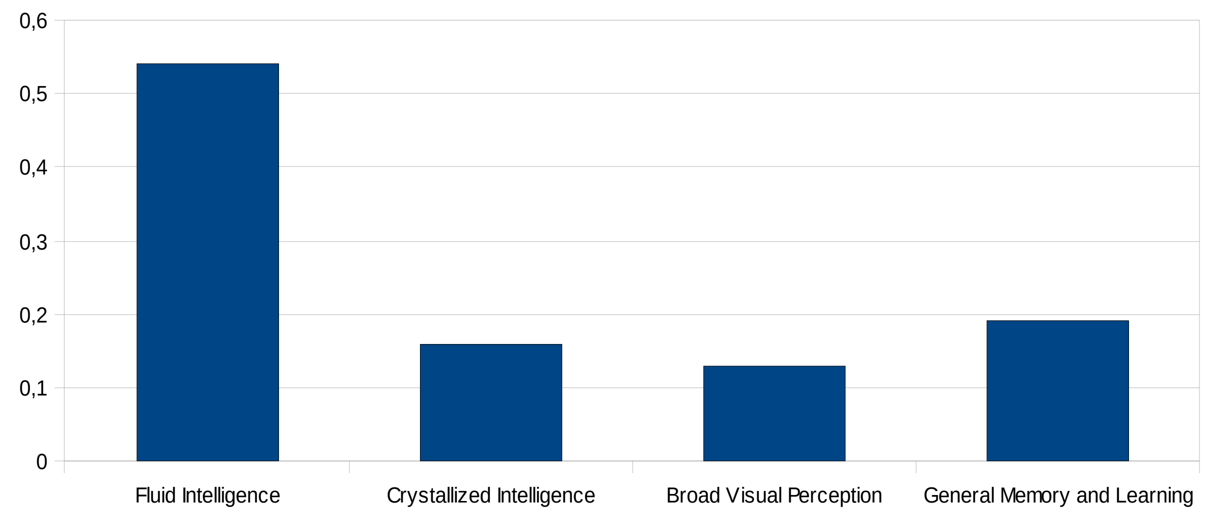


Diagram 3. Effect sizes for differences between children prenatally exposed to cocaine and control groups on tests of Fluid Intelligence, Crystallized Intelligence, Broad Visual Perception, and General Memory and Learning.

Conclusion

The goal of this analysis was to explore the correlation of differences in IQ profile between children prenatally exposed to cocaine and control groups. Based on a bare-bones meta-analysis of two data points with a sample size of 215 we obtained a rho of -.23 with 17% of observed variance explained by sampling error. Clearly, this result does not indicate a strong relationship between general intelligence and IQ impairment due to prenatal cocaine exposure. A further analysis of differences on broad cognitive abilities showed that effect sizes for subtests of Fluid Intelligence are much larger than effect sizes of Crystallized Intelligence, Broad Visual Perception, and General Memory and Learning. We therefore conclude that prenatal cocaine exposure has a differential effect on broad cognitive abilities, with Fluid Intelligence as the cognitive ability on which differences are most strong. The studies included in this meta-analysis both have shortcomings. One study only reported six IQ subtests. We stated earlier to include only studies with at least seven subtests. We deviated from this rule, because there was only a small amount of studies available. However, the correlation in this study could be different if seven subtests were reported. Participants in the other study were also more likely to be exposed to other drugs than cocaine, for instance, marijuana and alcohol. Therefore, the effects on IQ subtests scores are not solely due to the use of cocaine.

Study 5c: Fetal Alcohol Syndrome

To explore the correlation *d* x *g* between the magnitude of *g* loadings and IQ subtest scores of individuals who suffered from fetal alcohol syndrome, an analysis was performed on the data from a study on subjects who suffered from FAS. If we observe a negligible correlation, we will explore whether observed differences between IQ are related to broad dimensions of cognitive abilities. Furthermore, we test whether there is a correlation *d* x *g* close to zero between the magnitude of *g* loadings and IQ subtest scores of individuals with different degrees of severity of FAE/ FAS. If we observe a negligible correlation, we will explore whether observed differences between IQ are related to broad dimensions of cognitive abilities.

Method[b]

Searching and screening studies.[23] Te Nijenhuis, de Pater, van Bloois, and Geutjes (2009) employed several searches. First, an electronic search for published research using PsycINFO, PiCarta, Academic search premier, Web of science, and PubMed was conducted. The following combinations were used to conduct the searches for studies concerning alcohol: the keywords alcohol, alcoholism, alcoholic, Korsakoff, and fetal alcohol syndrome in combination with the keywords IQ, intelligence, intellectual, cognitive, cognition, Wechsler, WAIS, and WISC. Also, the

[b] Verbatim use of formulation of method from te Nijenhuis & Franssen (2010)
[23] Paragraph taken and adapted from te Nijenhuis & Franssen (2010, p. 22)

book by Wechsler (1958) was scanned for suitable studies. Second, the reference lists of significant articles were analyzed in search of additional studies. Last, cited reference searches were conducted using Web of Science to identify the newest articles, citing already included key studies. This search yielded one study. (te Nijenhuis & Franssen, 2010)

Specific criteria for inclusion. Since FAE are lower in severity than FAS 1st° and 2nd° we expected that participants with FAE have a higher full scale IQ than participants with FAS 1st° and 2nd°. However, this is not the case. The FAE group has a Full Scale IQ range of 46 to 117 with a mean of 77, the FAS 1st° group has a Full scale IQ score range from 44 to 132 with a mean of 79, and the FAS 2nd° group has a Full Scale IQ range from 61 to 94 with a mean of 78. Therefore we left comparisons between FAE, FAS 1st°, and FAS 2nd° out of the study.

Computation of score differences between an FAS/ FAE group and a control group. Score differences between an FAS/ FAE group and a control group (*d*) were computed by subtracting the mean score of the FAS/ FAE group from the mean score of the control group, and then dividing the result by the *SD* of the control group. To use the subtest scores and the *SD* of a standardized group is also a theoretical option. However, the manual of the IQ battery was not available. *g* loadings of the HAWIK/E-R were not available. *g* loadings of the HAWIK/E were used instead.

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

Results[c]

The results of the study on the correlation between *g* loadings and score differences between FAE/ FAS and a control group are presented in Table 33. The Table gives data derived from one study, with participants numbering a total of 110. It also lists the reference for the study, the cognitive ability test used, the correlation between *g* loadings and *d*, and the sample size. The correlation is positive and small in magnitude. The results of the study on the correlation between *g* loadings and score differences between different degrees of FAE/ FAS are presented in Table 34.

[c] Verbatim use of formulation of reporting of results from te Nijenhuis & Franssen (2010)

The Table gives data derived from one study, with participants numbering a total of 125. It also lists the reference for the study, the cognitive ability test used, the correlations between *g* loadings and *d*, and the sample size. The correlations are small and positive as well as small and negative in sign. Table 35 presents the results of the bare-bones meta-analysis of the three data points. It shows the number of correlation coefficients (*K*), total sample size (*N*), the true correlation (rho) and their standard deviation (SD_r). The last column presents the percentage of variance explained by sampling errors (%VE). The analysis of all data points yields an estimated correlation (rho) of .12, with 83.04% of the variance in the observed correlations explained by sampling errors.

Table 33

Studies of Correlations Between g Loadings and Fetal Alchohol Syndrome/Fetal Alcohol Effects

reference	test	*r*	*N*
Juretko (2006)	HAWIE-R/ HAWIK-R	.16	110

Note. *N* = sample size; *r* = correlation *d* x *g*. The HAWIE-R is the German version of the WAIS-R. The HAWIK-R is the German version of the WISC-R.[d]

Table 34

Studies of Correlations Between g Loadings and Different Degrees of Fetal Alchohol Syndrome/Fetal Alcohol Effect

reference	test	*r*	*N*
Juretko (2006) FAE - FAS III°	HAWIE-R/ HAWIK-R	-.07	51
Juretko (2006) FAS I° - FAS III°	HAWIE-R/ HAWIK-R	.19	42
Juretko (2006) FAS II° - FAS III°	HAWIE-R/ HAWIK-R	.34	32

Note. *N* = sample size; *r* = correlation *d* x *g*. The HAWIE-R is the German version of the WAIS-R. The HAWIK-R is the German version of the WISC-R.[d]

[d] Format of table adapted from te Nijenhuis & Franssen (2010). Verbatim use of formulation of table descriptions from te Nijenhuis & Franssen (2010)

Table 35

Exploratory Bare-bones Meta-analytical Results for Correlations between g Loadings and different Degrees of Fetal Alcohol Syndrome/Fetal Alcohol Effect

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

Note. Bare-bones meta-analytical results: Score differences between different degrees of FAE/ FAS, and *g* loadings. *K* = number of correlations; *N* = total sample size; rho = true correlation (observed correlation corrected for sample size); SD_{rho} = standard deviation of true correlation; %VE = percentage of variance accounted for by sampling errors.[d]

Analysis of broad cognitive abilities. The results of both studies revealed a small positive correlation *d* x *g*. Such a result does not indicate a strongly positive relationship between differences in IQ profile caused by fetal alcohol syndrome and general intelligence. To explore differences on broad abilities we computed average *d* scores for all subtests with regard to the broad ability they measure. Table 36 shows the mean *d* scores and sample sizes of the study of differences between FAE/ FAS groups and a control group for the broad abilities Fluid intelligence, Crystallized Intelligence, Broad Visual Perception, and General Memory and Learning. These results are also shown in Diagram 5. Table 37 shows the mean *d* scores and sample sizes of the study of differences between degrees of severity of FAS for the broad abilities Fluid intelligence, Crystallized Intelligence, Broad Visual Perception, and General Memory and Learning. These differences are also depictured in Diagram 6.

Table 36

Mean d Scores in a Study of Differences between an Fetal Alcohol Syndrome/Fetal Alcohol Effect Group and a Control Group

Broad ability	mean *d* score
Fluid Intelligence	1.49
Crystallized Intelligence	1.59
Broad Visual Perception	1.11
General Memory and Learning	1.59

[d] Format of table adapted from te Nijenhuis & Franssen (2010). Verbatim use of formulation of table descriptions from te Nijenhuis & Franssen (2010)

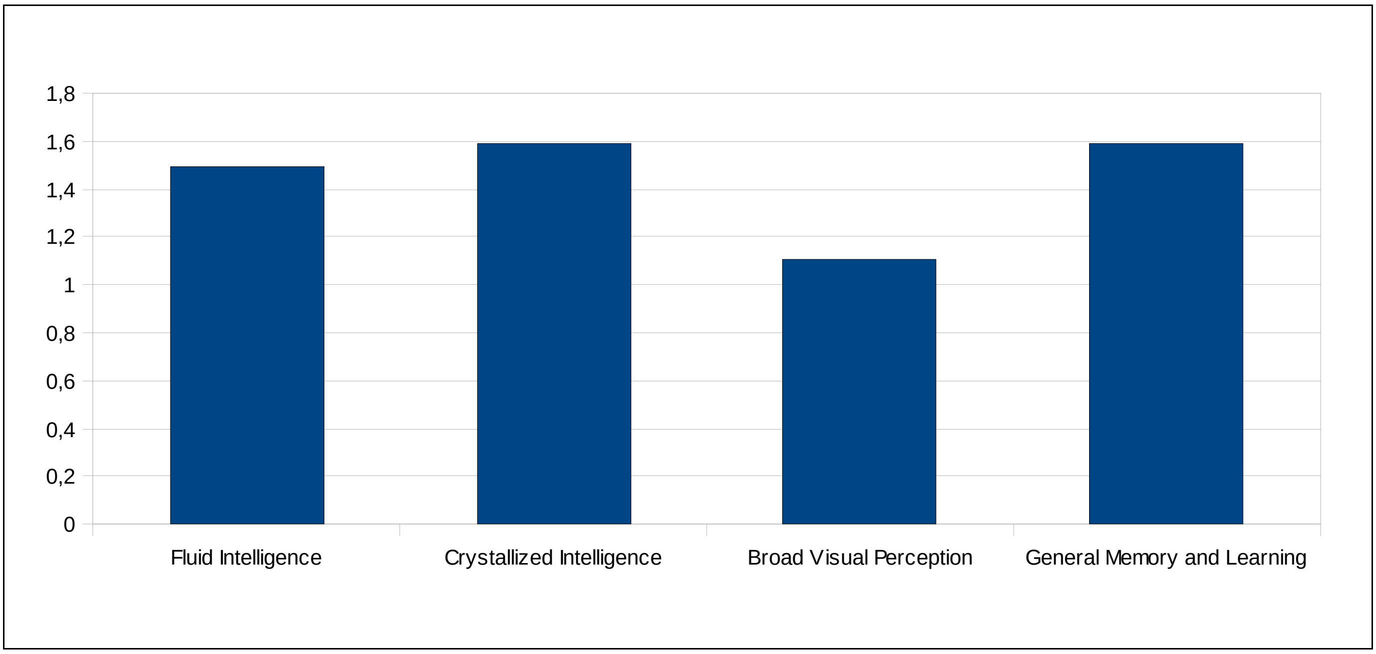


Diagram 5. Effect sizes for differences between an FAE/ FAS group and a control group on tests of Fluid Intelligence, Crystallized Intelligence, Broad Visual Perception, and General Memory and Learning.

Table 37

Mean d Scores in Studies of Differences Between Degrees of Severity of Fetal Alcohol Syndrome/Fetal Alcohol Effect

Broad ability	mean *d* score
Fluid Intelligence	.51
Crystallized Intelligence	.98
Broad Visual Perception	.46
General Memory and Learning	.91

Note. The *d* scores are based on the weighted average of three comparisons between FAE/ FAS of different severity.

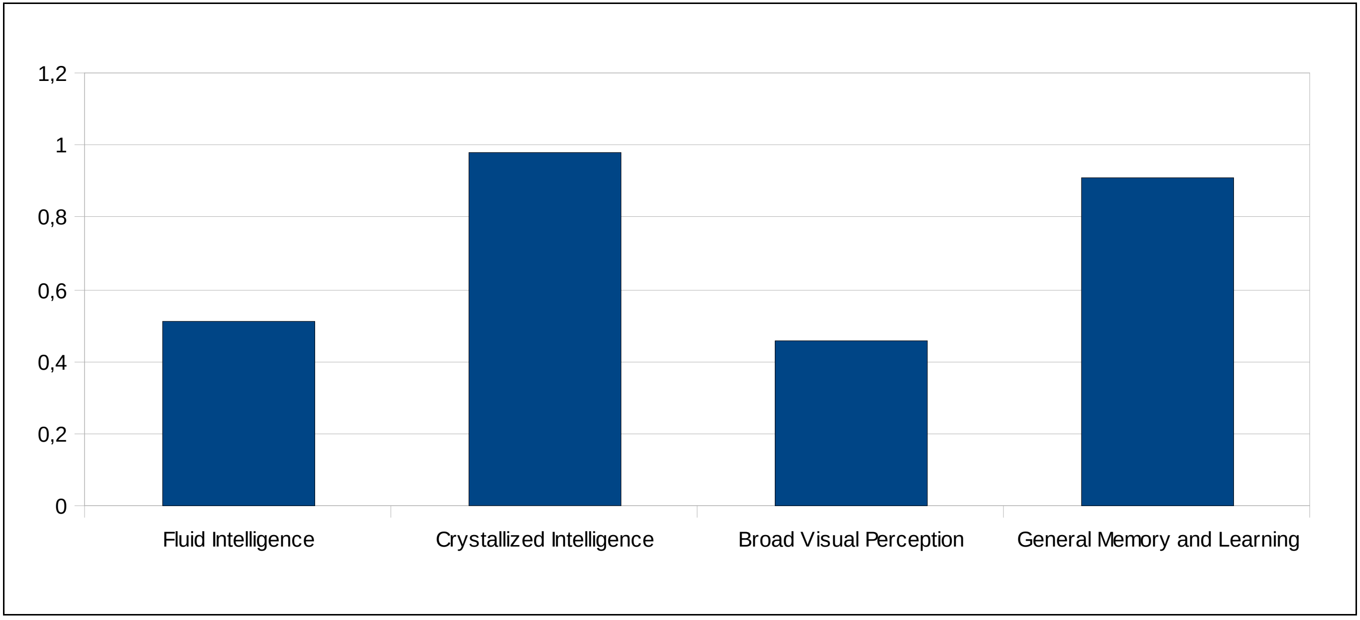


Diagram 6. Effect sizes for differences between Fetal Alcohol Syndrome/Fetal Alcohol Effect groups of different severity on tests of Fluid Intelligence, Crystallized Intelligence, Broad Visual Perception, and General Memory and Learning

Conclusion

The goal of this study was to explore the correlation of differences in IQ profile between individuals that suffer from fetal alcohol syndrome and control groups. Results showed a correlation *d* x *g* of .16 for differences between an FAE/ FAS group and a control group, and a meta-analytic correlation *d* x *g* of .12 for differences between degrees of severity of FAE/ FAS. Based on a relatively small sample size we can conclude that differences in IQ profile between an FAE/ FAS group and a control group as well as differences between degrees of severity of FAE/ FAS are nearly completely unrelated to general intelligence, or *g*. A further analysis of broad cognitive abilities revealed that differences between an FAE/ FAS group and a control group are stronger for tests of Fluid Intelligence, Crystallized Intelligence, and General Memory and Learning, and less strong for tests of Broad Visual Perception. Concerning the difference in IQ profile between different degrees of severity of FAE/ FAS we found that differences on test of Crystallized Intelligence and General Memory and Learning are stronger than differences on tests of Fluid Intelligence and Broad Visual Perception. We therefore conclude that a) IQ impairment due to fetal alcohol syndrome does affect the broad cognitive abilities Fluid Intelligence, Crystallized Intelligence, and General Memory and Learning more strongly than Broad Visual Perception. b) IQ impairment due to a more severe degree of FAE/ FAS does affect the cognitive abilities Crystallized Intelligence and General Memory and Learning more strongly than Broad Visual Perception and Fluid Intelligence. This study has several shortcomings. First, a higher degree of severity of FAE/ FAS does not necessarily correspond to a lower IQ score. The Full scale IQs of FAE, FAS 1st°, and

FAS 2ndº were nearly the same. Only FAS 3rdº had a lower Full scale IQ score than the other degrees. Since we could only make a comparison between FAE, FAS 1stº, FAS 2ndº, and FAS 3rdº, the implications of the results of the correlation *d* x *g* between different degrees of severity are rather limited. We should also consider that a diagnosis with FAE/ FAS does not lead to a pronounced IQ profile at all. The FAE group had a Full Scale IQ range of 46 to 117 with a mean of 77 and the FAS 1stº group had a Full scale IQ score range from 44 to132 with a mean of 79. Therefore, these groups comprised individuals that can be considered severely mentally retarded on the one hand and individuals that can be considered gifted on the other hand. Although we can expect that there are individuals that would have had a rather low IQ score without the effect of FAE/ FAS and also individuals that would have been extremely gifted without the effects of FAE/ FAS, this is an enormous IQ range for a diagnosis.

Study 5d: Air Pollution

To explore the correlation between differences in IQ profile due to exposure with polluted air and general intelligence, or *g*, we computed the correlation *d* x *g* between children from a highly polluted and a lowly polluted city. If we observe a negligible correlation *d* x *g*, we will explore whether differences in intelligence lie on broad cognitive abilities.

Method[b]

Searching and screening studies. No structural search was conducted to identify articles on this topic. The article was obtained by coincidence when searching for articles on other topics that concern the influence of biological-environmental variables on *g*.

Computation of score differences between a group of children from an area with a high level of air pollution and a control group. To compute the correlation *d* x *g*, the subtest scores of the group with children from areas with high air pollution were subtracted from the subtest scores of the children from an area with low levels of air pollution (the control group). The difference is divided by the standard deviation of subtest scores that was obtained in the manual of the IQ battery.

Results[c]

The result of the study on the correlation between *g* loadings and the score differences between children exposed to high levels of air pollution and a control group (*d*) are shown in Table

[b] Verbatim use of formulation of method from te Nijenhuis & Franssen (2010)

[c] Verbatim use of formulation of reporting of results from te Nijenhuis & Franssen (2010)

38. The Table gives data derived from one study, with participants numbering a total of 55. It also lists the reference for the study, the cognitive ability test used, the correlation between *g* loadings and *d*, and the sample size. The correlation is small and negative in sign.

Table 38

Studies of Correlations Between g Loadings and Air Pollution

reference	test	*r*	*N*
Calderon-Garciduenas et al. (2008)	WISC-R	-.17	55

Note. *N* = sample size; *r* = correlation *d* x *g*.[d]

Analysis of broad cognitive abilities. We did only obtain one study on the effects of air pollution on IQ profile, so our result is based on a very small sample size and cannot be considered representational for the whole population of children exposed to air pollution. Nevertheless, we computed the correlation *d* x *g*, which resulted in a value of -.17. Since this correlation does not suggest a strong positive relationship between general intelligence and effects of air pollution on IQ subtest scores, we computed mean *d* scores for each broad cognitive ability. The results of this analysis are shown in Table 39. Overall, mean *d* scores are not very high. Still positive *d* scores were obtained for all broad cognitive abilities. The highest mean *d* score was observed for tests of Fluid Intelligence. Slightly lower difference score were obtained for tests of Crystallized Intelligence, Broad Visual Perception, and General Memory and Learning. These results are also shown in Diagram 7.

Conclusion

The study on the relationship between differences in IQ profile caused by air pollution and general intelligence is different to the previous studies on the effect of prenatal cocaine exposure, and fetal alcohol syndrome on general intelligence, respectively, because in the air pollution study, children were not exposed to a toxin prenatally. Still, we did not obtain results that suggest a strong positive relationship between general intelligence and IQ impairment due to the biological-environmental variable air pollution. Therefore, this study does suggest that the assumption of Spitz (1987), that biological-environmental variables mimic the pattern of genetic variables, is wrong. A further analysis of broad cognitive abilities indicated that the overall effect of air pollution on IQ impairment is rather weak. Nevertheless, the strongest impairment of IQ was found on tests of Fluid Intelligence and a slightly less strong impairment was found on tests of Crystallized Intelligence,

[d] Format of table adapted from te Nijenhuis & Franssen (2010). Verbatim use of formulation of table descriptions from te Nijenhuis & Franssen (2010)

Broad Visual Perception, and General Memory and Learning. Since our results are based on only one study we should not overestimate the theoretical implications. Still, in the context of the results of studies on other biological-environmental variables as iodine deficiency/ supplementation, prenatal cocaine exposure, and fetal alcohol syndrome, we can conclude that the results on the

Table 39

Mean d Scores of Broad Abilities

Broad ability	mean *d* score
Fluid Intelligence	.55
Crystallized Intelligence	.50
Broad Visual Perception	.46
General Memory and Learning	.46

Note. The *d* scores are based on one study.

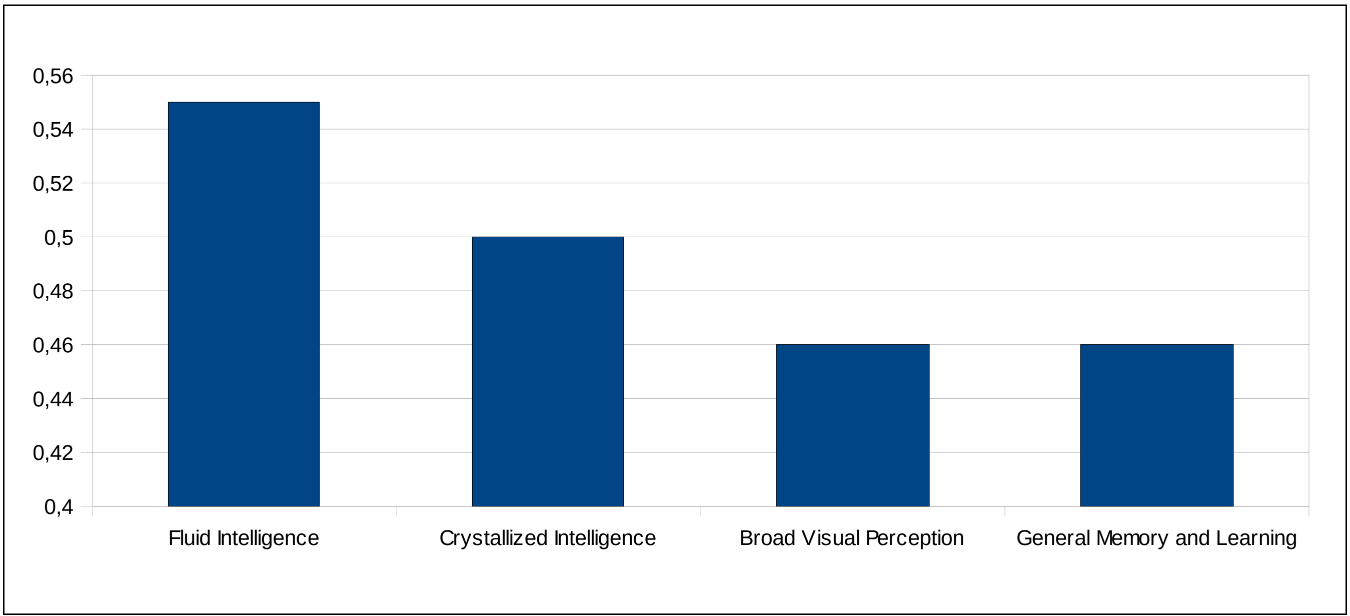


Diagram 7. Effect sizes for differences between children from an area with high levels of air pollution and children from an area with low levels of air pollution on tests of Fluid Intelligence, Crystallized Intelligence, Broad Visual Perception, and General Memory and Learning

impairment of IQ by air pollution fit into the pattern of a rather minor role of general intelligence as an explaining factor of IQ differences between average groups and groups exposed to biological-environmental variables that have a detrimental effect on IQ.

Study 5e: Traumatic Brain Injury

To explore the correlation *d* x *g* between the magnitude of *g* loadings and IQ subtest scores of individuals who suffered from TBI, an exploratory psychometric meta-analysis was performed on a number of studies that reported IQ scores from TBI subjects. If we observe a negligible correlation *d* x *g*, we will explore whether differences in intelligence lie on broad cognitive abilities.

Method[b]

Searching and screening studies.[24] To identify studies for inclusion in the meta-analysis, both electronic and manual searches were conducted for studies that contained cognitive ability data of TBI. Three methods were used to obtain scores of the traumatic brain injured from published studies for the present meta-analysis. First, an electronic search for published research using PsycINFO, Picarta, Academic search premier, Web of science, and PubMed was conducted. The following combinations were used to conduct the searches: any keyword that contains the word ‘traumatic brain injury’, or ‘brain trauma’ in combination with any keyword that contains one of the following words; IQ, g, general mental ability, GMA, cognitive ability, general cognitive ability, intelligence, Wechsler, cognitive ability test. Second, we browsed the tables of content of several major research journals with a strong focus on the traumatic brain injured: *Brain Injury* 2000-2010, *Applied Neuropsychology* 1999-2010, *Journal of the International Neuropsychological Society* 1997-2010, and *Journal of Neurotrauma* 1993-2010. Third, we checked the reference list of all currently included empirical studies to identify any potential articles that may have been missed by earlier search methods. This procedure resulted in 40 articles and reports on the concurrent topics of traumatic brain injury and mental ability. Eight studies met all criteria for inclusion in the meta-analysis.

Specific criteria for inclusion. For a study to be included in the meta-analysis, two additional criteria had to be met: First, only empirical studies reporting IQ scores of a TBI group were included. Second, the mean subtest scores had to be lower than the mean scores of the control group.

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

---

[b] Verbatim use of formulation of method from te Nijenhuis & Franssen (2010)

[24] Paragraph taken and adapted from te Nijenhuis & Franssen (2010, p. 22)

Results[c]

The results of the studies on the correlation between *g* loadings and the score differences between a TBI group and control/ standardized groups (*d*) are shown in Table 40. The Table reports data derived from nine studies, with participants numbering a total of 629. It also lists the reference for the study, the cognitive ability test used, the correlation between *g* loadings and *d*, and the sample size. The correlations show no clear pattern with regard to magnitude or sign. Table 41 presents the results of the bare-bones meta-analysis of the 14 data points. It shows the number of correlation coefficients (*K*), total sample size (*N*), the true correlation (rho) and their standard deviation (SD_r). The last column presents the percentage of variance explained by sampling errors (%VE). The analysis of 14 data points yields an estimated correlation (rho) of -.07, with 35.43% of the variance in the observed correlations explained by sampling error. Sample sizes were highly comparable, which most likely led to much lower %VE.

Table 40

Studies of Correlations Between g Loadings and Traumatic Brain Injury

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

Note. *N* = sample size; *r* = correlation *d* x *g*.[d]

[c] Verbatim use of formulation of reporting of results from te Nijenhuis & Franssen (2010)

[d] Format of table adapted from te Nijenhuis & Franssen (2010). Verbatim use of formulation of table descriptions from te Nijenhuis & Franssen (2010)

Table 41

Exploratory Bare-Bones Meta-Analytical Results for Correlations Between g Loadings and TBI Groups and Control/ Standardized Groups Score Differences

predictor	*K*	*N*	rho	SD_{rho}	%VE
Traumatic Brain Injury	14	629	-.07	.20	35.43%

Note. Bare-bones meta-analytical results: Score differences between a TBI group, control group, and *g* loadings. *K* = number of correlations; *N* = total sample size; rho = true correlation (observed correlation corrected for sample size); SD_{rho} = standard deviation of true correlation; %VE = percentage of variance accounted for by sampling errors.[d]

Analysis of broad cognitive abilities. The bare-bones meta-analysis of the correlation *d* x *g* between TBI groups and control groups did not yield a correlation that points to a substantial positive relationship of IQ impairment due to TBI and general intelligence. Therefore, we also conducted an analysis of difference scores on broad cognitive abilities. We averaged *d* scores of subtests for every broad ability in every study that used a Wechsler test and conducted a bare-bones meta-analysis of effect sizes with the average broad ability *d* score of each study as data points. Table 42 presents the results of the bare-bones meta-analyses of 11 data points for Fluid Intelligence, Crystallized Intelligence, Broad Visual Perception, and General Memory and Learning. It shows (from left to right): the number of *d* scores (*K*), total sample size (*N*), the true effect size (d_t) and their standard deviation (SD_d). The last column presents the percentage of variance explained by sampling errors (%VE). The analysis of 11 data points of differences on tests of Fluid Intelligence yields an estimated effect size (d_t) of .53, with 79.73% of the variance in the observed effect sizes explained by sampling errors. The analysis of 11 data points of differences on tests of Crystallized Intelligence yields an estimated effect size (d_t) of .49, with 132.90% of the variance in the observed effect sizes explained by sampling errors. The analysis of 11 data points of differences on tests of Broad Visual Perception yields an estimated effect size (d_t) of .52, with 77.34% of the variance in the observed effect sizes explained by sampling errors. Finally, the analysis of 11 data points of differences on tests of General Memory and Learning yields an estimated effect size (d_t) of .65, with 105.54% of the variance in the observed effect sizes explained by sampling errors. These results are also shown in Diagram 8.

The analysis of data points of Crystallized Intelligence and General Memory and Learning yield estimated effect sizes, with a percentage of variance explained by sampling errors larger than 100. This phenomenon is called "second-order sampling error", and results from the sampling of

[d] Format of table adapted from te Nijenhuis & Franssen (2010). Verbatim use of formulation of table descriptions from te Nijenhuis & Franssen (2010)

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).[25] (taken from te Nijenhuis & Franssen, 2010)

Table 42

Exploratory Bare-Bones Meta-Analytical Results for d Scores on Tests of Fluid Intelligence, Crystallized Intelligence, Broad Visual Perception, and General Memory and Learning for TBI/ Control groups

variable	*K*	*N*	d_t	SD_d	%VE
Fluid Intelligence	11	449	.53	.06	79.73%
Crystallized Intelligence	11	449	.49	0	132.90
Broad Visual Perception	11	449	.52	.17	77.34
General Memory and Learning	11	449	.65	0	105.54

Note. Bare-bones meta-analytical results: Score differences between TBI groups and control groups on subtests of fluid intelligence, crystallized intelligence, broad visual perception, and general memory and learning. *K* = number of d_t scores; *N* = total sample size; *d* = true effect size corrected for sample size; SD_d = standard deviation of true effect size; %VE = percentage of variance accounted for by sampling error.[d]

Conclusion

The study on the relationship between differences in IQ profile caused by TBI reveals further evidence against the assumption of Spitz (1987), that environmental-biological variables mimic the pattern of genetic variables. Our analysis resulted in a meta-analytic correlation *d* x *g* for differences between TBI groups and control groups of -.07. Therefore, general intelligence is nearly completely unrelated to differences in IQ profile between TBI and control groups. A further analysis of differences on broad cognitive abilities revealed higher difference scores on test of General Memory and Learning and roughly the same magnitude of difference scores on tests of Fluid Intelligence,

[25] Paragraph taken and adapted from te Nijenhuis & Franssen (2010, p. 35)

[d] Format of table adapted from te Nijenhuis & Franssen (2010). Verbatim use of formulation of table descriptions from te Nijenhuis & Franssen (2010)

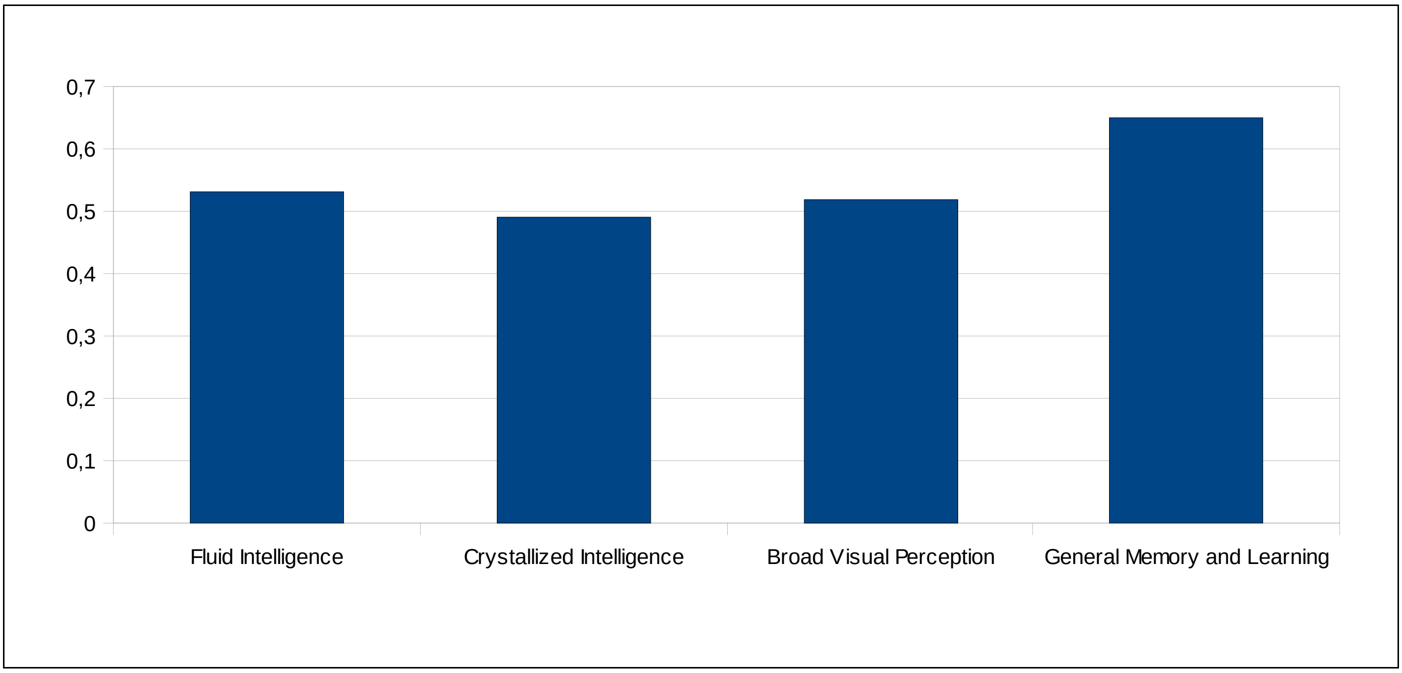


Diagram 8. Effect sizes for differences between TBI groups and control groups on tests of Fluid Intelligence, Crystallized Intelligence, Broad Visual Perception, and General Memory and Learning

Crystallized Intelligence, and Broad Visual Perception. Although TBI does clearly lead to an impairment of all cognitive abilities, the impairment of General Memory and Learning is most severe. These results present further evidence against a major role of general intelligence in IQ impairment due to biological-environmental variables.

Study 5f: Malnutrition

To explore the correlation between the magnitude of *g* loadings and IQ scores of malnourished individuals, an exploratory meta-analysis was performed on all studies that reported IQ scores from malnourished individuals. If we obtain a negligible correlation *d* x *g*, we will explore whether differences lie on broad cognitive abilities.

Method[b]

Searching and screening studies.[26] To identify studies for inclusion in the meta-analysis both electronic and manual searches for studies that contained cognitive ability data of malnourished children or adults were conducted. Four methods were employed. First, an electronic search for published work was conducted, using PsycINFO, ERIC, MEDLINE, PiCarta, Academic

[b] Verbatim use of formulation of method from te Nijenhuis & Franssen (2010)
[26] Paragraph taken and adapted from te Nijenhuis & Franssen (2010, p. 22)

search premier, Web of science, Google Scholar, and PubMed. The following keyword combinations were used to conduct searches: malnutrition, malnourishment* (* is a truncation symbol to represent multiple spellings or endings), in combination with the keywords IQ, intelligence, intellectual development, *g*, GMA, general mental ability, cognitive development, cognitive ability, and general cognitive ability. Second, we browsed the content tables of several major research journals of development, genetics, and of intelligence, such as *Child Development* 1930-2007*, and Developmental Psychology* 1969-2007. Third, we browsed the tables of contents of several major research journals of genetics, development, and of intelligence, such as *Behavior Genetics* 1970-2007*, Intelligence* 1977-2007, and *The Journal of Pediatrics* 1966-2000. Fourth, 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, cited reference searches were conducted using Web of Science to identify the newest articles, citing key studies already included.

This procedure resulted in seven articles, book chapters, and reports on the concurrent topics of malnourishment and mental ability. Only one study met all criteria for inclusion in the MA, comprising all published research on the subject published in English-language research journals and books.

Specific criteria for inclusion. For a study to be included in the meta-analysis the mean subtest scores had to be lower than the mean scores of the control group.

Computation of score differences between a malnourished group and a comparison group. Score differences between a malnourished group and a control group (*d*) were computed by subtracting the mean score of the malnourished group from the mean score of the control group, and then dividing the result by the *SD* of the standardization sample from the manual of the IQ battery.

Results[c]

The result of the study on the correlation between *g* loadings and the score differences between a malnourished group and a control group (*d*) are shown in Table 43. The Table gives data derived from one study, with participants numbering a total of 51. It also lists the reference for the study, the cognitive ability test used, the correlation between *g* loadings and *d*, and the sample size. The correlation is small and negative in sign.

Analysis of broad and narrow abilities. The correlation *d* x *g* between a malnourished group and a control group does not point to an either pronounced positive or negative relationship between the IQ profile of a malnourished group and general intelligence. Therefore, we did also compute mean *d* scores for all broad abilities. Table 44 shows the results of this computation.

---

[c] Verbatim use of formulation of reporting of results from te Nijenhuis & Franssen (2010)

Table 43

Studies of Correlations Between g Loadings and Malnutrition

reference	test	*r*	*N*
Gilberg, Rastam, Wentz, & Gillberg (2007)	WAIS-R	-.15	51

Note. *N* = sample size; *r* = correlation *d* x *g*.[d]

Overall, *d* scores are not very high. However, Fluid Intelligence and Broad Visual Perception seem slightly more affected by malnutrition than Crystallized Intelligence and General Memory and Learning. These results are also shown in Diagram 9.

Conclusion

The study on the relationship between differences in IQ profile caused by malnutrition reveals further evidence against the assumption of Spitz (1987), that environmental-biological variables mimic the pattern of genetic variables. Our analysis resulted in a correlation *d* x *g* for IQ differences between a malnourished group and a control group of -.15. Therefore, general intelligence is nearly completely unrelated to differences in IQ profile between a malnourished and a control group. A further analysis of differences on broad cognitive abilities revealed higher difference scores on test of Fluid Intelligence and Broad Visual Perception and roughly the same lower magnitude of difference scores on tests of Crystallized Intelligence, and General Memory and Learning. Overall, the magnitude of difference scores is rather low. These results present further evidence against a major role of general intelligence in IQ impairment due to biological-environmental variables.

Table 44

Mean d Scores of Broad Abilities

Broad ability	mean *d* score
Fluid Intelligence	.22
Crystallized Intelligence	.17
Broad Visual Perception	.22
General Memory and Learning	.18

[d] Format of table adapted from te Nijenhuis & Franssen (2010). Verbatim use of formulation of table descriptions from te Nijenhuis & Franssen (2010)

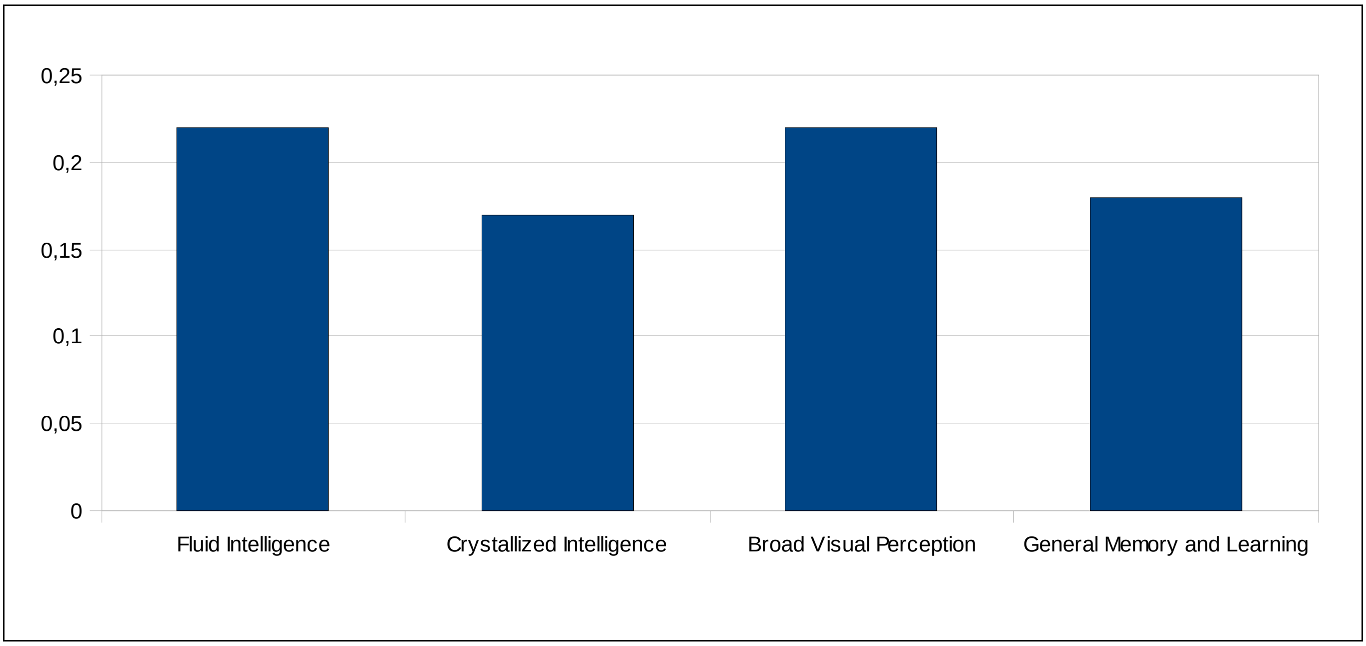


Diagram 9. Effect sizes for differences between a malnourished group and a control group on tests of Fluid Intelligence, Crystallized Intelligence, Broad Visual Perception, and General Memory and Learning.

General Conclusion

The goal inherent to all studies in this section was to test whether differences in IQ due to biological-environmental variables show no strong correlation with general intelligence. Table 45 gives an overview of all correlations between *g* loadings and differences on biological-environmental variables we obtained in our studies. With one exception, mean correlations are very close to zero. Based on the results of seven biological-environmental variables we can conclude that Spitz assumption is wrong, and biological-environmental variables do not mimic the pattern of genetic variables. Group differences due to biological-environmental variables are rather unrelated to general intelligence, or *g*. The effects of biological-environmental variables should therefore be studied at the level of broad abilities.

Table 45

Correlations of g Loadings with Biological-Environmental Variables

variable	*K*	*N*	*r*	rho
Iodine supplementation	1	72	-.54	
Iodine deficiency	6	196		.01
Prenatal cocaine exposure	2	215		-.23
Fetal alcohol syndrome/ Fetal alcohol effects	1	110	.16	
Air pollution	1	55	-.17	
Traumatic brain injury	14	629		-.07
Malnutrition	1	51	-.15	

Note. *K* = Number of correlations; *N* = Total sample size or harmonic *N*; *r* = mean observed correlation (sample size weighted); rho = true correlation (observed correlation corrected for unreliability, range restriction, and imperfectly measuring the construct of *g*).[d]

Study 6: Aging and Autism

Study 6a: Aging

To explore the correlation between the magnitude of *g* loadings and the decline on IQ subtest scores of individuals due to aging, an exploratory analysis was performed on data of a longitudinal study (Schaie and Willis, 1993) and the standardization study of the Spanish WAIS-III (Juan-Espinosa, 2002). If we observe a strong correlation *d* x *g* this would indicate that general intelligence declines over the course of a lifetime.

Measures of Narrow and Broad Abilities in the Study of Schaie and Willis (1993)

The primary mental ability (PMA) battery in the longitudinal study of Schaie and Willis (1993) includes the tests *PMA Reasoning*, *PMA Space*, *PMA Number*, *PMA Verbal* (Thurstone et al., 1948), *Cube Comparison*, *Finding As*, *Number Comparison*, *Identical Pictures*, *Addition*, *Subtraction & Multiplication*, *Vocabulary 2*, *Vocabulary 4* (Ekstrom et al., 1976), ADEPT *Letter Series (Form A)* (Blieszner, Willis, & Baltes, 1981), *Word Series*, *Object Rotation* (Schaie, 1985), *Alphanumeric Rotation* (Willis & Schaie, 1983), and *Number Series* (Ekstrom, French, Hartman, & Derman, 1976). We will explain which of the subtests is a measure of Crystallized Intelligence and

[d] Format of table adapted from te Nijenhuis & Franssen (2010). Verbatim use of formulation of table descriptions from te Nijenhuis & Franssen (2010)

which is a measure of other broad cognitive abilities. Two different sources for this differentiation are shown in Table 46. First, we use the indication given by Willis and Schaie (1993). Second, we compare the test to the taxonomy of Carroll (1993).

Method[b]

Searching and screening studies.[27] Four methods were used to identify studies that contained correlations between aging and IQ subtest scores. First, an electronic search for published research using PsycINFO, ERIC, MEDLINE, PiCarta, Academic search premier, Web of science, Google Scholar, and PubMed was conducted. Keywords used were aging, old age, cognitive, mental*, intelligence, IQ, WAIS, Wechsler, and combinations of these concepts (* is a truncation symbol to represent multiple spellings or endings; AND is a Boolean operator that combines search terms so that the search result contains all of the terms). Second, the reference lists of significant articles were analyzed in search of additional studies. Third, cited reference searches were conducted using Web of Science, to search for articles citing significant articles. Fourth all articles of Werner Schaie that were available on Psycinfo were reviewed. This procedure resulted in three articles. (te Nijenhuis & Franssen, 2010)

Specific criteria for inclusion. A study by Ardila (2007) converted scaled scores of the WAIS-III into raw scores by using the mean of the raw score range that corresponded to a scaled score. Differences between age groups on subtest level are only visible if subtest score values fall into two different raw score ranges. We decided to leave this study out of the analysis, because real differences scores between groups were not obtainable in this dataset.

Computation of score differences between a group consisting of older individuals and a group consisting of younger individuals. Score differences between younger groups and older groups (*d*) were computed by subtracting the mean score of the old group from the mean score of the young group, and then dividing the result by the pooled *SD* of all age groups reported in the study.

Results[c]

The results of the longitudinal study on the correlation between *g* loadings and the score differences between different age groups (*d*) are shown in Table 47. The Table reports data derived from one study, with participants numbering a total of 2,677. It also lists the reference for the study, the cognitive ability test used, the correlation between *g* loadings and *d* (r_1), the correlation between *g* loadings and *d* when subtests that measure crystallized intelligence are left out (r_2), and the

[b] Verbatim use of formulation of method from te Nijenhuis & Franssen (2010)
[27] Paragraph taken and adapted from te Nijenhuis & Franssen (2010, p. 22)

[c] Verbatim use of formulation of reporting of results from te Nijenhuis & Franssen (2010)

Table 46

Underlying Factors Measured in the Primary Mental Abilities Battery

Subtests	*Primary ability*	*Stratum II Factor*	*Taxonomy of Carroll (1993)*
PMA Reasoning *ADEPT Letter Series (Form A)* *Word Series* *Number Series*	Inductive Reasoning	Fluid Intelligence	According to Schaie and Willis (1993) all four subtests are measures of Inductive Reasoning which is a measure of Fluid Intelligence. In the taxonomy of Carroll (1993) Inductive Reasoning represents a stratum I factor that belongs to the stratum II factor Fluid intelligence.
PMA Space *Object Rotation* *Alphanumeric Rotation* *Cube Comparison*	Spatial Orientation	Broad Visual Perception	According to Schaie and Willis (1993) these subtests measure Mental Rotation. In the taxonomy of Carroll (1993) this is represented in the Stratum I factor Visualization that belongs to the Stratum II factor Broad Visual Perception.
Finding As *Number Comparison* *Identical Pictures*	Perceptual Speed	Broad Visual Perception	According to Schaie and Willis (1993) these subtests are measures of Speed of Cognitive Processing. In the taxonomy of Carroll (1993) this is represented in the Stratum II factor Broad Visual Perception.
PMA Number *Addition* *Subtraction & Multiplication*	Numeric	Crystallized Intelligence	According to Schaie and Willis (1993) these subtests are measures of Crystallized Intelligence. In the taxonomy of Carroll the testing of basic arithmetic operations belongs to Crystallized Intelligence at stratum II.
PMA Verbal *Vocabulary 2* *Vocabulary 4*	Verbal	Crystallized intelligence	According to Schaie and Willis (1993) these subtests are measures of Crystallized Intelligence. In the taxonomy of Carroll (1993) measures of word meanings, vocabulary or verbal ability belong to Crystallized Intelligence at stratum II.

sample size. In general, correlations are substantially positive. Table 48 presents the results of the analysis of the standardization study of the Spanish WAIS-III, with participants numbering a total of 850. It also lists the reference for the study, the cognitive ability test used, the correlation between *g* loadings and *d* (r_1), the correlation between *g* loadings and *d* when subtests that measure crystallized intelligence are left out (r_2), and the sample size. The correlations are highly positive. Table 49 reports the results of the bare-bones meta-analysis of 23 data points from both studies. It shows the number of correlation coefficients (*K*), total sample size (*N*), the true correlation (rho) and their standard deviation (SD_r). The last column presents the percentage of variance explained by sampling errors (%VE). The table reports meta-analytical results for correlations *d* x *g* including all subtests and for correlations only including subtests that do not belong to the Crystallized Intelligence domain. The analysis of all data points yields an estimated correlation (rho) of .45 for all subtest, with 28% of observed variance explained by sampling errors, and a correlation of rho = .59 when subtests of crystallized intelligence are left out, with 15% of observed variance explained by sampling error.

Table 47

Studies of Correlations Between g Loadings and Different Age Groups in a Longitudinal Study

reference	test	r_1	r_2	*N*	*reference*	test	r_1	r_2	*N*
Schaie & Willis (1993) Age 29-39	PMA	.05	-.26	142	Schaie & Willis (1993) Age 46-67	PMA	.31	.32	201
Schaie & Willis (1993) Age 29-46	PMA	.22	.44	159	Schaie & Willis (1993) Age 46-74	PMA	.35	.39	193
Schaie & Willis (1993) Age 29-53	PMA	.28	.38	178	Schaie & Willis (1993) Age 46-81	PMA	.51	.58	145
Schaie & Willis (1993) Age 29-60	PMA	.23	.40	202	Schaie & Willis (1993) Age 46-88	PMA	.56	.54	65
Schaie & Willis (1993) Age 29-67	PMA	.29	.38	211	Schaie & Willis (1993) Age 53-60	PMA	.11	.33	223
Schaie & Willis (1993) Age 29-74	PMA	.32	.43	202	Schaie & Willis (1993) Age 53-67	PMA	.29	.32	233
Schaie & Willis (1993) Age 29-81	PMA	.45	.59	150	Schaie & Willis (1993) Age 53-74	PMA	.34	.41	222
Schaie & Willis (1993) Age 29-88	PMA	.51	.56	66	Schaie & Willis (1993) Age 53-81	PMA	.53	.61	161
Schaie & Willis (1993) Age 39-46	PMA	.34	.55	137	Schaie & Willis (1993) Age 53-88	PMA	.58	.55	68
Schaie & Willis (1993) Age 39-53	PMA	.39	.48	151	Schaie & Willis (1993) Age 60-67	PMA	.39	.22	276
Schaie & Willis (1993) Age 39-60	PMA	.29	.49	168	Schaie & Willis (1993) Age 60-74	PMA	.42	.38	261
Schaie & Willis (1993) Age 39-67	PMA	.35	.46	174	Schaie & Willis (1993) Age 60-81	PMA	.62	.63	180
Schaie & Willis (1993) Age 39-74	PMA	.37	.51	168	Schaie & Willis (1993) Age 60-88	PMA	.64	.55	71
Schaie & Willis (1993) Age 39-81	PMA	.52	.64	130	Schaie & Willis (1993) Age 67-74	PMA	.41	.54	275
Schaie & Willis (1993) Age 39-88	PMA	.57	.61	62	Schaie & Willis (1993) Age 67-81	PMA	.73	.76	186
Schaie & Willis (1993) Age 46-53	PMA	.32	.18	172	Schaie & Willis (1993) Age 67-88	PMA	.68	.59	72
Schaie & Willis (1993) Age 46-60	PMA	.23	.35	194					

Note. *N* = sample size; r_1 = correlation *d* x *g* when Crystallized Intelligence subtests are included; r_2 = correlation *d* x *g* when Crystallized Intelligence subtests are excluded.[d]

[d] Format of table adapted from te Nijenhuis & Franssen (2010). Verbatim use of formulation of table descriptions from te Nijenhuis & Franssen (2010)

Table 48

Studies of Correlations Between g Loadings and Different Age Groups in the Standardization Study of the Spanish WAIS-III

reference	test	r_1	r_2	*N*
Juan-Espinosa (2002) Age 16-19 & Age 70-94	SpanishWAIS-III	.29	.78	148
Juan-Espinosa (2002) Age 20-24 & Age 70-94	SpanishWAIS-III	.30	.78	144
Juan-Espinosa (2002) Age 25-34 & Age 70-94	SpanishWAIS-III	.32	.77	181
Juan-Espinosa (2002) Age 35-54 & Age 70-94	SpanishWAIS-III	.41	.79	204
Juan-Espinosa (2002) Age 55-69 & Age 70-94	SpanishWAIS-III	.49	.68	173

Note. *N* = sample size; r_1 = correlation *d* x *g* when Crystallized Intelligence subtests are included; r_2 = correlation *d* x *g* when Crystallized Intelligence subtests are excluded.[d]

Table 49

Exploratory Bare-Bones Meta-Analytical Results for Correlations Between g Loadings and Age Group Score Differences with a Cut-Off Point at Age 68 in both Studies

variable	*study*	*K*	*N*	rho	SD_{rho}	%VE
Age	All studies	23	3,527	.45	.10	27.57
Age (without subtest of crystallized intelligence)	All studies	23	3.527	.59	.13	15.14

Note. Bare-bones meta-analytical results: Score differences between different age groups, 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 sampling errors.[d]

Type of study as a moderator. Since the two studies included in the analyses were very different in nature, we did explore whether the meta-analytic correlation rho differs between studies. Table 50 reports the results of this analysis. It shows the number of correlation coefficients (*K*), total sample size (*N*), the true correlation (rho) and their standard deviation (SD_r). The last column presents the percentage of variance explained by sampling errors (%VE). The true correlation rho when subtests of Crystallized Intelligence are included and when they are excluded differed extremely between the longitudinal study and the standardization study of the WAIS-III. When subtests measuring Crystallized Intelligence are included in the analysis the true correlation rho for the WAIS-III is only .37 compared to a correlation of .48 in the longitudinal study. This result changes dramatically when subtest measuring Crystallized Intelligence are left out of the comparison. Then the true correlation rho of the standardization study of the WAIS-III rises to .76. The true correlation rho for the longitudinal study enhances to .53, only. Clearly, the effects of aging on general intelligence are very different in the longitudinal study and in the standardization study of the WAIS-III. The effect of the exclusion of subtests of Crystallized Intelligence on the true correlation rho is shown in Diagram 12.

Table 50

Type of Study as a Moderator

variable	*Moderator*	*K*	*N*	*rho*	*SD*rho	%VE
Age (including Crystallized subtests)	Longitudinal study	18	2,677	.48	.10	27.57
Age (excluding Crystallized subtests)	Longitudinal study	18	2,677	.53	.09	30.23
Age (including Crystallized subtests)	Standardization study WAIS-III	5	850	.37	.03	76.51
Age (excluding Crystallized subtests)	Standardization study WAIS-III	5	850	.76	.02	62.99

Note. *K* = number of correlations; *N* = total sample size; rho = true correlation (observed correlation corrected for sample size); SD_r = standard deviation of true correlation; %VE = percentage of variance accounted for by sampling errors.[d]

Age range as a moderator. A close inspection of the data from the longitudinal study reveals that when both groups in the comparison are above age 60 the correlations are much stronger. So, we tested the age range as a moderator. Table 51 reports the results of this analysis. It shows the number of correlation coefficients (*K*), total sample size (*N*), the true correlation (rho) and their standard deviation (SD_r). The last column presents the percentage of variance explained by sampling errors (%VE). When subtests that measure Crystallized Intelligence are included the true correlation rho for comparisons between groups younger than 60 is .24. For groups older than 60 this correlation rises to .55. When subtests that measure Crystallized Intelligence are excluded from the analysis the true correlation rho for comparisons between groups younger than 60 is .34. This

[d] Format of table adapted from te Nijenhuis & Franssen (2010). Verbatim use of formulation of table descriptions from te Nijenhuis & Franssen (2010)

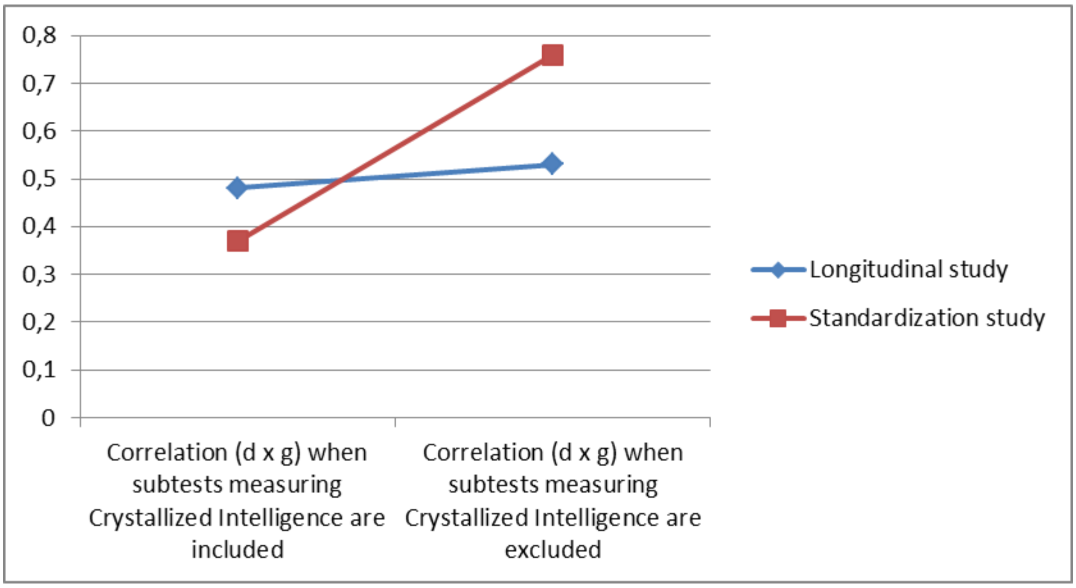


Diagram 10. Correlations *d* x *g* for age comparisons with a cut-off point at age 68 when crystallized subtests are included and when they are excluded in the longitudinal study and the standardization study of the WAIS-III.

Tabel 51

Comparisons Between Groups Younger Than 60 and Comparisons Between Groups Older Than 60 in the Longitudinal Study

variable	*K*	*N*	rho	SD_{rho}	%VE
Comparison between groups younger than 60 (including Crystallized subtests)	10	1,726	.24	.06	57.3
Comparison between groups younger than 60 (excluding Crystallized subtests)	10	1,726	.34	.19	11.12
Comparison between groups older than 60 (including Crystallized subtests)	10	1,634	.55	.15	11.16
Comparison between groups older than 60 (excluding Crystallized subtests)	10	1,634	.52	.19	8.4

Note. *K* = number of correlations; *N* = total sample size; rho = true correlation (observed correlation corrected for sample size); SD_r = standard deviation of true correlation; %VE = percentage of variance accounted for by sampling errors.[d]

[d] Format of table adapted from te Nijenhuis & Franssen (2010). Verbatim use of formulation of table descriptions from te Nijenhuis & Franssen (2010)

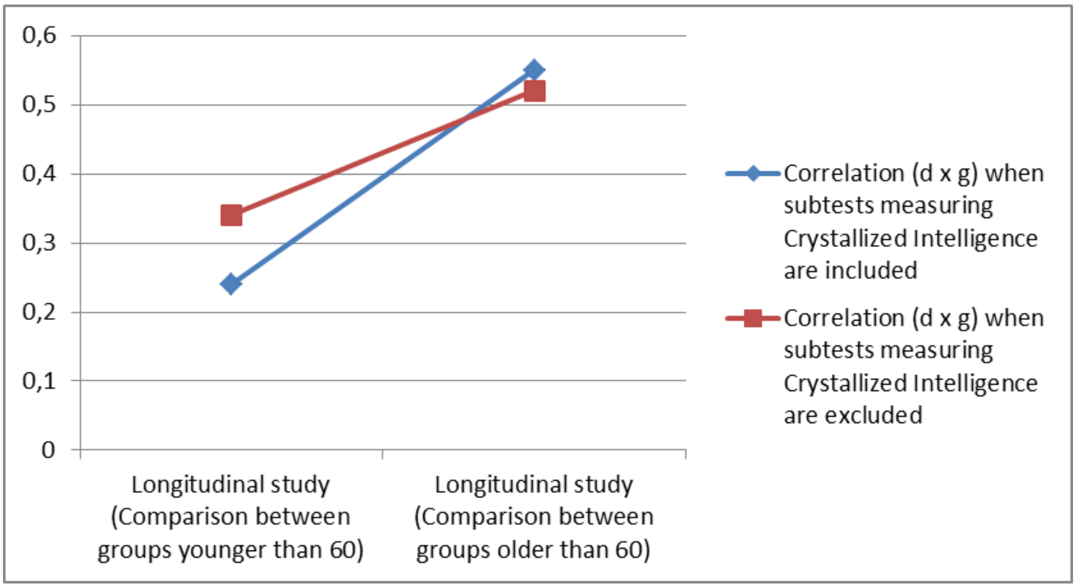


Diagram 11. Correlations *d* x *g* between groups younger than 60 and correlations *d* x *g* between groups older than 60.

correlation also shows a dramatic increase to .52 when computed for comparisons between groups older than 60. The effect of age range on the true correlation rho is shown in Diagram 11.

Age cut-off point as a moderator. A close inspection of the data of the longitudinal study revealed that the correlations *d* x *g* become higher when the age of the second group increased. To test whether the age cut-off point moderates the true correlation rho, eight cut-off points were selected. An age cut-off point is always one year above the age of the age groups reported in the study; except for the oldest age group, of course. The results of this analysis are shown in Table 52 and Diagram 12. Table 52 shows the number of correlation coefficients (*K*), total sample size (*N*), the true correlation (rho) and their standard deviation (SD_r). The last column presents the percentage of variance explained by sampling errors (%VE). Clearly, the true correlation rho increases when the age cut-off point increases. This is true for comparisons between age groups when tests of Crystallized Intelligence are included, as well as when they are excluded.

Conclusion

The goal of this study was to explore the correlation between the declines in IQ scores due to aging and general intelligence. Since score decline in old age on tests of Crystallized Intelligence is a lot lower than score decline on tests of Fluid Intelligence, we expected that the correlation *d* x *g* is not strongly positive. Based on two studies with a total sample size of 3,527 we obtained a meta-analytic correlation *d* x *g* of rho = .45 between groups younger than 68 and groups older than 68, when subtests that measure Crystallized Intelligence are included. When these subtests are removed from the analysis the true correlation rho increases to .59. These results indicate a modest positive relationship between decline in IQ scores due to aging and general intelligence, which becomes stronger when subtests of Crystallized Intelligence are excluded. A further analysis of moderator variables in the longitudinal study showed additional support for this conclusion. First, correlations

Table 52

Comparison Between Age Groups Using Different Cut-off Points in the Longitudinal Study

variable	*K*	*N*	rho	SD_{rho}	%VE
Cut-off Age 30 (including Crystallized subtests)	8	1,310	.28	.09	40.88
Cut-off Age 30 (excluding Crystallized subtests)	8	1,310	.36	.22	9.00
Cut-off Age 40 (including Crystallized subtests)	14	2,158	.34	.06	60.57
Cut-off Age 40 (excluding Crystallized subtests)	14	2,158	.48	.05	62.76
Cut-off Age 47 (including Crystallized subtests)	18	2832	.35	.06	54.76
Cut-off Age 47 (excluding Crystallized subtests)	18	2832	.44	.09	34.17
Cut-off Age 54 (including Crystallized subtests)	20	3,238	.35	.1	33.38
Cut-off Age 54 (excluding Crystallized subtests)	20	3,238	.45	.08	38.28
Cut-off Age 61 (including Crystallized subtests)	16	2,451	.39	.07	49.82
Cut-off Age 61 (excluding Crystallized subtests)	16	2,451	.46	.08	36.21
Cut-off Age 68 (including Crystallized subtests)	18	2,677	.48	.10	27.57
Cut-off Age 68 (excluding Crystallized subtests)	18	2,677	.53	.09	30.23
Cut-off Age 75 (including Crystallized subtests)	14	1,606	.61	.10	24.77
Cut-off Age 75 (excluding Crystallized subtests)	14	1,606	.64	.07	38.12
Cut-off Age 82 (including Crystallized subtests)	8	538	.60	0	128.77
Cut-off Age 82 (excluding Crystallized subtests)	8	538	.54	0	105.1

Note. *K* = number of correlations; *N* = total sample size; rho = true correlation (observed correlation corrected for sample size); SD_r = standard deviation of true correlation; %VE = percentage of variance accounted for by sampling errors.[d]

[d] Format of table adapted from te Nijenhuis & Franssen (2010). Verbatim use of formulation of table descriptions from te Nijenhuis & Franssen (2010)

d x *g* between groups older than 60 were substantially larger than correlations between groups younger than 60. Second, correlations *d* x *g* increase when the cut-off point for age comparisons increases. In conclusion, the correlation between IQ decline due to aging and general intelligence is not strongly positive. This is in line with our expectations. However, the correlation *d* x *g* becomes substantially high when test of Crystallized Intelligence are excluded.

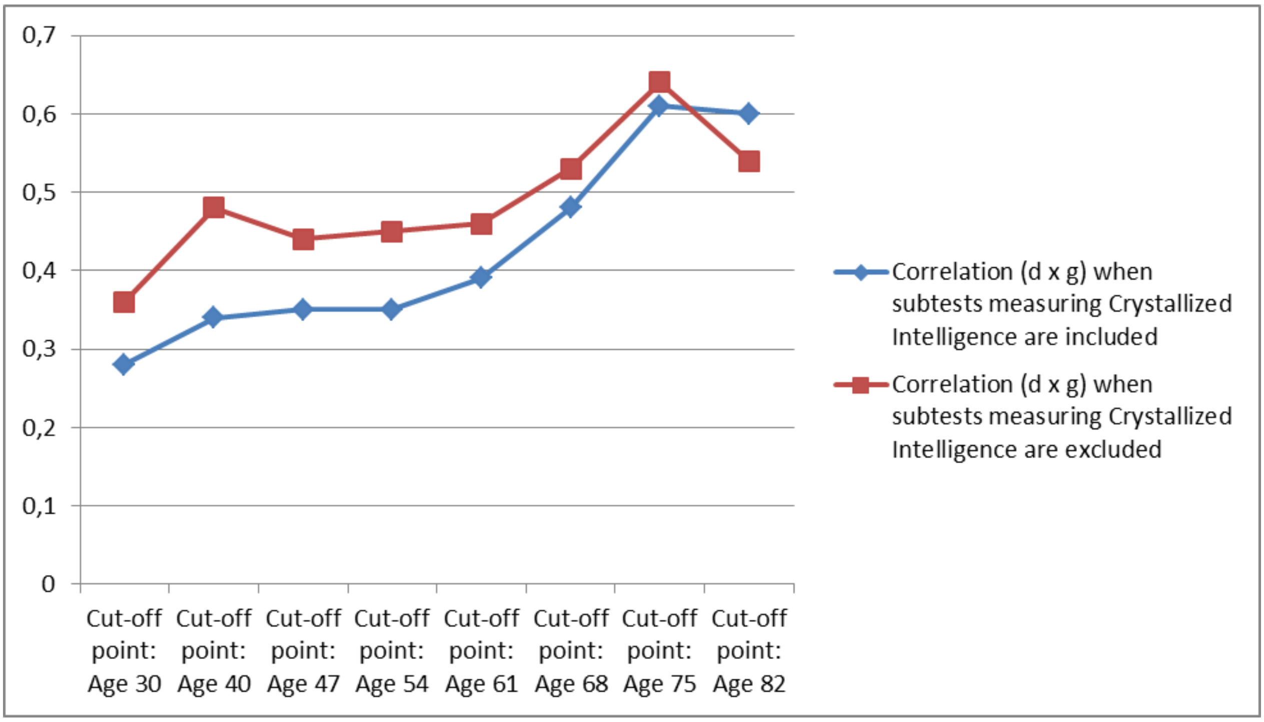


Diagram 12. Correlation *d* x *g* in the longitudinal study between age groups when different age cut-off points are used.

Study 6b: Autism

To test whether there is a strong positive correlation or a negligible correlation between the magnitude of *g* loadings and difference scores on IQ subtests between an autistic and a comparison group, an exploratory psychometric meta-analysis was performed on a number of studies that reported IQ scores of at least seven subtests from subjects with autism. If we do not find a strongly positive correlation, we will also explore whether differences lie on broad or narrow cognitive abilities. Autistic children tend to show less deviation on tests of perceptual motor organization skills and rote memory skills, but more deviation on tests of verbal abstraction and verbal comprehension (Lincoln, et al., 1988). In almost all studies we analyzed cognitive abilities were tested with a Wechsler test. In Wechsler tests Perceptual motor organization is administered with the tests: *Picture Completion*, *Picture Arrangement*, *Block Design*, and *Object Assembly*. Rote memory skills are measured with the test *Digit Span*. Verbal abstraction and verbal comprehension are measured with the tests *Vocabulary*, *Similarities*, *Information*, *Arithmetic* and *Comprehension*. If the above statement is true, we would expect less deviation on the subtests *Picture Completion*, *Picture Arrangement*, *Block Design*, *Object Assembly*, and *Digit Span* and more deviation on the subtests *Vocabulary*, *Similarities*, *Information*, *Arithmetic*, and *Comprehension*.

Method[b]

Searching and screening studies.[28] The starting point for this meta-analysis is the study by Siegel, Minshew, and Goldstein (1996). The reference list of this article was checked for studies that contained IQ scores of autistic individuals. Next, a fourfold search strategy was employed to identify additional studies containing IQ scores of autistic individuals. First, an electronic search for published research using PsycINFO, ERIC, MEDLINE, PiCarta, Academic search premier, Web of science, Google Scholar, and PubMed was conducted. Keywords used were autism*, autist*, and asperger*, combined with the words: cognitive, mental ability, intelligence, IQ, WISC, Wechsler, and combinations of these concepts (* is a truncation symbol to represent multiple spellings or endings; AND is a Boolean operator that combines search terms so that the search result contains all of the terms). Second, the reference lists of significant articles were analyzed in search of additional studies. Third, the Journal of Autism and Developmental Disorders 1971 – 2010 was searched for additional studies. Last, cited reference searches were conducted using Web of Science to identify the newest articles, citing already included key studies. (te Nijenhuis & Franssen, 2010)

Specific criteria for inclusion. Only empirical studies reporting IQ scores of an autistic group were included.

Computation of score differences between a group of autistic children and a comparison group. Score differences were computed between an autistic group and a comparison group; the comparison group could be 1) a standardization sample or 2) a control group. Score differences between an autistic group and the comparison group (*d*) were computed by subtracting the mean score of the autistic group from the mean score of the comparison group, and then dividing the result by the *SD* of the standardization sample.

Results[c]

The results of the studies on the correlation between *g* loadings and the score differences

---

[b] Verbatim use of formulation of method from te Nijenhuis & Franssen (2010)

[28] Paragraph taken and adapted from te Nijenhuis & Franssen (2010, p. 22)

between autistic individuals and non-autistic individuals (*d*) are shown in Table 53. The Table gives data derived from 12 studies, with participants numbering a total of 249. It also lists the reference for the study, the cognitive ability test used, the correlation between *g* loadings and *d*, and the sample size. The correlations range from substantially positive to substantially negative. Table 54 presents the results of the bare-bones meta-analysis of 12 data points. It shows the number of correlation coefficients (*K*), total sample size (*N*), the true correlation (rho) and their standard deviation (SD_r). The last column presents the percentage of variance explained by artifactual errors (%VE). The analysis of all data points yields an estimated correlation (rho) of .00, with 42.88% of the variance in the observed correlations explained by sampling errors. However, it is clear that the studies of Allen, Lincoln, and Kaufman (1991b), and Bartak, Rutter, and Cox (1975) are extreme outliers: taking the reduced sample of 10 studies, the values of r = -.67 and .52 are more than three *SD* below the average sample-sized weighted correlation of .04. Taking out these extreme outliers increased the percentage of variance to a value of 58.94%.

Table 53

Studies of Correlations Between g Loadings and Score Differences Between Autistic and Control Groups

Reference	test	*r*	*N*
Shah & Frith (1993)	WISC-R & WAIS-R	.38	18
Allen, Lincoln, & Kaufman (1991a)	WISC-R	.08	20
Allen, Lincoln, & Kaufman (1991b)	K-ABC	-.67	20
Szatmari, Tuff, Finlayson, & Bartolucci (1990)	WISC-R & WAIS-R	-.15	17
Lincoln, Courchesne, Kilman, Elmasian, & Allen (1988)	WISC-R & WAIS-R	.28	33
Lincoln, Courchesne, Kilman, Elmasian, & Allen (1988)	WISC-R	.17	13
Ohta (1987)	WISC Japan	-.34	16
Bartak, Rutter, & Cox (1975)	WISC	.52	9
Siegel, Minshew, & Goldstein (1996)	WISC-R	-.30	45
Siegel, Minshew, & Goldstein (1996)	WAIS-R	-.12	36
Happé (1994)	WISC-R & WAIS-R	.17	21
Happé (1994)	WISC-R & WAIS-R	.41	30

Note. *N* = sample size; *r* = correlation *d* x *g*.[d]

[c] Verbatim use of formulation of reporting of results from te Nijenhuis & Franssen (2010)
[d] Format of table adapted from te Nijenhuis & Franssen (2010). Verbatim use of formulation of table descriptions from te Nijenhuis & Franssen (2010)

Table 54

Exploratory Bare-Bones Meta-Analytical Results for Correlations Between g Loadings and Autistic/ Non-Autistic Subjects Score Differences

variable	*K*	*N*	rho	SD_{rho}	%VE
Autism[1]	12	278	.00	.25	42.88%
Autism (without extreme outliers)	10	249	.04	.17	58.94%

Note. [1]Bare-bones meta-analytical results: Score differences between an autistic group, non-autistic 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.[d]

Analysis of broad cognitive abilities. The bare-bones meta-analysis of the correlation *d* x *g* between autistic groups and control groups did yield a meta-analytic correlation of rho = .04. This implies that differences in IQ profile between autistic groups and control groups are completely unrelated to general intelligence. To test whether differences in IQ profile lie on lower levels of the intelligence hierarchy, we also conducted an analysis of difference scores on subtest level. We conducted a bare-bones meta-analysis of *d* scores for each Wechsler subtest reported in the studies. In all studies that used a Wechsler test the same subtests were reported. Table 55 presents the results of the bare-bones meta-analyses of 10 data points for 11 different Wechsler tests. It shows the number of *d* scores (*K*), total sample size (*N*), the true effect size (d_t) and their standard deviation (SD_d). The last column presents the percentage of variance explained by sampling errors (%VE). According to previous research (Lincoln, et al., 1988), we should find higher differences on the tests *Vocabulary*, *Similarities*, *Information*, *Arithmetic*, and *Comprehension*, and smaller differences on the tests *Picture Completion*, *Picture Arrangement*, *Block Design*, *Object Assembly*, and *Digit Span*. In general, these expectations are strongly confirmed. The results are visualized in Diagram 13.

[d] Format of table adapted from te Nijenhuis & Franssen (2010). Verbatim use of formulation of table descriptions from te Nijenhuis & Franssen (2010)

Table 55

Exploratory Bare-Bones Meta-Analytical Results for d Scores Between Autistic and Control Groups on Wechsler Subtests

Subtest	*K*	*N*	d_t	SD_d	%VE
Information	10	249	1.03	.78	23.41
Similarities	10	249	1.06	.54	39.53
Arithmetic	10	249	1.18	.64	32.13
Vocabulary	10	249	1.41	.79	24.63
Comprehension	10	249	1.67	.65	34.29
Digit Span	10	249	.69	.36	57.60
Picture Completion	10	249	.90	.34	60.67
Picture Arrangement	10	249	1.06	.52	40.78
Block Design	10	249	.01	.38	53.95
Object Assembly	10	249	.31	.27	69.43
Coding	10	249	1.34	.30	69.07

Note. Bare-bones meta-analytical results: Score differences between autistic groups and control groups on subtests of Wechsler scales *K* = number of *d* scores; *N* = total sample size; d_t = true effect size corrected for sample size; SD_d = standard deviation of true effect size; %VE = percentage of variance accounted for by sampling errors.[d]

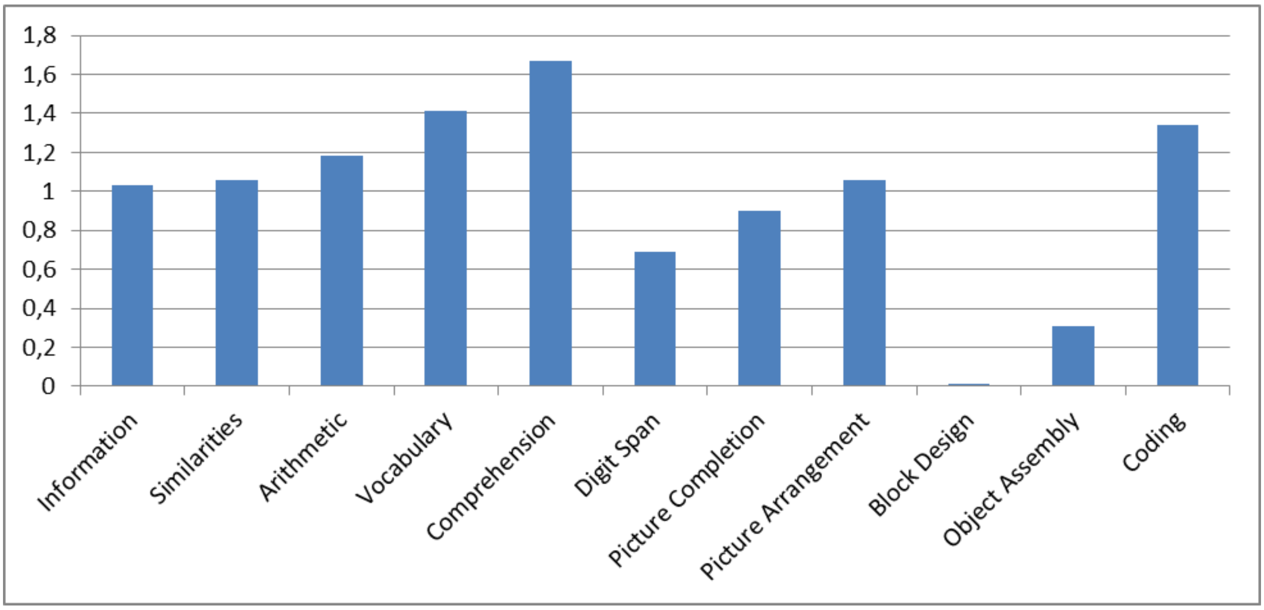


Diagram 13. Effect sizes for differences on subtests of Wechsler scales between autistic and control groups

[d] Format of table adapted from te Nijenhuis & Franssen (2010). Verbatim use of formulation of table descriptions from te Nijenhuis & Franssen (2010)

Conclusion

The goal of this study was to explore whether differences in cognitive profile between autistic groups and control groups are related to general intelligence. Our analysis revealed a meta-analytic correlation *d* x *g* of rho = .04. Therefore, we have to conclude that IQ differences between autistic groups and control groups are completely unrelated to general intelligence. A further analysis of differences on subtest level revealed that there are only minor differences between autistic and control groups on the tests Block Design and Object Assembly, which measure perceptual motor organization, and major differences on tests that measure verbal abstraction and verbal comprehension. We therefore conclude that differences in IQ profile between autistic and control groups are not related to general intelligence, but are due to differences in in verbal comprehension and verbal abstraction.

General Conclusion and Discussion

The huge IQ gap between non-Western immigrants and ethnic Dutch has emerged as one of the primary explanations for the large differences in school and work achievement between these groups (te Nijenhuis, et al., 2004). Is there a genetic component in the IQ gap between these immigrants and ethnic Dutch? Previous meta-analyses have shown that the group differences on IQ subtests correlate almost perfectly with the cognitive complexity of these subtests; moreover, the cognitive complexity correlates perfectly with heritability and moderately to strongly with physical characteristics of the brain.

The purpose of the present study was to test a fundamental hypothesis: Only variables under genetic influence are strongly and positively related to general intelligence, or *g*, and variables not under genetic influence are not. We conducted analyses on a manifold of variables to test this hypothesis (see Table 66). The hypothesis was strongly supported: variables under genetic influence, namely heritabilities and most group differences, showed moderate to strong, positive correlations with *g*; the meta-analytical correlation of brain volume with *g* was weak to modest. All other phenomena showed no strong positive correlation with *g*.

It is concluded that group differences are moderately to strongly positively related to general intelligence, heritabilities are moderately to strongly positively related to general intelligence, and a physical characteristic of the brain, namely brain volume, is weakly to moderately positively related to general intelligence. Biological-environmental factors show weakly negative to weakly positive correlations. These findings in combination suggest a strong genetic component in group differences with regard to general intelligence, and that biological-environmental variables presumably do not affect these differences. Therefore, group differences in general intelligence should been seen as rather stable over time. Previous research on Spearman's hypothesis showed that IQ differences between Whites and Blacks and IQ differences between non-Western immigrants and ethnic Dutch (te Nijenhuis & Dragt, 2010; te Nijenhuis & Repko, 2011; te Nijenhuis & Willigers, 2011) are primarily caused by differences in general intelligence. Concerning the role of IQ differences in debate on the integration of non-Western immigrants into the Dutch society, we can conclude, that first, there are large differences in IQ between the second immigrant generation and ethnic Dutch (te Nijenhuis, et al., 2004). Second, we showed that group differences in IQ are related to general intelligence, which is in line with the results of previous research (te Nijenhuis & Dragt, 2010; te Nijenhuis & Repko, 2011; te Nijenhuis & Willigers, 2011). Third, general intelligence is largely heritable (te Nijenhuis & Jongeneel-Grimen, 2007), and in the present study, we also showed that general intelligence as it is reflected in reaction time is heritable. Finally, we showed that environmental variables do only weakly affect general intelligence. In sum, differences in IQ between non-Western immigrants and ethnic Dutch are primarily related to general intelligence, which is under strong genetic influence, but only weakly affected by biological-environmental factors. Therefore, further IQ gains of the second generation of immigrants, should only constitute non-*g* related gains in IQ, which should leave the gap in school and work achievement unaffected. In consequence, I/O psychologists should find ways to deal with differences in general intelligence instead of ways of trying to change them.

An analysis of differences in reaction time between Whites and lower-IQ groups, and Whites and higher-IQ groups revealed that differences in reaction time between these groups are only weakly to moderately related to *g*. This finding could be due to different reasons. First, reaction time is only an indirect measure of general intelligence, and previously established moderate to substantial positive correlations between general intelligence and reaction time do not necessarily imply that also group differences in reaction time are moderately to strongly related to general intelligence. In addition, the reliability of the reaction time measures showed serious flaws, which works against finding strong results. An analysis of a study on differences in IQ profile between German and immigrant children showed a moderate to strong positive correlation *d* x *g*. Similarly, analyses of studies on differences in cognitive profile between Jews and non-Jewish Whites, and European Jews and Oriental Jews indicated a strong positive correlation *d* x *g*, too. However, an analysis on differences in IQ between Jews and Arabs residing in Israel showed a negative correlation *d* x *g*. In conclusion, we find modest support for the hypothesis that group differences in IQ are strongly related to *g* on reaction time measures, strong support that three other comparisons

Table 66

Results of Individual Studies on Group Differences, Subgroup Differences, Heritability, Physical Characteristics of the Brain, Biological Environmental Factors, Aging, and Autism

variable	*K*	*N*	*r*	*rho*	*rho-5*	%VE
Group differences						
Differences in reaction time between Whites and lower-IQ groups	4	762		.18	.37	1.07
Differences in reaction time between Whites and higher-IQ groups	5	1094		.20	.48	2.73
Differences between Germans and immigrants	1	218	.68			
Differences between Jews and non-Jewish Whites	2	128		.80		82.29
Differences between European Jews and Oriental Jews	4	870		.70		40.81
Differences between European Jews and Arabs	5	1443		-.25		216.96
Subgroup differences						
Explorative comparison of school types	31	8617		.29		5.85
Explorative comparisons of religious groups	6	1913		-.21		1.31
Heritability						
h^2 reaction times	2	389		.51		65.22
Physical characteristics of the brain						
Brain volume	3	156		.35		34.21
Biological-environmental factors						
Iodine deficiency/ supplementation	6	196		.01		51.09
Prenatal cocaine exposure	2	215		-.23		16.98
Fetal alcohol syndrome	3	125		.12		83.04
Air pollution	1	55	-.17			
Traumatic brain injury	14	629		-.07		35.43
Malnutrition	1	51	-.15			
Aging and autism						
Aging	23	3527		.45		27.57
Autism	10	249		.04		58.94

Note. *K* = number of correlations; *N* = total sample size; *r* = correlation *d* x *g*; *rho* = true correlation *d* x *g* (observed correlation corrected for sample size); *rho-5* = true correlation *d* x *g* (observed correlation corrected for five statistical artefacts) ; %VE = percentage of variance accounted for by artefactual error.[d]

between ethnic groups are strongly on the *g* factor, and the comparison involving Arabs in Israel shows the first recorded exploratory meta-analytic finding of a negative link between group

[d] Table adapted from te Nijenhuis & Franssen (2010). Verbatim use of table descriptions from te Nijenhuis & Franssen (2010)

differences. It should also be noted that immigrants represent a sub-population of an ethnic groups, in the sense, that migrants from a certain country, do not necessarily represent a random sample from this country's population. It could even be assumed that lower cognitive abilities of migrants compared to the average population of the home country, were a reason for low achievement in the home country, which led migrants to leave the home country. Although this is only speculation, the point is that migrants' cognitive abilities might be structurally lower than the cognitive abilities of the average home-country population. Therefore, it should be emphasized that the comparison between populations on country level are not necessarily comparable to comparisons between whole country populations and migrant groups. Still, the comparison of country level differences between Jews and non-Jewish Whites, the comparison of within country differences between Jews of European ancestry and Jews of Oriental ancestry, and the comparison of German and migrant children lead to similar results: (Ethnic) group differences in IQ are strongly and positively related to general intelligence. Only the group differences between Jews and Arabs residing in Israel could not be explained with differences in general intelligence.

The previous group comparisons concerned ethnic groups, and additionally we explored differences between subgroups within an ethnic group. The analyses on subgroups that differ with regard to religious belief and the school type the subgroups attend, respectively, did not yield strong positive correlations, as is generally the case in comparisons between ethnic groups. In particular, differences in IQ between religious groups, namely Catholics, Protestants, and atheists, showed a correlation d x g of close to zero. Differences in IQ between school types showed a small correlation d x g. Therefore, subgroup differences do not seem to have a particularly strong relationship with g. These findings can be construed as providing support for the hypothesis that when comparing samples only group differences and generally not subgroup differences are strongly and positively related to g. However, it should be mentioned that te Nijenhuis et al. (2009) reported a rho = +1 for differences between gifted persons and average persons. Clearly, more exploratory meta-analyses are required to see which subgroup differences act like ethnic group differences.

In previous studies, genetic variables were found to have a strong positive relationship with g. Therefore, we expected that also the heritability coefficients of reaction time measures would show a strong correlation with g loadings. A bare-bones meta-analysis on two studies revealed a correlation h^2 x g of .51. This finding provides modest support for our hypothesis.

A range of physical characteristics of the brain was found to have a substantial correlation with g. In the present study, we explored the relationship between brain volume and g. Results indicate a modest correlation *d* x *g*. Spitz (1987) hypothesized that biological-environmental variables mimic the pattern of genetic variables. Previous studies on this topic, however, did not indicate a pronounced relationship between the biological-environmental cocaine-, lead-, and smoke-exposure. Nonetheless, it was still possible to find support for Spitz' hypothesis in studies of other biological-environmental variables. Therefore, we explored the correlation *d* x *g* of the variables iodine supplementation/ deficiency, prenatal cocaine exposure, fetal alcohol syndrome, air pollution, traumatic brain injury, and malnutrition. The picture that emerged from all this studies is straightforward: Differences in IQ caused by these biological-environmental variables are virtually unrelated to general intelligence.

A further exploration of the psychological phenomena aging and autism revealed that IQ decline in aging has a substantial relationship with general intelligence, but the meta-analytical correlation is clearly not +1. The analyses showed that an analysis of gains on broad abilities is more likely to be helpful for understanding this phenomenon than the *g* factor. The IQ profile of autistic groups clearly does not correlate with general intelligence.

Limitations of the Studies

In general, many studies were based on a relatively low *N* (see Table 66). Clearly, more studies are needed to confirm our results. It should also be taken into consideration that the method of correlated vectors is a unidimensional approach to the nature-nurture debate in intelligence. The effect of other possible factors that might influence intelligence like SES cannot be determined by this method. The least we can conclude from the results of our analysis is that if only variables under genetic influence show a strong correlation with *g*, and non-genetic variables do not, the effect of non-genetic variables on *g* will be rather limited.

The meta-analysis on group differences in reaction time had several limitations. First, the tasks used in Jensen (1993), Jensen and Whang (1994), Ja-Song and Lynn (1992), Lynn and Holmshaw (1990), Lynn and Shegisa (1991), and Lynn, Chan, and Eysenck (1991) yielded two types of reaction time, namely reaction time (RT) and movement time (MT). The former measures the time interval between onset of stimulus and release of the home button and the latter measures the time interval between release of the home button and activation of the response button. A close inspection of reaction time differences between groups indicated that when one group scores better on the first measure it has as strong tendency to score worse on the second measure, and vice versa. Since the same pattern emerged in the reaction time and movement time measures of all three tasks, it is likely that the observed differences in reaction time are due to a response tendency, which makes the tasks less suitable as a measure of reaction time to be analyzed by the method of correlated vectors. A possible solution to this problem would be to add reaction and movement time for each group and to correlate differences between groups on these sum measures with appropriate *g* loadings. Although the resulting *d* vectors would have only six values instead of 12, confounds due to response tendency on reaction time measures would be removed. Second, we could not base the computation of *g* loadings on the correlation matrix of subtests scores. As a substitute procedure, we used the reaction time subtest's correlation with the SPM to estimate the subtests' loadings on *g*. Although, from a theoretical perspective, this procedure should yield a good estimate of *g* loadings, some of the subtests showed negative correlations with the SPM. Due to a relatively low sample size, this result is not surprising. Still, a negative *g* loading does not exist, and using a negative *g* loading in the computation of the correlation *d* x *g*, would have made the comparison useless. We decided to reverse the sign of the correlation and use the resulting positive correlation as the best estimate of the *g* loading. We acknowledge that this procedure is not optimal; still, it posed the best possibility to conduct the analysis under these circumstances. Better estimates of *g* loadings in future analyses on differences in reaction time are desirable.

Practical Implications

The integration of non-Western immigrants into the Dutch labor market is desired by all parties involved, but it is also a very challenging endeavor. Problems of integration have been linked to differences in culture between the country of origin and the host country, and to differences between the socio-economic status (SES) of the first immigrant generations and the average SES of the host country population. It has often been assumed that when the cultural and the socio-economic gap will disappear, the integration of non-Western immigrants into the Dutch labor market should not be a problem anymore. However, this assumption only holds if non-Western immigrants are not structurally different on any other factor that drives the success of the integration into the Dutch labor market. In the present study, we addressed such a factor, namely the huge IQ gap between the Western population and non-Western immigrants. Many studies have shown that job success is strongly related to IQ. Therefore, it is reasonable to assume that the IQ gap needs to be reduced to give non-Western immigrants the same job opportunities as the Western population. Results of our study indicate that IQ differences between ethnic groups are related to general intelligence, or *g*, and therefore are stable over time. Similarly, differences in reaction time between ethnic groups, and heritability coefficients of reaction time measures also are found to be strongly related to *g*. A second line of studies explored the relationship between IQ differences that are not related to group membership, but to biological-environmental causes, aging, and autism. The results of all this studies indicate that differences in IQ due to biological-environmental causes, aging, or autism are not related to general intelligence.

In sum, differences in IQ due to group membership are strongly related to *g*, and IQ differences due to any other causes are not. Therefore, it is reasonable to assume that there is a structural factor, namely general intelligence, on which non-Western immigrants and the Western population differ, and that this factor will stay quite stable over time. The implication of this is that closing the cultural and the socio-economic gap might only lead to improvements of integration up to a certain point, but that another crucial driver of integration into the labor market is unlikely to change. Even worse, since socio-economic status is not only an antecedent of integration of the labor market, but also a consequence of it, a closing of the socio-economic gap might be impeded by the differences in general intelligence between non-Western immigrants and the Western population. In conclusion, to integrate non-Western immigrants into the Dutch labor market, realistic plans have to be drawn, which take the differences in general intelligence into account.

Suggestions for Additional Research

The present study explored the relationship between IQ differences due to a vast range of different variables and general intelligence. Future research needs to replicate and extend our findings. Concerning group differences future research should explore whether differences between all kinds of ethnic groups are related to *g* or whether these effects are specific to the groups studied in this research. Also, it is not clear to which extent differences in IQ due to membership of other kinds of groups within a specific ethnic group are related to *g*.

Although the present study offers evidence for a genetic perspective of general intelligence, we need to take into consideration that the correlations between genetic variables and *g* are moderate to strong, but they are certainly not perfect. Therefore, methods need to be developed that determine the relative contribution of other factors like SES, when relating IQ differences to general intelligence.

In sum, we conclude that IQ differences between non-Western immigrants and the Dutch population are quite stable over time, and that practices to reduce these differences are well intended, but, after all, might not lead to a large increase in general intelligence. According to the results of this thesis, a more fertile approach would be to accept differences in IQ and try to deal with them.

Literature

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

Agresti, A. & Franklin, C.A.(2007). *Statistics: The art and science of learning from data*. New Jersey: Pearson Prentice Hall.

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

*Allen, M. H., Lincoln, A. J., & Kaufman, A. S. (1991). Sequential and simultaneous processing abilities of high-functioning autistic and language-impaired children. *Journal of Autism and Developmental Disorders, 21,* 483-502.

*Ardila, A. (2007). Normal aging increases cognitive heterogeneity: Analysis of dispersion in WAIS- III scores across aging. *Archives of Clinical Neuropsychology, 22,* 1003-1011.

Arendt, R.E., Short, E.J., Singer, L.T., Minnes, S., Hewitt, J., Flynn, S., Carlson, L., Min, M.O., Klein, N., & Flannery, D. (2004). Children prenatally exposed to cocaine: Developmental outcomes and environmental risks at seven years of age. *Developmental and Behavioral Pediatrics, 25,* 83-90.

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

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

*Backman, M.E. (1972). Patterns of mental abilities: Ethnic, socioeconomic, and sex differences. *American Educational Research Journal, 9,* 1-12.

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

*Bartak, L., Rutter, M., & Cox, A. (1975). A comparative study of infantile autism and specific developmental receptive language disorders. *British Journal of Psychiatry, 126,* 127-145.

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

Beeghly, M., Martin, B., Rose-Jacobs, R., Cabral, H., Heeren, T., Augustyn, M., Bellinger, D., & Frank, D.A. (2006). Prenatal cocaine exposure and children´s language functioning at 6 and 9.5 years: Moderating effects of child age, birth weight, and gender. *Journal of Pediatric*

Psychology, 31, 98-115.

Bennett, D.S., Bendersky, M., & Lewis, M. (2002). Children´s intellectual and emotional-behavioral adjustment at 4 years as a function of cocaine exposure, maternal characteristics, and environmental risk. *Developmental Psychology, 38,* 648-658.

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

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

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

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

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

Bouchard, T.J., & McGue, M. (1981). Familial studies of intelligence: A review. *Science, 212,* 1055-1059.

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

*Calderon-Garciduenas, L., Mora-Tiscareno, A., Ontiveros, E., Gomez-Garza, G., Barragan-Meija, G., Broadway, J., Chapman, S., Valenncia-Salazar, G., Jewells, V., Maronpot, R.R., Henriquez-Roldan, C., Perez-Guille, B., Torres-Jardon, R., Herrit, L., Brooks, D., Osnaya-Brizuela, N., Monroy, M.E., Gonzalez-Maciel, A., Reynoso-Robles, R., Villarreal-Calderon, R., Solt, A.C., & Engle, R.W. (2008). Air pollution, cognitive deficits and brain abnormalities: A pilot study with children and dogs. *Brain and Cognition, 68,* 117-127.

Capron, C., & Duyme, M. (1989). Assessment of effects of socio-economic status on IQ in a full cross-fostering study. *Nature, 340,* 552-554.

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

Carroll, J.B. (1997). The three-stratum theory of cognitive abilities. In D. P. Flanagan, J. L. Genshaft, et al.(Eds.) *Contemporary intellectual assessment: Theories, tests, and issues* (pp.122-130) New York: Guilford.

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

Cattell, R. B.(1963). Theory of fluid and crystallized intelligence: A critical experiment. *Journal of Educational Psychology, 54*, 1-22.

Chasnoff, I.J., Anson, A., Hatcher, R., Stenson, H., Iaukea, K., & Randolph, L.A. (1996). Prenatal exposure to cocaine and other drugs: Outcome at four to six years. *Annals of the New York Academy of Sciences, 846,* 314-328.

Cochran, G., Hardy, J., & Harpending, H. (2006). The natural history of Ashkenazi intelligence. *Journal of Biosocial Science, 38*, 659-693.

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

*Daseking, M., Lipsius, M., Petermann, F., & Waldmann, H.-C. (2008). Differenzen im intelligenzprofil bei kindern mit migrationshintergrund: Befunde zum HAWIK-4 [Differences in the intelligence profile of migrant children: Results of the HAWIK-4]. *Kindheit und Entwicklung, 17,* 76-89.

David, H., & Lynn, R. (2007). Intelligence differences between European and Oriental Jews in Israel. *Journal of biological Science, 39*, 465-473.

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

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

*van Dijk, H., & Tellegen, P. J. (1994). *GIVO: Groninger Intelligentietest voor Voortgezet Onderwijs* [GIVO: Groningen Intelligence Test for Continued Education]. The Netherlands, Lisse: Swets.

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

*Egan, V., Chiswick, A., Santosh, C., Naidu, K., Rimmington, J.E., & Best, J.J.K.(1994). Size isn´t everything: A study of brain volume, intelligence and auditory evoked potentials. *Personality and Individual Differences, 17,* 357-367.

*Ellis, C.E., & Hill, E. (1975). Growth, intelligence, and school performance in children with cystic fibrosis who have had an episode of malnutrition during infancy. *Brief clinical and laboratory observations, 87,* 565-568.

Erickson, F. (1987). Transformation and school success: The politics and culture of educational achievement. *Anthropology & Education, 18,* 335-356.

Feingold, A. (1988). Cognitive gender differences are disappearing. *American Psychologist, 43,* 95-103.

*Evers, A., & Lucassen, W. (1991). *Handleiding DAT '83: Differentiële Aanleg Testserie* [Manual DAT'83: Differential Aptitude Test Series]. Amsterdam: Swets.

*Flashman, L.A., Andreasen, N.C., Flaum, M., & Swayze, V.W. (1998). Intelligence and regional brain volumes in normal controls. *Intelligence, 25,* 149-160.

*Fokkema, S. D., & Dirkzwager, A. (1968). *Handleiding voor de Differentiële-Aanlegtests* [Manual DAT: Differential Aptitude Test Series]. Amsterdam: Swets.

*Fox, C., & Birren, J.E. (1950). The differential decline of subtest scores of the Wechsler-Bellevue intelligence scale in 60-69-year-old individuals. *The Journal of Genetic Psychology, 77,* 313-317.

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

*Garde, E., Mortensen. E., Krabbe, K., Rostrup, E., & Larsson, H.B.W. (2000). Relationship between age-related decline in intelligence and cerebral white-matter hyperintensities in healthy octogenarians: A longitudinal study. *The Lancet, 356*, 628-634.

*Gillberg, I.C., Rastam, M., Wentz, E., & Gillberg, C. (2007). Cognitive and executive functions in anorexia nervosa ten years after onset of eating disorder. *Journal of Clinical and Experimental Neuropsychology, 29,* 170-178.

Gottfredson, L.S. (2002). Why *g* matters: The complexity of everyday life. *Intelligence, 24,* 79-132.

Gottfredson, L. S. (2007). Innovation, fatal accidents, and the evolution of general intelligence. In M. J. Roberts (Ed.), *Integrating the mind: Domain general versus domain specific processes in higher cognition* (pp. 387-425). UK, Hove: Psychology Press.

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

*Hakstian, A.R., & Cattell, R.B. (1976). *Manual for the Comprehensive Ability Battery (CAB).* Champaign, IL: Institute for Personality & Ability Testing.

*Happé, F. G. E. (1994). Wechsler IQ profile and theory of mind in autism: A research note. *Journal of Child Psychology and Psychiatry, 35,* 1461-1471.

Horn, J. L., & Cattell, R. B. (1967). Age differences in fluid and crystallized intelligence. *Acta Psychologica, 26,* 107-129.

Hong, Y.-C., Lee, J.-T., Kim, H., & Kwon, H.J. (2002). Air pollution: A new risk factor in ischemic stroke mortality. *Stroke, 33,* 2165-2169.

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

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

*Ja-Song, M., & Lynn, R. (1992). Reaction times and intelligence in Korean children. *The Journal of Psychology, 126,* 421-428.

Jensen, A.R. (1980). *Bias in mental testing.* New York: Free Press.

*Jensen, A.R., (1993). Spearman's hypothesis tested with chronometric information-processing tasks. *Intelligence, 17*, 47-77.

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 and the design of education. In R. J. Sternberg & W. M. Williams (Eds.), *Intelligence, instruction, and assessment: Theory into practice.* (pp. 111–131). Mahwah, NJ: Erlbaum.

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

*Jensen, A.R., & Whang, P.A. (1993). Reaction times and intelligence: A comparison of Chinese-American and Anglo-American children. *Journal of Biosocial Science, 25,* 397-410.

*Jensen, A.R., & Whang, P.A. (1994). Speed of accessing arithmetic facts in long-term memory: A comparison of Chinese-American and Anglo-American children. *Contemporary Educational Psychology, 19,* 1-12.

*Juan-Espinosa, M., Garcia, L. F., Escorial, S., Rebollo, I., Colom, R., & Abad, F. J. (2002). Age dedifferentiation hypothesis: Evidence from the WAIS III. *Intelligence, 30,* 395-408.

Kaplan, B. J. (1972). Malnutrition and mental deficiency. *Psychological Bulletin, 78,* 321-334.

*Kugelmass, S., Lieblich, A., & Bossik, D. (1974). Patterns of intellectual ability in Jewish and Arab children in Israel. *Journal of Cross-Cultural Psychology, 5* , 184-198.

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

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

*Levinson, B.M. (1959). Traditional Jewish cultural values and performance on the Wechsler test. *Journal of Educational Psychology, 50,* 177-181.

*Levinson, B.M. (1960). A comparative study of the verbal and performance ability of monolingual and bilingual native-born Jewish preschool children of traditional parentage. *The Journal of Genetic Psychology, 97,* 93-112.

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

*Lieblich, A., & Kugelmass, S. (1981). Patterns of intellectual ability of Arab school children in Israel. *Intelligence, 5*, 311-320.

*Lieblich, A., Kugelmass, S., & Ehrlich, C. (1975). Patterns of intellectual ability in Jewish and Arab children in Israel II. *Journal of Cross-Cultural Psychology, 6,* 218-226.

*Lieblich, A., Ninio, A., & Kugelmass, S. (1972). Effects of ethnic origin and parental SES on WPPSI performance of pre-school children in Israel. *Journal of Cross-Cultural Psychology, 3,* 159-168.

*Lincoln, A. J., Courchesne, E., Kilman, B. A., Elmasian, IL, & Allen, M. (1988). A study of intellectual abilities in high-functioning people with autism. *Journal of Autism and Developmental Disorders, 18,* 505-23.

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

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

Lynn, R. (2004). The intelligence of American Jews. *Personality and Individual Differences, 36,* 201-206.

Lynn, R., Harvey, J., & Nyborg, H. (2008). Average intelligence predicts atheism rates across 137 nations. *Intelligence, 37,* 11-15.

Lynn, R., & Vanhanen, T. (2002). *IQ and the wealth of nations.* Westport: Praeger.

Lynn, R., & Vanhanen, T. (2006). *IQ and global inequality.* Augusta: Washington Summit .

*Lynn, R., Chan, J.W.C., & Eysenck, H.J. (1991). Reaction times and intelligence in Chinese and British children. *Perceptual and Motor Skills, 72,* 443-452.

*Lynn, R., & Holmshaw, M. (1990). Black-White differences in reaction times and intelligence. *Social Behavior and Personality, 18,* 299-308.

*Lynn, R., & Shigehisa, T. (1991). Reaction times and intelligence: A comparison of Japanese and British children. *Journal of Biosocial Science, 23*, 409-416.

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

te Nijenhuis, J., & Dragt, J. (2010). *Causes of group differences studied with the method of correlated vectors: A psychometric meta-analysis of Spearman's hypothesis.* Unpublished manuscript, University of Amsterdam, the Netherlands.

te Nijenhuis, J., & Franssen, D. B. (2010). *What is the significance of test-score differences? Five psychometric meta-analyses on g loadings and IQ scores: The relation of inbreeding, visual impairment, schizophrenia, epilepsy, and giftedness with general intelligence*. Unpublished manuscript, University of Amsterdam, the Netherlands.

te Nijenhuis, J., de Jong, M.-J., Evers, A., & van der Flier, H. (2004). Are cognitive differences between immigrant and majority groups diminishing? *European Journal of Personality, 18*, 405-434.

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, the Netherlands.

Te Nijenhuis, J., & Repko, J. (2011). *Spearman's hypothesis tested with Raven's Progressive Matrices: A psychometric meta-analysis*. Unpublished manuscript, University of Amsterdam, the Netherlands.

te Nijenhuis, J., de Pater, I. E., van Bloois, R., & Geutjes, L.L. (2009). *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.* Unpublished manuscript, University of Amsterdam, the Netherlands.

te Nijenhuis, J., & Smit, J. (2010). *Are differences in IQ scores also differences in general intelligence? Four psychometric meta-analyses on the relation of lead level, breastfeeding, and prenatal cocaine and smoke exposure with general intelligence.* Unpublished manuscript, University of Amsterdam, the Netherlands.

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

Maheswaran, R., Pearson, T., Campbell, M.J., Haining, R.P., McLeod, C.W., Smeeton, N., & Wolfe, C.D.A. (2006). A protocol for investigation of the effects of outdoor air pollution on stroke incidence, phenotypes and survival using the South London Stroke Register. *International Journal of Health Geographics, 5,* 1-7.

McDaniel, M.A. (2005). Big-brained people are smarter: A meta-analysis of the relationship between in vivo brain volume and intelligence. *Intelligence, 33*, 337-346.

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

*Ohta, M. (1987). Cognitive disorders of infantile autism: A study employing the WISC, spatial relationship conceptualization, and gesture imitations. *Journal of Autism and Developmental Disorders, 17,* 45-62.

*Petrill, S. A., Thompson, L. A., & Detterman, D. K. (1995). The genetic and environmental variance underlying elementary cognitive tasks. *Behavior Genetics, 25,* 199-209.

Plomin, R. & DeFries, J.C. (1979). Multivariate behavioral genetic analysis of twin data on scholastic abilities. *Behavior Genetics, 9*, 505-517.

Prenzel, M., Artelt, C., Baumert, J., Blum, W., Hammann, M., Klieme, E., & Pekrun, R. (2007). *Pisa 2006. Die Ergebnisse der dritten internationalen Vergleichsstudie. Zusammenfassung* [Pisa 2006. The findings of the third international comparison study. Summary]. Retrieved January 10, 2012, from http://pisa.ipn.unikiel.de/zusammenfassung_PISA2006.pdf.

Ramm, G., Prenzel, M., Heidemeier, H., & Walter, O. (2004). Soziokulturelle Herkunft: Migration. In M. Prenzel, J. Baumert, W. Blum, R. Lehmann, D. Leutner, M. Neubrand, R. Pekrun, H.-G. Rolff, J. Rost & U. Schiefele, *Pisa 2003. Der Bildungsstand der Jugendlichen in Deutschland – Ergebnisse des zweiten internationalen Vergleichs* [The educational achievement of adolescents in Germany – Findings on the second international comparison] (pp. 254-282). Germany, Münster: Waxmann.

Richardson, G.A., Conroy, M.L., & Day, N.L. (1996). Prenatal cocaine exposure: Effects on the development of school-age children. *Neurotoxicology and Teratology, 18,* 627-634.

Rim, Y. (1983). Eye-color, ethnic origin or family size? *Personality and Individual Differences, 4,* 101-102.

Rosenberg, L. E. (1974). Vitamin-responsive inherited diseases affecting the nervous system. In *Brain Dysfunction in Metabolic Disorders*, F. Plum (pp. 263-270). New York: Raven Press.

Rushton, J. P., Čvorović, J., & Bons, T. A. (2007). General mental ability in South Asians: Data from three Roma (Gypsy) communities in Serbia. *Intelligence, 35*, 1-12.

Rutter, M. (1978). Diagnosis and definition of childhood autism. *Journal of Autism and Developmental Disorders, 8,* 139-161.

Salgado, J.F., Anderson, N., Moscoso, S., Bertua, C., de Fruyt, F., & Rolland, J.P. (2003). A meta-analytic study of general mental ability: Validity for different occupations in the European community. *Journal of Applied Psychology, 88,* 1068-1081.

*Schaie, K.W., & Willis, S.L. (1993). Age difference patterns of psychometric intelligence in adulthood: Generalizability within and across ability domains. *Psychology and Aging, 8,* 44-55.

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

*Shah, A., & Frith, U. (1993). Why do autistic individuals show superior performance on the block design task? *Journal of Child Psychology and Psychiatry, 34,* 1351-1364.

*Shresta, R. M. (1994). *Supplementation with iodine and iron improves mental development in Malawian children*. Doctoral dissertation, University of Wageningen, the Netherlands.

Scheffer, P.J. (2011). *Immigrants nations*. Cambridge: Polity Press.

*Siegel, D. J., Minshew, N. J., & Goldstein, G. (1996). Wechsler IQ profiles in diagnosis of high-functioning autism. *Journal of Autism and Developmental Disorders, 26,* 389-406.

*Singer, L.T., Minnes, S., Short, E., Arendt, R., Farkas, K., Lewis, B., Klein, N., Russ, S., Min, M.O., & Kirchner, H.L. (2004). Cognitive outcomes of preschool children with prenatal cocaine exposure. *Journal of the American Medical Association, 291,* 2448-2455.

Schmidt, F.L., & Hunter, J.E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research. *Psychological Bulletin, 124,* 262-274.

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

Schmidt, F.L., & Hunter, J.E. (2004). General mental ability in the world of work: Occupational attainment and job performance. *Journal of Personality and Social Psychology, 86,* 162-173.

Spitz, H.H. (1987). Variations in Wechsler interscale IQ disparities at different levels of IQ, *Intelligence, 13,* 157-167.

Szatmari, P., Jones, M.B., Zwaigenbaum, L., & MacLean, J.E. (1998). Genetics of autism: Overview and new directions. *Journal of Autism and Developmental Disorders, 28,* 351-368.

Szatmari, P., Tuff, L., Finlayson, M. A. J., & Bartolucci, G. (1990). Asperger's syndrome and autism: Neurocognitive aspects. *Journal of the American Academy of Child and Adolescent Psychiatry, 29,* 130-136.

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

*Verhage, F. (1964). Intelligentie en Kerkelijke Gezindte [Intelligence and church denomination]. *Nederlands Tijdschrift voor de Psychologie, 19,* 247-254.

*Vernon, P. A. (1988). The heritability of measures of speed of information-processing. *Personality and Individual Differences, 10,* 573-576.

*Vernon, P.A., & Jensen, A.R. (1984). Individual and group differences in intelligence and speed of information processing. *Personality and Individual Differences, 5,* 441-423.

Weber, M. (1934). *Die protestantische Ethik und der Geist des Kapitalismus* [The protestant ethic and the spirit of capitalism]. Germany, Tübingen: J.C.B. Mohr.

Wechsler, D. (1939). *Wechsler-Bellevue Intelligence Scale.* New York: The Psychological Corporation.

Wechsler, D. (1949). *Wechsler Intelligence Scale for Children*. New York: The Psychological Corporation.

Wechsler, D. (1955). *Manual for the Wechsler Adult Intelligence Scale*. Oxford: Psychological Corporation.

Wechsler D. (1967). *Manual for Wechsler Preschool and Primary Scale of Intelligence*. San Antonio, TX: The Psychological Corporation.

Wechsler, D. (1974). *Manual for the Wechsler Intelligence Scale for Children - Revised*. New York: The Psychological Corporation.

Wechsler, D. (1981). *WAIS-R manual: Wechsler Adult Intelligence Scale-Revised*. New York: Psychological Corporation.

Wechsler, D. (1989). *Wechsler Preschool and Primary Scale of Intelligence-Revised*. San Antonio, TX: Psychological Corporation.

Wechsler, D. (1991). *The Wechsler Intelligence Scale for Children - Third Edition*. San Antonio, TX: The Psychological Corporation.

Wechsler, D. (1997). *Wechsler Adult Intelligence Scale – Third Edition*. San Antonio, TX: The Psychological Corporation.

Wechsler D. (2002). *Wechsler Preschool and Primary Scale of Intelligence – Third Edition*. San Antonio, TX: Psychological Corporation.

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

Wechsler, D. (2008). *Wechsler Adult Intelligence Scale - Fourth Edition*. San Antonio, TX: Pearson Assessment.

White, K.R. (1982). The relationship between socioeconomic status and academic achievement. *Psychological Bulletin, 91,* 461-481.

*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.

*Wickett, J.C., Vernon, P.A., & Lee, D.H. (2000). Relationships between factors of intelligence and brain volume. *Personality and Individual Differences, 29,* 1095-1122.

*Zeidner, M. (1987). Validity of college admission indices for Jews and Arabs in Israel. *Personality and Individual Differences, 8,* 587-589.