The relation between fluid intelligence and the general factor as a function of cultural background: A test of Cattell’s Investment theory

Ann Valentin Kvist *, Jan-Eric Gustafsson

Department of Education, Göteborg University, Sweden

Received 1 March 2007; received in revised form 16 August 2007; accepted 30 August 2007
Available online 30 October 2007

Abstract

According to Cattell’s [Cattell, R.B. (1987). Intelligence: Its structure, growth and action. New York: North-Holland.] Investment theory individual differences in acquisition of knowledge and skills are partly the result of investment of Fluid Intelligence (Gf) in learning situations demanding insights in complex relations. If this theory holds true Gf will be a factor of General Intelligence (g) because it is involved in all domains of learning. The purpose of the current study was to test the Investment theory, through investigating the effects on the relation between Gf and g of differential learning opportunities for different subsets of a population. A second-order model was fitted with confirmatory factor analysis to a battery of 17 tests hypothesized to measure four broad cognitive abilities. The model was estimated for three groups with different learning opportunities (N=2358 Swedes, N=620 European immigrants, N=591 non-European immigrants), as well as for the total group. For this group the g–Gf relationship was .83, while it was close to unity within each of the three subgroups. These results support the Investment theory.

© 2007 Elsevier Inc. All rights reserved.

Keywords: Structure of intelligence; Cattell’s Investment theory; Fluid intelligence; General intelligence

1. Introduction

Ever since Spearman (1904, 1927) introduced his “Theory of Two Factors”, issues concerning the structure of human intelligence have been the focus of attention of much research. While there certainly are differences in opinion regarding a wide range of issues, consensus has been achieved that a hierarchical representation of the structure of cognitive abilities is required to capture the complexities of the phenomenon (e.g., Carroll, 1993; Gustafsson, 1988; Jensen, 1998; Messick, 1992). The currently most widely accepted hierarchical model is the Carroll (1993) “Three-Stratum Model”. Since this model may be regarded as an extension of the Cattell and Horn “Gf–Gc” model (see e.g., Horn & Cattell, 1966) it is also referred to as the Carroll–Horn–Cattell (CHC) model (McGrew, 2005).

The CHC model includes factors of three degrees of generality (Carroll, 1993; McGrew, 2005). At the lowest level (stratum I) there are at least some 60 narrow factors, many of which correspond to factors previously identified by Thurstone (1938), Guilford (1967) and other researchers working in the tradition of multiple

* Corresponding author. Department of Education, Göteborg University, P. O. Box 300, S-405 30 Goteborg, Sweden. Tel.: +46 31 786 22 77.
E-mail address: ann.valentin.kvist@ped.gu.se (A. Valentin Kvist).

0160-2896/$ - see front matter © 2007 Elsevier Inc. All rights reserved.
doi:10.1016/j.intell.2007.08.004
factor analysis. At stratum II some ten broad factors are identified, and among these, a few are seen as especially prominent, primarily because of the attention they have been given in the research conducted by Cattell and Horn (see, e.g., Cattell, 1963, 1971, 1987; Horn, 1968; Horn & Cattell, 1966). One is Fluid Intelligence (Gf), which is interpreted as the capacity to solve novel, complex problems, using operations such as inductive and deductive reasoning, concept formation, and classification. Another factor is Crystallized Intelligence (Gc), which represents individual differences in breadth and depth of knowledge of the language, information and concepts of a culture. Gc is acquired through education and experience and it primarily reflects verbal knowledge and skills, as well as declarative knowledge in wide areas. Another important factor is Broad Visual Perception (Gv), which is an ability to generate, retain, retrieve and transform visual images. Cognitive Processing Speed (Gs) is a broad ability to fluently perform relatively easy or overlearned tasks, particularly when attention and focused concentration is required.

At the third stratum the CHC model includes a factor of General Intelligence (g). This factor relates most highly to complex reasoning tasks while it has lower relations to the stratum II factors involving simpler speeded tasks. According to Carroll (1993) this suggests that the g-factor involves complex higher-order cognitive processes.

However, even though there is consensus at a general level that such a hierarchical arrangement of factors represents a useful taxonomy of human cognitive abilities, there are substantial differences in opinion concerning fundamental theoretical issues between the three researchers after which the CHC model has its name. The most striking locus of differences concerns the need for a g-factor at the apex of the hierarchy, and, if such a factor is accepted, the nature of this factor (cf. McGrew, 2005).

1.1. The general factor in hierarchical models of intelligence

According to Carroll (1993) the empirical evidence strongly supports the existence of a g-factor, and Jensen (1998), along with many others (e.g., Gustafsson, 1988), has arrived at the same conclusion. However, Horn (see, e.g., Horn & Blankson, 2005; Horn & Noll, 1997) has strongly objected to the idea of a general factor, favoring instead a hierarchical model with broad correlated factors at stratum II. The main reason for Horn’s resistance against a stratum III g-factor is that he regards such a factor as a hybrid factor, which is a composite of different stratum II factor. Since the nature of the g-factor is determined by the composition of the test battery, it lacks factorial invariance, and g is therefore not a meaningful scientific concept (Horn & Noll, 1997).

Undheim (1981) and Gustafsson (1984) suggested one approach to solving the problem of the potential non-invariance of g. They argued that the characteristics of the g-factor as described by Spearman (1904, 1927) agree so well with the characteristics of the Gf-factor as described by Horn and Cattell (1966), that g and Gf should be considered to be one and the same factor. The equality of g and Gf also has been demonstrated empirically in a series of studies in which a higher-order g-factor has been shown to have a perfect relationship with the Gf-factor (e.g., Gustafsson, 1984, 1988, 1994, 2002; Undheim, 1981; Undheim & Gustafsson, 1987). Since Gf is identified in an invariant manner, it follows that g too is invariantly defined as an apex factor in the CHC model.

Horn and Blankson (2005, p. 53) rejected this line of reasoning, arguing that Gf does not account for the interrelationships among other variables indicative of intelligence. However, if Gf is equivalent to a stratum III g-factor in the CHC model which accounts for the intercorrelations among the stratum II factors this statement is incorrect. This issue thus could and should be determined on the basis of empirical research.

While the g = Gf relationship has been observed in many other studies as well (e.g., Keith, 2005; Reynolds & Keith, 2007), all attempts at replication have not been successful. Carroll (1993) reanalyzed the Gustafsson (1984) data, and failed to find the perfect relationship between g and Gf. One reason for this may be that Carroll (1993) relied on exploratory factor analysis, and with this technique he failed to identify the inductive factor, which in turn caused him difficulties separating Gf and Gv. However, in another study of the relationship between Gf and g, Carroll (2003) used confirmatory factor analysis (CFA), without quite being able to show the identity. It thus must be concluded that the empirical support for the equivalence between Gf and g is strong, but not unanimous. The results presented by Carroll (1993) show, however, that Gf is the stratum II factor which has the highest loading on the stratum III g-factor.

However, strong opposition also has been voiced against the idea that the g-factor is equivalent with Gf, and instead it has been argued that measures of crystallized abilities are better indicators of g (e.g., Gignac, 2006; Robinson, 1999). One of the bases for this argument is the observation that the verbal subtests (e.g., Vocabulary and Information) in the Wechsler
batteries have the highest loadings on the first principal factor (Gignac, 2006). However, this need not necessarily be because these tests have higher g-loadings. If verbal tests are over-represented in the test battery this may cause the g-factor to be verbally biased, because the g-factor is confounded with a verbal group factor.

Gignac (2006) devised a procedure which aims to control for such bias. In the so called single-trait correlated uniqueness (STCU) CFA procedure a one-factor model is fitted first, and any further common variance is accounted for through allowing pair-wise covariances of the residuals of groups of tests. Applying this procedure to five datasets, Gignac (2006) found that there still remained a difference in g-loadings for verbal and reasoning tasks. He therefore concluded that the crystallized subtests are the best indicators of g, and rejected the Gustafsson (1984) conclusion that Gf equals g, arguing that this result was caused by a methodologically flawed selection of tests.

There are, however, several problems with Gignac’s (2006) analyses and conclusions. As was acknowledged by Gignac (2006, p. 43) the lack of fluid intelligence subtests within the batteries analyzed made this study less than ideally suited to investigate the differences in g-loadings for crystallized and fluent subtests. Ashton and Lee (2006) also showed that the STCU CFA procedure cannot be relied upon to produce unbiased estimates of g-loadings, because alternative models with equally good fit may be fitted to the same data. They furthermore demonstrated that different results were obtained with large test batteries than with small batteries. Applying the STCU CFA procedure to two large batteries they concluded that there was no difference in the size of the g-loadings for verbal and reasoning tasks.

There is yet another problem with the STCU modeling procedure that may cause it to produce biased results. The problem is that this procedure does not take into account the fact that the amount of truly unique variance may be different for different tests, and for different categories of tests. The test uniqueness attenuates the estimated loading of the test on g, but it does not influence the estimated loading of a lower-order factor on a higher-order factor. This circumstance may explain the seeming contradiction that there is a perfect relation between g and Gf, while at the same time the tests measuring Gf do not load more highly on g, or even lower, than they do on tests measuring crystallized abilities.

To summarize the discussion so far it may be concluded that there is considerable evidence in favor of a g-factor as an apex factor of the hierarchical model of the structure of intelligence. The available evidence also provides some, but far from unanimous, support for the idea that the g-factor is equivalent with Gf. Thus, more empirical evidence is needed to settle the issue about the nature of the g-factor. However, in order to arrive at a deeper understanding we also need stronger theory. There is reason, therefore, to go somewhat more deeply into a theoretical model that aims to explain the nature of the g-factor.

1.2. The Investment theory

Cattell is often ascribed the same negative position with regard to the g-factor as that taken by Horn. However, this is incorrect because there is a profound difference between the positions of these two researchers with regard to the meaningfulness of introducing a higher-stratum g-factor. While Horn rejected such a factor as meaningless because it is non-invariant, Cattell (1971, 1987) was not hostile to the idea that there is a higher-order g-factor. Cattell (1987) argued that according to conventional wisdom it would be expected that the stratum II factor Gc should have a loading of unity on the stratum III g-factor. However, four early third-order factor analytic studies involving Gf and Gc along with personality variables demonstrated that Gf had higher loadings on g than had Gc. Cattell interpreted the third-stratum general factor to be the “historical” g factor, and he asked “How is one to explain this tendency of the historical g (i.e., g₁) to load g more than it does g₂?” (p. 138). The answer proposed by Cattell was the Investment theory.

The Investment theory postulates that in the development of the individual there is initially a single, general, relation-perceiving ability which is connected with the maturation of the brain. This ability, which was labeled Gf by Cattell, is thus primarily associated with genetic factors and neurological functioning. It can be applied to any sensory, motor or memory area, and Cattell argued that a child’s rate of learning of different tasks (e.g., spatial, numerical, conceptual) depends on this general ability. In particular the child’s:

... rate of learning in fields demanding insights into complex relations – and these fields include especially the problems of reading, arithmetic, and abstract reasoning with which he struggles at school – will depend appreciably on his level of fluid intelligence (though motivation, goodness of teaching, etc., will still play their part, as with the acquisitions of low relational complexity). (Cattell, 1987, p. 139).

Thus, through practice and experience children develop knowledge and skills and according to the Investment theory these developed abilities (i.e., Gc)
are influenced by Gf and by effort, motivation and interest, and also by previous levels of Gc. The reason why Gf is a general ability is that:

... in all kinds of relation-eduction in new material requiring fluid ability, the child high in one manifestation will be high in another, and from correlations rooted in such observations eventually we obtain the fluid ability factor. But as a result of the fluid ability being invested in all kinds of complex learning situations, correlations among these acquired, crystallized abilities will also be large and positive, and tend to yield a general factor. (Cattell, 1987, p. 139).

The Investment theory thus can provide an explanation for the observation that Gf tends to have a perfect relation with g.

The Investment theory has been much discussed and investigated, but the empirical evidence in support of the theory has largely been missing. For example, the hypothesis derived from the theory that Gf should have higher heritability than Gc has not generally been supported, even though Cattell (1987) reports some studies showing this to be the case. Longitudinal studies investigating cross-lagged effects of Gf on Gc also generally have failed to identify the hypothesized relations (Gustafsson & Undheim, 1992).

It would carry too far to discuss the possible reasons why these different tests of the Investment theory have failed to provide support. It is, however, interesting to note that the connection between Gf and a higher-order g-factor has not been a focus of attention in discussions of the Investment theory after Cattell’s (1971) formulations. One reason for this may be that the link between the effects of Gf on individual differences in acquisition of knowledge and skill in different areas and Gf as a general, higher-order, factor is not immediately apparent. However, this link is easy to understand once it is realized that a higher-order factor exerts an influence on a greater number of manifest variables than does a factor below it. Thus, the g-factor has a wider breadth of influence than other factors of intelligence, but it does not necessarily exert a particularly strong influence on performance on any single task (see Coan, 1964; Gustafsson, 2002; Humphreys, 1962).

Thus, according to Cattell’s line of reasoning, the Gf-factor develops into a general factor because it influences acquisition of knowledge and skills in different domains. For example, most new words are learned by inferring their meanings from the contexts in which the words are embedded (Lohman, 2004). Similarly, Landauer and Dumais (1997) argued that most knowledge development occurs through inductive inference of partial information encountered in different contexts. In support of Cattell’s reasoning, this makes vocabulary tests and other tests of knowledge reflect the efficiency of past reasoning processes. But if different subgroups within a population have had different opportunities to acquire the knowledge tested, for example because the language of the test is the mother tongue for some, and the second or third language for others, the simple relationship between Gf and amount of knowledge acquired will break down. This should apply not only to acquisition of verbal knowledge and skills, but also to development of abilities in other domains, such as the spatial one.

This suggests a way to test both the Investment theory and the hypothesis that g equals Gf, namely through investigating the effect of differential learning opportunities for different subsets of a population on the relation between Gf and g. From the Investment theory follows the prediction that within populations which are homogeneous with respect to learning opportunities there should be a perfect relationship between Gf and g, while for populations which are composed of subgroups who have had different learning opportunities, the relation between Gf and g should be lower. This implies that the validity of the Investment theory may be tested through investigating the strength of the relationship between Gf and g in homogeneous and heterogeneous populations. A similar suggestion was made by Carroll (1996, p. 16).

2. Method

For the purpose of our study we need a fairly large number of subjects with different cultural backgrounds, who have been subjected to the same testing procedure. Such a group was made available to us through the Swedish National Labor Market Board. The agency offers vocational training to persons who are unemployed or at risk of becoming unemployed. In selecting candidates a procedure called “RA”, which is a Swedish acronym for “Directed Aptitude Testing”, is used. The procedure includes the use of traditional psychometric tests, and it has been administered by two County Labor Boards to a large number of job applicants during the years 1993–2003. Over time and over applicants to different training courses the tests have differed somewhat, but a core of 15–20 tests have been used with great frequency. These form the basis of the analyses in this study.

2.1. The test battery

Most of the tests used in the RA procedure are based on tests designed to measure the seven Primary Mental Abilities of Thurstone (1938). Tests were imported from the United States and adapted for use in Sweden, or they were developed in Sweden on the same schema. The tests will be interpreted within the framework of the CHC model, which entails attributing the Thurstone abilities to the stratum II abilities using the Carroll (1993) and McGrew (2005) findings.
The following tests were included:

1. **Raven’s Standard Progressive Matrices**. This test was developed to measure the eductive component of g as defined by Spearman (1923). The test requires the completion of a matrix pattern, and is non-verbal (Raven, Court & Raven, 1998). The test consists of 60 items. It is expected to be influenced primarily by Gf, even though a small relationship with Gv (Gustafsson, 1984; Lynn, Allik & Irving, 2004) is also expected.

2. **Aros Number Series**. Originally constructed by Thurstone the object of the test is to measure mathematical-inductive ability. The test consists of number series, and the object is to identify the mathematical basis of the series and then add the next two numbers in the series. There are 20 such items. The test is expected to be influenced primarily by Gf.

3. **USTM Number Series**. The principle is the same as for “Aros Number Series”, but the items are less complicated and the subjects are asked to add a single number only. The test has 38 items, and is expected to be influenced primarily by Gf.

4. **WIT Numbers**. The WIT tests were developed in Sweden on the basis of the Thurstone model. In WIT Numbers the task is to create a mathematical statement from given numbers, using simple arithmetic principles. An example is “2; 2; 4”. Here the correct answers are anyone of: “2+2=4”; “4–2=2”; “2×2=4”; or 4/2=2”. The test has 20 items, and is expected to be influenced primarily by Gf.

5. **R16A**. This test consists of mathematical tasks, where the problems are presented in written form. An example is: “Per had 3 apples and Anders had 7 apples. How many more apples did Per have, as compared to Anders?” The test has 28 items. It is primarily expected to be influenced by Gf, but also by Gc, because of the verbal instructions and since it presupposes some mathematical knowledge.

6. **Instructions II**. A number of instructions with verbal, numerical and spatial content demanding working memory capacity are to be carried out. The output is a written statement, a number, a drawing, or a combination of these. An example is: “If there are more than 50 centimeters to a meter, then underline “No”. If this is not the case, then circle “100”.” The test has 39 items and is expected to be influenced primarily by Gf, and to some extent by Gc.

7. **SP2A**. Simple drawings illustrate different technical situations, such as heating systems, vehicles in motion, or electric circuits. A written statement poses three alternative outcomes, and the correct one should be indicated. An example is: “Which pair of scissors would you use to cut wire?” The illustration shows scissors with different proportions. The test has 45 items and is expected to be influenced by Gc and Gv.

8. **DLS Reading**. This is a test of reading speed and reading comprehension in Swedish intended for grades 7–9. In the text parentheses are inserted, each containing three words or expressions. Only one fits the content of the story, and this should be underlined. An example is: “The largest herd living animal of the tundra is the musk ox. … The musk ox is well equipped for life (in the desert on the tundra in the forest), and is often forced to dig up its feed from under the snow.” The test has 34 items, and is expected to be influenced primarily by Gc.

9. **WIT Antonyms**. Part of the WIT battery, WIT Antonyms is a vocabulary test. On each line five words are presented. The subject should find the two that are antonyms. An example is “Beautiful Old Sad Fast Young”. The test has 29 items, and is expected to be influenced primarily by Gc.

10. **WIT Puzzle**. Part of the WIT battery, the WIT Puzzle is a test of two-dimensional spatial ability. The subject should indicate the parts that together with a given figure form a square. The test has 20 items and is expected to be influenced primarily by Gc.

11. **Aros Metal Folding**. The drawing of a sheet of metal is indicated with solid and dotted lines. The solid lines should be imagined cut, and the dotted folded into a sharp crease. This creates a figure that should be indicated among four choices. The test thus requires the mental transformation of two-dimensional figural representations into three-dimensional ones. The test has 40 items, and is expected to be influenced primarily by Gv.

12. **Wire**. This is a manual test of spatial ability. The subject is presented a large two-dimensional figure made of coarse wire, and should reproduce this, but on a smaller scale, using a straight piece of wire 75 centimeters long, and with a diameter of 1 millimeter. The result is graded directly on a stanine scale. It is expected to be influenced primarily by Gv.

13. **Stockholm box**. A manual test of spatial ability. The subject is presented with mechanical models, and a box with mechanic parts, that should be assembled to copy the models. A point is given for each correct part and the maximum score in the rescaled version used here is 21.3. This test is expected to be influenced primarily by Gv.

14. **Crawford Pins**. The task is to enter thin metal pins into small holes with the use of tweezers. A maximum of 36 can be obtained. The test is expected to be influenced primarily by Gv.

15. **P-numbers, P-words, P-figures**. These are tests of perceptual speed, with numerical, verbal and figural content, respectively. The subject is presented with two columns, each with 5 groups of numbers, letters or figures. The numbers consist of four digits, e.g. “2212”. The letter combinations are made up of three letters, e.g. “Hhp”. In the left hand column one, two or three items are indicated by being crossed out. The task is to cross out the matching items in the right hand column. The maximum score is 150 on each. The tests are expected to be influenced primarily by the general speed factor, Gs.
2.2. Subjects

The subjects in this study were all registered at the Employment office and participated in the RA procedure in connection with their employment officer suggesting vocational training. The group consists of \( N = 3570 \) subjects, of whom 86.1% were men and 13.9% were women. The predominance of men could be caused by selection of the employment officers, but most likely it is the result of self-selection, based on the type of training the courses offered. Many of these are oriented towards traditionally male areas of work.

The ages of the subjects range from 18 to 60, with a mean of 33.6 years and a standard deviation of 8.8. A little more than 10% of the subjects have a coded disability, typically involving aspects of mobility. The disability is often the reason for the need to change area of work.

In connection with the RA procedure all subjects were interviewed about their school backgrounds. The country where the applicant had received the basic schooling (approximately the first 9 years) was registered. This measure was chosen as an alternative to citizenship or country of birth, since primary schooling was considered to have a more direct influence on learning opportunities, even though it is realized that this too is a crude measure. For the purpose of our study countries were grouped into larger entities.

Those who received their primary schooling in another country than Sweden were regarded as immigrants. This group consisted of persons with a multitude of reasons for migrating to Sweden. The immigrants \( (N = 1211) \) had spent a mean of 8.2 years in Sweden, with a range from 1 to 31 years \( (sd = 5.5) \).

Language background is likely to affect performance on the test battery, because this has affected the opportunity to acquire knowledge and skills focused upon in the tests. Furthermore, the instructions in the RA procedure are given in Swedish. Being a native Swedish speaker with a Swedish school background may thus be an advantage in the test situation. However, cultural background can also influence the development of other than verbal aspects of intelligence (Sternberg & Kaufman, 1998), and it may affect the experiences of and attitudes towards psychometric testing. Thus, in Western societies, tests of different kinds are abundantly used, and being subjected to tests is a fairly common, if not an altogether relaxed, experience. In other cultures, the concept of psychometric testing may be more or less unknown, which could make the testing situation more obscure. Familiarity with this kind of procedure differs with cultural and educational background, as does the experience of its validity. Such cultural factors may affect test performance over and above the influence of language. It should be noted though, that the majority of studies conducted on immigrant testing in Europe show that only verbal tests are strongly biased against immigrants, while for other kinds of tests only little or no bias has been found (Evers, te Nijenhuis, & van der Flier, 2005; te Nijenhuis & van der Flier, 1999, 2003).

In a broad classification it can be expected that immigrants from Western countries are more adequately prepared and have a more familiar relation to the tasks involved in the test battery than immigrants from non-Western countries. In this material immigrants from the European group of countries, with the addition of USA, Australia and New Zealand, were therefore brought together in one category. The remaining non-European group consisted largely of immigrants from the Middle East and northern Africa. The analysis thus focuses on three groups of subjects: Swedish non-immigrants \( (N = 2358) \), European immigrants \( (N = 620) \) and non-European immigrants \( (N = 591) \).

The applicants were asked about their educational backgrounds, and even though this information may not be perfectly reliable, it should be useful for judging the comparability of the groups with respect to level of education. Among the Swedish non-immigrants 43% had 12 or more years of theoretical education, while for European immigrants and non-European immigrants the corresponding figures were 36% and 43%, respectively. The three groups thus were quite similar with respect to level of education.

2.3. Analytical procedures

The data collected from the RA procedure were available in a data base. No single participant took every test in the test battery. Instead different subgroups of participants were administered different subsets of tests. The test battery used in each case was put together by the psychologist responsible for the RA procedure, primarily on the basis of what training program the applicant had expressed interest. While there is likely to be a certain amount of self-selection to different programs, there is no reason to expect that this would cause any serious threat to the MAR assumption.

The missing-data procedure applied makes the assumption that the data is ‘missing at random’ (MAR), which implies that the procedure yields unbiased estimates when the missingness is random given the information in the data. This is a much less restrictive assumption than the assumption that the data is ‘missing completely at random’. Even though we cannot guarantee that the assignment of tests to applicants yields data that completely satisfy the MAR assumption, the fact that there are high interrelations among observed variables which are exchangeable indicators of a limited set of latent variables implies that there is much information in the data, which should allow for good possibilities to satisfy the MAR assumption. Furthermore, a major source of variability in choice of tests was that partly different test batteries were relied upon by the two participating sites, which is not likely to cause any threat to the validity of the MAR assumption. Another major factor that determined the composition of the test battery was for which program the applicant had expressed interest. While there is likely to be a certain amount of self-selection to different training programs, there is no reason to expect that this would cause any serious threat to the MAR assumption.

The tests listed above were used as manifest variables in a confirmatory factor analysis model. A higher-order model was fitted to the data for the whole group, and this model was then tested on the three subgroups of subjects. The modeling was done with the Mplus Version 3 program (Muthén & Muthén, 2004),
under the STREAMS 3.0 modeling environment (Gustafsson & Stahl, 2005).

Even though the three-level CHC model served as the conceptual framework, the model was set up as a higher-order model with factors at two levels, the stratum II factors being identified as first-order factors. The reason for this was that the test battery included too few tests to allow identification of the stratum I factors. At the level of first-order factors $G_f$ was set to relate to tests requiring general novel problem solving capacity; $G_c$ was set to relate to tests measuring verbal knowledge and skills such as reading speed and vocabulary, and also to tests measuring numerical skills; $G_v$ was hypothesized to relate to tests involving two-and three-dimensional tasks, as well as some manual tests of spatial skills; and $G_s$ was hypothesized to relate to tests of perceptual speed and accuracy. On the second-order level the stratum III general factor, $g$, was hypothesized to relate to all the stratum II latent variables. The hypothesized model is presented in Fig. 1.

Even though the multiple-group CFA procedure was used as the main analytical tool, an additional analysis was conducted with the method of correlated vectors developed by Jensen (1998), which involves computing a correlation between the factor loadings of the tests and the observed differences in means. This method was originally developed as a tool for investigating the hypothesis that the black–white performance difference on cognitive tests in the U.S. may be accounted for by a group difference in $g$, which Jensen refers to as the Spearman hypothesis. This method has been widely adopted as a method for investigating factors associated with the $g$-factor, and it has been applied to test if the Spearman hypothesis can account for performance differences between other groups, such as non-immigrant and immigrant groups. Using the method of correlated vectors te Nijenhuis and van der Flier (1997, 2003) concluded that performance differences between immigrants and non-immigrants in the Netherlands can be explained by the Spearman hypothesis in combination with language bias in tests with a strong verbal component.

However, the method of correlated vectors has been criticized on methodological grounds. Ashton and Lee (2005) observed that the $g$-loadings, and therefore the outcome of the analysis, varied as a function of which tests were included in the factor analysis. Dolan (2000) and Dolan, Roorda and Wicherts (2004) argued that the method of correlated vectors suffers from several weaknesses, which primarily are due to the fact that the method is not based on an explicit and testable model. Instead they argued that multiple-group CFA should be used to investigate the nature of group differences in cognitive performance. Thus, for both substantive and methodological reasons it is interesting to compare the results from the multiple-group CFA and the method of correlated vectors.

3. Results

In the first step of the analysis, descriptive results for the total group of subjects, as well as for the three subgroups of subjects, are presented, and in the second step results from the model fitting are reported. In the third and final step results from application of the method of correlated vectors are reported.

3.1. Descriptive results

Table 1 presents descriptive statistics for the total group of participants.

Most of the tests in the battery were taken by more than 50% of the participants, and for some tests there is data for almost all subjects (e.g., Instructions, WIT Puzzle and WIT Antonyms). For a few tests, the number of participants was smaller (e.g., USTM Number Series and Raven). The proportion of cases having observations on each possible combination of tests also was satisfactory, even though for one combination of tests (Crawford Pins/USTM Number Series) there was no observation. As has been shown by Kaplan (1995) a lack of observations for some combinations of variables does not threaten the usefulness of the missing-data modeling procedure.

Table 2 presents descriptive statistics for the three subgroups of cases.

The proportion of cases within each of the subgroups who received the different tests was roughly the same. It may be observed, however, that there were substantial differences in level of performance between the three different groups, the Swedish non-immigrants generally performing highest, and the non-European immigrants performing lowest. The performance differences were largest for the tests hypothesized to measure $G_c$, and somewhat smaller for the tests measuring $G_f$ and $G_v$. The smallest group differences were observed for the tests measuring psychomotor skills (Wire, Crawford Pins).

3.2. Modeling results

In the first modeling step the hypothesized model was fitted to the data for the whole group. This model had a fit which was not quite acceptable ($\chi^2 = 1182.96$, $df=110$, $p<0.00$, RMSEA 0.053, with a 90% confidence interval of 0.052–0.055, SRMR 0.068). After three modifications of this model (addition of a path from $G_v$ to P-figures, and addition of paths from $G_f$ to WIT Puzzle and to Aros Metal Folding), the fit was judged acceptable, even though the overall goodness-of-fit test was still highly significant ($\chi^2 = 814.16$, $df=107$, $p<0.00$). This, however, was due to the large sample size, as shown by a Root Mean Square Error of Approximation (RMSEA) value of 0.043, with a 90% confidence interval of 0.040–0.046. The Standardized Root Mean Residual (SRMR) value was 0.044, which also indicates good fit. The standardized loadings of the manifest variables on the first-order factors are presented in Table 3.

The sizes of the loadings for the different tests generally agreed very well with expectations. One exception was the Instructions test, which was hypothesized to have a major loading on $G_f$ and a minor loading on $G_c$. However, for this test $G_c$ was found to account for the largest part of the variance, while the contribution from $G_f$ was smaller. Performance on this test thus seems to be more dependent on acquired knowledge than on reasoning ability.
The model also estimated the relations between the four first-order factors and the second-order g-factor. The loadings were found to be 0.83, 0.80, 0.55 and 0.61 for \( G_f \), \( G_c \), \( G_v \) and \( G_s \), respectively. Even though the \( G_f \)-factor had a loading which was marginally higher than the loading for \( G_c \), this loading was far from unity. A formal test of the hypothesis that the loading of \( G_f \) on \( g \) is 1.0 had to be rejected \( (\Delta \chi^2 = 57.76, \Delta df = 1, p < 0.00) \). It must therefore be concluded that the data for the total group of subjects did not support the hypothesis of equivalence between \( G_f \) and \( g \).

In the next step of modeling the model was fitted to each of the three subgroups of cases separately, which yielded the model fit values presented in Table 4.

The model fitted excellently within all three groups, with RMSEA- and SRMR-values similar to those observed for the total group of cases. It may be observed that there was a slight variation in numbers of degrees of freedom over groups. This is due to the fact that in separating the material into three subgroups there were a few more combinations of tests without any observations, and the number of these varied for the three subgroups.

Table 5 presents the estimates of the standardized relations between the manifest variables and the first-order factors.

The sizes of the loadings were generally highly similar over the three subgroups, and they also agreed very well with the results obtained for the pooled group of cases (see Table 3). There were a few exceptions, however.

Comparing the results for the three subgroups with the results for the total group of cases it can be noted that the loadings on the tests measuring \( G_c \) (i.e., Instructions, DLS

![Diagram](image-url)  
Fig. 1.
Reading, and WIT Antonyms) were lower within all three subgroups than they were within the pooled group. This is likely to be due to the fact that there were large differences in level of performance on the Gc-tests between the three subgroups, and these differences appeared as individual differences when the subgroups were pooled.

The technical test SP2A, where each problem is presented with a picture and a text, was in all groups a measure of both Gc and Gv. However, this test had its highest loading on Gv for the Swedish group, while for the immigrant groups the highest loadings were observed for Gc. One reason for this may be that the written instruction makes the test more Gc-loaded for those who do not have Swedish as their first language. A similar pattern of differences was observed for the mathematical test R16A. This test had loadings on both Gf and Gc, and for the Swedish and European groups the highest loading were observed for Gf, while for the non-European group loadings on Gf and Gc were equal.

For the manual dexterity test Crawford Pins there was a more pronounced influence from Gv in the non-European group than in the other groups. However, for the other Gv-tests the standardized loadings on Gv generally were highly similar over the three groups.

The main conclusion from this comparison of relations between manifest variables and the stratum II factors is that the relations generally were highly similar. However, in some cases the demands for Swedish in tests designed to measure Gv and Gf seemed to make these tests measure Gc to a higher extent in immigrant groups.

Table 6 presents the loadings of the first-order factors on the second-order g-factor.

The standardized loadings for the three subgroups were quite similar to one another, but there were some striking differences compared to the results obtained in the analysis of the pooled group of cases. For all three groups the observed loadings of Gf on g were so high that they cannot be regarded as being different from unity. Statistical tests of the hypothesis that the four first-order factors had a perfect relation with g are also presented in Table 6. For Gf this hypothesis could not be rejected for any of the three groups, while for all the other factors it was rejected for all groups. Thus, in contrast to the analysis of the pooled group of cases the results from the analysis of the three subgroups provide strong support for the hypothesis that Gf equals g.

The models considered so far have, for simplicity, been fitted within one of the subgroups at a time. However, such
one-group models do not allow estimation of latent variable means and they do not allow statistical tests of differences in model parameters over groups. Therefore a series of three-group models were also fitted, with the primary purpose of investigating group differences in latent variable means. The means are more easily interpretable within a model with correlated first-order factors than in a higher-order model, so the model compared over the three groups was the oblique model with four correlated stratum II factors (i.e., the model compared over the three groups was the oblique means are more easily interpretable within a model with group models were also fitted, with the primary purpose of model parameters over groups. Therefore a series of three-group models do not allow estimation of latent variable means and they do not allow statistical tests of differences in one-group models do not allow estimation of latent variable means and they do not allow statistical tests of differences in model parameters over groups. Therefore a series of three-group models were also fitted, with the primary purpose of investigating group differences in latent variable means. The means are more easily interpretable within a model with correlated first-order factors than in a higher-order model, so the model compared over the three groups was the oblique model with four correlated stratum II factors (i.e., \( Gf, Gc, Gv, \) and \( Gs \)).

In the first step a model (Model 1) was fitted in which each and every parameter was constrained to be equal over groups. As may be seen in Table 7 this model fitted poorly, indicating that there were differences between the groups for one, several or all of the model parameters.

Next a series of models was therefore fitted in which the constraints on model parameters over groups were successively relaxed. In Model 2 the constraints on the latent variable means were relaxed, and since this model fitted considerably better than Model 1 it may be concluded that the differences between the groups with respect to the latent variable were highly significant. In Model 3 the constraints on the manifest variable intercepts were relaxed, which implies that differences were allowed between the groups with respect to manifest variable means over and above what is accounted for by the latent variables. For this model a slightly improved fit was obtained. In Model 4 the constraints of equality over groups with respect to the residual variances of the manifest variables were relaxed. This too brought about a slight improvement of fit. In Model 5 the constraints on the covariances for the latent variables were relaxed, which caused a considerable improvement of fit. In Model 6, finally, the equality constraints on the factor loadings were relaxed, which also caused a slight improvement of fit. In Model 6 no constraints of equality remain over groups, and as may be seen in Table 7 this model fitted excellently.

Some of the differences between the groups have already been commented upon above, so here only the differences in latent variable means will be focused upon. These were estimated from a model in which the observed variable means were added to Model 6. This model fitted well (\( \chi^2 = 1191.43, \) \( df = 345, p < 0.00, \) RMSEA = 0.045, with a 90% confidence interval of 0.043–0.048) and estimates are presented in Table 8.

The means are expressed in terms of standard deviation (sd) units. The means for the Swedish non-immigrants have been set to zero, so the estimates shown for the other groups are differences in latent variable means as compared to the reference group of Swedish non-immigrants.

It must be observed that the three subgroups are not representative samples of any well-defined populations so simple generalizations cannot be made. However, in this case the profile of performance over the four latent variables is of the greatest interest, and these may be meaningfully compared over the groups. The largest differences between the groups were observed for \( G_c \), for which factor the non-European immigrants had a level of performance 3.4 sd units below the Swedish group, while the European immigrants performed 1.8 sd units below. The smallest differences were observed for \( G_f \), and for this factor the non-European group only performed .4 sd unit below the Swedish non-immigrants. The pattern of differences observed for \( G_c \) and \( G_f \) indicates that the two immigrant groups have had less opportunity to acquire the knowledge and skills measured by the \( G_c \)-tests than have the Swedish non-immigrants. The pattern of differences for \( G_s \) was similar to that observed for \( G_f \), while for \( G_v \) the differences were of an intermediate size. The fact that there were quite substantial differences also with respect to \( G_v \) indicates that spatial-figural knowledge and skills also are culturally determined.

### 3.3. Results from the method of correlated vectors

The Spearman hypothesis, as formulated by Jensen (1998), basically states that group differences in performance on cognitive tests are a function of the \( g \)-loading of the tests. However, the results from the multiple-group CFA analyses do not lend any support to the hypothesis that the locus of differences between the Swedish non-immigrant group and the immigrant groups is in the \( g \)-factor, if this factor is taken to be the same factor as the \( G_f \)-factor. The major source of
differences rather is the Gc-factor. It is, therefore, of great interest to investigate which results are obtained with the method of correlated vectors.

Table 9 presents the basic ingredients of this analysis. The first two columns present the standardized group differences between the Swedish non-immigrants and the European and non-European immigrants, respectively. These differences were computed as Cohen’s d (i.e., the observed mean difference divided by the pooled within-group standard deviation). The column labeled gp presents standardized factor loadings from a one-factor CFA model, which was estimated from a three-group model in which the observed variable means were allowed to vary, but all other parameters of the measurement model were constrained to be equal over the group. This model had poor fit (χ² = 8238.88, df = 44 8, p < 0.00, RMSEA = 0.121, with a 90% confidence interval of 0.119 – 0.123), because it imposes a one-dimensional model on multidimensional data. The next four columns present estimated factor loadings from a so called nested-factor model (Gustafsson & Balke, 1993), in which four orthogonal factors were fitted to data. One factor (Gf = g) was related to all tests, while the others (Gc, Gv and Gs) were residual factors which were related to subsets of tests in the same manner as in the higher-order CFA model presented above. It should be observed that in this model there was no Gf-factor for the Gf subset of tests, which is due to the fact that the equivalence of Gf and g causes this residual factor to disappear (see Gustafsson, 2001, 2002). The four-factor model was estimated in the same manner as the one-factor model, and it had acceptable fit (χ² = 1804.73, df = 42 8, p < 0.00, RMSEA = 0.052, with a 90% confidence interval of 0.050 – 0.055).

Table 10 presents the correlations among the standardized mean differences and the estimated factor loadings. The correlations between the mean differences and the Gf-factor loadings estimated from the one-factor model amounted to .84 for the SNI – EI difference and to .86 for the SNI – NEI difference, so these results provided strong support for Spearman’s hypothesis. However, the correlations with the factor loadings estimated from the four-factor model provided a different pattern of results. There were no significant correlations between the Gf = g-factor and the mean

Table 5
Standardized factor loadings for the tests for the subgroups

<table>
<thead>
<tr>
<th>Test label</th>
<th>Gf</th>
<th>Gc</th>
<th>Gv</th>
<th>Gs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Raven</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aros Number Series</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>USTM Number Series</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WIT Numbers</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R16A</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Instructions</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DLS Reading</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WIT Antonyms</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SP2A</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WIT Puzzle</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aros Metal Folding</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wire</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stockholm box</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Crawford Pins</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P-numbers</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P-letters</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P-figures</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The groups are Swedish non-immigrants (SNI), European immigrants (EI), and non-European immigrants (NEI).

Table 6
Standardized loadings of first-order factors on the g-factor and χ² tests of the hypothesis that the loading is unity

<table>
<thead>
<tr>
<th>Factor</th>
<th>Swedish non-immigrants</th>
<th>European immigrants</th>
<th>Non-European immigrants</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>r</td>
<td>χ²</td>
<td>r</td>
</tr>
<tr>
<td>Gf</td>
<td>0.98</td>
<td>0.32</td>
<td>0.99</td>
</tr>
<tr>
<td>Gc</td>
<td>0.80</td>
<td>73.92</td>
<td>0.67</td>
</tr>
<tr>
<td>Gv</td>
<td>0.37</td>
<td>556.36</td>
<td>0.49</td>
</tr>
<tr>
<td>Gs</td>
<td>0.58</td>
<td>514.46</td>
<td>0.56</td>
</tr>
</tbody>
</table>

The χ² statistics all have 1 df, and values larger than 3.84 are significant at the 5% level.

Table 7
Fit statistics for multiple-group models

<table>
<thead>
<tr>
<th>Model Description</th>
<th>χ²</th>
<th>df</th>
<th>RMSEA</th>
<th>Δχ²</th>
<th>Δdf</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>6833</td>
<td>445</td>
<td>0.110</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>3744</td>
<td>437</td>
<td>0.080</td>
<td>3089</td>
<td>8</td>
</tr>
<tr>
<td>3</td>
<td>3306</td>
<td>411</td>
<td>0.077</td>
<td>438</td>
<td>26</td>
</tr>
<tr>
<td>4</td>
<td>3035</td>
<td>377</td>
<td>0.077</td>
<td>271</td>
<td>34</td>
</tr>
<tr>
<td>5</td>
<td>1186</td>
<td>359</td>
<td>0.044</td>
<td>1848</td>
<td>18</td>
</tr>
<tr>
<td>6</td>
<td>967</td>
<td>319</td>
<td>0.041</td>
<td>220</td>
<td>40</td>
</tr>
</tbody>
</table>
Table 8
Estimated latent variable means for the subgroups

<table>
<thead>
<tr>
<th></th>
<th>Gf</th>
<th>Gc</th>
<th>Gv</th>
<th>Gs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Swedish non-immigrants</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>European immigrants</td>
<td>−0.14</td>
<td>−1.84</td>
<td>−0.63</td>
<td>−0.31</td>
</tr>
<tr>
<td>Non-European immigrants</td>
<td>−0.40</td>
<td>−3.38</td>
<td>−1.28</td>
<td>−0.72</td>
</tr>
</tbody>
</table>

The latent variable means for the Swedish non-immigrants have been set to 0, to identify the differences between latent variable means over groups.

differences, while there were strikingly high correlations between the mean differences and the $G_c$-factor (.94 and .88, respectively).

These results thus indicate fairly good agreement between the results of the method of correlated vectors and the multiple-group CFA analysis, when the factor loadings from the well-fitting four-factor model were used in the analysis. However, the correlated vectors analysis yielded incorrect results when it was based upon factor loadings estimated from the one-factor model. The reason for this is that the $g$-factor estimated from the one-factor model is biased in such a way that the loadings for $G_c$-tests are overestimated, while the loadings for $G_f$-tests are underestimated.

4. Discussion and conclusion

The main aim of the current study was to test a prediction derived from Cattell’s Investment theory,

Table 9
Standardized mean differences and factor loadings from a one-factor and a four-factor model with nested factors

<table>
<thead>
<tr>
<th>Test label</th>
<th>$\Delta$SNI–EI</th>
<th>$\Delta$SNI–NEI</th>
<th>$g$</th>
<th>$G_f = g$</th>
<th>$G_c$</th>
<th>$G_v$</th>
<th>$G_s$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Raven</td>
<td>0.29</td>
<td>0.74</td>
<td>0.50</td>
<td>0.65</td>
<td>0.26</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aros Number Series</td>
<td>0.25</td>
<td>0.44</td>
<td>0.51</td>
<td>0.80</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WIT Numbers Series</td>
<td>0.31</td>
<td>0.70</td>
<td>0.61</td>
<td>0.84</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R16A</td>
<td>0.47</td>
<td>0.84</td>
<td>0.64</td>
<td>0.81</td>
<td>0.13</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Instructions</td>
<td>1.40</td>
<td>2.07</td>
<td>0.85</td>
<td>0.77</td>
<td>0.47</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DLS Reading</td>
<td>1.49</td>
<td>2.09</td>
<td>0.72</td>
<td>0.51</td>
<td>0.50</td>
<td></td>
<td></td>
</tr>
<tr>
<td>WIT Antonyms</td>
<td>1.32</td>
<td>1.46</td>
<td>0.72</td>
<td>0.61</td>
<td>0.39</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SP2A</td>
<td>1.05</td>
<td>1.85</td>
<td>0.68</td>
<td>0.47</td>
<td>0.29</td>
<td>0.46</td>
<td></td>
</tr>
<tr>
<td>WIT Puzzle</td>
<td>0.50</td>
<td>1.06</td>
<td>0.57</td>
<td>0.58</td>
<td>0.55</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aros Metal</td>
<td>0.57</td>
<td>1.12</td>
<td>0.58</td>
<td>0.52</td>
<td>0.65</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aros Folding</td>
<td>0.57</td>
<td></td>
<td>0.58</td>
<td>0.52</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wire</td>
<td>−0.04</td>
<td>0.15</td>
<td>0.20</td>
<td>0.24</td>
<td>0.29</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stockholm box</td>
<td>0.39</td>
<td>0.66</td>
<td>0.36</td>
<td>0.26</td>
<td>0.62</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Crawford Pns</td>
<td>0.08</td>
<td>0.26</td>
<td>0.16</td>
<td>0.16</td>
<td>0.25</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P-numbers</td>
<td>0.34</td>
<td>0.64</td>
<td>0.41</td>
<td>0.46</td>
<td>0.65</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P-letters</td>
<td>0.26</td>
<td>0.60</td>
<td>0.44</td>
<td>0.50</td>
<td>0.75</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P-figures</td>
<td>0.39</td>
<td>0.67</td>
<td>0.45</td>
<td>0.46</td>
<td>0.26</td>
<td>0.63</td>
<td></td>
</tr>
</tbody>
</table>
The groups are Swedish non-immigrants (SNI), European immigrants (EI), and non-European immigrants (NEI).

namely that the $G_f$-factor would be equal to the $g$-factor in populations which are homogeneous with respect to opportunity to having learned the knowledge and skills measured, but that this relationship would not hold in heterogeneous populations where subgroups differ with respect to opportunity to learn. Using a set of data consisting of Swedish non-immigrants, European immigrants, and non-European immigrants who had been tested with a Swedish test battery, clear-cut support was obtained for the prediction: the relationship between $G_f$ and $g$ was only 0.83 when all subjects were treated as a single group, but it was unity within each of the three subgroups of cases. This result provides support for the Investment theory, and for the hypothesis that $G_f$ is equivalent to $g$. However, the results of this study also imply that the hypothesis of $G_f$–$g$ equivalence only holds true when the subjects have had approximately equally good, or equally poor, opportunities to develop the knowledge and skills measured.

The results from the current study also provide a possible explanation for why some studies have failed to establish the equivalence of $G_f$ and $g$. Thus, in studies based upon heterogeneous populations the perfect relation cannot be expected to appear, even though a high relationship between $G_f$ and $g$ is expected. For example, in the study by Carroll (2003) previously referred to, which failed to find the perfect relation between $G_f$ and $g$, the matrices analyzed were pooled across the ages from kindergarten to adulthood, and this may have caused a population heterogeneity which prevented the perfect relation to appear. These data could be reanalyzed with the data organized into homogeneous age groups to test this hypothesis.

The results of the current study thus indicate that $G_f$ is a causal factor in determining individual differences in the full range of knowledge and skills measured by cognitive tests, presumably because $G_f$ is involved in at

Table 10
Correlations between the difference in mean scores for European immigrants (EI) and non-European immigrants (NEI) as compared to the Swedish non-immigrant (SNI) group with factor loadings estimated from a one-factor model and a nested-factor model with four factors

<table>
<thead>
<tr>
<th>SNI–EI</th>
<th>SNI–NEI</th>
<th>G</th>
<th>$G_f = g$</th>
<th>$G_c$</th>
<th>$G_v$</th>
<th>$G_s$</th>
</tr>
</thead>
<tbody>
<tr>
<td>SNIEI</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SNI–EI</td>
<td>0.97*</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>G</td>
<td>0.84*</td>
<td>0.86*</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$G_f = g$</td>
<td>0.22</td>
<td>0.24</td>
<td>0.68*</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$G_c$</td>
<td>0.94*</td>
<td>0.88*</td>
<td>0.74*</td>
<td>0.21</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>$G_v$</td>
<td>−0.11</td>
<td>0.00</td>
<td>−0.21</td>
<td>−0.51*</td>
<td>−0.30</td>
<td>1.00</td>
</tr>
<tr>
<td>$G_s$</td>
<td>−0.20</td>
<td>−0.21</td>
<td>−0.17</td>
<td>−0.27</td>
<td>−0.23</td>
<td></td>
</tr>
</tbody>
</table>

* Correlation is significant at the 0.05 level (2-tailed).
least the early phases of acquisition of knowledge and skills in all domains. To the extent that learning opportunities systematically differ between different subgroups of the population the Gf-factor will no longer take the role of being the g-factor. It must be observed, however, that if such differential learning opportunities only affect knowledge and skill represented by a single stratum II ability dimension, the stratum III g-factor will still be equal to Gf. This is because in this situation only the residual variance of the single ability dimension will be affected by the differential learning opportunities. For the covariances among abilities to be affected in such a way that the g=Gf relationship is disturbed, the differential learning opportunities must influence two or more of the abilities in the model.

As was shown in the analysis of group differences in latent variable means the two immigrant groups had a much lower level of performance on the Gc-factor. This factor was primarily measured by tests of reading and vocabulary, and it is quite obvious that the immigrant groups had not had the same opportunities to acquire the Swedish language proficiency needed to perform well on these tests as had the Swedish non-immigrants.

It is also interesting to observe that there were quite substantial group differences in level of performance on the Gv factor, which indicates that cultural background exerts an influence on visual-spatial performance as well. This is also indicated by the late and prolonged maturation rate of the Gv factor (McArdle, Ferrer-Caja, Hamagami, & Woodcock, 2002). This influence may at least partially be mediated by the educational system, as is suggested by a recent study by Cliffordson and Gustafsson (in press), which demonstrated differential effects of high school educational track on development of spatial ability.

While the verbal area is subject of direct training in most cultures, the visual-spatial areas of performance are presumed to develop more spontaneously and indirectly, while the growing individual is engaged in play or other motor activities. Thus it has traditionally been thought of as relatively more “culture free”. However, Maruyama (1999) has studied cultural differences with respect to attitudes towards spatial experiencing and processing, and has described distinctly different ways of relating to space and using spatial constructs. Such cultural differences can be assumed to influence also the development of individual abilities.

Nevertheless, it is necessary to be cautious in making conclusions about cultural differences, since the groups investigated are not representative samples from any well-defined cultural groups. Becoming an immigrant involves processes of selection and self-selection, as does the process leading up to an application for a vocational training course. It would thus be of great interest to have the current study repeated on other groups, which should be more clearly defined in this respect and preferably also should be more balanced with respect to gender composition.

As has already been noted the empirical evidence in support of the Investment theory has largely been missing, while the results of the current study are in agreement with predictions based the theory. It may also be noted that the notion of Gf as a biologically and genetically determined ability which has been associated with the Investment theory does not agree with findings of a strong environmental determination of Gf as evidenced by the Flynn effect (Dickens & Flynn, 2001), effects of schooling (Cliffordson & Gustafsson, in press) and recent findings of the fluidity of the human brain, particularly in the early years (Blair, 2006). It is obvious that further research is needed to resolve these contradictory and paradoxical results. One interesting approach to be elaborated in this research is the “mutualism” dynamical model developed by van der Maas, Dolan, Grasman, Wicherts, Huizenga, and Raijmakers (2006), both as a vehicle to investigate alternative models for possible interrelations between Gf and Gc in development, and as a general framework for understanding mutual influences among abilities.

One interesting methodological finding of the current study is that the method of correlated vectors was shown to yield different results than the CFA procedure when factor loadings were estimated with a simplified one-dimensional model, but that results were in agreement over the methods when factor loadings were estimated with a well-fitting four-dimensional model. This suggests that the method used for estimating the g-factor may be of greater importance than is usually recognized. According to the conventional wisdom of the field very much the same g-factor is estimated, whether this is done via a sum of scores on a heterogeneous test battery, via a principal factor or principal component solution, or via a hierarchical factor model. Even though Jensen (1998) favored the latter method, it seems that the first principal factor is the most commonly used method for identifying the g-factor. However, as was observed by Ashton and Lee (2005) the pattern of loadings on the first principal factor is sensitive to the composition of the test battery, and they observed a tendency for the first principal factor to be biased in favor of Gc-tests, as was also the case with the one-factor CFA model fitted here. This effect does not seem to be caused by there being an excessive number of Gc-tests in the batteries, but rather by the fact that a larger proportion of the systematic variance in the Gc-tests is turned into common variance than is the case...
for Gf-tests. If this hypothesis is correct it implies that much of the research thought to focus on the g-factor has in fact focused on Gc.

The results obtained in the current study conflict with those reported in other European studies on immigrant testing, which show the g-factor to be the main locus of differences in performance between immigrants and non-immigrants (e.g., te Nijenhuis & van der Flier, 1997, 2003). However, these studies have relied on the method of correlated vectors with principal factor estimates of the g-loadings, which suggests that the different results are due to methodological differences between the studies. This hypothesis can easily be investigated through reanalysis of the previous studies with the methods used in the current study.

It may, finally, be noted that there is considerable confusion in the literature concerning the meaning and nature of the g-factor. Blair (2006) discussed relations between fluid intelligence and general intelligence, and rejected the idea that these are identical because of an obvious lack of agreement in many studies. However, the g-factors investigated in the studies reviewed by Blair (2006) were typically defined by scores on IQ-tests or as the first principal factor, which explains why only relatively low relations were found with measures of Gf.

Acknowledgement

The research reported here has been supported financially by the Institute for Labor Market Policy Evaluation in Sweden.

References


of fluid and crystallized intelligence. *Journal of Educational
Psychology, 57*, 253–270.

theory. In D. P. Flanagan, J. L. Genshaft, & P. L. Harrison (Eds.),
*Contemporary intellectual assessment* (pp. 53–91). New York,
NY: Guilford Press.


patterns on goodness-of-fit tests in factor analysis. *Journal of
Educational and Behavioral Statistics, 20*(1), 69–82.

Keith, T. Z. (2005). Using confirmatory factor analysis to aid in
understanding the constructs measured by intelligence tests. In D. P.
Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual
New York, NY: Guilford Press.

problem: The latent semantic analysis theory of acquisition,
induction, and representation of knowledge. *Psychological Review,
104*(2), 211–240.


Lynn, R., Allik, J., & Irving, P. (2004). Sex differences on three
factors identified in Raven’s standard progressive matrices. *Intelli-
gerence, 32*, 411–424.

selection: two new directions in counseling. In P. Pedersen (Ed.),
*Multiculturalism as a fourth force* (pp. 37–72). Philadelphia, PA:
Taylor & Francis Group.

McArdle, J. J., Ferrer-Caja, E., Hamagami, F., & Woodcock, R. W.
(2002). Comparative longitudinal structural analyses of the growth
and decline of multiple intellectual abilities over the life span.
*Developmental Psychology, 38*(1), 115–142.

abilities. past, present and future.* In D. P. Flanagan, & P. L.
Harrison (Eds.), *Contemporary intellectual assessment: Theories,
tests and issues* (pp. 136–181). (2nd edition). New York, NY:
Guilford Press.

Messick, S. (1992). Multiple intelligences or multilevel intelligences?
Selective emphasis on distinctive properties of hierarchy: On
Gardner’s frames of mind and Sternberg’s beyond IQ in the
context of theory and research on the structure of human abilities.

equation modeling with data that are not missing completely at


Ltd.

Reynolds, M. R., & Keith, T. Z. (2007). Spearman’s law of
diminishing returns in hierarchical models of intelligence for

Roberts, R. D., Goff, G. N., Anjou, F., Kyllonen, P. C., Pallier, G., &
Stankov, L. (2000). The Armed Services Vocational Aptitude
Battery (ASVAB). Little more than acculturated learning (Gc)?
*Learning and Individual Differences, 12*, 81–103.

theory and measurement. *Personality and Individual Differences,
27*, 715–735.

Spearman, C. (1904). General intelligence: objectively determined and

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


GATB scores for immigrants and majority group members: Some

te Nijenhuis, J., & van der Flier, H. (1999). Bias research in The
Netherlands: Review and implications. *European Journal of
Psychological Assessment, 15*(2), 165–175.

differences in cognitive performance: Jensen effects, cultural
effects, or both. *Intelligence, 31*, 443–459.

Monographs, 1.*

replace Cattell’s theory of fluid and crystallized intelligence.

organization of cognitive abilities: Restoring general intelligence
through the use of linear structural relations (LISREL). *Multi-

van der Maas, H. L. J., Dolan, C. V., Grasman, R. P. P., Wicherts, J.
model of general intelligence: The positive manifold of intelligence