ABSTRACT

An Examination of the Flynn Effect in the National Intelligence Test in Estonia

William Shiu, Ph.D.

Chairperson: A. Alexander Beaujean, Ph.D.

This study examined the Flynn Effect (FE; i.e., the rise in IQ scores over time) in Estonia from Scale B of the National Intelligence Test using both classical test theory (CTT) and item response theory (IRT) methods. Secondary data from two cohorts (1934, n = 890 and 2006, n = 913) of students were analyzed, using both classical test theory (CTT) and item response theory (IRT) methods. CTT analysis compared the summed score for each subtest between the cohorts. IRT analysis examined item invariance across the time period and then, for each subtest, linked the latent variable (θ) scores between the two cohorts using the invariant items.

IRT analyses revealed that each subtest displayed invariance on over 50% of the items (i.e., partial measurement invariance). Additionally, results from the current study found positive score gains but also revealed reverse FEs. CTT analysis showed three subtests had a FE (Vocabulary [.74], Analogies [1.09], and Comparisons [1.71]), while two subtests had a reverse FE (Computation [-.33] and Information [-.03]). The IRT analysis found that four subtests had a positive FE (Information [.44], Vocabulary [.79], Analogies [1.02], and Comparisons [1.51]), with only the Computation (-.10) subtest displaying a negative FE.

The results confirm previous research that the FE continues in Estonia. Using CTT methods, Must, te Nijenhuis, Must, and van Vianen (2009) found positive gains

on the Estonian NIT subtests Computation (.15), Information (.94), Vocabulary (.65), Analogies (1.81), and Comparison (2.34).

An implication of the current study shows the viability of IRT to supplement CTT when analyzing the FE. The IRT procedures demonstrated in the current study provides a counter argument that the rise in IQ scores is a psychometric artifact, at least in the domains of Comprehension-Knowledge, Fluid Reasoning, and Visual Processing. As this study was unable to examine causative factors involved in the FE, future studies should examine if the score gains might be attributed to some environmental cause (e.g., nutrition, education) or biological cause (e.g., heterosis).

An Examination of the Flynn Effect in the National Intelligence Test in Estonia

by

William Shiu, B.B.A., M.S.Ed.

A Dissertation

Approved by the Department of Educational Psychology

Marley W. Watkins, Ph.D., Chairperson

Submitted to the Graduate Faculty of
Baylor University in Partial Fulfillment of the
Requirements for the Degree
of
Doctor of Philosophy

Approved by the Dissertation Committee

A. Alexander Beaujean, Ph.D., Chairperson

Tonya N. Davis, Ph.D.

Terrill F. Saxon, Ph.D.

Randall L. Vaughn, Ph.D.

Jim H. Patton, Ph.D.

Accepted by the Graduate School
December 2012

J. Larry Lyon, Ph.D., Dean

Page bearing signatures is kept on file in the Graduate School.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

Throughout my graduate studies I have received so much encouragement and support from so many people. First and foremost my Lord and Savior Jesus Christ who's words encouraged and bolstered my resolve to finish my degree. I have to thank my family, my dad Johnny who financially supported me when it counted most, my mom Chi-Lan who hugged me and never gave up on me, my sister Dr. Alex Shiu who patiently relived this process and stuck by me, and my brother Bebo, his music was a blissful respite from writing. When Jesus carried me, my family dragged me the rest of the way. I also could not have finished this degree without the patience and tenacious attention to detail from my lifelong friend and mentor Dr. Alex Beaujean. His openness to questions, constant encouragement through revision after revision, and wisdom on the writing process showed me that a little writing everyday adds up very quickly. To Dr. Terrill Saxon, Dr. Tonya Davis, Dr. Randall Vaugh, Dr. Jim Patton, I am highly honored you agreed to serve on my dissertation committee. Thank you for your time and encouragement. A very special thank you goes to Dr. Susan Johnsen for her guidance and encouragement. I also have to thank the faculty and staff in the Educational Psychology and Curriculum and Instruction departments along with the staff in the School of Education Deans Office at Baylor University. I would also like to thank my friend Shannon Edwards for providing the perfect work-study job in the Media Center for the 10 years I was under her employment. I could not have finished my dissertation and juggle a full time job without the understanding and encouragement from my current supervisor Dr. Ben Cox, the staff in the department of Institutional Effectiveness, Research, and Planning, and my friends at Texas State Technical College. Lastly I cannot forget my graduate school colleagues and friends Tracey, David, Greg, Cindy, Sonia, Paige, Ted, Karen... there are too many to list, thank you so much for everything, we are not alone in this. I would like to especially thank Olev Must for providing the item-level data used in my study.

CHAPTER ONE

Introduction

The Flynn Effect (FE) is the rise in IQ scores over time (approximately 3 IQ points per decade or .3 points per year (Neisser, 1998). The cause of the FE remains inconclusive, although several theories have been put forth, such as an increase in nutrition, a proliferation of electronic media, urbanization, education, and genetics (Colom, Lluis-Font, & Andrés Pueyo, 2005; Flynn, 1998; Greenfield, 1998; Lynn, 1989, 1990; Mingroni, 2007; Schooler, 1998; Sigman, 2000; Sigman & Whaley, 1998; Williams, 1998). Some researchers argue that the cause of the FE is not strictly environmental, but that the rise is a psychometric artifact that gives the impression that there is a change in the examinees intelligence scores. For example, Sundet, Barlaug, and Torjussen (2004) theorize that IQ increases may merely reflect measurement artifacts, such as heightened test sophistication or altered test-taking strategies. Wicherts et al. (2004) found evidence from five data sets that IQ scores are not measurement invariant over time, meaning that test properties are changing over time. In other words, the intelligence instrument itself is changing in some way, either in its construct, interpretation, and measurement (Campbell, Barry, Joe, & Finney, 2008).

Statement of the Problem

In order to analyze the FE accurately, the method instruments measure constructs should remain consistent across time. This issue is called invariance (Bontempo & Hofer, 2007; Millsap, 2011). The issue of invariance is important because if an instrument's scores have changed the construct(s) it measures, or the way it measures these construct(s), then researchers cannot make accurate judgments about any observed score changes. One method to examine if an instrument's scores have changed over time (i.e., is not invariant) is to analyze the instrument's items to see if they are operating the same way across time (Borsboom, Mellenbergh, & van Heerden, 2002).

One way to investigate changes in an instrument's item properties is through the use of Item Response Theory (IRT)(Lord & Novick, 1968). IRT is a type of latent variable modeling that concurrently estimates both a tests items properties and the test respondents ability. Thus, IRT is uniquely able to differentiate changes taking place in the test versus changes taking place in the respondents over time.

Purpose of the Current Study

The purpose of the current study was to conduct an IRT analysis of the National Intelligence Test (NIT; Haggerty, Terman, Thorndike, Whipple, & Yerkes, 1920) over time in Estonia. The NIT is a measure of cognitive ability developed in the early twentieth century to assess the cognitive abilities of children in a group setting. The NIT subtests were designed to cover a range of cognitive abilities, such as language skills, reading comprehension, problem solving, and mathematical tasks. The NIT consists of two scales, each containing five subtests. Each scale, however, can stand independently as a measure of cognitive ability. The current study analyzed Scale B comprised of five subtests: Computation, Information, Vocabulary, Analogies, and Comparisons. The Estonian data are unique because they span a large period of time (1933-2006), and provide an opportunity to examine the invariance of test items. Although previous research has examined invariance in the population (e.g., Must, te Nijenhuis, Must, & van Vianen, 2009), to date no study has done so at the item level. Thus, the current study's IRT analysis of the NITs items over time is novel, and adds to the growing body of knowledge concerning the FE. There have been several factors that have been considered as causes of the FE such as nutrition, education, genetics, or that gains in scores are a psychometric artifact specifically, changing instrument properties (Beaujean & Osterlind, 2008;

Colom et al., 2005; Lynn, 1989; Mingroni, 2004). The current study shows that IRT is a viable method for controlling one of these factors, changing instrument properties. The use of IRT is a step in the right direction for determining a cause of the FE.

Research Questions

(1) Do the subtests on the B Scale of the National Intelligence Test (Computation, Information, Vocabulary, Analogies, and Comparison) exhibit (partial) measurement invariance over time in Estonia?

(2) If there is at least partial measurement invariance, then do the subtests, individually, exhibit a Flynn Effect?

Results

Item response theory analysis revealed that all five subtests displayed partial measurement invariance. Using IRT comparisons the Information, Vocabulary, Analogies, and Comparisons subtests displayed positive FE's, but the Computation subtest displayed a small, negative (i.e., reverse) Flynn Effect. CTT comparisons showed that Vocabulary, Analogies, and Comparisons had Flynn Effects, but Computation and Information had reverse Flynn Effects. Of interest was the CTT mean and variance comparison on the Information subtest. There were virtually no mean differences, means decreased -.18 from 1934 to 2006 and the variance became smaller over time with a decrease of 2.43 SD. Another interesting finding was the Flynn Effect exhibited in the Information subtest even after DIF items were removed. The Information subtest measures Comprehension-Knowledge (i.e., Crystallized Intelligence) and previous research suggested that the Flynn Effect is predominantly found among fluid intelligence abilities (Colom, Andrés Pueyo, & Juan-Espinosa, 1998; Flynn, 1987, 1998, 2006; Lynn, 2009; Teasdale & Owen, 2000).

One implication is the results confirm previous research that the Flynn Effect continues in Estonia. Must, te Nijenhuis, Must, and van Vianen (2009) found positive gains on the Estonian NIT subtests Computation (.15), Information (.94), Vocabulary (.65), Analogies (1.81), and Comparison (2.34). The current study similarly found positive gains on the Information, Vocabulary, Analogies, and Comparison subtests except for a reverse Flynn Effect on Computation. In another study, Must, Must, and Raudik (2003b) also found both positive score gains and reverse Flynn Effects among Estonian children (see Table A.1).

Another implication of the current study shows the viability of IRT to supplement CTT when analyzing the Flynn Effect. The viability of IRT provides a counter argument that the rise in IQ scores is a psychometric artifact. The IRT procedures used in the current study controlled for non-invariance and displayed evidence that the gains in scores are not psychometric artifacts and can be attributed to some other environmental cause such as nutrition or education.

CHAPTER TWO

Literature Review

The Flynn Effect (FE) is the rise in psychometric Intelligence Quotient (IQ) scores, a trend that has been evident since at least the early twentieth century (Flynn, 2009b). Herrnstein and Murray (1996) first coined the term, naming it after New Zealand sociologist James R. Flynn who, although he did not discover the IQ gains, did more than anyone before him to systematically gather and publish data in this area. He gathered IQ data from 14 developed nations and analyzed their rates of IQ gain per year. He found that culturally reduced tests along with tests of fluid intelligence displayed some of the larger IQ gains.

Historically, the FE can trace its beginning to studies that indicated changes in IQ scores across time (e.g., Smith, 1942; Tuddenham, 1948). Richard Lynn (1982) was one of the first scholars to publish about a rise in IQ scores, systematically, looking at scores in both Japan and Great Britain (Lynn, Allik, & Must, 2000; Lynn & Hampson, 1986, 1989; Lynn & Pagliari, 1994; Lynn, Pullman, & Allik, 2003), although it was James Flynn's (1984, 1987) report of an average rate of gain of approximately 0.3 IQ points per year (3 points per decade) that brought the effect more international attention.

Theories About the Flynn Effect

In the 20 years that the FE has been systematically studied, many researchers have developed theories as to the underlying causes of the FE, but a general agreement has not been reached. The explanations typically fall into one of two categories: either the FE represents a real rise in cognitive ability due to factors such as a proliferation of electronic media, urbanization, education, nutrition, and genetics (Greenfield, 1998; Flynn 1998; Lynn & Hampson, 1989; Lynn 1990; Mingroni, 2007;

Schooler, 1998; Williams, 1998), or the FE is a psychometric artifact (i.e., properties of items change over time) meaning that the intelligence tests are changing and not an individuals intelligence ability (Beaujean & Osterlind, 2008; Campbell, et al., 2008; Wicherts et al., 2004) and does not represent a real rise in intelligence. Ultimately, Jensen (1998) even suggested that the cause of the FE could be a combination of small, multiple unidentified factors.

A Genuine Rise in Cognitive Ability

Those who think the FE represents a genuine rise in cognitive ability have suggested a number of theories. These theories draw focus to changes in nutrition, health care, education levels, and tendencies to live in a city, heterosis, and genetics. The most prominent theory is that the rise in IQ is due to better nutrition available in the twentieth century (Colom et al., 2005; Greenfield, 1998; Sigman, 2000; Sigman, & Whaley, 1998; but see Flynn 2009a). Specifically, increases in intelligence are because of advances in nutrition, but also that nutrition has increased height, head size, and brain size as well (i.e., secular trend; Lynn, 1989; Lynn, 2009).

Two hypotheses exist on how head size is related to intelligence. The first stems from head size trends among animals in that a large head contains a large brain, which may indicate more intelligence (Jerison, 1982; Mackintosh, Wilson, & Boakes, 1985). The second theorizes that nutrition may affect the internal neurological development of the brain and thus correlate but not cause (Dobbing, 1984; Winick, Ross, & Waterlow, 1970). Correlation studies found positive relationships ranging from .26 to .56 between brain volume and IQ (Andreasen et al., 1993; Colom, Jung, & Haier, 2006; Wilke, Sohn, Byars, & Holland, 2003). McDaniel (2005), in a meta-analysis, analyzed 37 studies on the relationship between brain volume and intelligence and found, across both sex and age, there was strong evidence that brain size was related to intelligence, with an average of approximately $r = +.33$.

Studies have shown that malnourished children tend to have low IQs (Winick, Meyer, & Harris, 1975) and that low birth weight affects intelligence as well (Churchill, 1965; Willerman & Churchill, 1967). Historical figures from Britain, the United States, and Japan show that in the time between the two world wars there was widespread poor nutrition (Corry Mann, 1926; Orr, 1936; Palmer, 1935; Takahashi, 1966). After World War 2 economic improvements allowed populations to purchase more nutritious food. This resulted in increases in height along with increases in head size (van Wieringen, 1978; Whitehead & Paul, 1988). Brandt (1978) found a .80 correlation between head size and brain size. Lynn (1989) conducted a study of 310 children and found additional evidence supporting the theory that nutrition is one determining factor of intelligence and head size (cf., Arija et al., 2006; Colom et al., 2005; Jensen, 1998). Lynn (2009) found that improvements in nutrition in the quality of food eaten by pregnant mothers and the quality of food given to babies after birth is the most likely cause for increases in Development Quotients (DQ) of infants along with IQ gains among both adults and children. DQs are similar to IQs, but for infants two years old and younger, DQs consist of scores from the Bayley Scales (Bayley, 1993) and Griffiths Test (Griffiths, 1954).

Another possible explanation for the FE has been increased cognitive stimulation, whether it is due to more complex technology available to more people or changes in educational systems, such as the direction of school education shifting from rote learning to problem solving (Schooler, 1998; Teasdale & Owen, 1987, 1989). Daley, Whaley, Sigman, Espinosa, and Neumann (2003) studied children in rural Kenya over a fourteen-year period from 1984 to 1998 and examined common theories of the FE. The results indicated a gain of 26.3 IQ points over the 14 years. They found that the 1998 cohort had better nutrition, more access to media (TV, newspapers), smaller family sizes, and mothers were more literate. The cohort also had an increase in nursery and Sunday school attendance. In response to a possible argument, the instrument used was similar to any other Kenyan standardized exam and was familiar to teachers. They ruled out teaching to the test as an explanation for the observed FE.

Other researchers theorize that the FE is due to the change in family structure and fertility patterns (Blair et al., 2005; Wicherts et al., 2004). People born in the later half of the twentieth century were more educated, have more often lived in cities, and were typically raised in smaller families than those born in the earlier twentieth century. Mingroni (2004) put forth his hypothesis that heterosis (out breeding) may be a cause of the FE. Heterosis is also commonly referred as "hybrid vigor." Heterosis is a genetic result of mating between members of genetically distinct subpopulations due to urbanization and greater population mobility. This is evidenced by the increasing measurements of the following traits: height, growth rate, myopia, asthma, autism, ADHD, head circumference, and head breadth, which all have a heritability estimate greater than .6. Heritability (h^2) is defined as genetically influenced traits, which represents broad based change and have largely paralleled the IQ trend in their timing, location, and relative magnitudes (Mingroni, 2004, p. 68). It is the percentage of a traits (i.e. height or IQ) variability from a population that is due to effects from genes and the environment. Heritability is scaled from 0 to 1 where 0 indicates that none of the variance is due to genes and 1 indicates that all the variability is due to genes Devlin, Daniels, & Roeder, 1997; Herrnstein & Murray, 1996. However, heterosis is not without its critics. One criticism of Mingroni's (2007) simulation study of genetic outbreeding is the .3 IQ point gain over 50 years is far less than the 27 point gain found in the United States over an 80 year period (Flynn, 2009a; Lynn, 2009; Woodley, 2011). If anything, heterosis may affect the FE through development factors that may enhance abilities (Woodley, 2009).

Further theories credit a real gain in cognitive ability to genetics. In contrast, if gains in IQ scores were due to environmental factors, then the environmental factors should have also represented the introduction of a new source of IQ variance, which ultimately would lower IQ heritability estimates over time (Sundet, Borren, & Tambs, 2008). One might even say that the so called IQ paradox is only created when one insists that the IQ trend be environmental in origin (Mingroni, 2004, p. 67). One difficulty in resolving this conflict is that environmental factors and quality of education can vary across different populations (Colom et al., 2005; Greenfield, 1998; Jensen, 1998; Sigman, 2000; Sigman, & Whaley, 1998).

An additional hypothesis is the improvement of health care, which states that because mothers who have better prenatal care and children who have better health care now than in the past, leads to a reduction in severe childhood illness. This might result in more time spent in educational pursuits and overall cognitive growth and cognitive growth (Lynn, 1998). The prenatal period is critical because the intrauterine environment can influence neurological development (Smotherman & Robinson, 1995). Parents who are aware of the importance of nutrition and neural stimulation may be a factor in higher IQ (Blair, Gamson, Thorne, & Baker, 2005; Jensen, 1998; Lynn, 1998; Neisser, 1998; Williams, 1998). As a result of improved education, parents who are more cognizant of the importance of early intellectual stimulation in their children have been discussed as a possible explanation of the FE (Belmont & Marolla, 1973). Elley (1992) hypothesizes that this is due to the greater prevalence of books at home in 1968 than in 1936. On the other hand, Dickens and Flynn (2001) stated that a relatively small environmental factor could lead to a large change in IQ, and there is a narrow window of time in which environmental effects on IQ are seen. For example, improving a child's IQ does not translate into a higher IQ in adulthood. To improve adult IQ, one would need enrichment at the adult level. However, Dickens and Flynn (2001) support at least some enrichment programs that provide a stimulating environment because the temporary IQ boosts may play a role in long-term achievement.

An Artificial Rise in Cognitive Ability

In contrast to theories of education, genetics, or nutrition, some scholars postulate that the rise in IQ scores is a psychometric artifact and that the way the test is measured has changed. Researchers have examined this theory through the use of latent variable models to determine if IQ changes are caused by real changes in ability (Beaujean & Osterlind, 2008; Beaujean & Sheng, 2010). Still, the question remains if general intelligence (g) has changed and the source of this rise in IQ scores (Flynn 1987; Flynn, 2006; Jensen, 1998; Kane & Oakland, 2000; Must, Must & Raudik, 2003b).

Findings from Flynn Effect Studies

United States

The FE has been examined extensively in the United States with varying results.

Between 1932 and 1978 Flynn (1984), discovered that American children gained 14.3 IQ points at a rate of .311 IQ points per year (Flynn, 1984; Flynn 1987). Kanaya, Scullin, and Ceci 2003, found that the magnitude of the FE can vary between WISC (Wechsler, 1949) norms. Their findings reported a 4 IQ point increase between the WISC (Wechsler, 1949) and WISC-R (Wechsler, 1974), and about a 6 point increase between the WISC-R (Wechsler, 1974) and WISC-III (Wechsler, 1991). In a different study, Lynn and Pagliari (1994) established that the mean IQ of children was increasing at a rate of .3 IQ points per year between 1972 and 1989. Similar findings were present among children from a national database who exhibited the FE at a consistent increase of .3 IQ points per year between 1978 and 2000 (Rodgers & Wänström, 2007). While the FE had been established in the average range of the IQ spectrum, other researchers sought to establish evidence of the FE among those with learning disabilities or intellectual disability.

In examining for a FE among children with intellectual disability, Kanaya, Scullin, and Ceci (2003), found that the children scored 5 to 6 IQ points higher on the WISC-R (Wechsler, 1974), than the WISC-III (Wechsler, 1991). Sanborn, Truscott, Phelps, and McDougal (2003) focused on children with a learning disability. Their results indicated a positive FE rate of 4.2 IQ points per decade between the WISC-R (Weschsler, 1974) and WISC-III (Weschsler, 1991).

The Flynn Effect Studied in Other Countries

The FE has been identified in developed and undeveloped countries with varying magnitudes (Colom, Andrés Pueyo & Juan-Espinosa, 1998; Colom, Juan-Espinosa, & García, 2001; Daley, Whaley, Sigman, Espinosa & Neumann, 2003; Flynn, 1987; Lynn, 1982; Lynn & Hampson, 1986; Rodgers & Wänström, 2007; Sundet, Barlaug & Torjussen, 2004; Teasdale & Owen, 1989; Teasdale & Owen, 2000). In Australia, children displayed gains around 3.7 IQ points per decade (Cocodia et al., 2003; Lynn, 2009; Nettelbeck & Wilson, 2004), French children gained 6 IQ points over 23 years (Bocéréan, Fischer, & Flieller, 2003; Bradmetz & Mathy, 2006), children from the Netherlands demonstrated an increase of .3 IQ points per year over a 25 year span (Resing & Tunteler, 2007), in Israel men and women displayed rates of gain of .6 IQ points per year on the Ravens Matrices from 1971 to 1984 (Flynn, 1998), and Sudanese, men displayed an IQ gain of 2.9 points per decade between 1964 and 2006 (Khaleefa, Abdelwahid, Abdulradi, & Lynn, 2008). One meta-analysis summarized the major findings of 508 studies on the FE among German-speaking countries. The studies consisted of 48,147 participants with a mean age of 42 years and covered a span of 37 years. They examined two vocabulary instruments that measure crystallized intelligence. The results indicated that the FE was observed among the crystallized domain with consistent IQ gains of approximately 3.5 points per decade. Additionally, the results suggest that the observed IQ gains are similar to the fluid intelligence gains reported in Flynn (2009) (Pietschnig, Voracek, & Formann, 2010).

Reverse Flynn Effect

Despite the evidence of gains in IQ scores, several countries such as Denmark, Norway, and Britain have seen IQ decrease over time. In Denmark, Teasdale and Owen (2005) investigated the FE using data gathered on Danish conscripted men tested on the Børge Priens Prøve (BPP, Rasch, 1980). The BPP is an instrument used to assess for cognitive abilities and appropriateness for conscription and involves four paper-and-pencil subtests (Teasdale, 2009). The test was administered in 1959 to 2004 to 18 year old male conscripts. For several decade time points, IQ gains averaged 3 points with the peak of IQ gains from 1988 to 1998. Interestingly, the results showed that the rate of gain in test performance slowed between 1959 and 1998 to about 1.3 IQ points and eventually diminished by 2000. A possible explanation of the decline in scores is the changing trends in the Denmark educational system. Teasdale and Owen (2005) analyzed the attendance rate at an Advanced Level College (ALC). Attendance rate at the college correlated with the decline in test performance (r = .82). According to the researchers this would suggest that the decline in attendance might contribute to the decline in test scores. Teasdale and Owen (2005) attributed other possible causes to the cessation of the FE such as the age of the BPP, unadulterated since its creation, and the fact that the BPP is a paper and pencil test that may be unfamiliar to those whose education has been dominated by computer use.

Other countries have shown diminished gains over time. Among Norwegian male conscripts were large IQ gains between the mid-1950s to early 1970 and then a decrease and stop in gains by the mid-1990s (Flynn, 1987; Flynn, 2009a; Sundet, Barlaug, & Torjussen, 2004). Between 1975 and 2000, British children displayed a decrease in scores on the Volume & Heaviness, a Piagetian instrument (Fitzgerald, Gray, & Snowden, 2007; NFER, 1979; Shayer, 2008; Shayer, Ginsburg, & Coe, 2007).

Estonia

History

Estonia is a small country located in northeastern Europe. The country covers 17,462 square miles, including islands, and a population of 1.34 million. Estonia is mainly industrial including machinery, building materials, and textiles. Estonia had been under Germanic rule until the end of the 15th century when Sweden took control of the country. Estonia came under Russian control when Sweden relinquished control at the Peace of Nystad in 1721. During the 1800s serfdom was abolished and peasants were allowed to own private property. Near the end of the 19th century during Russification, Estonia began to prosper and its first university, University of Tartu, was built in 1893. Estonia gained independence with the Russian Revolution in 1917 during World War I. Independence continued until 1939 with the German-Soviet Nonagression Pact. This agreement allowed Russia to incorporate Estonia into the U.S.S.R. Russian occupation continued until 1991 when Estonian independence was recognized by the Soviet Union (Encyclopædia Britannica, 2011; History of Nations, 2004).

Previous Work on Intelligence in Estonia

Lynn, Allik, Pullman and Leider (2002) examined whether the average intelligence of Estonia is similar to other European countries or if the average intelligence would match the level of IQ of the United States during the 1950s. The Standard Progressive Matrices was standardized on 2689 examinees with average ages of 12.4, 14.4, 16.1, and 17.8 years. The results showed that the average IQ was 100.2 for Estonia, which is similar to the British IQ of 100. The significance of this result shows that decades of low living standards do not impair the intelligence of a European population.

In 2003, Lynn, Pullman and Allik examined the IQs of junior school children from Estonia and compared them to IQ levels of adolescents. Participants had average ages of 7 though 11 years. The Standard Progressive Matrices was standardized on 1857 examinees. The results showed that the mean IQ of the junior school children was 98. The researchers deemed this close to the adolescent IQ of 100.2. This result displays that the IQ in Estonia is similar to the IQs of Britain and other European countries.

A year later in 2004, Pullman, Allik and Lynn examined IQ scores for 4874 Estonian children who were between 7 and 19 years old and compared them to scores of children in Britain and in Iceland. The Standard Progressive Matrices test was used. The Estonian IQ scores were higher at the youngest ages and older ages, but lower from first grade to age 12 compared with the other two countries. Possible explanations included schooling or sampling error, but it was most likely differences in intellectual development in the Estonian populations compared with the other countries.

Previous Work on the Flynn Effect in Estonia

The few studies on the FE can trace their basis to Tork's (1940) work on the examination of mental abilities of children in Estonia.

In 1934 Tork adapted the National Intelligence Test (NIT; Haggerty et al., 1920) to Estonia as part of his doctoral dissertation (Must, Must, & Raudik, 2003b). The purpose of his study was to examine the mental abilities of children and their influences (Strenze, 2006). His study was the first to systematically examine Estonian school children on a large scale (Pullman, 2005). Participants numbered about 3000 school children with ages ranging from 12 to 14 years old in grades 5 and 6 and came mainly from the city of Tartu. His results reported correlations of .5 - .6 between ability test scores and average school grades for 13 year old children (Tork, 1940; Pullman, 2005). Lynn, Allik, and Must (2000) reported that Torks study is relatively unknown to the intelligence literature due to the occupation of Russia. Russia at that time forbade academic work on intelligence and removed most copies of the book from circulation. Few copies currently exist because it was outlawed. As a result, the Estonia NIT was never translated to any other major language.

Must, Must, and Raudik (2003a) examined literacy scores of 522 Estonian school children who were 9 and 14 years old in 1999 and compared them to 1994 literacy scores from Estonia. They used the International Association for the Evaluation of Educational Achievement (IEA) literacy tests Binkely, Rust, and Wingless 1994, which is made of three subscales: Narratives, Expository, and Documents. The results showed that the 1999 cohort performed better than the 1994 cohort. The researchers determined that the literacy gains were due to other environmental and educational factors (Elley, 1992; Munch & Lundberg, 1994) and were not an increase in *g*. This was determined by a principle component analysis where the subtests were rank-ordered based on highest to lowest *g*-loading. The subtest Narratives ranked highest on *g*, but displayed a minimal change in score gain. The subtest Documents was ranked lowest, but displayed the greatest score gains.

Must et al. (2003b) investigated the rise in IQ scores in Estonia school children from 1934 to 1997. The instrument used was an updated version of Torks 1933 adaptation of the NIT (Haggerty et al., 1920). The sample consisted of 381 Estonian school children that were split into two groups consisting of 12 to 13 year olds in the 5th grade and 13 to 14 year olds in the 6th grade. The results indicated that there was an overall increase in scores from 1934 to 1997, in particular, verbal subtests such as Analogies and Sentence Completion and the Symbol-Number and Comparisons subtests. These subtests were not highly g-loaded, but displayed the greatest changes in scores that contributed to a FE.

The 1997 female sample and the 1997 12 to 13 year olds displayed the greatest test score increases (See Table A.1). The researchers hypothesize that one possible explanation for the increase in the females scores is that the Soviet socialist system put in place in the 1940s favored womens rights and education for women (Jacobs, 1996). An explanation for the gains among the younger students is that students had to mature at an earlier age when the Soviet system was in power. A second explanation is that the NIT changed between 1934 and 1997. A factor analysis of the factor structure in 1934 displayed a distinct single factor whereas the 1997 factor structure displayed an additional smaller second factor. Additionally, the researchers conducted a principle component analysis and found that when comparing the first principle components, the hypothesis of secular changes in IQ being on the g factor was not supported. The first principle component congruence coefficient was +.996 for the time period of 60 years, which was evidence of the general factors of the two matrices being almost identical. This means that there was not a change in the structure of g.

This conclusion was additionally supported by negative correlations between the two matrices that consist of the secular changes on the subtests and the rank order of the PCA. The average of these correlations was calculated and the mean Spearmans rank order correlation coefficient was -.40. Therefore, the FE was mostly evident in the least g-loaded tests similarly to Rustons (1999) study.

Must, te Nijenhuis, Must, Annelies, and van Vianen (2009) continued this line of study by examining the FE of Estonian school children. The school children were collected in three waves from 1933, 1997, and 2006 and were divided into age/grade homogenous samples for comparison. The 1933 cohort consisted of 899 school children in grades 4-6 with a mean age of 13.4 years. The 1997 cohort consisted of 361 school children in grades 7-8 with a mean age of 13.2 years. The 2006 cohort consisted of 913 school children in grades 6-8 with a mean age of 13.5 years. The instrument used was an updated version of Torks 1933 adaptation of the National Intelligence Test (Haggerty et al., 1920). The researchers used multi-group confirmatory factor analysis to investigate measurement invariance. In the cohort comparison from 1933 to 2006 only configural invariance held (i.e., factor structure was same for all groups) and between the cohorts from 1997 to 2006 only weak measurement invariance held (i.e., factor structure was same for all groups and the relationships between factors and items were the same for all groups). This is consistent with the results found in Wicherts et al. (2004). The researchers posit that the age/grade heterogeneity may be the cause for a lack of strict measurement invariance.

The results showed that from 1933 to 2006 both age groups displayed IQ gains of about 1.65 IQ points per decade. The youngest group showed a greater amount of gain than the older group. Interestingly, from 1997 to 2006 there was a gain of about 3 IQ points per decade almost double the rate from 1933 to 2006, with no differences between age groups. The results also showed that the interpretation of the test-scores changed over time.

Measuring the Flynn Effect

The FE is typically measured in one of two ways. First, two or more versions of the same instrument (or two different instruments normed at different time points) is administered to a single sample at one time point. Figure A.1a displays a path diagram of how this design can be visually conceptualized. An example of this design is when Covin (1976) administered both the WISC (Wechsler, 1949) and WISC-R (Wechsler, 1974) to children with learning difficulties. The second way to assess the FE is to administer one single instrument to similar samples at two or more different time points. Figure A.1b displays a graphic representation of this design.

An example of this design is when Flynn (1987) analyzed the Scholastic Aptitude Test (SAT, Educational Testing Service, 1926) scores from high school seniors from 1973 to 1985.

The FE is typically measured in, or converted to, the IQ metric (mean = 100; SD = 15) (e.g., Breslau, Dickens, Flynn, Peterson, & Lucia, 2006; Ceci, Scullin, & Kanaya, 2003; Dickens & Flynn, 2006; Flynn, 1984; Flynn, 1985; Flynn, 1987; Flynn, 1998; Flynn, 2006; Flynn & Weiss, 2007; Ceci & Kanaya, 2010; Kane & Oakland, 2000; Light & Chambers, 1958; Rodgers & Wänström, 2007; Sanborn, Truscott, Phelps, & McDougal, 2003; Simon & Clopton, 1984; Spitz, 1983; Spitz, 1986; Truscott & Frank, 2001; Zhou, Zhu, & Weiss, 2010). The most common method researchers use to examine for a FE is to compare mean differences in aggregated scores (e.g., Full scale IQ, Verbal IQ). In doing so, there is an implicit assumption that the compared test scores are measuring the same construct(s) the same way. If test scores are measuring differently across time though (i.e., they are non-invariant), then these score comparisons are futile. Beaujean and Sheng ((2012) make the following analogy: comparing means from non-invariant test scores is akin to comparing average temperatures at two different geographic locations with thermometers that use different scales. While mean differences could be due to different temperatures, they could also be the result of the scales having different origins (e.g., Fahrenheit vs. Rankine), different units (e.g., Kelvin vs. Rankine), or both (e.g., Fahrenheit vs. Kelvin).

Measurement Theories

The classical test theory (CTT) model is concerned about an examinee's true score, but investigates it using examinee's observed scores and error in measurement (i.e., Observed Score = True Score + Error; Crocker & Algina, 1986). While it does have methods for item analysis, they are all post hoc and not part of the CTT model

(Allen & Yen, 1979). Consequently, in CTT models item properties such as difficulty and discrimination are interwoven with examinee ability from which they cannot be separated (McDonald, 1999). That is, the sample of examinees taking a test has a direct influence on that tests item parameter estimates and, conversely, the items making up a given test have a direct influence on the examinee ability estimates. Consequently, from a CTT perspective, it is extremely difficult, if not impossible, to separate examinee ability changes (i.e., becoming higher or lower) from changes in the tests items properties (i.e., becoming more or less difficult and/or more or less discriminating).

Latent variable models (LVM) are an alternative to CTT. LVMs concurrently examine how examinee responses (either at the item or aggregate level) and the scores that comprise the test (either at the item or aggregate level) relate to the latent trait the test is designed to measure (Borsboom, 2005). The advantage of the latent variable perspective is the model involves estimating item parameters and examinee ability, but does so independently. Thus such models are able to differentiate changes that might occur in the latent trait from changes that might occur in the test itself. If an instrument is invariant over time, then the construct the test is designed to measure, as well as the way the test measures the construct, does not change and manifest scores (e.g., summed scores, norm-referenced scores) are directly comparable (Beaujean & Osterlind, 2008; Rupp & Zumbo, 2006).

Measurement Invariance

When researching the FE, investigators have typically only compared mean differences, but they take on the assumption that test scores allow them to make fair comparisons across groups. However, the validity of the tests scores is threatened if items operate differently among groups (Kane, 2006; Messick, 1988; Messick, 1989, Ployhart & Vandenberg, 2010).

Before comparing groups, it is important to assess that the instruments are measuring the same construct, the same way (Yoo, 2002; Campbell et al., 2008). If a researcher is not sure that the same construct is being measured at both time periods, one cannot be sure that the observed group mean differences is due to measuring different constructs or if there is a true difference between the two groups (Little, 1997, Steenkamp & Baumgartner, 1998; Thompson & Green, 2006; Vandenberg & Lance, 2000).

In order to determine if the FE was due to changes in test properties (as opposed to, or in addition to, change in the examinees), the instruments items would need to be examined if they operate the same way across time. This issue is called invariance (Bontempo & Hofer, 2007). If an item functions the same way for all examinees this means that invariance holds. If the item does not function the same way it is non-invariant (Borsboom, Mellenbergh, & van Heerden, 2002).

Invariance can be assessed through CTT or LVM. The CTT framework examines non-invariance post hoc whereas LVM can measure item non-invariance through the model (Borsboom, 2005). The CTT framework is limited in that it cannot differentiate if the FE is a true change in cognitive ability or if the FE is due to the intelligence tests changing over time (Chan, 1998). This is because the CTT framework cannot adequately represent a measured attribute. Additionally, theoretical concepts from the CTT model are similar to the functions that are used to analyze them, such as observed scores and true scores. However, the way the true score is defined within CTT, there is no mention or assumption of a latent construct. In order to determine if the FE is a true change in ability examining the latent construct and item functioning is better suited for latent variable modeling (Borsboom, 2005). Additionally, there are some questions that the CTT model cannot answer: Do people from different cultures interpret an item conceptually the same way? Does race, sex, ethnicity or some other individual difference impede people from responding similarly (Vandenberg & Lance, 2000)?

LVM has several advantages over CTT, when investigating invariance in an instruments items. First, LVM allows the differentiation between the observed test score and the latent trait they are intended to measure. When investigating for a FE, this allows the examiner to equate two unequal groups (Zimowski, (2003). Take, for example, an achievement test, and its developmentally appropriate versions, that are administered to a group of 3rd graders and 5th graders. These groups are unequal due to the amount and level of general information taught in the respective grades. To compare achievement test scores from these two groups the most difficult items on the 3rd grade test are mixed in with the 5th grade tests least difficult items and the 5th grade tests least difficult items are mixed with the 3rd grade tests most difficult items. The scores on these two tests can now be compared because of the common items that measure the same construct, are on both tests.

In the case of the FE, if two groups from two different time points take the same test and score differently on the latent trait, the items properties should remain the same after converting to the same scale. LVM can be used to take two unequal groups scores and convert them to a common scale that can then be directly compared (Kolen & Brennan, 2004; Osterlind & Everson, 2009).

Investigating Invariance Through Latent Variable Models

From a LVM perspective, multi- group confirmatory factor analysis (MG-CFA) has been used to test for measurement invariance (Horn & McArdle, 1992). CFA is a very general LVM that can handle both continuous indicators (traditional factor analysis) and categorical indicators (IRT) (Bartholomew & Knott, 1999). If there is (strict) invariance on the instrument among the groups, then any group differences on the instruments scores is due to group differences in the latent constructs of a test and not a result of measurement artifacts or cultural differences. The CFA process of measurement invariance consists of nested models that gradually become more constrained. The first step in this method is to determine if the same general factor structure is held for all groups. To determine this, a CFA is run on all samples. If the CFA model is held for all groups three additional invariance tests are conducted to determine if the factor structure functions differently for all groups. Three levels of measurement invariance that are tested consist of, configural, metric, and scalar invariance. These three levels are nested in an increasingly restrictive process (Cheung & Rensvold, 2002; Little, 1997).

Configural invariance tests if the overall factor structure is the same for all groups. It is concerned with the basic factor model consisting of the same number of factors and items functioning the same way across groups. This invariance model is established after the separate CFAs have been conducted. After configural invariance has been established, the next test is metric invariance (Campbell, Barry, Joe, & Finney, 2008).

Metric invariance examines if the relationships between factors and items are the same for all groups. This would indicate that all groups are interpreting items similarly (Byrne, 1998). Additionally, metric non-invariance could indicate items that one group may interpret differently than another (Chan, 1998). After metric invariance has been established, the last test is scalar invariance.

Scalar invariance examines if the factor loadings and intercepts are equivalent, meaning if the factor loadings and intercepts are invariant (do not change). This also checks that items are calibrated the same where item scales begin at the same zero point. Meaning, when the intercept is set at zero for the first group, this centers the common factor and sets all groups to a common scale.

Item Response Theory. IRT is a form of CFA (Bartholomew & Knott, 1999). Whereas traditional CFA models predict continuous indicators, IRT models predict item performance through item parameters that are related by the IRT model. While there are multiple IRT models (van der Linden & Hambleton, 1996), one of the most common models has two item parameters: item discrimination and item difficulty (see Equation 1). The item discrimination parameter (a) indicates how well an item can discriminate among examinees ability levels (Hambleton, Swaminathan, & Rogers, 1991; McDonald, 1999). Item difficulty (b) is the parameter that identifies where the question is in relation to the ability scale. In other words, the larger the difficulty value is, the more difficult an item, hence, the greater amount of the ability is needed to correctly answer the question (Hambleton, Swaminathan, & Rogers, 1991; McDonald, 1999; Brown, 2006).

IRT models are typically unidimensional, meaning that there is one latent ability or construct that takes items and measures their relationship with the measured ability or construct θ (Hambleton, Swaminathan, & Rogers, 1991). Examinees who have a high value on θ have a higher probability of answering the item correctly. The most basic IRT model is a one-parameter model/Rasch (1960) model. In this model only the difficulty parameter is allowed to differ across items; the discrimination parameter (a) is held equal across all items (Hambleton, Swaminathan, & Rogers, 1991). In the two-parameter model, item difficulty and discrimination relate to the single ability the test is measuring.

$$P_i(i = 1 \mid \theta_j) = \frac{1}{1 + e^{-a_1(\theta_j - b_i)}} \tag{2.1}$$

Where

$P_i(i = 1 \mid \theta_j)$ is the probability that a randomly chosen examinee, j, with ability θ_j answers item i correctly.

b_i is the item i's difficulty parameter.

a_i is the item i's discrimination parameter.

θ_j is the latent trait score for respondent j.

Examining Change. In examining change in a construct over time, researchers typically find one of three types of changes: alpha (α), beta (β), and gamma (γ) change (Millsap & Hartog, 1988; Chan, 1998; Bontempo & Hofer, 2007). Alpha change or true change is due to an actual change in the level of the construct over time. This would mean that a higher IQ score signifies an actual increase in cognitive ability.

Beta change is due to a different way of measuring the construct, while the construct itself does not change. Perhaps a test was re-normed, and the mean is now higher or lower than before. A higher score does not necessarily reflect better cognitive ability or more knowledge; rather, it is due to a change in measurement.

Gamma change is similar to non-invariance in that the structure of the construct, or meaning of the construct changes over time in addition to the number of factors that make up the construct changes (Bontempo & Hofer, 2007). This would be similar to a student taking a post-test that measured something different than the pre-test. It would be difficult to compare these scores.

Once again, the importance of invariance is that if the instrument is not invariant over time, or the way the items are interpreted change over time, or if the way the instrument is measured changes over time, the analysis can be misinterpreted as a true change instead of how the instrument is measured or the construct itself changes over time. The results would not be comparable across groups or across time if invariance were not tenable.

Current Study

The purpose of the current study is twofold. First, it will supplement the existing body of knowledge of the FE in Estonia. Second, it will supplement the research on the FE by analyzing the FE using an IRT instead of merely examining mean differences.

Research Questions

(1) Do the subtests on the B Scale of the National Intelligence Test (Computation, Information, Vocabulary, Analogies, and Comparison) exhibit (partial) measurement invariance over time in Estonia?

(2) If there is at least partial measurement invariance, then do the subtests, individually, exhibit a Flynn Effect?

CHAPTER THREE

Methods

Research Questions

(1) Do the subtests on the B Scale of the National Intelligence Test (Computation, Information, Vocabulary, Analogies, and Comparison) exhibit (partial) measurement invariance over time in Estonia?

(2) If there is at least partial measurement invariance, then do the subtests, individually, exhibit a Flynn Effect?

Research Design

This research project is an observational study using secondary data collected from the National Intelligence Test (Haggerty et al., 1920) in Estonia in 1934 and 2006.

Sample

The data comes from two samples of Estonian school children. The first sample comes from students ($n = 899$) who were part of the original Estonian NIT (Tork, 1940) norming sample. These students were gathered in 1933/36 (mean age: 13.4 years, SD age: 1.31). The second sample was gathered in 2006 ($n = 913$) (mean age: 13.5 years, SD age: .93). The second sample came from the same region as the first sample (Must, te Nijenhuis, Must, & van Vianen 2009). An additional sample was collected in 1997/98, but due to the small sample size ($n = 361$) was excluded from this analysis.

Instrument

In 1920, Haggerty, Terman, Thorndike, Whipple, and Yerkes developed the National Intelligence Test (NIT). Whipple (1921) reports that the NIT was created due to the success of the Army Alpha/Beta Intelligence Exams (Yerkes, 1921). Haggerty et al. (1920) wanted to apply the method of group intelligence examination used in the military during World War I to school children. The goal of the American NIT was to create a diverse set of tests in a single booklet that could be administered to any child who could read, write and participate in a group examination.

The NIT (Haggerty et al., 1920) is comprised of 10 subtests in two scales, Scale A and Scale B, each scale containing 5 subtests. The scales are designed to be diverse enough to stand independently as an intelligence exam. Scale A consists of the Arithmetic, Sentence completion, Synonyms-Antonyms, Symbol-digit, and Logical Selections subtests, while Scale B consists of the Computation, Information, Vocabulary, Analogies, and Comparisons subtests. The items of each subtest are arranged progressively from least difficult to most difficult (See Figure A.2 for sample items). Each NIT subtest is timed (taking 2 - 4 minutes) (Whipple, 1921). For the purpose of this study only Scale B was examined. This was done for three reasons. First, the scale measures a diversity of tasks, both strongly g-loaded (e.g., Comprehension-Knowledge, Fluid Reasoning) and weakly g-loaded (e.g., Visual-Processing). Second, due to the complexity involved with item-response modeling, the relatively computationally intensive methods involved in this analysis require much time. Third, once the results from Scale B are completed, an analysis of Scale A can serve as a later cross validation of the current study.

Juhan Tork (1940) adapted the American NIT to Estonia as part of his doctoral dissertation. Although Torks Estonian NIT norming sample used 6000 school children, Must et al. (2003b) claimed that Torks sample was not fully descriptive of the Estonian population due to 75% of his sample coming from only within the Tartumaa county, mostly in the city of Tartu. Additionally, Must et al. (2003b), asserts that the city of Tartu is not representative of a typical Estonian city since the countrys oldest university exists in Tartu.

Tork (1940) made some modifications to every subtest except the Symbol-Number subtest for his Estonian translation to include content specific to Estonia. Similar to the original NIT, the Estonian version comes with practice items on separate sheets. Test takers first complete the practice items as a group, and then complete the actual test items independently (Must et al., 2003b; Must et al., 2009).

Data Analysis

Software. Except where indicated, the data was analyzed using NOHARM (Fraser & McDonald, 1988, Mplus (Muthén & Muthén, 2010) and R (R Development Core Team, 2012), specifically the psych (Revelle, 2012), difR (Magis, Beland, & Raiche, 2012), and ltm (Rizopoulos, 2011) packages. All the relevant syntax is included in Appendices B-E.

Procedures

Classical Test Theory. I calculated CTT estimations of item difficulty and item discrimination for all items separately for each group as well as for all groups together. I used these in conjunction with statistical tests of differential item functioning (DIF) to determine if an item was invariant or non-invariant over time.

Item Difficulties. To examine how frequently the respondents were able to answer each item correctly, I calculated item difficulties for all five subtests, singly, both for all groups combined as well as within each grouping year (1934 and 2006).

Item Discriminations. The relationship between an item and the test from which it comes is one measure of how well the item is measuring the tests construct as well as how effectively the item is discriminating between individuals with high or low ability. To examine the relationship between a single item and the total subtest score, I calculated item total correlations for all subtests, both within each grouping year as well as for all groups combined. I calculated the item total correlations by correlating the item (scored as either correct, 1, or incorrect, 0) with its respective summed test score (McDonald, 1999).

Reliability. Cronbach's (1951) alpha was used as a CTT-measure of score reliability, and was estimated for each subtest in the 1934 and 2006 groups.

Flynn Effect. From the CTT perspective, the FE is just the difference in mean scores across time. Thus, I calculated the average summed score in each domain between the 1934 and 2006 groups.

Item Response Theory. FE estimation was done for each subtest using four steps. First, I determined which IRT model best fit a given subtest's data: Rasch, one-parameter (1P), or two-parameter (2P). Second, I found the items exhibiting differential item functioning (DIF) between the 1934 and 2006 groups in the difficulty parameter, discrimination parameter, or both. Third, I linked the items not exhibiting DIF between the two group, and the used the linked items to equate the 1934 and 2006 θ estimates (i.e., the trait the subtest is measuring). Last, I examined the θ scores between groups to see if there were any differences in the means, variances, or precession.

Dimensionality. Traditional IRT models assume that each test is measuring one construct (i.e., is unidimensional). To examine each subtests dimensionality, I conducted an exploratory factor analysis (EFA) for each of the subtests with the Normal Ogive Harmonic Analysis Robust Method (NOHARM; Fraser & McDonald, 1988) program using tetrachoric correlations (Harris, 1988) as the input. NOHARM is a free computer program used in nonlinear factor analysis to determine the fit of a unidimensional normal ogive model to the data (de Ayala, 2009; McDonald, 1999). McDonald (1999) compared NOHARM with other factor analysis methods (e.g., heuristic, weighted least squares) and found the methods used in the NOHARM program to the most robust method.

NOHARM uses two indices to determine model fit: the Tanaka and Huba 1985 goodness of fit index (GFI) and the Root Mean Square Residual (RMSR). The GFI acts like a coefficient of determination to explain the amount of variance of a given model. GFI values closer to one indicate a better fitting model, with values of .90 being considered acceptable, .95 being considered good, and a value of 1 considered a perfect fit of the data to the model (McDonald, 1999). The RMSR is an indicator of unconditional fit where a value of zero is a perfect fit, but values less than .08 are considered a good fit (Hu & Bentler, 1998).

Model Fit Comparison. I used Mplus (Muthén & Muthén, 2010) to fit 2-parameter (2P), 1-parameter (1P), and Rasch models to the item data for each subtest. The 2P model allowed the difficulty and discrimination parameters to be freely estimated, the 1P model constrained the discrimination parameters to be a single value across all subtest items while estimating all the difficulty parameters, and the Rasch model constrained the all discrimination parameters to unity, but estimated the difficulty parameters.

For each model, I used both Robust Maximum Likelihood (MLR), which produces a logistic IRT model, and robust weighed least squares (WLSMV; Flora & Curran, 2004), which produces a normal ogive IRT model to estimate the parameters. Model comparison was examined through several fit statistics: Comparative

Fit Index (CFI; Bentler, 1990), Tucker Lewis Index (TLI; Tucker & Lewis, 1973) and the root mean square error of approximation (RMSEA; Steiger & Lind, 1980). These indices were chosen due to their general robustness and relative insensitivity to sample size (but see Wirth & Edwards, 2007). The CFI and TLI are fit indices that range from 0 to 1 where values closer to 1 indicate better model fit. RMSEA is an indication of how well a model approximately fits the data, instead of examining the exact fit of a model (Brown, 2006). Studies suggest the following threshold cut-offs for good model fit: CFI $>$ .95, TLI $>$ .95, RMSEA $<$ .06 (Hu & Bentler, 1995; Brown, 2006).

Differential Item Function Analysis. Following Millsap (2011), I used multiple measures to determine if an item exhibited DIF: Mantel-Haenszel (MH; Holland & Thayer, 1988), logistic regression (Swaminathan & Rogers, 1990), transformed item difficulties (TID; Angoff & Ford, 1973), standardization (Std; Dorans & Kullick, 1983), and Breslow-Day (BD; Aguerri et al., 2009; Penfield, 2003). For the purposes of this study, I defined an item to exhibit DIF if at least three of the five DIF statistics indicated DIF was present.

DIF occurs when an item's parameters (difficulty or discrimination) differ across groups (Embretson & Reise, 2000). Several factors can influence DIF analysis (Osterlind & Everson, 2009), such as sample size, test length, α (i.e., pre-set Type I error rate for experiment), and the analysis method itself, as they can lead to inflation of the false positives (i.e., indicating that an item exhibits DIF when it dos not). Roussos and Stout (1996) show that the Mantel-Hanzel (Holland & Thayer, 1988) and logistic regression (Swaminathan & Rogers, 1990) procedures work well in DIF analysis while preserving typical Type 1 error rates. Güler and Penfield (2009), suggest that a minimum sample size be at least 200-250 respondents when using the Mantel-Hanzel and logistic regression methods to assess DIF.

Test length is important because the number of items in a test can influence the parameter estimate's precision (Li, Brooks, & Johanson, 2012; Narayanan & Swaminathan, 1996). Fidalgo, Mellenbergh, and Muñiz (2000) report that a longer test may improve the accuracy of the MH method, and consider test lengths of 20 items (short), 40 items (moderate), and 60 items (long). Magis and Facon (2012b) report that Type I errors generally increase as test length increases. On the other hand, Fidalgo, Mellenbergh, and Muñiz (2000) found that test length modestly affects power as test length increases. However there are conflicting reports. Fidalgo, Mellenbergh, and Muñiz (2000) cite several studies (Rogers & Swaminatham, 1993; Uttaro & Millsap, 1994) that simulated DIF analysis that used 20 to 40 items and found that test length had a slight effect on DIF identification, but (Uttaro & Millsap, 1994) only applied DIF analysis on a single item.

The Angoff and Ford (1973) TID has several advantages in DIF analysis such as not requiring a large sample size and is not overly complicated to compute. Magis and Facon (2012a) report that they have recently updated the procedure and found it superior to the original in that it is more efficient in controlling for Type I errors and DIF items. However, the recent updated method still requires more research (Magis & Facon, 2012b).

Osterlind and Everson (2009) report that the standardization method (Dorans & Kulick, 1983) was developed as an alternative to the MH method to be more interpretable instead of only statistical power like the MH. Dorans and Kulick (1986), state that the standardization method controls for differing ability leveles and item discrimination. The standardization indicates that one of the variables has been controlled for prior to group comparisons. Dorans and Kulick (1986) report that a downfall of the method is the large sample sizes required. However, it performed well on detecting DIF when applied to large SAT (Educational Testing Service, 1926) samples in their study.

Penfield (2003) applied the BD method to a simulated study of 40 items with sample sizes of 200, 500, and 1000. They compared the results of BD alone, in combination with MH in a combined decision rule (CDR), and compared the CDR against the logistic regression (Swaminathan & Rogers, 1990) method. Penfield (2003) found that both BD and CDR outperformed the logistic regression (Swaminathan & Rogers, 1990) method with alpha at .05. He recommends if using both BD and MH to lower alpha to at least .025.

In significance testing alpha is conventionally set at .05 because most researchers are unable to specify the tradeoffs of Type I and Type II errors. If these tradeoffs were known, a more specific alpha could be set. An alpha at .05 is used for researchers to initially determine if chance is a possible explanation for the observed results (Kirk, 2007). As this study involved assessing for DIF on multiple items across five subtests, I pre-set α to.005 in order to reduce chance of a Type I error.

Test Equating. The purpose of test equating is to convert item and ability estimates from different tests (or the same test given to different populations) to a common scale to be able to compare the examinees abilities (Baker, 1984). While one property of IRT is item and ability estimates are non-invariant across samples, the non-invariance only holds up to a linear transformation. That is, two groups could take the same test and have different IRT based estimates of the item and ability estimates; the estimates from these groups, however, are linearly related to each other. For example, say in group one item one has a discrimination of $\alpha 11 = 0.50$, and difficulty of $\delta 11 = 0.70$, but in group two, the same item has a discrimination of $\alpha 12 = 0.56$, and difficulty of $\delta 12 = 1.23$. Further, say there is an examinee whose ability is $\theta 1 = 0$ in group one, and an examinee whose ability is $\theta 2 = 0.60$ in group two. Plugging the value sets for each group into Equation 1 yields the same result: the probability of correctly answering the item is 0.41. Thus, while it may look like the parameter estimates are different, they are linearly related: $\theta 2 = 0.90(\theta 1){+}0.60$, $\alpha 12 = \alpha 11/0.90$, and $\delta 12 = 0.90(\delta 11){+}0.60$. The goal of test equating is the to find the linking values (i.e., 0.90 and 0.60 in this example) to put the parameter estimates on the same scale.

One type of equating is vertical equating, which is done when examinees from different groups take tests that measure the same construct, but differ in the contents difficulty. Typically, vertical equating is done by having a common set of items (i.e., anchor items) across all test versions, and then using those common items to determine the link needed to make all the test scores on the same scale (Baker, 1984; Kolen & Brennan, 2004).

For the current study, since the same test was administered to the groups at both time points, potentially all the items could be used as anchor items within a subtest. To determine if an item could be used as an anchor item, though, it had to exhibit no DIF (i.e., be invariant) across the two time groups (Byrne, Shavelson, & Muthén, 1989).

Thus, for this project I obtained item and ability estimates for each subtest, separately, for both time groups. I then examined the items for DIF, and categorized those not exhibiting DIF as anchor items. The items exhibiting DIF were used to within a time group to aid in estimating θ, whereas the anchor items were used to estimate θ as well as form the link needed to put the θ estimates from the two groups onto the same scale.

Differences in Cognitive Ability. I tested to see if there was any difference in cognitive ability between the two time groups by examining the average and variability of the subtest scores. From the CTT perspective, cognitive ability was estimated as the sum score for all the items in a subtest, while from the IRT perspective cognitive ability was the equated θ value.

CHAPTER FOUR

Results

This chapter is organized in two parts, the first part addresses the CTT and IRT analysis and the second part answers the research questions of this study.

Classical Test Theory

Inter-Item Correlations

Tables A.2-6 show the inter-item tetrachoric correlations for the 1934 group, Tables A.7-11 show the correlations for the 2006 group, and Tables A.12-16 show the correlations for the combined group. For the 1934 group, the inter-item correlations for the Computation subtest ranged from -.25 to .78, Information -.17 to .91, Vocabulary -.28 to .85, Analogies -.22 to .85, and Comparisons -.60 to 1.00. For the 2006 group the inter-item correlations for Computation ranged from -.50 to .67, Information -.46 to .57, Vocabulary -.37 to .76, Analogies -.39 to .89 and Comparisons -.34 to .98. Among the combined group, inter-item correlations for Computation ranged from -.21 to .68, Information, -.42 to 73, Vocabulary -.35 to .75, Analogies -.17 to .86, and Comparisons -.31 to .99.

A closer examination of these correlations showed that a pattern emerged where most of the inter-item correlations are positive, but some items especially those near the end of a subtest, have a negative correlation with other items. One possible reason for this is that examinee ran out of time and either did not answer the items or rapidly guessed at them before time elapsed (Lord, 1956; Lu & Sireci, 2007). Another possibility is that these items are more difficult than items that come earlier in the subtest, so examinees were, again, more likely to not answer the item or guess at the answer. As the dataset only contained item responses (i.e., correct or incorrect), it is not possible at this time to test these hypotheses.

Item-Total Correlations and Item Difficulties

Table A.17 gives the item-total correlations and percent correct for each item, i.e., CTT item discrimination and difficulties. All item-total correlations were positive except for three: For the combined group, item 14 of the Comparisons test (r = -.11), and in the 2006 group item 7 from the Information subtest (r = -.06) and item 1 from the Vocabulary subtest (r = -.004). As the magnitudes of the negative correlations are small, it is likely that these items did not function well within the subtest for the given sample, as opposed to measuring a construct bipolar to the one the items were intending to measure.

Item Response Theory

Dimensionality

Exploratory factor analysis performed in NOHARM (Fraser & McDonald, 1988) found that one factor fit the best (see Table A.18). While it is recommended that GFI values be over .95 as a good fit, a value of at least .90 is still considered acceptable. Recall that Hu and Bentler (1998) recommend RMSR values less than .08 for a good fitting model. The Computation and Analogies subtests had the best fitting values with GFI of .952 and .957, respectively and RMSR values of .007 and .012, respectively. The Information, Vocabulary, and Comparisons subtests did not fit as well as Computation and Analogies. Their GFI values were high enough to be deemed acceptable and their RMSR values were below the .08 threshold. Therefore, the GFI and RMSR values on all five subtests indicated good or acceptable model fit for one factor.

Model Fit

Three models were compared for data fit: the 2P, 1P, and Rasch models. The fit indices for the three models are in Table A.19. Based on the values of the CFI, TLI, and RMSEA the 2P model appears to fit the data better than the 1P and Rasch models for all the tests. Consequently, it was the model used for the subsequent IRT-based analyses.

Differential Item Functioning

The items that meet the criteria for DIF can be seen in Table A.20. As defined in sections 3.5.1.4 and 3.5.1.4.3, an item that displays DIF if its difficulty or discrimination parameter is different across groups (Embretson & Reise, 2000). I will illustrate this use the CTT item estimates (actual DIF estimation was done using IRT). Item 7 in the Information subtest has similar difficulties for both years 1934 (.92) and 2006 (.99), but has quite different discriminations (item-total correlations): r = .33 for the 1934 group and r = -.06 for the 2006 group. Conversely, item 18 in the Computation subtest is an item that functions well across both groups. This item has similar difficulties for 1934 (.02) and 2006 (.00), as well as similar discriminations: r = .56 for the 1934 group and r = .57 for the 2006 group.

Test 1: Computation

Nine of the 22 items (41%) for the Computation subtest (items 3, 8, 9, 10, 13, 14, 17, 19, 22) exhibited DIF. Specifically, seven items (items 8, 9, 10, 13, 14, 17, 19, and 22) were flagged for DIF by three of the five methods. Two items (items 3 and 10) were flagged for DIF by four of the five methods.

Test 2: Information

Nineteen of the 40 items (48%) in the Information subtest exhibited DIF (items 2, 4, 5, 8, 10, 11, 13, 14, 15, 18, 21, 22, 23, 24, 25, 27, 29, 30, 34). Four methods flagged the majority of items for DIF where the BD method appeared to be the most conservative flagging only three items for DIF. Four items (items 14, 18, 24, 30) were flagged for DIF by three of the five methods while 13 items (items 2, 4, 8, 10, 11, 13, 15, 21, 22, 25, 27, 29, 34) were flagged by four of the five methods; however, all five methods identified two items (items 5 and 23) as displaying DIF.

Test 3: Vocabulary

Fourteen of the 40 items (35%) in Vocabulary exhibited DIF (items 4, 14, 16, 19, 20, 22, 24, 25, 26, 29, 35, 37, 38, 40). Five items (items 4, 26, 35, 37, 38) were identified for DIF by three of the five methods while eight (items 14, 16, 19, 20, 22, 24, 25, 29) were identified by four of the five and one item (item 40) was identified by all five methods for DIF.

Test 4: Analogies

Fifteen of the 32 items (47%) in the Analogies subtest exhibited DIF (items 4, 7, 8, 9, 10, 14, 16, 17, 21, 22, 24, 27, 28, 30, 31). Seven items (items 10, 14, 17, 22, 24, 27, 31) were flagged for DIF by three of the five methods and eight items (items 4, 7, 8, 9, 16, 21, 28, 30) were flagged by four of the five methods. No items were flagged by all five DIF procedures.

Test 5: Comparisons

Eleven of the 50 (20%) items in the Comparisons subtest exhibited DIF (item 2, 12, 13, 14, 25, 36, 37, 38, 39, 40, 45). Three items (items 2, 25, 45) were flagged by three of the five methods while the remaining eight (items 12, 13, 14, 36, 37, 38, 39, 40) were flagged by four of the five methods. No items were flagged for DIF by all five methods.

Answering the Research Questions

Question One

The first research question asked: *Do the subtests on the B Scale (Computation, Information, Vocabulary, Analogies, and Comparison) of the National Intelligence Test (Haggerty et al., 1920) exhibit (partial) measurement invariance over time in Estonia?*

The DIF results showed that at least 52% of the items across all subtests had invariant items. The Comparisons subtest had the highest percentage of invariant items (80%), followed by Vocabulary (65%), Computations (59%), Analogies (53%), and Information (52%). Consequently, the answer is: *yes, all the NIT Scale B subtests have a majority of items that are non-invariant over time, and thus exhibit partial measurement invariance across the 1934 and 2006 time groups.*

Question Two

The second research question asked: *If there is at least partial measurement invariance, then do the subtests, individually, exhibit a Flynn Effect?*

Table A.21 displays the subtests score average and variability for the 1934 and 2006 groups using both the CTT and IRT parameter estimates. In addition, the table shows the CTT-version of score reliability (α) and the IRT-version of reliability (θ standard error and the standard error of the θ standard error) for the two groups. The θ standard error (θSE) is an information index. It determines an item's precision at any trait level and is calculated where the standard error on θ equals the inverse of the square root of the standard error of measurement ($\frac{1}{\sqrt{SEM}}$). Additionally, it is a comparable substitute to CTT reliability indicators, meaning that a smaller standard error of theta indicates a higher reliability (Flannery, Reise & Widaman, 1995).

For each subtest and for each parameter estimation method, I calculated Hedges (1981) *g*. Like Cohens (1992) *d*, Hedges (1981) *g* is an effect size (ES) estimator measured in standard deviation units. To calculate Hedges (1981) *g*, I subtracted the 1934 scores from the 2006 scores. This means that a positive number signified a gain over time (i.e., a Flynn Effect) and a negative number indicated a decrease over time (i.e., a reverse Flynn Effect). Table A.22 shows the ES per year (ES/Year), i.e., the FE per year between the 72 years between 1934 and 2006.

Computation

The average CTT-based summed scores showed a decrease over time from (ES = -.33 and ES/Year = .00). The average IRT-based θ scores showed a decrease as well, but at a smaller magnitude (ES = -.10 and ES/Year = .00). Both the CTT score variability and the IRT score variability changed very little from 1934 to 2006 (.18 and .07 SDs, respectively). Likewise, from both the CT and IRT perspective, the scores reliability stayed relatively constant from 1934 to 2006.

Information

The CTT-score average remained about the same from 1934 to 2006 (ES = -.03 and ES/Year = .00). However, variance decreased 2.43 SDs over the 72 years, meaning the 2006 participants were more homogenous in their general knowledge ability than those in 1934. IRT mean comparison showed that scores increased .44 SDs (ES/Year = .01), in addition to a decrease in variability of 0.49 SDs. The reliability of both the CTT and IRT scores remained relatively over the time.

Vocabulary

There was an increase in average CTT scores of 3.47 points from 1934 to 2006 or (ES = .74 and ES/Year = .01), with a moderate decrease in variability of 0.92 SDs. The IRT average scores increased .68 points (ES = .79 and ES/Year = .01), with only a small decrease in variability of 0.22 SDs. The IRT score reliability estimates was virtually identical across both time periods, while the CTT score reliability index showed a slight decrease of .10, going from .83 in 1934 to .73 in 2006.

Analogies

The CTT scores showed a large increased over the 72-year period (ES = 1.09 and ES/Year = .02), but showed only a slight change in variability of 0.28 SDs. Likewise, the IRT scores showed a large average score increase (ES = 1.02 and

ES/Year = .01), with only a small change in variability of .01 SDs. Both the θSE and α estimates showed that the score reliability for both time periods was the same.

Comparisons

The average CTT scores increased 10.53 points from 1934 to 2006 (ES = 1.71 and ES/Year = .02), that largest FE of all the subtests. The score variability showed a small decrease of .40 SDs. The IRT average scores increased 1.35 points over time (ES = 1.51 and ES/Year = .02), while the variability decreased slightly .26 SDs. The θSE showed that the Comparisons subtests IRT-based scores became more reliable over time, with a decrease of .14- the largest change in reliability among the subtests. Cronbach's (1951) α however, remained the same across the two time periods.

CHAPTER FIVE

Discussion

The purpose of the current study was to examine the FE through IRT methods on the NIT in Estonia using respondents from 1934 and 2006 by asking research questions:

(1) Do the subtests on the B Scale of the National Intelligence Test (Computation, Information, Vocabulary, Analogies, and Comparison) exhibit (partial) measurement invariance over time in Estonia?

(2) If there is at least partial measurement invariance, then do the subtests, individually, exhibit a Flynn Effect?

To answer the first question, the DIF analysis results shows that all the NIT Scale B subtests have a majority of items that are non-invariant over time and thus exhibit partial invariance. DIF analysis identified items whose item parameters were non-invariant between groups. To answer the second question, the results show that three of the individual subtests exhibited a FE (Vocabulary, Analogies, and Comparisons), one exhibited a reverse FE (Computation), and the last (Information) showed mixed results: a FE using IRT methods, but a reverse FE using CTT methods (see Table A.21).

Other Findings of Interest

Score Variability

In addition to examining mean differences, this study also examined differences in score variability. The Information subtest had the largest decrease in variance across time, the Vocabulary and Comparisons subtests showed small-to-moderate decreases in variability, and the Analogies and Computation subtests displayed little change in their variability.

One possible explanation for the decrease in variances across the subtests is that the people of Estonia are becoming more homogenous, meaning there are less distinct groups of people. Another possible explanation is the difference in sampling methods used in the 1934 and 2006 groups. One of the major criticisms of Torks (1940) work was that his sample was not an accurate representation of the population and was biased toward one region of the country (Must, Must, & Raudik, 2003b). Perhaps the sampling used for the 2006 data collection was more inclusive and allowed for a better representation of Estonian youths.

Score Reliability

Cronbach's (1951) α for 1934 and 2006 scores along with the standard error of theta can be seen in Table A.21. One should note that while α is a CTT-based index of score reliability, it is an average for an entire instrument for all examinees (i.e., it is the same value for all examinees completing the instrument at a given time). IRT recognizes that reliability is not the same throughout an entire instrument, so estimates score reliability using information (de Ayala, 2009) (in the Fisher sense of the term; Fisher, (1956) of the θ estimates. Thus, reliability (more specifically, standard error) is estimated separately for ability level (Flannery, Reise, & Widaman, 1995), with a lower standard error (θSE) indicating greater parameter precision and thus higher reliability. In Table A.21 the pattern of θ SE values generally maps onto the pattern of α values. For example the 2006 Computation subtest θSE = .61, the α = .56 compared to the 2006 Vocabulary subtest θSE = .37, α increased to .73.

Effect Sizes

The largest effect sizes were for the Analogies and Comparisons subtests. The magnitude of change on the Analogies and Comparisons subtests fall in step with previous research where the magnitude of the FE has been much larger on tests of fluid intelligence (Flynn & Weiss, 2007).

An interesting finding from the current study was the FE exhibited in the subtests that measure Comprehension-Knowledge: Vocabulary and Information. Knowledge-Comprehension is the knowledge of the culture that is incorporated by individuals vis-a‘-vis a process of acculturation (Newton & McGrew, 2010, p. 623). It is how skills and knowledge are applied to solving problems. Previous research suggests that the FE is predominantly found among measures of Fluid Intelligence (Colom, Andres Pueyo, & Juan-Espinosa, 1998; Flynn, 1987, 1998, 2006; Lynn 2009; Teasdale & Owen, 2000), with FE magnitude being much smaller on tests of Knowledge-Comprehension (Flynn, 2007). Thus, a positive FE in this area could mean that there was a large influx of information during the time between the assessments that has not typically been seen with more Western societies, such as could have occurred after the end of the USSRs annexation of Estonia, the growth of the Internet to anyone with a portable computer, and/or the joining the European Union to name a few.

While the results from the Vocabulary subtest are robust (i.e., similar effects sizes across both CTT and IRT estimations), the results from the Information subtest should be interpreted with some caution, as there was a stark contrast in CTT and IRT results. The CTT-based scores show a small decrease (-.03), but the IRT-based scores show a moderate-to-large increase (.44). One possible reason for the disparity between CTT and IRT method results for the Information subtest is the large number of items that were identified as exhibiting DIF for this subtest (19/40 items, approximately 48%).

Synthesis with Previous Research

The results from the current study have several similarities and differences with previous research. Must et al. (2009) found positive FEs for the Computation (.15), Information (.94), Vocabulary (.65), Analogies (1.81), and Comparisons (2.34) subtests. The current studys IRT results showed a similar pattern of positive gains (except for the small decrease on the Computation subtest) where the largest gains were on Analogies and Comparisons. CTT results displayed similar gains among the Vocabulary, Analogies, and Comparisons subtests with the greatest gains on Analogies and Comparisons.

There were differences between the current study and previous research, as well. Where Must et al. (2009) found positive gains on all subtests from Scale B, the current study found negative gains on the CTT analysis of the Computation and Information subtests, albeit Information had a smaller reverse FE. Additionally, the rates of gain from Must et al. (2009) were much larger than the results from the current study. This could be due to the smaller sample sizes for 1934 ($n = 270$) and 2006 ($n = 243$) that Must et al. (2009) analyzed, whereas the current study analyzed all examinees in 1934 ($n = 899$) and in 2006 ($n = 913$). However, the gains that Must et al. (2009) found used Hedges (1981) g and were only CTT calculations, whereas the current study used Hedges (1981) g for both IRT and CTT analysis. Ultimately, the findings of the current study confirm the findings of Must et al. (2009) that the FE is continuing in Estonia.

Study Strengths

The largest strength of this study is that this is the first implementation of IRT to investigate the FE in large sample of individuals on a multiple-subtest scale specifically developed to measure cognitive ability. IRT procedures allowed scores from unequal groups to be equated and directly compared. Since 1934 and 2006 items were equated and DIF methods removed non-invariant items, any FE or increase in scores can be attributed to some other explanation and not to a psychometric artifact or changes in an instruments properties.

While IRT has previously been used to examine the FE (Beaujean & Osterlind 2008; Beaujean & Sheng 2010; Flieller, 1988), they have all used instruments designed to measure a single aspects of cognitive ability (e.g., the Vocabulary test in the GSS, the Peabody Picture Vocabulary Test), not subtests from a larger battery designed to measure multiple aspects of cognitive ability.

Limitations

The study's main limitation was that because the test was administered in Estonia, I could not examine the stems for each item. Having this information could have aided in understanding why there were some negative inter-item correlations and item discriminations, as well as it would have helped to understand the reason why some items exhibited DIF and others did not.

An additional limitation of this was not knowing which items students guessed the answers to when the allotted time was running out, which ones they did not answer, and which ones they thought they knew the answer to but answered incorrectly.

A third limitation of this study was the DIF methods used. While I used multiple methods to determine if an item exhibited DIF, there are many methods for detecting DIF that I did not employ [e.g. Lord 1980, Raju (Raju, 1990), Likelihood-ratio (Thissen, Steinberg, & Wainer, 1988), Generalized MH (Penfield, 2001), Generalized logistic (Magis, Raiche, Beland, & Gerard, 2011), Generalized Lord (Kim, Cohen, & Park, 1995)]. The MH and Logistic methods identified the greatest amount of DIF items across all subtests, and perhaps were too stringent (i.e., exhibited too much Type I error) in determining DIF. For instance, the MH and Logistic methods flagged all nine items from the Computation subtest for DIF. In Information, MH, and Std methods flagged 18 and 19 items respectively while the Logistic method flagged 20 items for DIF. In the Vocabulary subtest the MH and Logistic method flagged all 14 items for DIF while the Std method flagged all but one for DIF and the BD identified only one item for DIF. In Analogies the MH, Logistic, and Std methods flagged the most items while the BD method flagged only three items for DIF. In Comparisons, MH and Logistic flagged all 11 items while

TID, and BD flagged 8 and 7 items, respectively, but the Std method flagged only four items.

Future Studies

The results of this study provide additional insight into the use of IRT in the role of analyzing the theory that the FE is due to a psychometric artifact. However, this research barely scratches the surface of what needs to be accomplished. Potential studies should try to analyze more FE longitudinal studies using item level data, or reanalyze previous research that uses longitudinal data. Item level data allows researchers a more precise look at item properties. Future studies should also consider a less stringent alpha in DIF analysis. While it is important to control for Type I error, a combination of stricter thresholds and liberal DIF methods may identify too many items as DIF. The removal of too many DIF items could affect item estimations. Additionally, future studies should consider using different DIF methods in IRT analysis. As discussed in previous sections, some methods could be too aggressive in identifying DIF items.

As mentioned before, IRT has been used in previous studies to analyze the FE those instruments have all measured single aspects of cognitive ability. Future studies should continue to use IRT to analyze instruments with large batteries that measure more than a single aspect. While the findings from the current study show that IRT adds a robust method to the analysis of the FE. Prospective studies should continue to use IRT to remove the possibility that a cause of the FE is due to a psychometric artifact. One future study could be an IRT analysis of Scale A as a cross-validation of the current study results. Additional studies that would be beneficial to the field could replicate this study using different data sets with different populations using IRT to compare findings of the FE.

APPENDICES

APPENDIX A

Figures and Tables

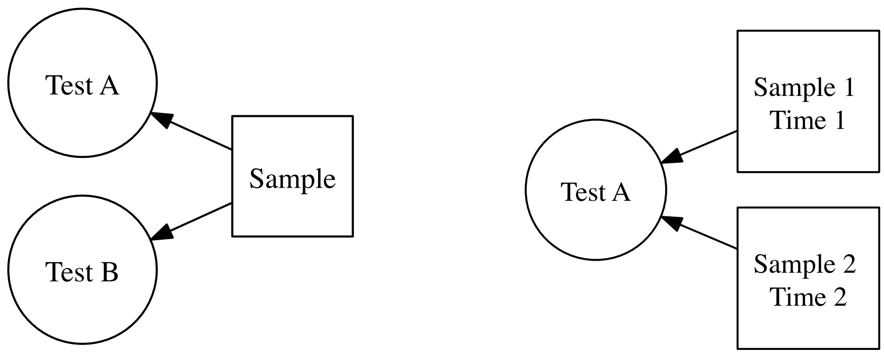

Figure A.1. Models illustrating the two ways to measure for a Flynn Effect *(a) A group of examinees are administered two versions of the same instrument. (b) A repeated measures model of administering the same instrument to similar groups at different time points.*

Test 1: Computation

Do this work in arithmetic as quickly as you can without making mistakes.

(2) Multiply 2 x 4 =

Test 2: Information

In each sentence draw a line under the one word that makes the sentence true.

(10) A bird of prey is the: pheasant hawk thrush plover

Test 3: Vocabulary

Draw a line under the right answer to each question.

(14) Do dogs hunt rabbits? Yes No

Test 4: Analogies

Read carefully the first three words in each line. Then read the last four and draw a line under the right one.

(1) gun – shoot ---- knife – run cut hat bird

Test 5: Comparisons

If the two things in a pair are the same, write S. If they are different, write D.

17360 16370

Figure A.2. Example items from the American *National Intelligence Test* (Haggerty et al., 1920).

Table A.1

Changes in the IQ test scores: The Flynn Effect, 1934/1936-1997/1998

Subtest	1934/1936 younger ($n = 72$) Mean	SD	1997/1998 younger ($n = 240$) Mean	SD	1934/1936 older ($n = 235$) Mean	SD	1997/1998 older ($n = 141$) Mean	SD
Computation	21.89	4.24	22.02	5.64	25.23	5.3	22.67	4.81
Information	19.22	5.29	18.72	5.69	27.11	5.96	20.02	4.9
Vocabulary	23.08	4.98	24.22	6.72	26.44	5.42	25.89	5.97
Analogies	10.75	4.65	16.7	6.25	14.81	5.76	18.27	5.37
Comparisons	23.17	6.40	32.5	8.66	28.3	8.59	33.86	8.53

Note. Table taken from Must, Must, and Raudik (2003b, p. 464).

Table A.2

Inter-Item Correlation 1934 National Intelligence Test Scale B Test 1: Computation

Item	1	2	3	4	5	6	7	8	9	10	11
1											
2	0.45										
3	-0.04	0.10									
4	0.24	0.27	0.35								
5	0.28	0.15	0.21	0.35							
6	0.06	0.16	0.23	0.09	0.07						
7	0.00	0.02	0.29	0.21	-0.06	0.26					
8	0.09	0.12	0.64	0.36	0.12	0.17	0.18				
9	0.01	0.08	0.11	0.22	-0.01	0.09	0.11	0.04			
10	-0.17	0.19	0.01	0.11	-0.09	0.01	0.22	0.10	0.06		
11	0.05	0.02	0.14	0.16	0.11	0.18	0.08	0.17	0.07	0.11	
12	0.03	0.35	0.17	0.12	0.02	0.06	0.08	0.06	0.63	0.19	0.18
13	0.02	0.06	0.16	0.01	-0.03	0.02	0.15	0.12	0.05	0.13	0.15
14	0.23	-0.04	-0.05	-0.03	0.00	0.06	0.14	-0.03	0.08	0.23	0.15
15	0.01	-0.03	-0.02	0.10	0.04	0.02	0.13	0.00	0.02	0.20	0.15
16	0.27	0.09	0.02	0.26	0.01	0.04	0.12	0.03	0.11	0.14	0.07
17	-0.07	0.11	-0.01	0.19	0.00	0.03	0.03	0.01	0.19	0.16	0.01
18	-0.23	-0.25	-0.20	0.06	0.03	0.21	-0.10	0.13	0.02	0.03	0.13
19	-0.10	0.15	0.11	0.09	-0.16	0.13	-0.11	0.17	0.22	0.00	0.07
20	-0.19	-0.07	0.24	-0.03	-0.09	0.01	0.08	0.17	0.19	0.20	-0.15
21	-0.12	-0.14	-0.13	0.03	0.13	0.02	-0.13	-0.07	0.02	0.10	0.08
22	-0.11	0.02	0.04	0.32	0.11	-0.03	-0.04	-0.07	0.08	-0.05	-0.01

(continued)

Item	12	13	14	15	16	17	18	19	20	21	22
13	0.12										
14	0.17	0.25									
15	0.11	0.21	0.49								
16	0.27	0.12	0.38	0.55							
17	0.26	0.15	0.33	0.52	0.78						
18	0.03	0.15	0.54	0.47	0.69	0.71					
19	0.20	-0.03	0.08	0.11	0.36	0.50	0.34				
20	0.21	0.31	0.19	0.22	0.65	0.66	0.47	0.58			
21	0.00	-0.09	0.10	0.07	0.29	0.49	0.20	0.63	0.32		
22	0.16	0.02	0.05	0.36	0.41	0.46	0.06	0.40	0.53	0.57	

Table A.3

Inter-Item Correlation 1934 National Intelligence Test Scale B Test 2: Information

Item	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1																				
2	0.36																			
3	0.38	0.20																		
4	0.30	0.29	0.39																	
5	0.30	0.26	0.03	0.00																
6	0.40	0.14	0.21	0.11	-0.05															
7	0.25	0.08	-0.17	0.08	0.29	0.30														
8	0.06	0.38	0.17	0.29	0.09	0.19	0.18													
9	-0.17	0.37	0.37	0.27	0.31	0.23	0.18	0.32												
10	0.24	0.08	0.31	0.30	0.13	0.16	0.22	0.25	0.38											
11	0.19	0.09	-0.17	0.22	-0.08	0.23	0.13	0.15	0.25	0.15										
12	0.25	0.35	0.54	0.14	0.22	0.36	0.31	0.18	0.27	0.17	0.22									
13	0.08	0.17	0.22	0.07	0.11	0.15	0.01	0.04	0.11	0.16	0.25	0.21								
14	0.13	0.13	0.36	0.12	0.04	0.33	0.04	0.17	0.14	0.19	0.06	0.23	0.05							
15	0.20	0.22	0.33	0.33	-0.02	0.17	0.13	0.21	0.29	0.24	0.22	0.29	0.17	0.06						
16	0.23	0.43	0.22	0.26	0.17	0.23	0.11	0.19	0.27	0.09	0.24	0.58	0.04	0.23	0.20					
17	0.22	0.25	0.53	0.23	0.20	0.19	0.01	0.31	0.39	0.31	0.40	0.42	0.21	0.19	0.32	0.24				
18	-0.06	0.16	0.08	0.23	0.06	0.16	0.14	0.20	0.28	0.14	0.23	0.23	0.28	0.11	0.25	0.36	0.38			
19	0.27	0.16	0.35	0.23	0.23	0.24	0.23	0.20	0.21	0.40	0.32	0.52	0.28	0.25	0.41	0.44	0.53	0.62		
20	0.33	0.24	0.27	0.13	0.03	0.25	0.07	0.19	0.27	0.09	0.28	0.32	0.12	0.14	0.27	0.29	0.40	0.31	0.59	
21	0.34	0.20	0.33	0.23	0.19	0.19	0.11	0.25	0.19	0.32	0.34	0.50	0.24	0.26	0.34	0.33	0.46	0.42	0.87	0.46
22	0.18	0.35	0.15	0.15	0.09	0.29	0.13	0.28	0.26	0.26	0.32	0.43	0.23	0.30	0.23	0.44	0.38	0.39	0.72	0.45
23	0.30	0.36	0.36	0.17	0.03	0.30	0.16	0.28	0.27	0.27	0.26	0.50	0.16	0.27	0.32	0.39	0.49	0.40	0.80	0.51
24	0.10	0.10	0.35	0.03	-0.14	0.07	0.08	0.12	0.16	0.09	0.31	0.18	0.17	0.03	0.20	0.17	0.31	0.26	0.53	0.28
25	0.30	0.14	0.08	0.17	0.05	0.21	0.26	0.26	0.21	0.25	0.22	0.39	0.17	0.10	0.26	0.27	0.31	0.21	0.54	0.40
26	0.31	0.21	0.35	0.15	0.17	0.20	0.20	0.20	0.21	0.22	0.14	0.47	0.18	0.23	0.26	0.29	0.31	0.32	0.72	0.37

(continued)

Item	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
27	0.26	0.19	0.43	0.13	0.17	0.27	0.22	0.28	0.19	0.26	0.22	0.51	0.10	0.29	0.24	0.35	0.41	0.34	0.73	0.37
28	0.49	0.28	0.35	0.20	0.15	0.31	0.31	0.25	0.17	0.33	0.23	0.35	0.14	0.44	0.15	0.42	0.34	0.32	0.54	0.33
29	0.44	0.27	0.40	0.21	0.08	0.33	0.17	0.23	0.24	0.23	0.23	0.51	0.17	0.21	0.17	0.38	0.32	0.27	0.51	0.31
30	0.34	0.30	0.27	0.11	0.13	0.28	0.32	0.19	0.25	0.27	0.31	0.48	0.20	0.23	0.17	0.39	0.43	0.31	0.66	0.43
31	0.37	0.28	0.40	0.03	0.23	0.33	0.29	0.15	0.19	0.24	0.31	0.43	0.24	0.20	0.22	0.36	0.38	0.33	0.67	0.51
32	0.24	0.30	0.20	0.19	0.37	0.19	0.13	0.37	0.25	0.11	0.22	0.31	0.25	0.15	0.12	0.21	0.35	0.25	0.50	0.26
33	0.40	0.29	0.22	0.23	0.15	0.37	0.14	0.24	0.24	0.22	0.27	0.41	0.21	0.22	0.19	0.28	0.34	0.33	0.54	0.40
34	0.22	0.35	0.38	0.05	0.34	0.31	0.13	0.15	0.15	0.19	0.25	0.34	0.17	0.27	0.21	0.31	0.37	0.25	0.52	0.43
35	0.11	0.30	0.31	0.13	0.37	0.38	0.18	0.13	0.33	0.17	0.42	0.24	0.34	0.24	0.27	0.27	0.43	0.36	0.40	0.39
36	0.19	0.32	0.36	0.07	0.15	0.32	0.31	0.31	0.31	0.34	0.39	0.48	0.23	0.19	0.29	0.37	0.45	0.37	0.55	0.50
37	0.08	0.19	0.32	0.03	0.12	0.33	0.18	0.20	0.21	0.27	0.30	0.23	0.13	0.30	0.19	0.36	0.40	0.30	0.51	0.46
38	0.27	0.29	0.34	0.08	0.17	0.35	0.31	0.29	0.22	0.28	0.22	0.34	0.17	0.27	0.17	0.39	0.37	0.31	0.48	0.45
39	0.22	0.21	0.30	0.16	0.25	0.30	0.24	0.14	0.20	0.24	0.37	0.42	0.17	0.27	0.14	0.33	0.41	0.23	0.52	0.52
40	0.10	0.27	0.18	0.04	0.13	0.20	0.09	0.11	0.05	0.19	0.16	0.16	0.03	0.24	0.01	0.38	0.37	0.17	0.46	0.35

(continued)

Item	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
21																				
22	0.69																			
23	0.77	0.67																		
24	0.46	0.35	0.47																	
25	0.60	0.41	0.49	0.29																
26	0.63	0.51	0.69	0.43	0.58															
27	0.64	0.55	0.72	0.37	0.54	0.71														
28	0.45	0.50	0.55	0.25	0.26	0.50	0.58													
29	0.48	0.36	0.54	0.25	0.37	0.52	0.56	0.39												
30	0.63	0.60	0.68	0.40	0.47	0.66	0.69	0.58	0.69											
31	0.63	0.51	0.66	0.49	0.49	0.57	0.65	0.55	0.57	0.91										
32	0.54	0.44	0.56	0.22	0.37	0.29	0.46	0.40	0.37	0.80	0.67									
33	0.51	0.56	0.57	0.26	0.44	0.53	0.54	0.40	0.50	0.82	0.77	0.54								
34	0.53	0.51	0.61	0.37	0.37	0.58	0.62	0.37	0.55	0.83	0.82	0.46	0.71							
35	0.44	0.57	0.70	0.42	0.37	0.61	0.54	0.38	0.41	0.75	0.75	0.40	0.68	0.72						
36	0.53	0.59	0.66	0.35	0.47	0.62	0.69	0.46	0.47	0.85	0.84	0.52	0.76	0.81	0.76					
37	0.46	0.49	0.56	0.36	0.41	0.48	0.63	0.48	0.38	0.78	0.78	0.42	0.67	0.77	0.75	0.87				
38	0.40	0.47	0.55	0.27	0.35	0.51	0.53	0.43	0.47	0.76	0.75	0.40	0.61	0.70	0.62	0.83	0.84			
39	0.49	0.60	0.59	0.36	0.37	0.49	0.53	0.44	0.41	0.80	0.76	0.43	0.67	0.74	0.70	0.83	0.86	0.78		
40	0.37	0.34	0.56	0.19	0.23	0.47	0.36	0.34	0.35	0.61	0.65	0.31	0.45	0.73	0.39	0.63	0.65	0.62	0.72	

Table A.4

Inter-Item Correlation 1934 National Intelligence Test Scale B Test 3: Vocabulary

Item	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1																				
2	0.51																			
3	0.61	0.37																		
4	0.55	0.51	0.42																	
5	0.32	0.82	0.02	0.16																
6	0.21	0.85	0.07	0.37	0.76															
7	0.24	0.19	0.10	0.24	0.00	0.05														
8	0.09	0.75	0.08	0.09	0.60	0.63	0.43													
9	0.43	0.18	0.08	0.23	0.31	0.29	0.42	0.12												
10	0.12	0.07	0.11	0.12	0.00	-0.08	0.25	0.17	0.31											
11	0.38	0.17	0.08	0.22	0.40	0.17	0.38	0.40	0.19	0.34										
12	0.13	0.72	0.12	0.13	0.70	0.68	0.27	0.59	0.18	-0.04	0.64									
13	0.31	0.10	0.32	0.15	0.04	0.09	0.13	-0.05	-0.05	0.14	0.27	0.16								
14	0.62	0.22	0.28	0.27	0.18	0.08	0.11	0.26	0.38	0.30	0.54	0.31	0.16							
15	0.24	0.17	-0.12	0.24	0.26	-0.04	0.00	-0.20	0.32	0.31	-0.02	-0.15	-0.02	0.26						
16	0.17	0.69	-0.12	0.36	0.56	0.49	0.14	0.38	0.17	-0.27	0.11	0.45	-0.24	0.18	0.20					
17	0.19	0.63	0.19	0.04	0.58	0.46	-0.13	0.46	0.06	0.00	0.14	0.48	0.05	0.21	0.30	0.54				
18	-0.10	0.13	-0.28	-0.10	0.27	0.07	0.05	0.33	0.00	-0.21	0.10	0.26	-0.12	-0.11	0.02	0.22	0.23			
19	0.25	0.51	0.16	0.39	0.42	0.56	0.02	0.24	0.27	-0.05	0.18	0.39	0.20	0.07	0.13	0.49	0.58	0.18		
20	0.31	-0.08	0.24	-0.02	0.23	0.14	0.12	0.12	0.12	0.14	0.34	0.20	0.15	0.18	0.00	0.30	0.20	0.18	0.33	
21	0.35	0.47	0.21	0.35	0.57	0.37	-0.03	0.26	0.13	0.04	0.34	0.34	0.30	0.23	0.27	0.41	0.35	0.21	0.31	
22	0.28	0.21	0.36	0.09	0.50	0.28	0.25	0.08	0.18	0.12	0.36	0.35	0.18	0.10	0.28	0.30	0.19	0.21	0.26	
23	0.16	0.09	0.24	0.30	0.40	0.14	-0.07	0.23	0.23	0.04	0.39	0.20	0.20	0.09	0.23	0.16	0.17	0.15	0.20	
24	0.40	0.69	0.34	0.26	0.55	0.43	-0.13	0.36	0.10	-0.05	0.45	0.44	0.27	0.28	0.06	0.41	0.36	0.23	0.39	
25	0.16	0.38	0.21	0.16	0.31	0.36	0.03	0.26	0.07	0.11	0.34	0.25	0.11	0.17	0.09	0.25	0.26	0.09	0.28	
26	0.20	-0.24	0.24	0.03	-0.09	0.01	-0.05	0.05	0.17	0.19	0.13	-0.04	0.08	-0.13	0.03	0.03	0.03	0.19	-0.05	

(continued)

Item	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
27	0.16	0.15	0.20	0.44	0.19	0.16	0.01	0.00	0.08	0.05	0.30	0.08	0.19	0.03	0.27	0.22	0.17	0.12	0.26
28	0.34	0.41	0.19	0.34	0.34	0.22	0.13	0.25	0.09	0.08	0.35	0.35	0.33	0.01	0.14	0.33	0.32	0.09	0.22
29	0.35	-0.11	0.39	0.22	-0.15	0.04	-0.02	0.05	0.06	0.13	0.15	-0.09	0.21	-0.17	0.14	0.11	0.15	0.04	0.07
30	0.18	0.03	0.30	0.18	0.10	-0.08	0.06	0.03	0.16	0.11	0.36	0.11	0.22	0.19	0.17	0.17	0.26	0.15	0.23
31	0.43	-0.11	0.26	0.43	0.10	-0.02	0.09	0.07	0.03	0.03	0.13	0.03	0.18	-0.07	0.18	0.14	0.17	0.12	0.24
32	0.38	0.32	0.04	0.38	0.43	0.37	-0.04	0.16	0.07	0.06	0.47	0.30	0.25	0.18	0.13	0.22	0.20	0.17	0.13
33	0.31	0.39	0.23	0.18	0.48	0.40	0.10	0.29	0.25	0.04	0.38	0.39	0.34	-0.10	0.19	0.23	0.34	0.13	0.25
34	0.24	0.30	0.40	0.07	0.21	0.19	0.14	0.10	0.18	0.03	0.10	0.27	0.23	-0.08	0.18	0.06	0.17	0.06	0.17
35	0.27	0.04	0.18	0.27	0.15	0.08	-0.05	0.34	0.13	0.37	0.45	0.15	-0.04	0.05	0.18	0.33	0.30	0.12	0.11
36	0.11	0.44	0.22	0.11	0.38	0.17	0.11	0.19	0.31	0.09	0.29	0.27	0.13	0.12	0.25	0.16	0.23	0.12	0.22
37	0.22	0.27	0.25	0.09	0.46	0.31	0.28	0.10	0.21	0.31	0.07	0.29	-0.02	-0.02	0.13	0.13	0.13	0.06	0.27
38	-0.03	-0.15	0.13	0.11	0.11	0.05	0.16	0.13	0.14	0.01	-0.07	0.07	0.26	-0.24	0.15	0.16	0.28	0.09	0.25
39	0.18	0.07	0.11	0.00	0.34	0.05	0.15	0.21	0.14	0.07	0.39	0.25	0.15	0.00	0.15	0.07	0.34	0.06	0.29
40	0.17	-0.06	0.44	-0.01	0.34	0.04	0.36	0.08	0.17	0.02	0.39	0.19	0.27	-0.16	0.20	0.06	0.21	0.12	0.19

(continued)

Item	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40
20																					
21	0.39																				
22	0.35	0.59																			
23	0.33	0.53	0.73																		
24	0.33	0.53	0.57	0.56																	
25	0.23	0.27	0.31	0.42	0.55																
26	0.23	0.23	0.23	0.31	0.25	0.30															
27	0.36	0.55	0.44	0.45	0.44	0.26	0.28														
28	0.20	0.43	0.36	0.38	0.39	0.32	0.14	0.38													
29	0.19	0.20	0.20	0.21	0.31	0.2	0.24	0.44	0.25												
30	0.29	0.32	0.36	0.40	0.36	0.22	0.24	0.45	0.37	0.36											
31	0.31	0.24	0.36	0.30	0.36	0.14	0.24	0.35	0.32	0.42	0.43										
32	0.07	0.15	0.26	0.30	0.30	0.24	0.03	0.21	0.36	0.22	0.27	0.31									
33	0.27	0.40	0.44	0.43	0.40	0.33	0.22	0.44	0.45	0.41	0.50	0.55	0.56								
34	0.21	0.32	0.33	0.41	0.33	0.36	0.29	0.33	0.33	0.34	0.37	0.45	0.40	0.72							
35	0.08	0.27	0.25	0.44	0.26	0.25	0.16	0.39	0.25	0.37	0.38	0.45	0.32	0.66	0.50						
36	0.19	0.35	0.35	0.51	0.37	0.34	0.23	0.40	0.31	0.38	0.43	0.44	0.43	0.80	0.73	0.57					
37	0.11	0.27	0.36	0.32	0.29	0.24	0.17	0.27	0.20	0.34	0.23	0.32	0.36	0.62	0.53	0.38	0.71				
38	0.13	0.30	0.32	0.59	0.35	0.21	0.25	0.32	0.19	0.27	0.40	0.42	0.20	0.59	0.53	0.31	0.64	0.47			
39	0.23	0.28	0.29	0.41	0.30	0.18	0.11	0.33	0.28	0.33	0.37	0.39	0.34	0.74	0.51	0.40	0.74	0.64	0.68		
40	0.25	0.35	0.43	0.52	0.29	0.19	0.18	0.37	0.25	0.34	0.44	0.39	0.34	0.73	0.51	0.47	0.72	0.55	0.68	0.84	

Table A.5

Inter-Item Correlation 1934 National Intelligence Test Scale B Test 4: Analogies

Item	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1																			
2	0.16																		
3	0.53	0.16																	
4	0.06	0.32	0.01																
5	0.48	0.32	0.24	0.47															
6	0.50	0.33	0.38	0.17	0.35														
7	0.54	0.13	0.22	0.20	0.41	0.38													
8	0.09	0.19	0.14	0.31	0.29	0.18	0.27												
9	0.25	0.31	0.22	0.21	0.20	0.34	0.19	0.12											
10	0.36	0.33	0.22	0.29	0.42	0.32	0.34	0.33	0.32										
11	0.14	0.00	0.22	-0.03	0.13	0.10	0.05	0.06	0.05	0.09									
12	0.35	0.45	0.20	0.47	0.56	0.46	0.42	0.32	0.48	0.52	0.03								
13	0.31	0.23	0.25	0.33	0.41	0.35	0.24	0.36	0.23	0.32	0.21	0.43							
14	0.37	0.35	0.33	0.39	0.52	0.46	0.33	0.40	0.19	0.40	0.14	0.59	0.52						
15	0.40	0.29	0.15	0.41	0.48	0.38	0.30	0.38	0.25	0.39	0.00	0.55	0.34	0.67					
16	0.44	0.31	0.38	0.33	0.46	0.48	0.34	0.33	0.19	0.32	0.10	0.47	0.53	0.73	0.57				
17	0.52	0.33	0.20	0.36	0.53	0.42	0.30	0.40	0.23	0.42	-0.01	0.54	0.39	0.63	0.70	0.59			
18	0.46	0.35	0.17	0.40	0.40	0.44	0.32	0.37	0.27	0.38	0.02	0.56	0.44	0.55	0.56	0.60	0.67		
19	0.23	0.33	0.25	0.30	0.39	0.23	0.22	0.36	0.25	0.20	0.01	0.50	0.47	0.56	0.52	0.55	0.50	0.55	
20	0.32	0.38	0.20	0.35	0.51	0.41	0.39	0.33	0.25	0.34	-0.17	0.64	0.47	0.63	0.63	0.60	0.67	0.64	0.63
21	0.50	0.18	0.17	0.12	0.29	0.36	0.30	0.23	0.18	0.21	-0.09	0.29	0.24	0.32	0.32	0.43	0.44	0.35	0.34
22	0.43	0.05	0.22	0.17	0.21	0.17	0.24	0.27	0.10	0.34	0.12	0.32	0.33	0.41	0.26	0.44	0.33	0.46	0.30

(continued)

Item	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
23	0.38	0.32	0.13	0.39	0.53	0.42	0.44	0.30	0.26	0.36	-0.10	0.66	0.47	0.58	0.54	0.53	0.58	0.54	0.44
24	0.36	0.22	0.08	0.29	0.33	0.36	0.27	0.26	0.31	0.32	-0.01	0.52	0.40	0.47	0.44	0.50	0.54	0.44	0.46
25	0.46	0.21	0.15	0.27	0.36	0.45	0.34	0.37	0.18	0.31	-0.07	0.54	0.43	0.52	0.50	0.48	0.51	0.51	0.44
26	0.41	0.25	0.10	0.36	0.41	0.36	0.30	0.36	0.22	0.37	0.04	0.65	0.44	0.55	0.48	0.49	0.52	0.47	0.45
27	0.16	0.25	0.11	0.31	0.36	0.42	0.41	0.34	0.24	0.38	-0.08	0.50	0.42	0.60	0.54	0.50	0.51	0.49	0.46
28	0.33	0.06	0.16	0.03	0.16	0.35	0.23	0.13	0.11	0.16	0.06	0.08	0.31	0.28	0.15	0.42	0.13	0.31	0.22
29	0.33	0.13	0.05	0.16	0.17	0.42	0.16	0.19	0.13	0.14	-0.08	0.24	0.34	0.30	0.30	0.36	0.28	0.30	0.28
30	0.02	0.09	-0.22	-0.05	0.11	0.29	-0.15	0.06	0.27	0.26	0.08	-0.02	0.13	0.09	0.40	0.23	0.11	0.26	0.21
31	0.30	-0.04	0.02	0.00	0.12	0.20	0.25	0.04	0.12	0.01	0.06	0.01	0.33	0.14	0.20	0.30	-0.01	0.08	0.15
32	0.34	0.01	0.08	-0.10	-0.03	0.21	0.00	0.05	0.01	-0.06	0.07	-0.12	0.14	0.12	0.17	0.19	0.08	0.15	0.10

(continued)

Item	20	21	22	23	24	25	26	27	28	29	30	31	32
20													
21	0.57												
22	0.45	0.27											
23	0.85	0.49	0.50										
24	0.66	0.41	0.37	0.71									
25	0.66	0.42	0.40	0.69	0.68								
26	0.66	0.36	0.45	0.71	0.65	0.78							
27	0.68	0.31	0.43	0.61	0.60	0.67	0.75						
28	0.24	0.44	0.27	0.34	0.34	0.43	0.57	0.56					
29	0.32	0.35	0.37	0.46	0.46	0.48	0.52	0.66	0.75				
30	0.00	0.56	-0.02	0.22	0.16	0.07	0.12	0.14	0.73	0.61			
31	0.13	0.30	0.21	0.31	0.32	0.35	0.33	0.35	0.67	0.58	0.53		
32	0.05	0.27	0.18	0.15	0.21	0.24	0.17	0.28	0.63	0.60	0.75	0.63	

Table A.6

Inter-Item Correlation 1934 National Intelligence Test Scale B Test 5: Comparisons

Item	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1																			
2	0.63																		
3	0.87	0.45																	
4	0.50	-0.05	0.17																
5	0.29	0.24	0.16	0.43															
6	0.31	0.07	0.19	0.18	0.00														
7	0.05	0.29	0.21	0.04	0.25	0.62													
8	0.56	0.02	0.41	0.05	0.05	0.93	0.66												
9	0.38	0.05	0.25	0.08	0.08	0.96	0.77	0.98											
10	0.17	0.34	0.04	-0.08	0.32	0.72	0.79	0.78	0.84										
11	0.52	-0.03	0.19	0.10	0.10	0.72	0.47	0.78	0.81	0.58									
12	0.13	0.24	0.07	0.00	0.26	0.45	0.46	0.48	0.49	0.42	0.54								
13	0.40	0.37	0.29	0.21	0.29	0.48	0.40	0.47	0.51	0.38	0.65	0.55							
14	0.25	-0.02	0.12	0.16	0.15	0.65	0.39	0.70	0.74	0.52	0.89	0.43	0.55						
15	0.32	0.15	0.19	0.19	0.25	0.73	0.45	0.78	0.82	0.61	0.92	0.61	0.74	0.92					
16	0.57	0.14	0.24	0.07	-0.09	0.52	0.12	0.53	0.56	0.32	0.52	0.12	0.29	0.56	0.72				
17	0.36	0.30	0.24	-0.11	0.26	0.42	0.21	0.39	0.45	0.29	0.31	0.23	0.27	0.34	0.46	0.58			
18	0.35	0.24	0.33	0.04	0.22	0.33	0.29	0.20	0.34	0.29	0.21	0.27	0.30	0.29	0.43	0.44	0.48		
19	0.30	0.04	0.17	0.16	-0.03	0.31	0.16	0.37	0.41	0.11	0.37	0.07	0.18	0.44	0.58	0.80	0.38	0.61	
20	0.22	0.08	0.09	0.12	-0.03	0.20	0.13	0.32	0.31	0.08	0.30	-0.05	0.08	0.35	0.50	0.68	0.16	0.47	0.93
21	0.07	0.25	-0.07	-0.03	0.21	0.57	0.64	0.68	0.72	0.71	0.49	0.33	0.26	0.46	0.58	0.53	0.29	0.37	0.78
22	0.13	-0.04	0.17	0.06	-0.07	0.60	0.26	0.72	0.75	0.36	0.50	0.11	0.13	0.46	0.61	0.61	0.26	0.28	0.84
23	0.07	-0.07	-0.06	0.02	-0.08	0.60	0.24	0.68	0.76	0.32	0.47	0.06	0.07	0.40	0.56	0.54	0.15	0.17	0.80
24	0.01	0.08	-0.13	-0.06	0.06	0.55	0.49	0.65	0.73	0.45	0.47	0.17	0.13	0.40	0.54	0.50	0.15	0.24	0.72
25	0.19	-0.04	-0.16	0.05	0.08	0.51	0.12	0.67	0.70	0.18	0.40	0.04	0.10	0.33	0.52	0.56	0.12	0.15	0.73
26	0.24	0.05	0.00	0.04	0.06	-0.15	0.07	0.01	-0.06	0.00	-0.23	0.01	-0.01	-0.22	-0.07	0.23	-0.01	0.10	0.39
27	0.07	-0.11	-0.12	0.07	-0.03	-0.31	-0.12	-0.10	-0.26	-0.15	-0.18	-0.09	-0.16	-0.27	-0.18	0.16	-0.13	-0.10	0.40

(continued)

Item	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
28	0.26	-0.03	0.25	0.05	0.03	-0.13	0.05	-0.04	-0.12	0.01	-0.10	0.04	0.08	-0.12	-0.11	0.12	0.01	0.18	0.41
29	0.12	-0.19	-0.12	0.05	-0.16	-0.33	-0.21	-0.17	-0.29	-0.25	-0.17	-0.10	-0.17	-0.19	-0.22	0.08	-0.25	-0.19	0.36
30	0.07	-0.16	-0.18	0.03	-0.14	-0.26	-0.16	-0.16	-0.26	-0.24	-0.17	-0.10	-0.16	-0.22	-0.25	-0.02	-0.26	-0.23	0.26
31	0.18	-0.09	-0.07	0.05	-0.03	-0.15	-0.04	-0.18	-0.26	-0.14	-0.15	-0.02	0.00	-0.17	-0.19	0.05	-0.13	-0.05	0.20
32	0.16	-0.15	-0.19	-0.01	-0.22	-0.29	-0.14	-0.17	-0.28	-0.21	-0.18	-0.06	-0.11	-0.13	-0.25	0.07	-0.19	-0.20	0.18
33	0.24	-0.07	0.07	-0.05	-0.13	-0.38	-0.12	-0.29	-0.40	-0.20	-0.27	0.03	-0.08	-0.24	-0.28	-0.02	-0.17	-0.08	0.12
34	0.04	-0.09	-0.22	-0.02	-0.20	-0.34	-0.17	-0.26	-0.36	-0.18	-0.22	-0.04	-0.19	-0.20	-0.27	0.00	-0.21	-0.22	0.01
35	0.00	-0.18	-0.15	0.04	-0.23	-0.28	-0.24	-0.20	-0.29	-0.25	-0.11	-0.03	-0.16	-0.13	-0.23	0.00	-0.22	-0.24	0.00
36	-0.01	-0.14	-0.17	-0.13	-0.12	-0.13	-0.07	0.06	-0.04	-0.05	0.13	0.11	-0.01	0.10	0.12	0.04	-0.12	0.02	0.06
37	-0.02	-0.27	-0.18	-0.08	-0.17	-0.11	-0.19	0.05	-0.05	-0.13	0.11	0.05	-0.09	0.05	0.05	-0.03	-0.10	-0.06	-0.04
38	-0.08	-0.24	-0.11	-0.03	-0.11	-0.15	-0.12	0.04	-0.06	-0.10	-0.01	0.11	-0.02	-0.07	0.04	-0.04	-0.12	-0.01	0.02
39	0.08	-0.21	0.06	0.02	-0.15	-0.06	-0.24	0.11	0.00	-0.16	0.04	-0.02	-0.11	-0.03	0.03	-0.04	-0.11	-0.08	0.02
40	0.07	-0.18	0.06	0.01	-0.15	-0.07	-0.26	0.04	-0.07	-0.19	0.10	-0.08	-0.17	0.00	0.03	-0.05	-0.15	-0.11	0.06
41	-0.21	0.22	-0.25	0.07	0.20	0.05	0.12	0.01	-0.18	0.07	-0.10	0.04	0.09	-0.02	-0.10	-0.01	-0.17	-0.12	-0.08
42	-0.02	-0.22	-0.05	0.07	-0.25	0.05	-0.40	0.09	0.05	-0.22	-0.03	-0.16	-0.23	-0.09	-0.04	-0.03	-0.16	-0.13	0.07
43	-0.04	-0.25	-0.07	0.04	-0.24	0.02	-0.39	0.06	0.03	-0.23	-0.06	-0.16	-0.17	-0.13	-0.06	0.05	-0.17	-0.18	-0.03
44	-0.18	0.13	-0.07	0.22	-0.04	-0.07	-0.18	-0.12	-0.15	-0.13	-0.17	-0.04	-0.09	-0.09	-0.07	-0.13	-0.12	-0.03	-0.05
45	-0.18	-0.01	-0.22	0.10	-0.04	-0.17	-0.26	-0.12	-0.15	-0.07	-0.17	-0.12	-0.14	-0.09	-0.18	-0.13	-0.16	-0.15	-0.05
46	-0.13	-0.24	-0.02	0.16	-0.33	-0.53	-0.49	-0.56	-0.60	-0.48	-0.41	-0.36	-0.42	-0.39	-0.50	-0.17	-0.29	-0.30	0.03
47	-0.17	-0.25	-0.06	0.23	-0.34	-0.52	-0.50	-0.54	-0.57	-0.51	-0.43	-0.33	-0.43	-0.33	-0.48	-0.23	-0.28	-0.36	-0.03
48	-0.34	-0.06	-0.24	0.05	0.05	-0.50	-0.43	-0.54	-0.57	-0.41	-0.40	-0.13	-0.32	-0.32	-0.41	-0.17	-0.19	-0.18	0.05
49	-0.35	-0.08	-0.25	0.03	0.04	-0.42	-0.39	-0.47	-0.49	-0.44	-0.30	-0.24	-0.18	-0.35	-0.43	-0.18	-0.38	-0.13	0.03
50	-0.22	-0.26	-0.12	0.17	-0.20	-0.45	-0.43	-0.44	-0.48	-0.42	-0.23	-0.39	-0.43	-0.30	-0.40	-0.03	-0.34	-0.29	0.17

(continued)

Item	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39
20																				
21	0.76																			
22	0.82	0.85																		
23	0.77	0.75	0.96																	
24	0.68	0.86	0.92	0.96																
25	0.65	0.63	0.89	0.95	0.92															
26	0.40	0.39	0.48	0.60	0.64	0.72														
27	0.44	0.32	0.50	0.63	0.58	0.73	0.86													
28	0.43	0.29	0.42	0.53	0.47	0.58	0.73	0.92												
29	0.44	0.17	0.36	0.52	0.42	0.60	0.72	0.95	0.84											
30	0.40	0.13	0.31	0.45	0.35	0.54	0.66	0.92	0.78	0.99										
31	0.32	0.15	0.28	0.39	0.35	0.50	0.66	0.86	0.79	0.96	0.96									
32	0.30	0.06	0.25	0.36	0.27	0.47	0.60	0.83	0.65	0.94	0.95	0.95								
33	0.23	0.02	0.16	0.26	0.20	0.34	0.58	0.79	0.68	0.88	0.91	0.90	0.97							
34	0.19	-0.07	0.14	0.21	0.14	0.29	0.51	0.76	0.54	0.87	0.88	0.84	0.96	0.93						
35	0.10	-0.10	0.12	0.16	0.05	0.25	0.47	0.74	0.47	0.82	0.86	0.79	0.95	0.88	0.97					
36	0.15	0.14	0.29	0.24	0.24	0.30	0.45	0.57	0.41	0.63	0.69	0.67	0.80	0.78	0.86	0.92				
37	0.10	-0.06	0.21	0.20	0.09	0.28	0.30	0.54	0.34	0.63	0.68	0.62	0.76	0.75	0.86	0.91	0.95			
38	0.06	0.03	0.20	0.19	0.15	0.24	0.33	0.50	0.31	0.58	0.61	0.52	0.69	0.68	0.77	0.85	0.93	0.95		
39	0.06	-0.06	0.27	0.25	0.08	0.24	0.29	0.44	0.26	0.59	0.60	0.56	0.70	0.64	0.77	0.84	0.90	0.95	0.95	
40	0.09	-0.06	0.26	0.24	0.08	0.21	0.25	0.45	0.25	0.59	0.60	0.54	0.67	0.63	0.76	0.83	0.88	0.93	0.93	1.00
41	-0.17	0.32	0.12	0.09	0.25	0.02	0.28	0.34	0.17	0.41	0.46	0.31	0.62	0.56	0.66	0.76	0.86	0.75	0.83	0.90
42	0.05	-0.18	0.22	0.30	-0.02	0.29	0.22	0.44	0.22	0.58	0.57	0.51	0.69	0.55	0.72	0.77	0.77	0.87	0.81	0.96
43	-0.03	-0.18	0.19	0.22	-0.04	0.22	0.15	0.38	0.19	0.52	0.53	0.44	0.63	0.48	0.66	0.72	0.71	0.83	0.78	0.94
44	-0.05	0.00	0.15	0.13	0.13	0.06	0.12	0.31	0.16	0.35	0.33	0.37	0.53	0.50	0.57	0.61	0.71	0.72	0.72	0.82
45	-0.05	0.00	0.15	0.13	0.07	-0.04	0.26	0.26	0.09	0.35	0.37	0.37	0.57	0.54	0.66	0.67	0.77	0.72	0.72	0.88
46	0.03	-0.47	-0.18	-0.10	-0.28	-0.08	0.16	0.39	0.22	0.52	0.57	0.40	0.62	0.56	0.69	0.76	0.51	0.68	0.60	0.76
47	-0.12	-0.49	-0.18	-0.10	-0.29	-0.06	0.06	0.33	0.12	0.51	0.47	0.32	0.58	0.50	0.62	0.69	0.45	0.61	0.56	0.74
48	-0.01	-0.37	-0.29	-0.22	-0.15	-0.11	0.36	0.36	0.15	0.47	0.39	0.48	0.68	0.73	0.61	0.79	0.59	0.39	0.58	0.59
49	0.11	-0.40	-0.21	-0.15	-0.18	-0.14	0.00	0.20	0.11	0.32	0.27	0.28	0.31	0.46	0.51	0.55	0.34	0.43	0.49	0.56
50	0.14	-0.42	-0.15	-0.08	-0.22	-0.07	-0.01	0.39	-0.01	0.44	0.43	0.22	0.45	0.41	0.55	0.66	0.38	0.58	0.50	0.68

(continued)

Item	40	41	42	43	44	45	46	47	48	49	50
40											
41	0.92										
42	0.97	0.71									
43	0.95	0.73	1.00								
44	0.83	0.70	0.92	0.93							
45	0.91	0.70	0.95	0.96	0.94						
46	0.80	0.39	0.86	0.88	0.77	0.83					
47	0.77	0.32	0.84	0.84	0.75	0.77	0.99				
48	0.65	0.17	0.68	0.65	0.68	0.79	0.92	0.96			
49	0.63	0.19	0.66	0.68	0.70	0.76	0.91	0.93	0.94		
50	0.71	0.19	0.75	0.77	0.69	0.76	0.95	0.95	0.86	0.95	

Table A.7

Inter-Item Correlation 2006 National Intelligence Test Scale B Test 1: Computation

Item	1	2	3	4	5	6	7	8	9	10	11
1											
2	0.06										
3	0.06	0.15									
4	0.01	0.24	0.42								
5	-0.13	-0.04	-0.04	0.03							
6	0.00	0.02	-0.25	0.21	0.13						
7	0.09	-0.05	0.11	0.03	0.14	0.21					
8	0.00	0.10	0.39	0.31	0.03	-0.01	-0.14				
9	0.36	0.01	0.12	0.19	0.24	0.25	0.22	0.18			
10	0.13	0.03	0.03	0.04	0.18	0.23	0.22	0.10	0.31		
11	0.01	0.41	0.19	-0.01	0.10	0.19	0.12	0.12	0.27	0.43	
12	0.21	0.18	0.07	0.25	0.30	0.19	0.08	0.11	0.46	0.45	0.37
13	-0.27	0.05	0.13	-0.02	0.17	0.11	0.14	0.03	0.08	0.07	0.10
14	0.05	0.35	0.09	0.17	0.02	0.08	0.13	0.15	0.18	0.38	0.24
15	0.06	0.36	-0.09	0.09	0.03	0.11	0.07	-0.13	0.18	0.23	0.36
16	0.06	0.13	-0.03	0.16	-0.07	0.09	0.07	0.04	0.18	0.20	0.24
17	0.01	0.26	0.26	0.10	-0.23	0.43	0.06	0.06	0.10	0.04	0.09
18	-0.29	-0.37	-0.37	-0.32	-0.19	0.45	-0.06	-0.06	0.26	0.40	0.12
19	0.15	0.03	0.18	-0.08	0.03	0.07	-0.06	-0.06	0.02	-0.15	-0.11
20	-0.43	-0.50	-0.50	-0.46	-0.34	0.38	-0.21	-0.21	0.44	0.03	0.12
21	-0.11	-0.19	-0.19	-0.14	-0.14	0.23	0.13	0.13	-0.21	-0.30	-0.25
22	-0.19	0.10	0.10	0.16	-0.13	0.20	-0.02	-0.02	0.13	-0.08	0.03

(continued)

Item	12	13	14	15	16	17	18	19	20	21	22
13	0.07										
14	0.37	0.08									
15	0.21	0.12	0.51								
16	0.21	0.03	0.28	0.40							
17	0.19	-0.03	0.25	0.28	0.58						
18	0.21	0.09	0.28	0.27	0.30	0.47					
19	-0.01	-0.07	0.15	0.11	0.24	0.27	0.11				
20	0.39	0.29	0.26	0.44	0.28	0.24	0.67	0.32			
21	-0.26	-0.24	-0.30	0.07	0.05	0.06	0.58	0.34	0.53		
22	-0.05	0.01	0.08	0.13	0.18	0.37	0.19	0.34	0.39	0.24	

Table A.8

Inter-Item Correlation 2006 National Intelligence Test Scale B Test 2: Information

Item	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1																			
2	0.20																		
3	0.46	0.06																	
4	0.44	-0.28	0.46																
5	0.25	0.29	0.11	-0.24															
6	0.14	-0.07	0.12	0.14	0.10														
7	0.39	-0.07	0.25	0.39	-0.30	0.42													
8	-0.32	0.09	0.04	0.17	0.13	-0.09	-0.18												
9	0.35	0.25	0.21	0.35	0.01	0.04	0.30	-0.08											
10	-0.15	0.09	-0.07	-0.33	0.06	0.10	0.00	0.01	-0.19										
11	0.26	0.29	0.25	0.26	0.49	-0.11	0.20	-0.03	0.01	0.20									
12	0.29	-0.05	0.14	0.29	0.01	0.56	0.24	-0.02	0.20	0.25	-0.07								
13	-0.08	0.10	-0.02	-0.22	0.02	0.18	0.15	0.13	0.08	0.14	0.03	0.26							
14	0.28	0.17	0.16	-0.06	0.14	0.13	-0.44	0.09	-0.16	0.10	0.24	0.18	0.02						
15	0.05	0.09	-0.10	0.05	0.07	-0.03	0.12	-0.07	-0.12	0.07	0.08	0.01	-0.02	0.15					
16	0.23	0.27	-0.06	-0.38	0.33	0.10	-0.17	-0.02	-0.01	0.09	0.16	0.33	0.19	0.13	0.05				
17	0.17	0.21	0.17	0.02	0.17	0.09	0.02	-0.05	0.12	0.05	0.25	-0.01	0.08	0.09	0.19	0.19			
18	-0.12	0.13	0.02	0.07	0.33	0.02	-0.31	0.05	0.14	0.07	0.13	0.00	0.06	0.06	0.07	0.27	0.15		
19	0.09	-0.07	-0.06	0.09	0.16	0.33	0.19	-0.01	0.00	0.00	0.33	0.34	0.11	0.03	0.14	-0.08	0.08	0.14	
20	0.02	0.20	0.03	0.02	0.28	-0.07	-0.22	0.07	0.12	0.04	0.37	-0.14	0.14	0.10	0.11	0.19	0.15	0.17	-0.05
21	-0.10	0.12	0.03	-0.26	0.22	0.11	-0.02	0.09	0.21	0.17	0.18	-0.04	0.10	0.11	0.07	0.13	0.04	0.15	0.11
22	0.28	0.20	0.20	0.15	0.17	0.02	0.05	0.12	0.27	0.03	0.21	0.03	0.12	0.13	0.11	0.17	0.19	0.15	0.23
23	0.33	0.10	0.15	0.33	0.12	0.10	-0.24	0.04	-0.09	0.09	0.11	-0.17	-0.05	0.19	0.12	0.02	0.04	0.14	0.08
24	0.20	0.14	-0.02	-0.10	0.28	0.26	0.05	0.07	-0.26	0.17	0.26	0.09	0.18	0.13	-0.02	0.25	0.13	0.16	0.15
25	0.03	0.17	0.33	0.18	0.14	-0.09	0.23	0.08	-0.13	-0.03	0.35	-0.01	0.19	0.05	-0.01	0.09	0.15	0.17	-0.01
26	-0.11	0.05	0.03	0.23	0.16	0.13	-0.01	0.00	0.05	0.11	0.05	-0.08	0.03	0.16	0.00	0.17	0.14	0.15	0.21
27	0.35	0.23	-0.13	-0.11	0.27	0.05	-0.23	0.06	0.08	-0.02	0.20	-0.03	0.10	0.19	0.04	0.24	0.18	0.12	0.22

(continued)

Item	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
28	0.26	0.20	-0.02	0.12	0.20	-0.01	-0.34	0.05	0.08	0.16	0.22	-0.15	0.01	0.13	-0.03	0.10	0.14	0.14	-0.05
29	0.17	0.13	0.16	-0.13	0.26	0.15	-0.10	0.06	0.12	0.14	0.33	0.19	0.09	0.05	-0.23	0.29	0.15	0.22	0.16
30	0.23	0.25	0.16	0.08	0.19	0.30	0.20	0.09	0.09	0.12	0.19	0.12	0.13	0.12	-0.11	0.12	0.04	0.13	0.06
31	0.38	0.17	0.27	-0.10	0.35	0.22	0.00	-0.01	-0.22	0.09	0.35	0.24	0.13	0.24	-0.11	0.37	0.15	0.28	0.20
32	0.14	0.22	0.09	0.14	0.17	-0.02	-0.21	0.02	0.26	0.03	0.29	-0.12	0.12	0.24	0.07	0.07	0.21	0.21	0.14
33	0.19	0.04	0.25	0.01	0.08	0.05	-0.03	0.02	-0.06	0.05	0.13	0.49	0.13	0.07	-0.06	0.08	0.03	0.06	0.11
34	0.15	0.36	0.31	0.15	0.27	0.13	-0.09	0.02	-0.04	0.11	0.31	0.20	0.09	0.16	0.05	0.34	0.27	0.16	0.07
35	-0.02	0.04	0.14	-0.17	0.11	-0.03	0.03	0.15	-0.46	-0.46	0.34	0.34	-0.01	0.18	0.00	0.12	-0.05	0.10	0.03
36	0.15	0.29	0.01	0.28	0.36	0.06	-0.20	0.11	0.11	0.11	0.29	0.29	0.15	0.20	-0.06	0.25	0.19	0.23	0.14
37	0.20	0.15	-0.10	0.06	0.29	-0.06	0.12	0.12	0.29	0.29	0.14	0.14	0.11	0.20	0.02	0.23	0.09	0.20	-0.01
38	0.09	0.14	0.13	-0.25	0.03	-0.02	-0.17	-0.01	-0.24	-0.24	0.17	0.17	-0.01	0.11	0.07	0.02	0.09	0.12	0.05
39	0.30	0.16	0.45	0.30	0.20	-0.09	0.07	0.03	-0.09	-0.09	0.16	0.16	0.08	0.16	0.09	0.11	0.10	0.17	-0.02
40	0.02	0.08	0.21	0.02	0.21	-0.07	0.22	0.08	-0.03	-0.03	0.14	0.14	-0.01	-0.02	-0.01	0.22	0.05	0.09	-0.03

(continued)

Item	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40
20																					
21	0.17																				
22	0.26	0.24																			
23	0.12	0.15	0.13																		
24	0.17	0.24	0.18	0.09																	
25	0.30	0.29	0.13	0.03	0.19																
26	0.11	0.09	0.10	0.13	0.33	0.15															
27	0.16	0.23	0.16	0.16	0.25	0.15	0.21														
28	0.21	0.21	0.16	0.23	0.22	0.16	0.22	0.22													
29	0.22	0.22	0.20	0.12	0.25	0.14	0.20	0.19	0.24												
30	0.05	0.25	0.24	0.14	0.29	0.19	0.12	0.21	0.15	0.26											
31	0.22	0.21	0.21	0.21	0.46	0.10	0.24	0.34	0.40	0.44	0.46										
32	0.12	0.24	0.29	0.20	0.18	0.24	0.19	0.26	0.20	0.25	0.16	0.41									
33	0.03	0.19	0.03	0.03	0.12	0.15	0.02	0.08	0.12	0.10	0.24	0.32	0.16								
34	0.25	0.17	0.22	0.13	0.26	0.19	0.24	0.26	0.20	0.30	0.27	0.57	0.31	0.25							
35	0.14	0.09	0.02	0.17	0.19	0.18	0.15	0.11	0.25	0.05	0.08	0.25	0.07	0.12	0.31						
36	0.20	0.27	0.29	0.15	0.27	0.13	0.15	0.29	0.18	0.36	0.31	0.48	0.30	0.22	0.37	0.20					
37	0.15	0.29	0.13	0.15	0.22	0.11	0.16	0.33	0.27	0.27	0.23	0.42	0.25	0.12	0.31	0.02	0.42				
38	-0.08	0.10	0.13	0.13	0.16	0.13	0.01	0.18	0.21	0.10	0.09	0.24	0.21	0.08	0.22	0.01	0.22	0.23			
39	0.10	0.26	0.14	0.05	0.27	0.14	0.18	0.21	0.22	0.08	0.19	0.31	0.21	0.16	0.26	0.06	0.34	0.50	0.26		
40	0.00	0.17	0.12	0.02	0.16	0.10	0.06	0.17	0.06	0.21	0.10	0.39	0.09	0.19	0.28	0.02	0.22	0.35	0.20	0.35	

Table A.9

Inter-Item Correlation 2006 National Intelligence Test Scale B Test 3: Vocabulary

Item	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1																			
2	0.76																		
3	0.76	0.49																	
4	0.59	0.26	0.26																
5	0.74	0.47	0.47	0.22															
6	0.68	0.37	0.37	0.11	0.34														
7	0.68	0.37	0.37	0.26	0.34	0.58													
8	0.49	0.29	0.13	-0.02	0.25	0.12	0.42												
9	0.69	0.38	0.38	0.28	0.35	0.25	0.41	0.44											
10	0.51	0.15	0.15	-0.12	0.27	0.00	0.00	0.24	0.16										
11	0.65	0.33	0.69	0.07	0.30	0.19	0.19	0.07	0.21	0.09									
12	0.68	0.37	0.54	0.11	0.34	0.23	0.23	0.12	0.25	0.00	0.53								
13	0.53	0.34	0.34	-0.10	0.14	0.03	0.03	-0.23	0.04	-0.21	0.12	0.47							
14	0.17	-0.09	0.12	-0.01	0.05	0.27	0.18	0.03	0.31	0.02	0.09	-0.01	0.03						
15	0.43	0.06	0.06	0.15	0.17	-0.09	-0.09	0.04	0.24	0.00	0.15	0.04	-0.07	0.09					
16	0.30	0.06	-0.09	0.34	-0.13	-0.12	-0.12	-0.01	-0.10	-0.13	-0.17	0.04	0.02	0.27	0.23				
17	0.52	0.32	0.53	0.17	0.12	0.15	0.01	0.31	0.57	0.03	0.27	0.15	-0.20	0.20	0.21	0.13			
18	0.03	-0.05	-0.05	0.00	0.19	0.00	0.00	0.05	0.14	-0.05	-0.03	-0.09	0.05	0.26	0.11	0.19	0.10		
19	0.15	0.08	0.08	0.06	0.02	-0.16	0.05	0.08	-0.01	0.07	-0.03	0.32	0.10	-0.10	0.08	0.11	0.15	0.15	
20	0.25	-0.01	0.20	0.03	0.14	0.09	-0.19	-0.24	-0.17	0.17	0.39	0.19	0.07	0.12	0.03	0.06	0.12	0.14	0.05
21	0.50	0.15	0.15	0.35	0.60	0.14	0.00	-0.26	0.16	0.11	0.26	0.44	0.04	0.07	-0.10	-0.08	0.03	0.28	-0.04
22	0.30	0.26	0.59	0.04	0.20	0.15	-0.13	0.07	0.38	-0.21	0.25	0.34	0.01	0.19	0.17	0.04	0.20	0.21	-0.12
23	0.48	0.28	0.63	0.23	0.24	0.11	0.11	-0.05	0.13	0.34	0.44	0.28	0.19	-0.02	0.02	0.02	0.09	-0.09	0.07
24	0.26	0.01	0.56	-0.09	0.16	-0.17	-0.01	0.05	0.01	0.01	0.41	0.38	-0.04	0.12	0.13	0.30	0.34	0.14	0.23
25	0.13	-0.14	0.07	-0.07	0.01	0.13	-0.06	-0.03	-0.30	0.17	-0.05	0.13	0.17	0.03	0.06	0.04	-0.10	0.02	0.06
26	0.09	0.02	0.38	-0.07	0.11	0.25	-0.36	0.04	0.12	-0.05	0.12	0.34	0.17	0.04	0.04	0.08	0.01	0.11	0.00

(continued)

Item	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
27	0.31	0.27	0.44	0.05	0.37	0.27	-0.12	-0.05	-0.10	-0.12	0.19	0.44	-0.03	0.17	0.08	0.05	0.22	0.21	0.07
28	0.31	0.60	0.60	-0.13	0.37	-0.24	-0.24	0.08	-0.10	0.03	0.27	0.26	-0.03	0.02	0.16	0.06	0.22	0.16	0.07
29	0.37	0.51	0.35	0.15	0.30	-0.03	-0.03	-0.21	-0.01	-0.18	0.07	0.25	0.18	0.13	0.07	0.12	0.21	0.20	0.09
30	0.09	-0.32	0.50	0.04	0.11	0.08	-0.01	-0.04	0.31	0.02	0.20	0.25	0.02	0.18	0.15	0.05	0.25	0.21	0.02
31	0.10	0.02	0.38	0.04	0.11	0.17	0.08	0.07	0.21	0.02	0.20	0.25	-0.09	0.09	0.01	-0.02	0.10	0.13	0.00
32	0.10	0.51	0.03	0.04	0.12	0.08	0.26	0.11	0.32	0.09	-0.09	0.26	-0.01	-0.02	0.06	0.07	0.26	0.09	0.19
33	0.23	0.18	0.35	-0.06	0.42	-0.21	0.07	0.03	0.29	-0.37	0.08	0.59	0.23	0.05	0.10	0.12	0.13	0.12	0.15
34	-0.04	0.04	0.24	0.38	-0.05	0.42	0.18	0.15	0.15	0.07	0.17	0.18	-0.03	0.09	0.01	-0.03	0.11	0.02	0.01
35	0.09	-0.18	-0.18	-0.02	0.11	0.16	-0.01	0.04	0.21	-0.11	0.19	0.25	0.11	0.10	0.01	0.06	-0.06	0.15	0.02
36	0.13	0.24	0.07	-0.07	0.16	0.22	0.04	0.00	0.26	-0.15	0.03	0.39	0.18	0.03	0.08	0.08	0.20	0.12	0.10
37	0.05	-0.23	0.34	-0.03	0.38	0.02	-0.07	-0.08	0.16	-0.12	-0.01	0.20	0.08	0.13	0.01	0.11	0.21	0.18	0.06
38	0.00	-0.29	0.28	0.09	0.00	0.13	-0.05	-0.16	0.19	0.01	0.14	0.04	0.13	0.07	0.05	-0.01	0.10	0.15	0.02
39	0.06	-0.02	-0.02	0.18	0.22	0.21	0.03	-0.10	0.27	0.11	0.08	0.12	0.12	0.06	0.14	0.11	0.20	0.18	0.05
40	-0.20	0.06	0.21	-0.02	0.11	0.10	-0.02	-0.04	-0.16	0.09	0.30	0.10	0.01	0.03	-0.13	-0.13	0.21	0.04	-0.10

(continued)

Item	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40
20																					
21	0.22																				
22	0.22	0.34																			
23	0.35	0.33	0.08																		
24	0.05	0.34	0.27	0.27																	
25	0.13	-0.03	0.05	0.16	0.03																
26	0.23	0.09	0.34	0.11	0.18	0.23															
27	0.26	0.29	0.33	0.29	0.27	0.08	0.32														
28	0.14	0.38	0.20	0.29	0.18	0.04	0.08	0.30													
29	0.25	0.47	0.36	0.25	0.34	0.05	0.27	0.39	0.40												
30	0.25	0.34	0.32	0.14	0.26	0.02	0.17	0.37	0.28	0.26											
31	0.30	0.12	0.34	0.29	0.02	-0.06	0.25	0.20	0.18	0.31	0.10										
32	0.10	0.16	0.22	0.02	0.20	0.04	0.10	0.17	0.23	0.33	0.21	0.15									
33	0.22	0.34	0.21	0.25	0.20	-0.02	0.12	0.34	0.31	0.44	0.26	0.24	0.36								
34	0.08	0.20	0.11	0.00	0.07	0.13	0.09	0.22	0.07	0.21	0.14	0.18	0.23	0.23							
35	0.21	0.26	0.27	0.04	0.12	0.03	0.10	0.24	0.11	0.29	0.17	0.23	0.23	0.32	0.39						
36	0.16	0.51	0.42	0.11	0.24	0.12	0.28	0.38	0.23	0.46	0.24	0.28	0.25	0.50	0.30	0.52					
37	0.10	0.31	0.26	-0.03	0.19	0.03	0.19	0.32	0.33	0.35	0.23	0.20	0.27	0.39	0.25	0.41	0.59				
38	0.24	0.23	0.37	0.06	0.19	-0.04	0.27	0.34	0.24	0.35	0.35	0.26	0.14	0.32	0.17	0.31	0.57	0.63			
39	0.10	0.38	0.29	-0.07	0.27	0.01	0.20	0.31	0.22	0.43	0.31	0.28	0.18	0.38	0.30	0.31	0.68	0.65	0.71		
40	0.08	0.03	0.32	0.08	-0.01	-0.02	0.07	0.26	0.01	0.15	0.12	0.18	-0.01	0.10	0.13	0.08	0.22	0.18	0.40	0.47	

Table A.10

Inter-Item Correlation 2006 National Intelligence Test Scale B Test 4: Analogies

Item	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1																			
2	0.39																		
3	0.35	0.24																	
4	0.17	0.46	0.14																
5	0.16	0.32	0.42	0.39															
6	0.54	0.31	0.47	0.22	0.36														
7	-0.15	0.34	0.24	0.39	0.47	0.37													
8	0.09	0.41	0.18	0.43	0.27	0.36	0.42												
9	0.15	0.46	0.35	0.37	0.38	0.40	0.42	0.47											
10	0.50	0.59	0.39	0.64	0.58	0.48	0.41	0.55	0.66										
11	0.37	0.08	0.26	0.03	0.11	0.03	-0.01	0.14	0.08	-0.04									
12	0.19	0.65	0.29	0.71	0.51	0.49	0.53	0.54	0.62	0.89	-0.11								
13	0.55	0.18	0.26	0.18	0.21	0.25	0.14	0.22	0.19	0.21	0.18	0.22							
14	0.04	0.42	0.33	0.47	0.49	0.27	0.40	0.45	0.43	0.72	0.06	0.76	0.40						
15	0.05	0.40	0.20	0.48	0.32	0.29	0.39	0.43	0.44	0.60	0.10	0.68	0.23	0.66					
16	0.15	0.47	0.28	0.54	0.49	0.19	0.39	0.45	0.41	0.65	0.10	0.70	0.31	0.52	0.55				
17	0.50	0.52	0.36	0.48	0.47	0.39	0.40	0.42	0.49	0.68	-0.03	0.71	0.28	0.53	0.49	0.49			
18	-0.01	0.35	0.26	0.36	0.42	0.34	0.31	0.38	0.42	0.48	0.08	0.49	0.30	0.45	0.43	0.38	0.41		
19	0.21	0.38	0.23	0.42	0.27	0.19	0.28	0.30	0.27	0.41	0.12	0.44	0.17	0.37	0.31	0.33	0.38	0.31	
20	0.11	0.49	0.23	0.59	0.54	0.30	0.35	0.41	0.46	0.70	-0.23	0.80	0.28	0.54	0.47	0.61	0.56	0.39	0.44
21	0.36	0.01	0.17	-0.08	0.00	0.10	-0.03	0.03	0.06	-0.02	0.24	-0.11	0.23	0.09	0.09	-0.07	0.07	0.17	0.02
22	0.06	0.29	0.25	0.31	0.28	0.24	0.27	0.34	0.33	0.44	0.19	0.41	0.30	0.37	0.27	0.34	0.35	0.26	0.21

(continued)

Item	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
23	0.26	0.57	0.37	0.61	0.52	0.43	0.52	0.36	0.51	0.74	0.07	0.8	0.31	0.65	0.55	0.58	0.62	0.43	0.42
24	-0.31	0.2	0.22	0.3	0.28	0.34	0.37	0.21	0.35	0.34	0.18	0.37	0.23	0.36	0.35	0.37	0.31	0.33	0.28
25	0.34	0.35	0.3	0.48	0.32	0.34	0.35	0.3	0.29	0.41	0.05	0.52	0.29	0.41	0.41	0.36	0.43	0.42	0.27
26	0.25	0.41	0.25	0.53	0.35	0.33	0.42	0.43	0.41	0.59	0.03	0.6	0.41	0.49	0.46	0.49	0.54	0.42	0.35
27	0.12	0.27	0.29	0.12	0.09	0.19	0.34	0.32	0.34	0.18	0.17	0.21	0.39	0.21	0.22	0.2	0.3	0.25	0.31
28	0.44	-0.14	0.11	-0.09	-0.05	0.15	0.06	0.04	-0.07	-0.29	0.14	-0.3	0.23	-0.17	-0.15	-0.08	-0.14	0.08	-0.04
29	0.24	0.18	0.21	0.27	0.19	0.31	0.24	0.22	0.24	0.16	0.13	0.24	0.26	0.21	0.15	0.25	0.28	0.34	0.24
30	0.01	-0.2	0.15	-0.26	-0.1	0.01	-0.09	-0.17	-0.09	-0.33	0.07	-0.39	0.15	-0.26	-0.24	-0.13	-0.25	-0.04	0.13
31	-0.08	0.42	0.34	0.46	0.38	0.3	0.53	0.26	0.42	0.44	0.28	0.49	0.33	0.5	0.49	0.31	0.28	0.27	0.35
32	-0.24	-0.09	-0.04	-0.19	-0.21	-0.01	0.12	-0.02	-0.03	-0.24	-0.04	-0.18	0.02	-0.15	-0.08	-0.04	-0.11	0.05	-0.04

(continued)

Item	20	21	22	23	24	25	26	27	28	29	30	31	32
20													
21	0.19												
22	0.35	0.23											
23	0.84	0.21	0.53										
24	0.41	0.23	0.40	0.53									
25	0.56	0.16	0.34	0.58	0.46								
26	0.65	0.14	0.46	0.71	0.57	0.66							
27	0.18	0.25	0.30	0.32	0.39	0.36	0.43						
28	-0.09	0.33	0.08	0.08	0.22	0.22	0.35	0.40					
29	0.27	0.27	0.32	0.41	0.43	0.34	0.52	0.49	0.62				
30	-0.26	0.26	0.02	-0.16	0.18	0.01	-0.03	0.24	0.59	0.31			
31	0.59	0.27	0.45	0.61	0.33	0.32	0.63	0.56	0.53	0.49	0.31		
32	-0.18	0.24	0.04	0.00	0.09	0.06	0.07	0.24	0.47	0.36	0.40	0.41	

Table A.11

Inter-Item Correlation 2006 National Intelligence Test Scale B Test 5: Comparisons

Item	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1																			
2	0.77																		
3	0.52	0.58																	
4	0.24	0.30	0.12																
5	0.37	0.42	0.28	0.67															
6	0.28	0.34	0.03	0.20	0.37														
7	0.11	0.17	-0.15	0.16	0.00	0.52													
8	0.31	0.36	0.20	0.06	0.08	0.73	0.56												
9	0.28	0.33	0.02	0.30	0.05	0.64	0.46	0.72											
10	0.35	0.40	0.10	-0.02	0.13	0.74	0.65	0.81	0.82										
11	0.23	0.28	-0.03	-0.03	0.30	0.02	-0.18	-0.08	-0.11	0.10									
12	0.73	0.77	0.52	0.41	0.37	0.28	0.11	0.31	0.45	0.52	0.61								
13	0.78	0.65	0.34	0.07	0.21	0.27	0.08	0.29	0.11	0.34	0.60	0.90							
14	0.20	0.06	-0.03	-0.12	0.01	-0.16	-0.07	0.01	-0.24	-0.03	0.24	0.42	0.41						
15	0.61	0.66	0.36	0.42	0.23	0.13	-0.04	0.16	0.13	0.36	0.63	0.91	0.75	0.48					
16	0.30	0.35	0.19	0.06	0.07	0.39	-0.09	0.30	0.38	0.19	0.04	0.47	0.28	0.05	0.30				
17	0.37	0.21	0.08	0.15	0.20	0.23	0.21	-0.24	-0.05	0.16	0.01	0.37	0.31	-0.01	0.24	0.54			
18	0.11	0.01	-0.21	0.20	0.17	-0.07	0.17	0.12	0.28	0.13	-0.09	0.34	0.18	-0.14	0.21	0.63	0.46		
19	0.25	0.30	0.13	0.15	0.16	0.32	0.09	0.07	0.20	0.13	-0.02	0.42	0.22	-0.11	0.25	0.82	0.45	0.57	
20	0.19	0.25	0.23	0.27	0.09	0.14	-0.08	0.01	0.24	0.07	-0.09	0.36	0.16	-0.12	0.19	0.79	0.35	0.56	0.84
21	0.15	0.21	0.02	0.13	-0.09	0.35	0.64	0.31	0.34	0.53	0.01	0.32	0.12	-0.01	0.14	0.59	0.26	0.51	0.62
22	0.23	0.28	0.27	0.12	-0.01	0.38	0.20	0.41	0.37	0.48	0.21	0.40	0.38	0.21	0.22	0.69	0.40	0.42	0.78

(continued)

Item	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
23	0.19	0.25	0.35	0.34	0.26	0.47	0.21	0.44	0.46	0.52	0.06	0.36	0.47	-0.01	0.19	0.65	0.38	0.37	0.67
24	0.18	0.24	0.33	0.17	0.08	0.45	0.41	0.49	0.44	0.68	0.05	0.35	0.33	-0.04	0.17	0.58	0.26	0.34	0.66
25	0.12	0.18	0.35	0.09	0.01	0.38	0.23	0.41	0.30	0.49	0.07	0.29	0.26	0.01	0.11	0.51	0.23	0.31	0.58
26	0.11	0.01	-0.20	0.06	-0.29	0.08	0.00	0.05	0.07	0.14	0.04	-0.05	0.19	-0.01	0.09	0.28	0.26	0.21	0.22
27	0.10	0.16	0.23	-0.04	-0.01	0.02	-0.05	0.05	0.01	-0.04	0.04	0.27	0.06	0.07	-0.06	0.31	-0.07	0.14	0.28
28	0.09	-0.01	0.18	0.03	0.30	-0.02	0.14	0.21	0.04	-0.09	-0.12	0.09	0.16	0.03	-0.23	0.25	0.22	0.38	0.27
29	0.10	0.16	0.12	0.07	-0.01	-0.12	-0.12	0.06	0.02	-0.04	0.12	0.27	0.06	0.10	-0.06	0.31	-0.10	0.09	0.28
30	0.07	0.13	0.09	0.03	-0.04	-0.01	0.02	0.13	0.08	0.09	0.08	0.24	0.03	0.08	-0.09	0.33	-0.11	-0.03	0.29
31	0.15	0.05	0.18	0.21	0.13	0.06	0.05	0.10	0.05	-0.02	-0.20	0.15	0.10	0.00	-0.04	0.38	0.14	0.24	0.35
32	0.13	0.02	-0.05	0.13	0.19	0.10	-0.02	0.14	0.09	0.06	-0.14	0.13	0.08	0.01	-0.07	0.44	-0.05	0.03	0.32
33	0.11	0.01	-0.07	0.05	0.25	-0.07	-0.05	-0.04	-0.18	-0.07	-0.16	0.11	0.05	-0.03	-0.09	0.32	-0.05	0.07	0.25
34	0.24	0.09	0.09	0.02	0.14	-0.11	-0.12	-0.07	-0.12	0.09	-0.05	0.45	0.17	0.06	0.21	0.39	-0.02	0.06	0.24
35	0.21	0.34	-0.02	0.03	0.18	-0.08	-0.16	-0.18	-0.23	0.06	0.08	0.55	0.24	0.05	0.29	0.36	-0.12	0.01	0.21
36	0.27	0.12	-0.04	0.11	0.17	-0.08	0.01	-0.21	-0.02	0.05	0.03	0.27	0.20	0.00	0.24	-0.13	-0.11	-0.09	0.03
37	-0.07	-0.21	-0.15	0.07	-0.09	-0.09	-0.27	-0.07	-0.16	-0.02	0.15	0.15	0.12	0.24	0.08	0.11	-0.13	-0.21	-0.13
38	0.19	0.04	-0.05	0.09	0.22	-0.11	-0.06	-0.29	-0.34	0.03	0.05	0.19	0.12	-0.01	0.36	-0.31	-0.06	-0.02	-0.07
39	-0.06	0.03	0.01	0.02	0.06	-0.07	-0.06	-0.15	-0.20	-0.07	0.13	0.17	-0.01	0.06	0.34	-0.10	-0.14	-0.08	-0.01
40	-0.06	0.02	0.14	0.13	0.13	-0.02	-0.09	-0.23	-0.21	-0.15	0.02	0.38	-0.01	0.04	0.34	-0.24	-0.13	-0.11	-0.10
41	0.00	0.10	0.09	-0.01	-0.04	0.06	0.35	0.09	-0.09	0.32	0.07	0.16	0.22	-0.04	0.20	0.10	-0.03	-0.04	0.13
42	-0.23	-0.15	0.18	0.04	-0.01	0.03	-0.16	-0.04	-0.12	-0.08	0.02	0.36	0.09	0.02	0.27	-0.06	-0.09	-0.11	-0.05
43	-0.26	0.10	0.08	0.08	0.03	0.04	-0.18	-0.02	-0.07	-0.05	-0.02	0.33	-0.04	0.03	0.24	-0.05	-0.13	-0.13	-0.01
44	-0.32	0.04	0.15	0.07	0.19	0.21	0.03	0.06	-0.11	0.23	-0.07	0.28	-0.02	-0.06	0.32	-0.03	-0.07	-0.04	0.06
45	0.24	0.01	0.19	0.06	0.15	0.11	-0.15	0.01	-0.07	0.19	-0.05	0.24	-0.06	-0.07	0.29	-0.08	-0.14	-0.11	-0.07
46	-0.19	-0.04	0.23	0.04	0.00	0.05	-0.27	-0.05	-0.13	-0.03	-0.05	0.04	-0.21	0.04	0.24	0.15	-0.14	-0.13	0.03
47	-0.24	0.09	0.31	-0.02	-0.05	-0.06	-0.30	-0.11	-0.15	-0.08	-0.07	-0.01	-0.26	0.05	0.19	0.17	-0.19	-0.17	-0.03
48	-0.12	-0.02	0.05	0.12	0.17	0.05	-0.14	-0.07	-0.25	-0.06	0.07	-0.12	0.09	-0.03	0.20	0.26	-0.12	-0.03	0.04
49	-0.10	0.00	0.22	0.01	-0.07	0.00	-0.23	0.14	-0.12	-0.12	-0.02	-0.10	-0.05	0.02	0.22	0.05	-0.18	-0.08	0.06
50	-0.10	0.00	0.07	0.01	-0.25	0.00	-0.24	0.27	-0.06	0.21	-0.02	0.06	-0.05	0.04	0.22	0.05	-0.21	-0.15	0.00

(continued)

Item	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39
20																				
21	0.70																			
22	0.64	0.86																		
23	0.68	0.82	0.91																	
24	0.66	0.87	0.92	0.96																
25	0.51	0.75	0.89	0.87	0.91															
26	0.29	0.14	0.13	0.22	0.23	0.22														
27	0.41	0.17	0.18	0.29	0.31	0.29	0.67													
28	0.25	0.19	0.10	0.25	0.19	0.28	0.50	0.80												
29	0.38	0.30	0.25	0.34	0.36	0.33	0.65	0.95	0.83											
30	0.25	0.22	0.26	0.20	0.23	0.27	0.58	0.90	0.71	0.96										
31	0.34	0.25	0.18	0.27	0.27	0.28	0.52	0.79	0.63	0.89	0.91									
32	0.31	0.25	0.19	0.17	0.17	0.29	0.40	0.76	0.48	0.85	0.89	0.94								
33	0.25	0.16	0.07	0.18	0.18	0.18	0.35	0.74	0.48	0.85	0.88	0.92	0.98							
34	0.29	0.13	0.19	0.20	0.20	0.16	0.26	0.68	0.33	0.76	0.81	0.84	0.92	0.97						
35	0.26	0.04	0.12	0.07	0.07	0.14	0.23	0.58	0.30	0.72	0.77	0.79	0.88	0.96	0.95					
36	0.01	-0.08	-0.02	-0.17	-0.17	-0.13	0.03	0.10	0.02	0.22	0.32	0.47	0.48	0.54	0.64	0.76				
37	0.00	-0.16	0.04	-0.12	-0.12	-0.02	0.02	0.09	-0.05	0.13	0.17	0.17	0.23	0.27	0.29	0.35	0.62			
38	-0.03	-0.03	-0.07	-0.14	-0.14	-0.15	-0.06	0.09	-0.02	0.20	0.23	0.37	0.38	0.45	0.55	0.64	0.89	0.57		
39	-0.05	-0.05	-0.01	-0.09	-0.09	-0.20	-0.16	-0.05	-0.10	0.09	0.18	0.29	0.31	0.38	0.52	0.62	0.87	0.53	0.96	
40	-0.09	-0.09	-0.06	-0.17	-0.17	-0.18	-0.11	-0.01	-0.02	0.15	0.21	0.29	0.33	0.37	0.48	0.58	0.85	0.51	0.94	0.98

(continued)

Item	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39
41	0.05	0.05	0.29	0.22	0.22	0.16	-0.02	0.14	0.03	0.16	0.19	0.19	0.25	0.32	0.41	0.41	0.54	0.16	0.69	0.74
42	0.03	0.03	0.09	0.00	0.00	0.04	-0.08	0.05	-0.06	0.10	0.19	0.19	0.25	0.32	0.42	0.50	0.71	0.31	0.80	0.86
43	0.02	0.02	0.08	-0.04	-0.04	-0.01	-0.10	0.03	-0.07	0.06	0.15	0.15	0.24	0.29	0.42	0.48	0.67	0.31	0.77	0.84
44	0.08	0.08	0.15	0.02	0.02	0.08	-0.08	-0.03	-0.08	-0.04	0.07	0.09	0.18	0.24	0.34	0.43	0.63	0.22	0.74	0.81
45	0.03	0.03	0.14	0.03	0.03	0.09	-0.13	-0.05	-0.15	-0.04	0.02	0.03	0.16	0.19	0.30	0.41	0.61	0.20	0.73	0.79
46	0.03	0.03	0.03	-0.07	-0.07	-0.03	-0.02	0.02	-0.18	0.09	0.15	0.17	0.23	0.28	0.41	0.47	0.54	0.29	0.63	0.71
47	0.00	0.00	-0.03	-0.10	-0.10	-0.08	-0.10	-0.03	-0.16	0.02	0.03	0.07	0.15	0.20	0.31	0.41	0.49	0.29	0.61	0.65
48	0.01	0.01	-0.05	-0.08	-0.08	-0.06	-0.17	-0.11	-0.08	-0.09	-0.08	0.06	0.12	0.18	0.25	0.33	0.40	0.20	0.55	0.55
49	-0.01	-0.01	-0.07	-0.17	-0.17	-0.15	-0.07	-0.05	-0.05	0.01	0.07	0.14	0.21	0.22	0.31	0.37	0.39	0.20	0.56	0.56
50	-0.14	-0.14	-0.02	-0.14	-0.14	-0.10	-0.19	-0.08	-0.15	0.01	0.06	0.03	0.16	0.19	0.29	0.31	0.38	0.23	0.49	0.55

(continued)

Item	40	41	42	43	44	45	46	47	48	49	50
40											
41	0.78										
42	0.87	0.69									
43	0.85	0.62	0.98								
44	0.81	0.63	0.95	0.98							
45	0.79	0.52	0.91	0.95	0.96						
46	0.70	0.50	0.83	0.89	0.86	0.88					
47	0.64	0.44	0.79	0.84	0.82	0.82	0.98				
48	0.60	0.40	0.68	0.74	0.81	0.74	0.90	0.96			
49	0.56	0.41	0.71	0.82	0.82	0.76	0.94	0.96	0.90		
50	0.56	0.42	0.76	0.84	0.77	0.77	0.92	0.95	0.89	0.98	

Table A.12

Inter-Item Correlation Combined National Intelligence Test Scale B Test 1: Computation

Item	1	2	3	4	5	6	7	8	9	10	11
1											
2	0.22										
3	-0.09	0.12									
4	0.11	0.24	0.40								
5	0.08	0.15	0.22	0.30							
6	0.08	0.14	0.21	0.19	0.14						
7	0.09	-0.09	0.13	0.07	0.05	0.19					
8	0.14	0.17	0.62	0.37	0.13	0.13	0.02				
9	0.11	0.01	0.01	0.15	-0.01	0.12	0.16	-0.01			
10	0.02	0.10	-0.12	0.00	0.00	0.09	0.24	-0.01	0.26		
11	0.05	0.18	0.04	0.05	0.07	0.16	0.14	0.09	0.19	0.37	
12	0.10	0.22	0.11	0.14	0.07	0.11	0.11	0.02	0.53	0.35	0.26
13	-0.13	0.01	0.02	-0.04	0.00	0.04	0.20	0.02	0.12	0.19	0.17
14	0.10	0.06	-0.14	-0.06	-0.03	0.01	0.20	-0.08	0.18	0.41	0.26
15	0.02	0.11	-0.08	0.05	0.02	0.05	0.15	-0.08	0.11	0.28	0.26
16	0.17	0.07	0.03	0.21	0.03	0.10	0.10	0.03	0.07	0.11	0.13
17	0.00	0.18	0.04	0.15	-0.07	0.07	0.05	0.04	0.13	0.08	0.08
18	-0.12	-0.17	-0.21	0.04	0.04	0.22	0.07	0.10	0.15	0.29	0.13
19	-0.03	0.04	0.17	0.07	-0.01	0.20	-0.04	0.17	0.00	-0.16	-0.04
20	-0.11	-0.04	0.14	-0.07	-0.08	0.02	0.19	0.11	0.26	0.19	0.03
21	0.03	-0.02	-0.09	0.06	0.07	-0.03	0.10	0.00	-0.09	-0.06	-0.08
22	-0.13	0.12	0.11	0.33	0.00	0.04	-0.04	0.05	0.04	-0.08	0.04

(continued)

Item	12	13	14	15	16	17	18	19	20	21	22
13	0.12										
14	0.29	0.24									
15	0.19	0.20	0.51								
16	0.20	0.05	0.27	0.46							
17	0.23	0.06	0.25	0.40	0.68						
18	0.13	0.17	0.53	0.43	0.61	0.54					
19	0.01	-0.07	0.03	0.07	0.31	0.33	0.26				
20	0.25	0.36	0.27	0.32	0.54	0.46	0.44	0.43			
21	-0.11	-0.12	0.01	0.11	0.18	0.29	0.26	0.47	0.26		
22	0.01	-0.03	0.01	0.20	0.28	0.40	0.12	0.39	0.45	0.42	

Table A.13

Inter-Item Correlation Combined National Intelligence Test Scale B Test 2: Information

Item	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1																			
2	0.08																		
3	0.35	0.06																	
4	0.35	-0.18	0.31																
5	0.06	0.46	0.12	-0.42															
6	0.38	-0.18	0.18	0.29	-0.19														
7	0.29	-0.19	0.05	0.28	-0.19	0.43													
8	-0.12	0.29	0.12	0.03	0.23	-0.05	-0.03												
9	-0.08	-0.02	0.31	0.42	-0.14	0.34	0.36	0.07											
10	0.02	0.32	0.06	-0.16	0.32	-0.15	-0.08	0.26	-0.04										
11	0.25	-0.13	-0.06	0.40	-0.17	0.34	0.28	-0.10	0.38	-0.15									
12	0.21	-0.01	0.30	0.20	-0.06	0.39	0.29	0.02	0.27	-0.07	0.35								
13	-0.02	0.17	0.03	-0.04	0.10	0.06	0.00	0.11	0.04	0.17	0.11	0.21							
14	0.12	0.19	0.26	-0.01	0.19	0.15	-0.08	0.16	0.00	0.16	0.16	0.16	0.06						
15	0.02	0.40	0.13	-0.06	0.39	-0.14	-0.13	0.25	-0.01	0.37	0.37	0.07	0.13	0.18					
16	0.23	0.19	0.01	0.23	0.16	0.24	0.11	0.03	0.24	0.01	0.01	0.42	0.07	0.17	0.04				
17	0.26	0.07	0.13	0.26	0.06	0.18	0.06	0.08	0.34	0.02	0.02	0.20	0.09	0.12	0.09	0.21			
18	0.05	-0.06	0.02	0.37	0.00	0.26	0.22	-0.03	0.37	-0.10	-0.10	0.22	0.11	0.03	-0.07	0.33	0.29		
19	0.26	-0.02	0.19	0.28	0.03	0.30	0.27	0.03	0.25	0.05	0.05	0.44	0.16	0.16	0.17	0.23	0.30	0.44	
20	0.21	0.13	0.15	0.15	0.18	0.19	0.07	0.08	0.25	-0.01	-0.01	0.23	0.10	0.13	0.13	0.24	0.27	0.25	0.36
21	0.04	0.45	0.17	-0.26	0.53	-0.17	-0.22	0.36	-0.18	0.45	0.45	0.16	0.21	0.25	0.50	0.06	0.06	-0.04	0.45
22	0.09	0.35	0.10	-0.05	0.30	0.07	-0.02	0.29	0.10	0.21	0.21	0.24	0.18	0.26	0.33	0.24	0.19	0.13	0.51
23	0.18	0.27	0.18	-0.03	0.27	0.08	-0.04	0.20	0.03	0.21	0.21	0.15	0.08	0.28	0.34	0.11	0.15	0.10	0.46
24	0.20	-0.06	0.07	0.18	0.02	0.22	0.18	-0.04	0.22	-0.05	-0.05	0.21	0.09	0.07	-0.07	0.25	0.25	0.28	0.33
25	0.17	0.26	0.14	-0.01	0.26	0.00	0.13	0.26	0.04	0.21	0.21	0.21	0.20	0.14	0.29	0.15	0.18	0.05	0.28
26	0.25	-0.02	0.16	0.24	0.01	0.24	0.24	0.01	0.26	0.03	0.03	0.30	0.09	0.17	0.02	0.22	0.23	0.28	0.49

(continued)

Item	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
27	0.23	0.32	0.12	-0.05	0.35	0.08	0.04	0.23	0.02	0.19	0.19	0.19	0.13	0.27	0.27	0.22	0.23	0.09	0.48
28	0.48	0.21	0.10	0.14	0.17	0.16	0.14	0.15	0.09	0.19	0.19	0.18	0.09	0.29	0.11	0.23	0.22	0.22	0.24
29	0.40	-0.04	0.10	0.32	-0.01	0.38	0.22	-0.02	0.31	-0.07	-0.07	0.35	0.06	0.07	-0.19	0.32	0.23	0.33	0.35
30	0.29	0.21	0.16	0.13	0.13	0.27	0.30	0.09	0.22	0.11	0.11	0.33	0.13	0.18	0.01	0.26	0.21	0.21	0.37
31	0.41	0.03	0.31	0.17	0.10	0.37	0.34	-0.04	0.26	-0.06	-0.06	0.38	0.10	0.20	-0.08	0.37	0.28	0.35	0.48
32	0.24	0.09	0.14	0.25	0.06	0.17	0.13	0.07	0.31	-0.08	-0.08	0.14	0.11	0.20	-0.02	0.13	0.29	0.24	0.34
33	0.34	0.03	0.12	0.19	0.03	0.27	0.12	0.08	0.20	0.03	0.03	0.26	0.13	0.13	0.05	0.14	0.14	0.18	0.35
34	0.27	0.08	0.23	0.24	0.01	0.33	0.21	-0.05	0.24	-0.12	-0.12	0.29	0.04	0.16	-0.08	0.32	0.33	0.30	0.30
35	0.07	0.23	0.35	-0.03	0.25	0.18	0.08	0.20	0.08	0.18	0.18	0.02	0.17	0.26	0.28	0.16	0.19	0.16	0.19
36	0.17	0.26	0.14	0.06	0.30	0.22	0.20	0.18	0.24	0.14	0.14	0.31	0.19	0.22	0.15	0.30	0.29	0.25	0.36
37	0.10	0.17	0.15	-0.04	0.31	0.17	0.11	0.16	0.15	0.16	0.16	0.17	0.12	0.28	0.19	0.25	0.20	0.18	0.29
38	0.25	0.22	0.32	-0.03	0.14	0.18	0.15	0.16	0.09	0.18	0.18	0.13	0.06	0.23	0.20	0.21	0.21	0.15	0.25
39	0.28	0.09	0.46	0.18	0.16	0.19	0.21	0.05	0.16	0.03	0.32	0.23	0.09	0.26	0.07	0.21	0.24	0.18	0.28
40	0.11	0.01	0.30	0.15	0.10	0.15	0.21	0.02	0.11	-0.07	0.21	0.11	-0.04	0.10	-0.12	0.27	0.19	0.16	0.24

(continued)

Item	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40
20																					
21	0.17																				
22	0.31	0.52																			
23	0.25	0.53	0.44																		
24	0.27	0.02	0.15	0.11																	
25	0.29	0.48	0.36	0.30	0.13																
26	0.26	0.14	0.25	0.31	0.40	0.35															
27	0.23	0.48	0.40	0.47	0.18	0.42	0.44														
28	0.27	0.26	0.32	0.32	0.23	0.24	0.37	0.36													
29	0.27	-0.03	0.14	0.13	0.34	0.11	0.40	0.22	0.26												
30	0.25	0.27	0.38	0.32	0.34	0.32	0.41	0.44	0.35	0.43											
31	0.39	0.10	0.26	0.27	0.52	0.24	0.45	0.39	0.44	0.55	0.72										
32	0.22	0.12	0.25	0.24	0.26	0.24	0.26	0.29	0.26	0.37	0.41	0.57									
33	0.19	0.14	0.24	0.15	0.18	0.23	0.26	0.23	0.21	0.25	0.48	0.50	0.30								
34	0.36	-0.02	0.21	0.18	0.42	0.14	0.43	0.28	0.27	0.49	0.54	0.73	0.43	0.44							
35	0.26	0.35	0.39	0.41	0.24	0.37	0.38	0.38	0.31	0.16	0.46	0.52	0.24	0.43	0.49						
36	0.35	0.27	0.42	0.35	0.28	0.32	0.40	0.47	0.31	0.38	0.59	0.67	0.39	0.47	0.60	0.55					
37	0.30	0.33	0.34	0.34	0.24	0.31	0.32	0.46	0.37	0.27	0.51	0.60	0.32	0.37	0.52	0.52	0.70				
38	0.23	0.25	0.32	0.32	0.21	0.29	0.26	0.36	0.34	0.23	0.44	0.52	0.30	0.34	0.45	0.46	0.58	0.60			
39	0.31	0.21	0.33	0.23	0.32	0.21	0.33	0.29	0.32	0.26	0.47	0.57	0.34	0.38	0.53	0.44	0.58	0.68	0.55		
40	0.17	0.03	0.13	0.11	0.23	0.08	0.26	0.17	0.17	0.33	0.27	0.54	0.23	0.25	0.50	0.17	0.37	0.43	0.36	0.53	

Table A.14

Inter-Item Correlation Combined National Intelligence Test Scale B Test 3: Vocabulary

Item	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1																			
2	0.48																		
3	0.61	0.34																	
4	0.36	0.40	0.20																
5	0.35	0.69	0.06	0.09															
6	0.21	0.70	0.06	0.10	0.64														
7	0.23	0.13	0.08	0.12	-0.01	0.25													
8	0.04	0.57	0.01	-0.09	0.45	0.42	0.41												
9	0.45	0.17	0.12	0.25	0.31	0.24	0.40	0.23											
10	0.06	-0.04	0.03	0.15	0.13	-0.18	0.07	0.21	0.26										
11	0.36	0.12	0.35	-0.01	0.33	0.10	0.26	0.26	0.16	0.22									
12	0.18	0.61	0.30	-0.06	0.62	0.60	0.21	0.43	0.20	-0.12	0.61								
13	0.22	0.10	0.29	-0.15	-0.06	0.08	-0.03	-0.14	-0.13	0.03	0.25	0.24							
14	0.12	-0.03	0.00	0.08	-0.03	0.11	0.01	0.07	0.08	0.06	0.12	0.01	0.06						
15	0.23	0.19	-0.11	0.17	0.23	-0.10	-0.07	-0.04	0.33	0.15	0.06	0.00	-0.01	0.03					
16	0.01	0.39	-0.27	0.35	0.26	0.17	-0.06	0.16	0.04	-0.16	-0.07	0.20	-0.13	0.34	0.16				
17	0.17	0.51	0.31	0.03	0.46	0.33	-0.19	0.33	0.25	0.01	0.18	0.38	-0.07	0.12	0.25	0.33			
18	-0.11	0.08	-0.20	-0.01	0.20	0.05	0.02	0.11	0.07	-0.12	0.02	0.17	-0.01	0.15	0.09	0.15	0.13		
19	0.14	0.24	0.03	0.18	0.18	0.27	-0.06	0.10	0.05	0.05	0.05	0.25	0.09	0.10	0.06	0.24	0.32	0.10	
20	0.20	-0.07	0.20	0.04	0.22	0.10	0.01	-0.05	0.03	0.15	0.34	0.15	0.09	0.12	0.03	0.13	0.12	0.13	0.13
21	0.36	0.35	0.21	0.23	0.59	0.29	0.02	0.09	0.17	0.07	0.31	0.38	0.16	-0.11	0.19	0.11	0.29	0.24	0.07
22	0.10	0.13	0.32	0.10	0.33	0.09	0.12	0.03	0.13	0.01	0.24	0.20	0.09	0.22	0.16	0.15	0.14	0.22	0.12
23	0.22	0.17	0.40	0.20	0.38	0.14	0.06	0.14	0.31	0.13	0.41	0.26	0.12	-0.16	0.21	0.01	0.18	0.09	0.00
24	0.36	0.41	0.36	0.06	0.45	0.20	-0.09	0.16	0.04	0.00	0.40	0.35	0.12	0.18	0.08	0.31	0.33	0.14	0.29
25	0.06	0.15	0.14	-0.05	0.19	0.25	0.00	0.09	-0.06	0.11	0.14	0.17	0.11	0.06	0.05	0.11	0.09	0.01	0.12
26	0.19	-0.07	0.33	-0.10	0.05	0.16	-0.07	0.04	0.24	0.08	0.14	0.13	0.12	-0.15	0.09	-0.04	0.01	0.17	-0.09
27	0.33	0.21	0.34	0.09	0.28	0.19	0.04	-0.01	0.13	0.01	0.27	0.27	0.04	-0.10	0.26	0.00	0.16	0.16	-0.02

(continued)

Item	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
28	0.34	0.46	0.35	0.03	0.40	0.15	0.02	0.11	0.13	0.11	0.34	0.35	0.14	-0.14	0.20	0.06	0.25	0.12	0.04
29	0.46	0.15	0.41	-0.05	0.17	0.05	0.04	-0.01	0.20	-0.01	0.16	0.13	0.11	-0.35	0.19	-0.08	0.17	0.13	-0.16
30	0.16	-0.09	0.42	0.04	0.16	0.01	0.04	-0.01	0.25	0.07	0.31	0.22	0.15	0.07	0.20	0.04	0.27	0.19	0.03
31	0.38	0.00	0.33	0.12	0.16	0.09	0.12	0.06	0.11	0.03	0.18	0.16	0.04	-0.05	0.14	-0.03	0.10	0.14	-0.01
32	0.37	0.38	0.10	0.06	0.38	0.18	0.13	0.08	0.18	0.05	0.19	0.28	0.09	-0.19	0.16	0.01	0.24	0.14	0.05
33	0.29	0.29	0.30	0.01	0.50	0.17	0.16	0.14	0.30	-0.09	0.28	0.44	0.22	-0.11	0.21	0.06	0.22	0.12	0.09
34	0.14	0.13	0.35	0.22	0.17	0.24	0.16	0.10	0.18	0.10	0.15	0.27	0.09	0.00	0.13	-0.05	0.14	0.06	0.01
35	0.42	-0.02	0.14	-0.03	0.28	0.05	0.04	0.09	0.24	0.03	0.30	0.25	0.08	-0.18	0.15	-0.02	0.08	0.17	-0.09
36	0.25	0.32	0.23	-0.05	0.41	0.12	0.13	0.08	0.36	-0.05	0.19	0.34	0.16	-0.13	0.21	0.03	0.20	0.15	0.02
37	0.27	0.05	0.34	-0.06	0.40	0.15	0.11	-0.05	0.31	0.01	0.06	0.32	0.05	-0.12	0.10	-0.01	0.18	0.17	-0.02
38	0.22	-0.14	0.29	0.00	0.20	0.08	0.08	-0.05	0.31	0.00	0.10	0.19	0.12	-0.20	0.14	-0.08	0.16	0.14	-0.11
39	0.17	0.03	0.13	0.05	0.35	0.14	0.16	0.01	0.28	0.10	0.22	0.27	0.13	-0.12	0.18	0.00	0.25	0.14	-0.02
40	0.16	-0.01	0.42	-0.01	0.17	0.08	0.19	0.04	0.06	0.06	0.33	0.14	0.15	0.03	0.06	-0.04	0.17	0.06	0.01

(continued)

Item	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39
20																				
21	0.33																			
22	0.26	0.37																		
23	0.29	0.52	0.37																	
24	0.18	0.42	0.32	0.37																
25	0.15	0.13	0.19	0.28	0.25															
26	0.18	0.21	0.26	0.28	0.11	0.21														
27	0.25	0.48	0.20	0.46	0.23	0.10	0.37													
28	0.15	0.45	0.16	0.40	0.21	0.13	0.21	0.43												
29	0.12	0.39	0.13	0.35	0.10	0.03	0.41	0.53	0.41											
30	0.24	0.33	0.25	0.36	0.24	0.07	0.23	0.47	0.38	0.37										
31	0.27	0.25	0.34	0.34	0.10	0.04	0.30	0.34	0.31	0.42	0.30									
32	0.05	0.23	0.17	0.27	0.17	0.07	0.15	0.28	0.34	0.38	0.27	0.29								
33	0.23	0.42	0.26	0.41	0.25	0.12	0.25	0.43	0.44	0.45	0.43	0.46	0.51							
34	0.14	0.27	0.19	0.28	0.14	0.21	0.21	0.28	0.26	0.30	0.29	0.33	0.32	0.54						
35	0.12	0.35	0.22	0.33	0.07	0.06	0.29	0.41	0.29	0.51	0.30	0.39	0.37	0.53	0.44					
36	0.17	0.44	0.34	0.41	0.22	0.18	0.35	0.43	0.36	0.51	0.39	0.41	0.41	0.69	0.54	0.62				
37	0.07	0.36	0.21	0.25	0.11	0.09	0.30	0.36	0.35	0.54	0.29	0.30	0.37	0.55	0.40	0.51	0.70			
38	0.17	0.36	0.26	0.31	0.11	0.02	0.40	0.44	0.33	0.52	0.39	0.38	0.27	0.47	0.35	0.47	0.67	0.68		
39	0.14	0.38	0.23	0.30	0.21	0.08	0.30	0.39	0.34	0.48	0.37	0.38	0.33	0.60	0.43	0.46	0.75	0.70	0.74	
40	0.14	0.20	0.34	0.29	0.09	0.07	0.13	0.27	0.16	0.18	0.26	0.29	0.17	0.44	0.30	0.25	0.47	0.31	0.43	0.65

Table A.15

Inter-Item Correlation Combined National Intelligence Test Scale B Test 4: Analogies

Item	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1																			
2	0.27																		
3	0.52	0.24																	
4	0.28	0.45	0.18																
5	0.50	0.39	0.33	0.56															
6	0.55	0.37	0.43	0.33	0.42														
7	0.37	0.21	0.19	0.20	0.33	0.33													
8	0.10	0.27	0.15	0.25	0.23	0.22	0.35												
9	0.19	0.36	0.25	0.16	0.19	0.34	0.32	0.30											
10	0.39	0.48	0.30	0.41	0.43	0.36	0.37	0.44	0.50										
11	0.16	0.04	0.25	0.06	0.15	0.05	0.00	0.08	0.05	0.03									
12	0.39	0.58	0.28	0.59	0.57	0.52	0.44	0.40	0.51	0.73	-0.04								
13	0.42	0.29	0.31	0.48	0.46	0.38	0.17	0.26	0.15	0.30	0.23	0.40							
14	0.37	0.41	0.35	0.46	0.53	0.44	0.35	0.40	0.29	0.56	0.11	0.71	0.50						
15	0.42	0.37	0.23	0.52	0.48	0.41	0.32	0.37	0.30	0.49	0.07	0.64	0.40	0.68					
16	0.35	0.37	0.29	0.35	0.41	0.36	0.34	0.36	0.29	0.48	0.08	0.56	0.39	0.62	0.54				
17	0.55	0.40	0.28	0.41	0.50	0.46	0.34	0.40	0.36	0.53	-0.02	0.62	0.37	0.59	0.63	0.53			
18	0.35	0.31	0.18	0.30	0.38	0.39	0.21	0.28	0.25	0.32	0.04	0.41	0.32	0.39	0.40	0.43	0.45		
19	0.37	0.39	0.31	0.49	0.41	0.31	0.22	0.28	0.20	0.34	0.12	0.49	0.42	0.46	0.46	0.34	0.43	0.34	
20	0.45	0.47	0.28	0.60	0.60	0.46	0.29	0.30	0.26	0.49	-0.12	0.73	0.51	0.61	0.61	0.52	0.61	0.44	0.60
21	0.58	0.22	0.28	0.41	0.41	0.39	0.12	0.12	0.07	0.15	0.09	0.26	0.44	0.29	0.36	0.21	0.32	0.26	0.40
22	0.42	0.30	0.28	0.46	0.41	0.33	0.22	0.26	0.19	0.38	0.17	0.42	0.47	0.41	0.37	0.32	0.38	0.31	0.38

(continued)

Item	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
23	0.48	0.52	0.31	0.59	0.57	0.49	0.39	0.29	0.35	0.57	0.02	0.75	0.51	0.63	0.57	0.53	0.61	0.36	0.49
24	0.3	0.31	0.23	0.49	0.43	0.41	0.25	0.2	0.25	0.31	0.1	0.48	0.42	0.43	0.44	0.34	0.42	0.33	0.44
25	0.57	0.35	0.27	0.49	0.43	0.44	0.28	0.28	0.22	0.36	0.02	0.56	0.45	0.48	0.48	0.36	0.46	0.35	0.41
26	0.5	0.4	0.28	0.59	0.52	0.45	0.27	0.34	0.28	0.49	0.06	0.65	0.53	0.54	0.53	0.42	0.52	0.34	0.48
27	0.3	0.34	0.26	0.34	0.34	0.35	0.33	0.29	0.3	0.29	0.09	0.41	0.46	0.4	0.42	0.27	0.41	0.2	0.41
28	0.43	0.08	0.19	0.28	0.26	0.32	0.08	0.05	-0.03	-0.05	0.13	0	0.41	0.1	0.14	0.11	0.08	0.25	0.22
29	0.46	0.22	0.17	0.34	0.28	0.36	0.2	0.19	0.15	0.17	0.07	0.29	0.4	0.26	0.29	0.27	0.3	0.32	0.32
30	0.34	0.01	0.19	0.22	0.19	0.23	-0.11	-0.11	-0.09	-0.17	0.14	-0.17	0.29	-0.12	0	-0.07	-0.08	0.18	0.28
31	0.21	0.18	0.13	0.12	0.13	0.22	0.38	0.16	0.27	0.18	0.18	0.2	0.32	0.29	0.29	0.31	0.12	0.21	0.26
32	0.21	0.04	0.06	0.08	0.06	0.14	0.07	-0.03	-0.05	-0.17	0.03	-0.09	0.18	-0.04	0.06	0.03	-0.01	0.18	0.08

(continued)

Item	20	21	22	23	24	25	26	27	28	29	30	31	32
20													
21	0.59												
22	0.52	0.48											
23	0.86	0.51	0.59										
24	0.62	0.47	0.51	0.66									
25	0.66	0.43	0.46	0.66	0.61								
26	0.72	0.47	0.57	0.76	0.67	0.74							
27	0.50	0.39	0.41	0.51	0.50	0.53	0.65						
28	0.27	0.55	0.34	0.34	0.43	0.41	0.55	0.52					
29	0.38	0.44	0.40	0.48	0.48	0.45	0.57	0.57	0.73				
30	0.10	0.52	0.25	0.10	0.35	0.21	0.21	0.29	0.71	0.46			
31	0.23	0.26	0.31	0.41	0.31	0.33	0.43	0.45	0.57	0.52	0.33		
32	0.06	0.38	0.21	0.17	0.25	0.21	0.20	0.27	0.61	0.49	0.51	0.52	

Table A.16

Inter-Item Correlation Combined National Intelligence Test Scale B Test 5: Comparisons

Item	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1																			
2	0.59																		
3	0.69	0.35																	
4	0.31	-0.04	0.09																
5	0.21	0.29	0.15	0.54															
6	0.19	0.09	0.00	0.19	0.12														
7	0.14	0.32	-0.02	0.02	0.24	0.61													
8	0.39	0.04	0.26	0.09	0.10	0.88	0.65												
9	0.16	-0.03	0.10	0.11	-0.06	0.79	0.53	0.87											
10	0.12	0.40	-0.08	-0.05	0.31	0.74	0.78	0.80	0.73										
11	0.34	-0.01	-0.02	0.09	0.21	0.52	0.29	0.54	0.47	0.45									
12	0.24	0.44	0.04	0.06	0.33	0.38	0.47	0.38	0.26	0.51	0.44								
13	0.46	0.49	0.15	0.17	0.34	0.44	0.44	0.43	0.29	0.48	0.53	0.69							
14	0.15	-0.21	0.05	-0.01	-0.06	0.22	0.03	0.31	0.17	0.13	0.55	-0.04	0.11						
15	0.42	0.30	0.20	0.24	0.26	0.63	0.43	0.67	0.59	0.62	0.83	0.66	0.75	0.66					
16	0.39	0.20	0.17	0.09	-0.11	0.46	0.09	0.45	0.39	0.30	0.36	0.16	0.28	0.24	0.61				
17	0.41	0.35	0.21	0.06	0.29	0.33	0.24	0.23	0.22	0.28	0.19	0.28	0.32	0.07	0.42	0.56			
18	0.30	0.26	0.06	0.14	0.23	0.21	0.25	0.21	0.24	0.26	0.08	0.27	0.30	-0.03	0.39	0.55	0.47		
19	0.16	0.12	0.09	0.14	0.02	0.39	0.17	0.32	0.26	0.17	0.22	0.11	0.21	0.10	0.54	0.80	0.39	0.59	
20	0.10	0.14	0.12	0.16	0.01	0.28	0.11	0.27	0.23	0.14	0.18	0.05	0.15	0.06	0.49	0.73	0.25	0.52	0.90
21	0.00	0.32	0.00	0.07	0.19	0.54	0.67	0.63	0.49	0.73	0.33	0.40	0.34	0.07	0.54	0.55	0.29	0.43	0.69
22	0.05	0.08	0.19	0.09	-0.05	0.57	0.34	0.66	0.54	0.51	0.37	0.23	0.25	0.14	0.57	0.61	0.29	0.35	0.79
23	0.02	0.08	0.15	0.14	0.04	0.59	0.33	0.65	0.54	0.50	0.34	0.23	0.24	0.04	0.55	0.57	0.24	0.30	0.73
24	-0.02	0.21	0.10	0.04	0.11	0.55	0.53	0.63	0.50	0.61	0.34	0.34	0.31	0.00	0.54	0.52	0.22	0.31	0.68
25	0.10	0.11	0.11	0.10	0.14	0.50	0.26	0.62	0.46	0.41	0.28	0.24	0.28	-0.01	0.51	0.52	0.20	0.25	0.64
26	0.26	0.16	-0.08	0.06	0.08	-0.03	0.16	0.05	-0.09	0.10	-0.10	0.20	0.16	-0.14	0.07	0.22	0.15	0.20	0.30
27	0.03	0.08	-0.01	0.05	0.08	-0.17	0.03	-0.02	-0.15	0.03	-0.07	0.23	0.10	-0.25	0.01	0.20	-0.03	0.04	0.34

(continued)

Item	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
28	0.24	0.17	0.12	0.04	0.18	-0.03	0.20	0.09	-0.08	0.15	-0.04	0.34	0.29	-0.17	0.05	0.17	0.15	0.29	0.32
29	0.06	0.06	-0.05	0.07	0.00	-0.20	-0.01	-0.06	-0.15	-0.02	-0.04	0.28	0.16	-0.28	0.00	0.14	-0.10	-0.02	0.28
30	0.03	0.09	-0.09	0.06	0.03	-0.12	0.05	-0.04	-0.12	0.03	-0.05	0.31	0.18	-0.29	-0.02	0.09	-0.09	-0.07	0.26
31	0.13	0.16	-0.02	0.12	0.15	-0.03	0.12	-0.05	-0.10	0.10	-0.03	0.38	0.31	-0.29	0.06	0.15	0.03	0.13	0.23
32	0.11	0.12	-0.15	0.07	0.07	-0.11	0.05	-0.03	-0.11	0.07	-0.03	0.35	0.24	-0.26	0.01	0.18	-0.04	-0.02	0.21
33	0.16	0.20	-0.08	0.04	0.18	-0.19	0.08	-0.13	-0.17	0.07	-0.08	0.45	0.30	-0.31	0.01	0.10	-0.01	0.06	0.16
34	0.11	0.19	-0.11	0.04	0.10	-0.14	0.02	-0.13	-0.13	0.10	-0.04	0.40	0.22	-0.23	0.07	0.14	-0.02	-0.01	0.13
35	0.08	0.15	-0.11	0.06	0.09	-0.06	0.01	-0.08	-0.08	0.10	0.05	0.42	0.27	-0.21	0.14	0.13	-0.05	-0.03	0.13
36	0.19	0.20	-0.08	0.01	0.15	-0.01	0.13	0.01	-0.01	0.21	0.10	0.53	0.39	-0.24	0.31	0.00	0.04	0.02	0.05
37	0.04	-0.07	-0.13	0.03	-0.02	-0.08	-0.11	-0.02	-0.07	0.01	0.17	0.30	0.18	0.08	0.18	0.03	-0.03	-0.11	-0.06
38	0.13	0.13	-0.05	0.06	0.15	-0.06	0.09	-0.05	-0.05	0.16	0.05	0.52	0.37	-0.26	0.29	-0.08	0.03	0.02	-0.02
39	0.12	0.14	0.00	0.08	0.09	-0.01	0.04	-0.01	-0.03	0.12	0.11	0.45	0.30	-0.21	0.28	-0.03	0.01	-0.03	0.01
40	0.12	0.15	0.03	0.10	0.10	-0.02	0.03	-0.07	-0.04	0.08	0.08	0.42	0.28	-0.21	0.27	-0.09	0.00	-0.04	0.00
41	0.07	0.38	-0.05	0.00	0.11	0.10	0.34	0.13	0.05	0.36	0.02	0.40	0.34	-0.22	0.19	0.08	0.00	0.01	0.06
42	-0.01	0.06	0.03	0.07	0.01	0.11	-0.06	0.06	0.06	0.11	0.02	0.34	0.22	-0.20	0.22	-0.03	-0.03	-0.07	-0.01
43	-0.03	0.06	-0.03	0.10	0.04	0.10	-0.07	0.06	0.08	0.10	0.01	0.33	0.23	-0.20	0.19	-0.02	-0.05	-0.09	-0.03
44	-0.09	0.27	0.02	0.12	0.20	0.19	0.15	0.11	0.04	0.30	-0.04	0.47	0.31	-0.25	0.24	-0.03	0.02	-0.02	0.01
45	0.25	0.16	-0.01	0.10	0.20	0.11	0.04	0.08	0.06	0.30	-0.04	0.40	0.28	-0.24	0.17	-0.08	-0.05	-0.08	-0.08
46	-0.05	-0.02	0.12	0.05	-0.03	-0.17	-0.18	-0.24	-0.16	-0.08	-0.13	0.15	0.03	-0.19	-0.09	0.03	-0.11	-0.11	0.02
47	-0.08	0.04	0.16	0.03	-0.05	-0.20	-0.19	-0.24	-0.16	-0.09	-0.14	0.18	0.03	-0.17	-0.07	0.05	-0.12	-0.13	-0.01
48	-0.04	0.17	-0.05	0.20	0.30	-0.07	-0.05	-0.16	-0.17	0.01	-0.02	0.30	0.15	-0.21	0.07	0.18	-0.04	0.02	0.09
49	-0.03	0.29	0.06	0.12	0.14	-0.06	-0.10	-0.03	-0.07	0.00	-0.04	0.29	0.19	-0.17	0.08	0.06	-0.11	-0.01	0.11
50	-0.01	0.02	0.07	0.06	-0.10	-0.13	-0.16	-0.09	-0.08	-0.01	-0.05	0.11	0.02	-0.16	-0.02	0.08	-0.17	-0.12	0.08

(continued)

Item	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39
20																				
21	0.72																			
22	0.75	0.87																		
23	0.73	0.82	0.96																	
24	0.66	0.88	0.93	0.96																
25	0.60	0.73	0.90	0.94	0.93															
26	0.36	0.34	0.38	0.50	0.57	0.60														
27	0.41	0.34	0.45	0.57	0.58	0.66	0.82													
28	0.33	0.31	0.34	0.45	0.45	0.52	0.67	0.90												
29	0.38	0.28	0.36	0.49	0.49	0.57	0.69	0.95	0.87											
30	0.31	0.25	0.35	0.44	0.44	0.53	0.64	0.92	0.81	0.99										
31	0.30	0.26	0.32	0.41	0.43	0.48	0.60	0.86	0.79	0.95	0.97									
32	0.27	0.21	0.31	0.38	0.39	0.47	0.54	0.83	0.68	0.93	0.96	0.97								
33	0.23	0.17	0.24	0.32	0.34	0.40	0.51	0.80	0.69	0.91	0.94	0.95	0.99							
34	0.25	0.11	0.22	0.29	0.30	0.35	0.44	0.77	0.58	0.88	0.90	0.91	0.97	0.97						
35	0.20	0.10	0.22	0.25	0.25	0.34	0.41	0.72	0.54	0.85	0.89	0.87	0.95	0.95	0.98					
36	0.10	0.17	0.20	0.21	0.30	0.28	0.34	0.55	0.46	0.66	0.72	0.73	0.79	0.81	0.85	0.91				
37	0.06	-0.06	0.08	0.07	0.09	0.19	0.21	0.44	0.27	0.51	0.58	0.52	0.61	0.62	0.68	0.71	0.85			
38	0.03	0.09	0.14	0.17	0.24	0.23	0.23	0.48	0.38	0.60	0.64	0.65	0.70	0.74	0.78	0.83	0.95	0.80		
39	0.03	0.05	0.15	0.17	0.19	0.21	0.20	0.43	0.33	0.59	0.63	0.63	0.69	0.71	0.77	0.82	0.93	0.79	0.97	
40	0.02	0.03	0.12	0.14	0.17	0.19	0.20	0.44	0.36	0.59	0.63	0.61	0.67	0.69	0.74	0.79	0.92	0.77	0.96	0.99
41	0.02	0.24	0.27	0.25	0.37	0.23	0.15	0.35	0.26	0.43	0.47	0.44	0.52	0.56	0.60	0.63	0.75	0.38	0.80	0.83
42	0.07	0.00	0.17	0.19	0.18	0.24	0.13	0.38	0.26	0.51	0.55	0.53	0.60	0.60	0.67	0.72	0.83	0.59	0.86	0.93
43	0.03	-0.01	0.15	0.14	0.15	0.19	0.10	0.34	0.23	0.47	0.51	0.48	0.56	0.57	0.65	0.69	0.80	0.57	0.84	0.91
44	0.07	0.12	0.17	0.16	0.27	0.20	0.10	0.30	0.23	0.39	0.44	0.45	0.51	0.55	0.59	0.64	0.80	0.47	0.83	0.88
45	0.02	0.10	0.15	0.15	0.22	0.18	0.06	0.25	0.15	0.36	0.40	0.39	0.48	0.51	0.57	0.62	0.79	0.45	0.81	0.87
46	0.06	-0.14	0.04	0.05	0.03	0.08	0.11	0.30	0.14	0.44	0.49	0.45	0.53	0.55	0.62	0.66	0.66	0.45	0.71	0.78
47	0.03	-0.14	0.03	0.02	-0.01	0.07	0.06	0.25	0.12	0.40	0.41	0.38	0.48	0.49	0.56	0.61	0.63	0.45	0.68	0.75
48	0.09	-0.01	0.01	0.02	0.04	0.06	0.03	0.17	0.15	0.29	0.29	0.37	0.42	0.48	0.49	0.54	0.61	0.33	0.64	0.67
49	0.11	-0.05	0.05	0.02	0.05	0.04	0.04	0.20	0.17	0.32	0.35	0.40	0.42	0.47	0.51	0.54	0.57	0.33	0.66	0.68
50	0.02	-0.10	0.07	0.03	0.02	0.08	-0.05	0.21	0.06	0.33	0.36	0.29	0.40	0.42	0.49	0.52	0.53	0.38	0.61	0.68

(continued)

Item	40	41	42	43	44	45	46	47	48	49	50
40											
41	0.84										
42	0.94	0.77									
43	0.92	0.73	0.99								
44	0.88	0.73	0.97	0.98							
45	0.87	0.66	0.95	0.97	0.97						
46	0.79	0.59	0.88	0.91	0.87	0.90					
47	0.76	0.55	0.85	0.88	0.84	0.85	0.99				
48	0.71	0.52	0.77	0.80	0.84	0.82	0.92	0.97			
49	0.70	0.51	0.79	0.85	0.85	0.83	0.95	0.97	0.93		
50	0.68	0.52	0.80	0.85	0.79	0.80	0.94	0.95	0.90	0.98	

Table A.17

Item Total Correlation and Difficulty

Test	Item	Combined D	Combined ITC	1934 D	1934 ITC	2006 D	2006 ITC
Computation	1	0.98	0.27	0.98	0.27	0.98	0.25
Computation	2	0.98	0.39	0.98	0.39	0.99	0.46
Computation	3	0.94	0.35	0.89	0.47	0.99	0.35
Computation	4	0.97	0.43	0.95	0.49	0.98	0.42
Computation	5	0.94	0.28	0.91	0.27	0.96	0.33
Computation	6	0.85	0.42	0.81	0.38	0.87	0.51
Computation	7	0.93	0.46	0.95	0.42	0.91	0.43
Computation	8	0.91	0.34	0.86	0.44	0.96	0.36
Computation	9	0.35	0.52	0.44	0.49	0.22	0.58
Computation	10	0.58	0.57	0.74	0.44	0.46	0.61
Computation	11	0.68	0.56	0.76	0.45	0.62	0.61
Computation	12	0.31	0.61	0.36	0.60	0.27	0.64
Computation	13	0.52	0.44	0.64	0.47	0.40	0.36
Computation	14	0.30	0.61	0.44	0.58	0.20	0.62
Computation	15	0.25	0.63	0.29	0.63	0.22	0.61
Computation	16	0.38	0.63	0.34	0.72	0.40	0.62
Computation	17	0.24	0.61	0.23	0.71	0.22	0.55
Computation	18	0.01	0.59	0.02	0.56	0.00	0.57
Computation	19	0.12	0.35	0.06	0.49	0.16	0.35
Computation	20	0.01	0.62	0.02	0.64	0.00	0.74
Computation	21	0.01	0.28	0.01	0.42	0.01	0.09
Computation	22	0.09	0.39	0.06	0.47	0.11	0.38
Information	1	0.99	0.43	0.98	0.44	0.99	0.41
Information	2	0.72	0.40	0.91	0.46	0.56	0.47
Information	3	0.98	0.36	0.98	0.54	0.99	0.27
Information	4	0.93	0.26	0.83	0.33	0.99	0.07
Information	5	0.77	0.40	0.97	0.25	0.61	0.55
Information	6	0.89	0.37	0.80	0.46	0.96	0.24
Information	7	0.96	0.27	0.92	0.33	0.99	-0.06
Information	8	0.45	0.31	0.60	0.41	0.32	0.23
Information	9	0.94	0.38	0.89	0.45	0.99	0.07
Information	10	0.70	0.28	0.89	0.42	0.50	0.30
Information	11	0.82	0.41	0.67	0.45	0.95	0.55
Information	12	0.96	0.47	0.95	0.61	0.98	0.24
Information	13	0.81	0.31	0.85	0.34	0.77	0.31
Information	14	0.44	0.41	0.52	0.38	0.44	0.40
Information	15	0.38	0.33	0.63	0.41	0.21	0.17
Information	16	0.51	0.47	0.46	0.52	0.59	0.48
Information	17	0.46	0.46	0.38	0.59	0.52	0.38
Information	18	0.62	0.41	0.45	0.51	0.73	0.44
Information	19	0.91	0.63	0.88	0.85	0.95	0.25
Information	20	0.45	0.53	0.42	0.59	0.53	0.43
Information	21	0.47	0.51	0.79	0.76	0.24	0.50
Information	22	0.37	0.61	0.52	0.72	0.30	0.48
Information	23	0.58	0.56	0.72	0.81	0.50	0.36

(continued)

Test	Item	Combined D	Combined ITC	1934 D	1934 ITC	2006 D	2006 ITC
Information	24	0.40	0.47	0.27	0.50	0.55	0.54
Information	25	0.27	0.53	0.39	0.60	0.19	0.41
Information	26	0.68	0.57	0.59	0.70	0.74	0.40
Information	27	0.47	0.64	0.60	0.75	0.38	0.52
Information	28	0.26	0.57	0.28	0.65	0.27	0.49
Information	29	0.38	0.51	0.23	0.65	0.52	0.51
Information	30	0.54	0.70	0.51	0.85	0.59	0.49
Information	31	0.59	0.75	0.46	0.83	0.75	0.70
Information	32	0.21	0.54	0.15	0.64	0.29	0.53
Information	33	0.31	0.52	0.28	0.76	0.34	0.35
Information	34	0.47	0.64	0.31	0.78	0.66	0.61
Information	35	0.10	0.65	0.15	0.80	0.07	0.34
Information	36	0.27	0.76	0.28	0.86	0.29	0.64
Information	37	0.21	0.70	0.25	0.78	0.21	0.59
Information	38	0.13	0.62	0.17	0.78	0.13	0.36
Information	39	0.25	0.64	0.23	0.79	0.31	0.50
Information	40	0.12	0.46	0.08	0.64	0.18	0.40
Vocabulary	1	1.00	0.38	0.99	0.40	1.00	0.00
Vocabulary	2	0.99	0.36	0.99	0.48	1.00	0.22
Vocabulary	3	0.99	0.41	0.98	0.35	1.00	0.47
Vocabulary	4	0.98	0.17	0.99	0.35	0.98	0.19
Vocabulary	5	0.99	0.53	0.97	0.56	0.99	0.33
Vocabulary	6	0.98	0.32	0.98	0.41	0.99	0.23
Vocabulary	7	0.99	0.19	0.98	0.21	0.95	0.10
Vocabulary	8	0.95	0.22	0.96	0.38	0.95	0.13
Vocabulary	9	0.97	0.36	0.95	0.30	0.99	0.32
Vocabulary	10	0.96	0.16	0.96	0.22	0.96	0.11
Vocabulary	11	0.98	0.41	0.98	0.50	0.99	0.33
Vocabulary	12	0.98	0.47	0.96	0.45	0.99	0.43
Vocabulary	13	0.96	0.24	0.97	0.32	0.96	0.21
Vocabulary	14	0.85	0.06	0.98	0.16	0.72	0.31
Vocabulary	15	0.91	0.32	0.88	0.34	0.93	0.24
Vocabulary	16	0.88	0.20	0.94	0.42	0.84	0.27
Vocabulary	17	0.95	0.40	0.94	0.48	0.96	0.37
Vocabulary	18	0.49	0.35	0.46	0.31	0.54	0.41
Vocabulary	19	0.77	0.18	0.89	0.44	0.69	0.24
Vocabulary	20	0.82	0.37	0.84	0.45	0.80	0.41
Vocabulary	21	0.92	0.57	0.87	0.59	0.95	0.48
Vocabulary	22	0.84	0.48	0.90	0.60	0.84	0.52
Vocabulary	23	0.90	0.54	0.84	0.65	0.95	0.30
Vocabulary	24	0.86	0.42	0.89	0.62	0.81	0.44
Vocabulary	25	0.69	0.30	0.73	0.50	0.67	0.23
Vocabulary	26	0.46	0.51	0.31	0.43	0.62	0.45
Vocabulary	27	0.75	0.60	0.60	0.62	0.85	0.55
Vocabulary	28	0.76	0.54	0.65	0.55	0.85	0.43
Vocabulary	29	0.63	0.61	0.33	0.56	0.89	0.57
Vocabulary	30	0.57	0.59	0.47	0.63	0.62	0.52
Vocabulary	31	0.53	0.58	0.44	0.63	0.62	0.47
Vocabulary	32	0.51	0.55	0.37	0.57	0.63	0.45
Vocabulary	33	0.71	0.71	0.62	0.81	0.78	0.55

(continued)

		Combined		1934		2006	
Test	Item	D	ITC	D	ITC	D	ITC
Vocabulary	34	0.40	0.63	0.35	0.75	0.45	0.47
Vocabulary	35	0.44	0.67	0.25	0.70	0.62	0.53
Vocabulary	36	0.54	0.80	0.39	0.83	0.67	0.70
Vocabulary	37	0.38	0.73	0.20	0.72	0.57	0.65
Vocabulary	38	0.32	0.76	0.12	0.78	0.50	0.67
Vocabulary	39	0.43	0.78	0.29	0.78	0.58	0.70
Vocabulary	40	0.26	0.59	0.28	0.79	0.26	0.40
Analogies	1	0.98	0.72	0.95	0.72	1.00	0.53
Analogies	2	0.60	0.58	0.49	0.48	0.70	0.61
Analogies	3	0.79	0.48	0.71	0.39	0.85	0.52
Analogies	4	0.67	0.68	0.38	0.49	0.88	0.69
Analogies	5	0.86	0.71	0.75	0.69	0.95	0.65
Analogies	6	0.89	0.67	0.82	0.65	0.94	0.61
Analogies	7	0.78	0.48	0.79	0.55	0.76	0.60
Analogies	8	0.35	0.47	0.35	0.51	0.34	0.58
Analogies	9	0.44	0.45	0.48	0.44	0.41	0.62
Analogies	10	0.66	0.64	0.61	0.58	0.68	0.76
Analogies	11	0.64	0.20	0.59	0.15	0.67	0.24
Analogies	12	0.63	0.78	0.52	0.75	0.71	0.81
Analogies	13	0.72	0.66	0.54	0.65	0.85	0.50
Analogies	14	0.75	0.74	0.67	0.80	0.78	0.71
Analogies	15	0.72	0.72	0.60	0.73	0.79	0.67
Analogies	16	0.64	0.63	0.64	0.77	0.67	0.66
Analogies	17	0.62	0.69	0.56	0.75	0.67	0.68
Analogies	18	0.52	0.56	0.50	0.73	0.74	0.62
Analogies	19	0.34	0.64	0.17	0.64	0.48	0.53
Analogies	20	0.70	0.83	0.50	0.80	0.84	0.77
Analogies	21	0.65	0.60	0.40	0.56	0.86	0.30
Analogies	22	0.39	0.65	0.18	0.54	0.58	0.58
Analogies	23	0.41	0.82	0.23	0.77	0.55	0.81
Analogies	24	0.40	0.70	0.20	0.69	0.58	0.62
Analogies	25	0.35	0.72	0.20	0.73	0.47	0.65
Analogies	26	0.37	0.81	0.17	0.76	0.53	0.77
Analogies	27	0.17	0.67	0.09	0.72	0.22	0.54
Analogies	28	0.40	0.51	0.19	0.52	0.60	0.27
Analogies	29	0.22	0.62	0.12	0.56	0.34	0.57
Analogies	30	0.13	0.32	0.01	0.38	0.26	0.08
Analogies	31	0.07	0.52	0.06	0.40	0.08	0.76
Analogies	32	0.15	0.31	0.08	0.32	0.24	0.15
Comparisons	1	1.00	0.25	0.99	0.42	1.00	0.11
Comparisons	2	0.97	0.29	0.94	0.06	1.00	0.12
Comparisons	3	0.99	0.07	0.99	0.12	0.98	0.21
Comparisons	4	0.96	0.17	0.95	0.13	0.97	0.20
Comparisons	5	0.97	0.23	0.95	0.07	0.99	0.21
Comparisons	6	0.97	0.31	0.96	0.37	0.97	0.23
Comparisons	7	0.89	0.27	0.84	0.21	0.93	0.05
Comparisons	8	0.97	0.36	0.96	0.59	0.98	0.20
Comparisons	9	0.96	0.24	0.97	0.54	0.97	0.10
Comparisons	10	0.95	0.44	0.91	0.28	0.98	0.32
Comparisons	11	0.96	0.24	0.96	0.45	0.96	0.11

(continued)

Test	Item	Combined D	ITC	1934 D	ITC	2006 D	ITC
Comparisons	12	0.89	0.57	0.73	0.24	1.00	0.62
Comparisons	13	0.91	0.47	0.82	0.20	0.99	0.27
Comparisons	14	0.83	-0.11	0.94	0.34	0.73	0.13
Comparisons	15	0.98	0.56	0.96	0.53	0.99	0.39
Comparisons	16	0.97	0.39	0.97	0.53	0.98	0.48
Comparisons	17	0.83	0.17	0.78	0.13	0.86	0.09
Comparisons	18	0.81	0.20	0.77	0.20	0.84	0.16
Comparisons	19	0.96	0.43	0.95	0.68	0.97	0.41
Comparisons	20	0.94	0.46	0.93	0.65	0.96	0.41
Comparisons	21	0.90	0.48	0.85	0.56	0.94	0.33
Comparisons	22	0.93	0.60	0.89	0.81	0.96	0.49
Comparisons	23	0.91	0.63	0.85	0.82	0.96	0.42
Comparisons	24	0.89	0.65	0.81	0.74	0.95	0.46
Comparisons	25	0.87	0.61	0.78	0.78	0.93	0.35
Comparisons	26	0.78	0.50	0.67	0.69	0.84	0.25
Comparisons	27	0.82	0.74	0.67	0.80	0.92	0.53
Comparisons	28	0.68	0.62	0.47	0.69	0.82	0.31
Comparisons	29	0.76	0.82	0.54	0.81	0.93	0.65
Comparisons	30	0.73	0.82	0.48	0.79	0.91	0.66
Comparisons	31	0.65	0.83	0.38	0.79	0.87	0.70
Comparisons	32	0.63	0.85	0.36	0.83	0.85	0.73
Comparisons	33	0.59	0.85	0.29	0.77	0.84	0.76
Comparisons	34	0.52	0.85	0.24	0.78	0.76	0.76
Comparisons	35	0.49	0.86	0.20	0.79	0.74	0.79
Comparisons	36	0.52	0.87	0.19	0.84	0.78	0.72
Comparisons	37	0.34	0.66	0.18	0.79	0.41	0.42
Comparisons	38	0.45	0.85	0.15	0.77	0.71	0.76
Comparisons	39	0.44	0.86	0.14	0.81	0.70	0.77
Comparisons	40	0.43	0.85	0.14	0.80	0.69	0.76
Comparisons	41	0.14	0.76	0.03	0.67	0.25	0.67
Comparisons	42	0.29	0.86	0.09	0.77	0.49	0.81
Comparisons	43	0.27	0.86	0.08	0.73	0.46	0.82
Comparisons	44	0.22	0.86	0.04	0.67	0.38	0.82
Comparisons	45	0.19	0.83	0.04	0.69	0.34	0.77
Comparisons	46	0.16	0.78	0.05	0.47	0.29	0.82
Comparisons	47	0.14	0.76	0.04	0.43	0.24	0.79
Comparisons	48	0.08	0.77	0.02	0.45	0.15	0.73
Comparisons	49	0.09	0.80	0.01	0.40	0.17	0.78
Comparisons	50	0.09	0.73	0.03	0.43	0.16	0.75

Note. D: Difficulty; ITC: Item-Total Correlation.

Table A.18

NOHARM Dimensionality Results for a Single Factor

Subtest	RMSR	GFI
Computation	0.008	0.952
Information	0.012	0.914
Vocabulary	0.012	0.914
Analogies	0.012	0.958
Comparisons	0.015	0.912

Note. RMSR: root mean square residual, GFI: goodness of fit (Tanaka & Huba, 1985).

Table A.19

Model Fit Indices

Model	χ^2	*df*	CFI	TLI	RMSEA
Test 1: Computation					
2-Parameter	905.82	120	0.64	0.64	0.06
1-Parameter	1281.14	124	0.47	0.48	0.07
Rasch	5978.65	122	0.00	-1.67	0.15
Test 2: Information					
2-Parameter	4391.18	367	0.68	0.80	0.07
1-Parameter	5282.36	277	0.60	0.66	0.09
Rasch	13752.65	235	0.00	-0.07	0.16
Test 3: Vocabulary					
2-Parameter	1042.32	208	0.87	0.9	0.04
1-Parameter	3100.58	180	0.55	0.59	0.09
Rasch	7165.46	166	0.00	-0.07	0.14
Test 4: Analogies					
2-Parameter	3889.60	249	0.79	0.9	0.08
1-Parameter	5459.83	171	0.70	0.8	0.12
Rasch	4727.78	116	0.74	0.74	0.14
Test 5: Comparisons					
2-Parameter	4327.97	73	0.94	0.94	0.16
1-Parameter	10580.56	82	0.85	0.86	0.24
Rasch	16938.31	76	0.76	0.76	0.32

Note. *df*: degrees of freedom, CFI: Comparative Fit Index (Bentler, 1990), TLI: Tucker Lewis Index (Tucker & Lewis, 1973), RMSEA: Root Mean Square Error of Approximation (Steiger & Lind, 1980).

Table A.20

National Intelligence Test Scale B Differential Item Functioning Analysis

Subtest	Item	MH	Logistic	TID	Std	BD	#DIF
Computation	3	DIF	DIF	DIF	NoDIF	DIF	4/5
Computation	8	DIF	DIF	DIF	NoDIF	NoDIF	3/5
Computation	9	DIF	DIF	NoDIF	DIF	NoDIF	3/5
Computation	10	DIF	DIF	DIF	DIF	NoDIF	4/5
Computation	13	DIF	DIF	NoDIF	DIF	NoDIF	3/5
Computation	14	DIF	DIF	NoDIF	DIF	NoDIF	3/5
Computation	17	DIF	DIF	NoDIF	NoDIF	DIF	3/5
Computation	19	DIF	DIF	DIF	NoDIF	NoDIF	3/5
Computation	22	DIF	DIF	DIF	NoDIF	NoDIF	3/5
Information	2	DIF	DIF	DIF	DIF	NoDIF	4/5
Information	4	DIF	DIF	DIF	DIF	NoDIF	4/5
Information	5	DIF	DIF	DIF	DIF	DIF	5/5
Information	8	DIF	DIF	DIF	DIF	NoDIF	4/5
Information	10	DIF	DIF	DIF	DIF	NoDIF	4/5
Information	11	DIF	DIF	DIF	DIF	NoDIF	4/5
Information	13	DIF	DIF	DIF	DIF	NoDIF	4/5
Information	14	DIF	DIF	NoDIF	DIF	NoDIF	3/5
Information	15	DIF	DIF	DIF	DIF	NoDIF	4/5
Information	18	DIF	DIF	NoDIF	DIF	NoDIF	3/5
Information	21	DIF	DIF	DIF	DIF	NoDIF	4/5
Information	22	DIF	DIF	DIF	DIF	NoDIF	4/5
Information	23	DIF	DIF	DIF	DIF	DIF	5/5
Information	24	DIF	DIF	NoDIF	DIF	NoDIF	3/5
Information	25	DIF	DIF	DIF	DIF	NoDIF	4/5
Information	27	DIF	DIF	DIF	DIF	NoDIF	4/5
Information	29	DIF	DIF	DIF	DIF	NoDIF	4/5
Information	30	NoDIF	DIF	NoDIF	DIF	DIF	3/5
Information	34	DIF	DIF	DIF	DIF	NoDIF	4/5
Vocabulary	4	DIF	DIF	DIF	NoDIF	NoDIF	3/5
Vocabulary	14	DIF	DIF	DIF	DIF	NoDIF	4/5
Vocabulary	16	DIF	DIF	DIF	DIF	NoDIF	4/5
Vocabulary	19	DIF	DIF	DIF	DIF	NoDIF	4/5
Vocabulary	20	DIF	DIF	DIF	DIF	NoDIF	4/5
Vocabulary	22	DIF	DIF	DIF	DIF	NoDIF	4/5
Vocabulary	24	DIF	DIF	DIF	DIF	NoDIF	4/5
Vocabulary	25	DIF	DIF	DIF	DIF	NoDIF	4/5
Vocabulary	26	DIF	DIF	NoDIF	DIF	NoDIF	3/5
Vocabulary	29	DIF	DIF	DIF	DIF	NoDIF	4/5
Vocabulary	35	DIF	DIF	NoDIF	DIF	NoDIF	3/5
Vocabulary	37	DIF	DIF	NoDIF	DIF	NoDIF	3/5
Vocabulary	38	DIF	DIF	NoDIF	DIF	NoDIF	3/5
Vocabulary	40	DIF	DIF	DIF	DIF	DIF	5/5
Analogies	4	DIF	DIF	DIF	DIF	NoDIF	4/5
Analogies	7	DIF	DIF	DIF	DIF	NoDIF	4/5
Analogies	8	DIF	DIF	DIF	DIF	NoDIF	4/5

(continued)

Subtest	Item	MH	Logistic	TID	Std	BD	#DIF
Analogies	9	DIF	DIF	DIF	DIF	NoDIF	4/5
Analogies	10	DIF	DIF	NoDIF	DIF	NoDIF	3/5
Analogies	14	DIF	DIF	NoDIF	DIF	NoDIF	3/5
Analogies	16	DIF	DIF	DIF	DIF	NoDIF	4/5
Analogies	17	DIF	DIF	NoDIF	DIF	NoDIF	3/5
Analogies	21	DIF	DIF	DIF	DIF	NoDIF	4/5
Analogies	22	DIF	DIF	NoDIF	DIF	NoDIF	3/5
Analogies	24	DIF	DIF	NoDIF	NoDIF	DIF	3/5
Analogies	27	NoDIF	DIF	NoDIF	DIF	DIF	3/5
Analogies	28	DIF	DIF	NoDIF	DIF	DIF	4/5
Analogies	30	DIF	DIF	DIF	DIF	NoDIF	4/5
Analogies	31	DIF	DIF	DIF	NoDIF	NoDIF	3/5
Comparisons	2	DIF	DIF	DIF	NoDIF	NoDIF	3/5
Comparisons	12	DIF	DIF	DIF	DIF	NoDIF	4/5
Comparisons	13	DIF	DIF	DIF	DIF	NoDIF	4/5
Comparisons	14	DIF	DIF	DIF	DIF	NoDIF	4/5
Comparisons	25	DIF	DIF	NoDIF	NoDIF	DIF	3/5
Comparisons	36	DIF	DIF	DIF	NoDIF	DIF	4/5
Comparisons	37	DIF	DIF	NoDIF	DIF	DIF	4/5
Comparisons	38	DIF	DIF	DIF	NoDIF	DIF	4/5
Comparisons	39	DIF	DIF	DIF	NoDIF	DIF	4/5
Comparisons	40	DIF	DIF	DIF	NoDIF	DIF	4/5
Comparisons	45	DIF	DIF	NoDIF	NoDIF	DIF	3/5

Note. MH: Mantel-Haenszel (Holland & Thayer, 1988), Logistic: Logistic regression (Swaminathan & Rogers, 1990), TID: Transformed Item Difficulties (Angoff & Ford, 1973), Std: Standardization (Dorans & Kullick, 1986), BD: Brewlow-Day (Aguerri et al., 2009; Penfield, 2003), #DIF: The number of methods indicting that a given item displays DIF.

Table A.21

Ability Estimate Comparisons

			IRT Estimation					CTT Estimation			
			θ			θ Standard Error		Total Score			
Subtest	Group	n	Mean	SD	ES	Mean	SD	Mean	SD	ES	α
Computation	1934	890	0.04	0.84		0.57	0.09	11.72	2.47		0.58
	2006	913	-0.04	0.77	-0.10	0.61	0.06	10.93	2.29	-0.33	0.56
Information	1934	890	-0.21	1.14		0.33	0.07	22.26	7.38		0.90
	2006	913	0.20	0.65	0.44	0.31	0.05	22.08	4.95	-0.03	0.75
Vocabulary	1934	890	-0.34	0.97		0.36	0.06	28.54	5.07		0.83
	2006	913	0.34	0.75	0.79	0.37	0.07	31.95	4.15	0.74	0.73
Analogies	1934	890	-0.43	0.87		0.29	0.07	13.54	6.08		0.87
	2006	913	0.45	0.86	1.02	0.30	0.09	19.99	5.80	1.09	0.86
Comparisons	1934	890	-0.76	1.03		0.28	0.17	27.94	6.36		0.88
	2006	913	0.59	0.74	1.51	0.14	0.13	38.47	5.96	1.71	0.87

Note. n = sample size ES = Hedges g (Hedges, 1981), effect size, SD = standard deviation, IRT = item response theory, CTT = classical test theory, α = Cronbachs (1951) alpha measure of score reliability.

Table A.22

Effect Sizes Per Year

Subtest	IRT ES	CTT ES/Year	ES	ES/Year
Computation	-0.10	-0.00	-0.33	-0.01
Information	0.44	0.01	-0.03	0.00
Vocabulary	0.79	0.01	0.74	0.01
Analogies	1.02	0.01	1.09	0.02
Comparisons	1.51	0.02	1.71	0.02

Note. IRT: item response theory; CTT: classical test theory; ES: Hedges' (1981) g; ES/Year: $\frac{ES}{2006-1934}$

APPENDIX B

NOHARM Code

Unidimensionality of Data

```
Exploratory Analysis of Scale B Test 1 Combined Years
22 1 2164 1 1 0 1 1
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0.97
0.96 0.98
0.92 0.93 0.94
0.94 0.95 0.92 0.97
0.91 0.92 0.89 0.91 0.94
0.83 0.83 0.81 0.82 0.80 0.85
0.90 0.91 0.88 0.90 0.87 0.79 0.93
0.89 0.90 0.88 0.89 0.86 0.78 0.85 0.91
0.34 0.34 0.33 0.34 0.32 0.30 0.33 0.32 0.35
0.57 0.57 0.54 0.56 0.55 0.50 0.55 0.53 0.24 0.58
0.67 0.67 0.65 0.66 0.64 0.59 0.64 0.63 0.26 0.45 0.68
0.31 0.31 0.30 0.31 0.30 0.27 0.30 0.29 0.19 0.23 0.25 0.31
0.50 0.51 0.49 0.50 0.49 0.44 0.49 0.48 0.20 0.33 0.38 0.18 0.52
0.29 0.29 0.27 0.28 0.27 0.25 0.28 0.26 0.13 0.23 0.23 0.13 0.19 0.29
0.24 0.25 0.23 0.24 0.23 0.22 0.24 0.22 0.10 0.18 0.20 0.10 0.16 0.14 0.25
0.37 0.37 0.36 0.37 0.36 0.33 0.36 0.35 0.14 0.24 0.28 0.15 0.21 0.15 0.15 0.38
0.23 0.23 0.22 0.23 0.22 0.21 0.22 0.22 0.10 0.15 0.17 0.10 0.13 0.10 0.10 0.18 0.24
0.01 0.01 0.01 0.01 0.01 0.01 0.01 0.01 0.00 0.01 0.01 0.00 0.01 0.01 0.01 0.01 0.01 0.01
0.11 0.11 0.11 0.11 0.11 0.10 0.10 0.11 0.04 0.05 0.08 0.04 0.05 0.04 0.03 0.07 0.05 0.00 0.11
0.01 0.01 0.01 0.01 0.01 0.01 0.01 0.01 0.01 0.01 0.01 0.01 0.01 0.01 0.01 0.01 0.01 0.00 0.00 0.01
0.01 0.01 0.01 0.01 0.01 0.01 0.01 0.01 0.00 0.01 0.01 0.00 0.00 0.00 0.00 0.01 0.01 0.00 0.01 0.00 0.01
0.08 0.09 0.09 0.09 0.08 0.08 0.08 0.08 0.03 0.05 0.06 0.03 0.04 0.03 0.03 0.05 0.04 0.00 0.03 0.00 0.00 0.09
```

APPENDIX C

MPlus Code

Inter-Item Correlation, 1934 Estonia NIT Scale B, Computation subtest

```
TITLE: Estonian Data Analysis for FE, NIT Scale B
Data:
  FILE IS ItemDataBZeroForMissing-Mplus.csv;
VARIABLE:
NAMES ARE
  year IDNR grade sex age dateB
b1_1 b1_2 b1_3 b1_4 b1_5 b1_6 b1_7 b1_8 b1_9 b1_10 b1_11
  b1_12 b1_13 b1_14 b1_15 b1_16 b1_17 b1_18 b1_19
  b1_20 b1_21 b1_22;
USEVARIABLES ARE
b1_1 b1_2 b1_3 b1_4 b1_5 b1_6 b1_7 b1_8 b1_9 b1_10 b1_11
  b1_12 b1_13 b1_14 b1_15 b1_16 b1_17 b1_18 b1_19
  b1_20 b1_21 b1_22;
CATEGORICAL ARE
b1_1 b1_2 b1_3 b1_4 b1_5 b1_6 b1_7 b1_8 b1_9 b1_10 b1_11
  b1_12 b1_13 b1_14 b1_15 b1_16 b1_17 b1_18 b1_19
  b1_20 b1_21 b1_22;
USEOBSERVATIONS = year EQ 1934;
ANALYSIS:
  TYPE IS BASIC;
OUTPUT:
SAMPSTAT;
```

APPENDIX D

MPlus Code

Item Difficulty and Total Item Correlation, 1934 Estonia NIT Scale B, Computation subtest

```
TITLE: Estonian Data Analysis for FE, NIT Scale B
Data:
  FILE IS ItemDataBZeroForMissing-Mplus.csv;
VARIABLE:
NAMES ARE
  year IDNR grade sex age dateB
b1_1 b1_2 b1_3 b1_4 b1_5 b1_6 b1_7 b1_8 b1_9 b1_10 b1_11
  b1_12 b1_13 b1_14 b1_15 b1_16 b1_17 b1_18 b1_19
  b1_20 b1_21 b1_22;
USEVARIABLES ARE
b1_1 b1_2 b1_3 b1_4 b1_5 b1_6 b1_7 b1_8 b1_9 b1_10 b1_11
  b1_12 b1_13 b1_14 b1_15 b1_16 b1_17 b1_18 b1_19
  b1_20 b1_21 b1_22;
CATEGORICAL ARE
b1_1 b1_2 b1_3 b1_4 b1_5 b1_6 b1_7 b1_8 b1_9 b1_10 b1_11
  b1_12 b1_13 b1_14 b1_15 b1_16 b1_17 b1_18 b1_19
  b1_20 b1_21 b1_22;
USEOBSERVATIONS = year EQ 1934;
DEFINE:
b1_1 b1_2 b1_3 b1_4 b1_5 b1_6 b1_7 b1_8 b1_9 b1_10 b1_11
  b1_12 b1_13 b1_14 b1_15 b1_16 b1_17 b1_18 b1_19
  b1_20 b1_21 b1_22;
```

```
ANALYSIS:
  TYPE IS BASIC;
OUTPUT:
SAMPSTAT;
```

APPENDIX E

R Code

```
#Load R Packages
library(ltm)
library(psych)
library(nFactors)
library(MplusAutomation)
library(GPArotation)
library(difR)
library(lordif)

#Import Data
ScaleB<-data.frame(read.table("ItemdataBZeroForMissing-Mplus.
   csv", header=F, sep=","))
names(ScaleB)<-c("Group", "IDNR", "grade", "sex", "age", "
   dateB",
"b1_1", "b1_2", "b1_3", "b1_4", "b1_5", "b1_6",
   "b1_7", "b1_8", "b1_9", "b1_10", "b1_11",
   "b1_12", "b1_13", "b1_14", "b1_15", "b1_16",
   "b1_17", "b1_18", "b1_19", "b1_20", "b1_21",
   "b1_22")

#Subset Data by Group
ScaleB.1934<-ScaleB[which(ScaleB$Group==1934), ]
ScaleB.2006<-ScaleB[which(ScaleB$Group==2006), ]
```

```
#Subset Data by Tests
Test1 <- c("b1_1", "b1_2", "b1_3", "b1_4", "b1_5", "b1_6",
   "b1_7", "b1_8", "b1_9", "b1_10", "b1_11", "b1_12", "b1_13",
   "b1_14", "b1_15", "b1_16", "b1_17", "b1_18", "b1_19",
   "b1_20", "b1_21", "b1_22")
ScaleB.Test1 <- ScaleB[Test1]

#Subset Data by Group and Test
ScaleB.1934.Test1<-ScaleB.Test1[which(ScaleB$Group==1934), ]
ScaleB.2006.Test1<-ScaleB.Test1[which(ScaleB$Group==2006), ]

#Cronbach's Alpha
cronbach.alpha(ScaleB.1934.Test1)
cronbach.alpha(ScaleB.2006.Test1)

#Tetrachoric Correlations
ScaleB.Test1.tet<-tetrachoric(ScaleB.Test1)

#Data Sets for DIF analysis
#Scale B for just the 1934 and 2006 Groups
ScaleB.Test1.DIF<-rbind(ScaleB.1934.Test1, ScaleB.2006.Test1)

#Group Indicator Variable
ScaleB.Test1.DIF$Group<-c(rep(0,890), rep(1,913))

#DIF Testing Method
dichoDif(ScaleB.Test1.DIF, group="Group", focal.name=1,
   method=c("MH", "Logistic","TID", "Std", "BD" ),
```

```
   alpha=0.005, purify=TRUE)

#Vertical Linking
#Test1
Test1.names.comp<-paste("b1_", 1:22, sep="")
Test1.names.comp[3]<-"b1_3c"
Test1.names.comp[8]<-"b1_8c"
Test1.names.comp[9]<-"b1_9c"
Test1.names.comp[10]<-"b1_10c"
Test1.names.comp[13]<-"b1_13c"
Test1.names.comp[14]<-"b1_14c"
Test1.names.comp[17]<-"b1_17c"
Test1.names.comp[19]<-"b1_19c"
Test1.names.comp[22]<-"b1_22c"
names(ScaleB.2006.Test1)<-Test1.names.comp
Test1.Forms <- list(ScaleB.1934.Test1, ScaleB.2006.Test1)
Test1.all.items.equating<-testEquatingData(Test1.Forms)
Test1.equated<-ltm(Test1.all.items.equating~z1, control =
   list(GHk = 20, iter.em = 20))

#2 parameter, using ltm
summary(Test1.equated)

#Get Factor Scores
ltm::factor.scores(Test1.equated, method="EAP",
resp.patterns=Test1.all.items.equating)
ScaleB.Test1.DIF$theta<-ltm::factor.scores(Test1.equated,
```

```
    method="EAP", resp.patterns=Test1.all.items.equating)$score
        .dat[,"z1"]
ScaleB.Test1.DIF$theta.se<-ltm::factor.scores(Test1.equated,
    method="EAP", resp.patterns=Test1.all.items.equating)$score
        .dat[,"se.z1"]

##Effect Size
d.raw.data<-function(E.data,C.data){
    d<-(mean(E.data)-mean(C.data))/
    sqrt(((length(E.data)-1)*var(E.data)+
    (length(C.data)-1)*var(C.data))/
    (length(E.data)+length(C.data)-2))
    names(d)<-"effect size d"
    return(d)
}
```

REFERENCES

Agueeri, M. E., Galibert, M. S., Attorresi, H. F., & Maranon, P. P. (2009). Erroneous detection of nonuniform DIF using the Breslow-Day test in a short test. *Quality and Quantity*, *43*, 35-44. doi: 10.1007/s11135-007-9130-2

Allen, M. J., & Yen, W. M. (1979). *Introduction to measurement theory.* Long Grove, IL: Waveland Press.

Andreasen, N. C., Flaum, M., Swayze, V., O'Leary, D., Alliger, R., Cohen, G., ... Yuh, W. T. C. (1993). Intelligence and brain structure in normal individuals. *American Journal of Psychiatry*, *150*, 130-134.

Angoff, W. H., & Ford, S. F. (1973). Item-race interaction on a test of scholastic aptitude. *Journal of Educational Measurement*, *10*, 95-105. doi: 10.1111/j.1745-3984.1973.tb00787.x

Arija, V., Esparo, G., Fernandez-Ballart, J., Murphy, M. M., Biarnes, E., & Canals, J. (2006). Nutritional status and performance in test of verbal and non-verbal intelligence in 6 year old children. *Intelligence*, *34*, 141-149. doi: 10.1016/j.intell.2005.09.001

Baker, F. B. (1984). Ability metric transformations involved in vertical equating under item response theory. *Applied Psychological Measurement*, *8*, 261-271. doi: 10.1177/014662168400800302

Bartholomew, D., & Knott, M. (1999). *Kendall's library of statistics 7: Latent variable models and factor analysis.* London: Edward Arnold.

Bayley, N. (1993). *Bayley Scores of Infant Development.* San Antonio, TX: Psychological Corp.

Beaujean, A. A., & Osterlind, S. J. (2008). Using item response theory to assess the Flynn effect in the National Longitudinal Study of Youth 79 children and young adults data. *Intelligence*, *36*, 455-463. doi: 10.1016/j.paid.2009.10.019

Beaujean, A. A., & Sheng, Y. (2010). Examining the Flynn effect in the General Social Survey using item response theory. *Personality and Individual Differences*, *48*, 294-298. doi: 10.1016/j.paid.2009.10.019

Beaujean, A. A., & Sheng, Y. (2012). Assessing the Flynn effect in the Wechsler scales. Manuscript submitted for publication.

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, *107*, 238-246. doi: 10.1037/0033-2909.107.2.238

Binkely, M., Rust, K., & Wingless, M. (1994). *Methodological issues in comparative educational studies. The case of the IEA reading literacy study.* US Department of Education Office of Educational Research and Improvement.

Bocéréan, C., Fischer, J.-P., & Flieller, A. (2003). Long-term comparison (1921-2001) of numerical knowledge in three to five-and-a-half year-old children. *European Journal of Psychology of Education*, *18*, 405-424. doi: 10.1007/BF03173244

Bontempo, D. E., & Hofer, S. M. (2007). Assessing factorial invariance in cross-sectional and longitudinal studies. In A. D. Ong & M. van Dulmen (Eds.), *Handbook of methods in positive psychology* (pp. 153–175). New York: Oxford University Press.

Borsboom, D. (2005). *Measuring the mind: Conceptual issues in contemporary psychometrics.* Cambridge: Cambridge University.

Borsboom, D., Mellenbergh, G. J., & van Heerden, J. (2002). Different kinds of DIF: A distinction between absolute and relative forms of measurement invariance and bias. *Applied Psychological Measurement*, *26*, 433-450. doi: 10.1177/014662102237798

Bradmetz, J., & Mathy, F. (2006). An estimate of the Flynn effect: Changes in IQ and subtest gains of 10-yr.-old French children between 1965 and 1988. *Psychological Reports*, *99*, 743-746. doi: 10.2466/PR0.99.3.743-746

Brandt, I. (1978). Growth dynamics of low-birth weight infants with emphasis on the perinatal period. In F. Falkner & J. M. Tanner (Eds.), *Human growth: Postnatal growth* (Vol. 2, pp. 557–618). New York: Plenum Press.

Breslau, N., Dickens, W. T., Flynn, J. R., Peterson, E. L., & Lucia, V. C. (2006). Low birthweight and social disadvantage: Tracking their relationship with children's IQ during the period of school attendance. *Intelligence*, *34*, 351-362. doi: 10.1016/j.intell.2005.10.003

Brown, T. A. (2006). *Confirmatory factor analysis for applied research.* New York: Guilford Press.

Byrne, B. M. (1998). *Structural equation modeling with LISREL, PRELIS, and SIMPLIS: Basic concepts, applications, and programming.* Mahwah, NJ: Erlbaum.

Byrne, B. M., Shavelson, R. J., & Muthén, B. (1989). Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. *Psychological Bulletin*, *105*, 456-466. doi: 10.1037/0033-2909.105.3.456

Campbell, H. L., Barry, C. L., Joe, J. N., & Finney, S. J. (2008). Configural, metric, and scalar invariance of the modified achievement goal questionnaire across African American and White university students. *Educational and Psychological Measurement*, *68*, 988-1007. doi: 10.1177/0013164408315269

Ceci, S. J., & Kanaya, T. (2010). "Apples and oranges are both round": Furthering the discussion on the Flynn effect. *Journal of Psychoeducational Assessment*, *28*, 441-447. doi: 10.1037/h0059973

Ceci, S. J., Scullin, M., & Kanaya, T. (2003). The difficulty of basing death penalty eligibility on IQ cutoff scores for mental retardation. *Ethics and Behavior*, *13*, 11-17. doi: 10.1207/S15327019EB1301$_$03

Chan, D. (1998). The conceptualization and analysis of change over time: An integrative approach incorporating longitudinal mean and covariance structures analysis (LMACS) and multiple indicator latent growth modeling (MLGM). *Organizational Research Methods*, *1*, 421-483. doi: 10.1177/109442819814004

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*, *9*, 233-255. doi: 10.1207/S15328007SEM0902$_$5

Churchill, J. A. (1965). The relationship between intelligence and birth weight in twins. *Neurology*, *15*, 341-347. doi: 10.1212/WNL.15.4.341

Cocodia, E. A., Kim, J.-S., Shin, H.-S., Kim, J.-W., Ee, J., Wee, M. S. W., & Howard, R. W. (2003). Evidence that rising population intelligence is impacting in formal education. *Personality and Individual Differences*, *35*, 797-810. doi: 10.1016/S0191-8869(02)00284-2

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*, 155-159. doi: 10.1037//0033-2909.112.1.155

Colom, R., Andrés Pueyo, A., & Juan-Espinosa, M. (1998). Generational IQ gains: Spanish data. *Personality and Individual Differences*, *25*, 927-935. doi: dx.doi.org/10.1016/S0191-8869(98)00090-7

Colom, R., Juan-Espinosa, M., & García, L. F. (2001). The secular increase in test scores is a "Jensen effect". *Personality and Individual Differences*, *30*, 553-559. doi: dx.doi.org/10.1016/S0191-8869(00)00054-4

Colom, R., Jung, R. E., & Haier, R. J. (2006). Distributed brain sites from the *g*-factor of intelligence. *NeuroImage*, *31*, 1359-1365. doi: 10.1016/j.neuroimage.2006.01.006

Colom, R., Lluis-Font, J. M., & Andrés Pueyo, A. (2005). The generational intelligence gains are caused by decreasing variance in the lower half of the distribution: Supporting evidence for the nutrition hypothesis. *Intelligence*, *33*, 83-92. doi: 10.1016/j.intell.2004.07.010

Corry Mann, H. C. (1926). *Special reports series.* London: Medical Research Council.

Covin, T. M. (1976). Comparability of WISC and WISC-R full scale IQs for elementary school children with learning disabilities. *Psychological Reports*, *39*, 280.

Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory.* Belmont, CA: Wadsworth.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, *16*, 297-334. doi: 10.1007/BF02310555

Daley, T. C., Whaley, S. E., Sigman, M. D., Espinosa, M. P., & Neumann, C. (2003). IQ on the rise: The Flynn effect in rural Kenyan children. *Psychological Science*, *14*, 215-219. doi: 10.1111/1467-9280.02434

de Ayala, R. J. (2009). *The theory and practice of item response theory.* New York: Guilford Press.

Devlin, B., Daniels, M., & Roeder, K. (1997). The heritability of IQ. *Nature*, *388*, 468-471.

Dickens, W. T., & Flynn, J. R. (2001). Heritability estimates versus large environmental effects: The IQ paradox resolved. *Psychological Review*, *108*, 346-369. doi: 10.1037//0033-295X.108.2.346

Dickens, W. T., & Flynn, J. R. (2006). Black Americans reduce the racial IQ gap: Evidence from standardized sample. *Psychological Science*, *17*, 913-920. doi: 10.1111/j.1467-9280.2006.01802.x

Dobbing, J. (1984). Infant nutrition and later achievement. *Nutrition Review*, *42*, 1-7. doi: 10.1111/j.1753-4887.1984.tb02233.x

Dorans, N. J., & Kulick, E. (1983). *Assessing unexpected differential item performance of female candidates on SAT and TSWE forms administered in December 1977: An application of the standardization approach (ets research rep. rr-83-9).* Princeton, NJ: Educational Testing Service.

Dorans, N. J., & Kullick, E. (1986). Demonstrating the utility of the standardization approach to assessing unexpected differential item performance on the Scholastic Aptitude Test. *Journal of Educational Measurement*, *23*, 355-368. doi: 10.1111/j.1745-3984.1986.tb00255.x

Educational Testing Service. (1926). *College Board Scholastic Aptitude Test.* Princeton, NJ.

Elley, W. B. (1992). *How in the world do students read?* Hamburg: International Association for the Evaluation of Educational Achievement.

Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists.* Mahwah, NJ: Erlbaum.

Encyclopædia Britannica. (2011). *Estonia.* Retrieved from http://www.britannica.com/EBchecked/topic/193535/Estonia.

Fidalgo, A. M., Mellenbergh, G. J., & Muñiz, J. (2000). Effects of amount of DIF, test length, and purification type on robustness and power of Mantel-Haenszel procedures. *Methods of Psychological Research Online*, *5*, 43-53.

Fisher, R. A. (1956). *Statistical methods and scientific inference.* Edingurgh: Oliver and Boyd.

Fitzgerald, S., Gray, N. S., & Snowden, R. J. (2007). A comparison of WAIS-R and WAIS-III in the lower IQ range: Implications for learning disability diagnosis. *Journal of Applied Research in Intellectual Disabilities*, *20*, 323-330. doi: 10.1111/j.1468-3148.2006.00349.x

Flannery, W. P., Reise, S. P., & Widaman, K. (1995). An item-response theory analysis of the general and academic scales of the Self-Description Questionnaire II. *Journal of Research in Personality*, *29*, 168-188. doi: 10.1006/jrpe.1995.1010

Flieller, A. (1988). Application du modele de Rasch a un probleme de comparaison de generations [Applications of the Rasch model to a problem of intergenerational comparison]. *Bulletin de Psychologie*, *42*, 86-91.

Flora, D. B., & Curran, P. J. (2004). An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data. *Psychological Methods*, *9*, 466-491. doi: 10.1037/1082-989X.9.4.466

Flynn, J. R. (1984). The mean IQ of Americans: Massive gains 1932 to 1978. *Psychological Bulletin*, *95*, 29-51. doi: 10.1037/0033-2909.95.1.29

Flynn, J. R. (1985). Weschsler Intelligence Tests: Do we really have a criterion of mental retardation? *American Journal of Mental Deficiency*, *90*, 236-244.

Flynn, J. R. (1987). Massive IQ gains in 14 nations: What IQ tests really measure. *Psychological Bulletin*, *101*, 171-191. doi: 10.1037//0033-2909.101.2.171

Flynn, J. R. (1998). Israeli military IQ tests: Gender differences small; IQ gains large. *Journal of Biosocial Science*, *30*, 541-553. doi: 10.1017/S0021932098005410

Flynn, J. R. (2006). Tethering the elephant: Capital cases, IQ, and the Flynn effect. *Psychology, Public Policy, and Law*, *12*, 170-189. doi: 10.1037/1076-8971.12.2.170

Flynn, J. R. (2009a). Requiem for nutrition as the cause of IQ gains: Raven's gains in Britain 1938 to 2008. *Economics and Human Biology*, *17*, 18-27.

Flynn, J. R. (2009b). *What is intelligence? Beyond the Flynn effect (expanded ed.)*. Cambridge, England: Cambridge University Press.

Flynn, J. R., & Weiss, L. G. (2007). American IQ gains from 1932 to 2002: Subtests and educational progress. *International Journal of Testing*, *7*, 209-224. doi: 10.1080/15305050701193587

Fraser, C., & McDonald, R. P. (1988). *NOHARM.* Retrieved from http://people.niagaracollege.ca/cfraser/download/noharmdl.html

Greenfield, P. M. (1998). The cultural evolution of IQ. In U. Neisser (Ed.), *The rising curve: Long-term gains in IQ and related measures* (pp. 81–123). Washington, DC: American Psychological Association.

Griffiths, R. (1954). *The Abilities of Babies*. High Wycombe, UK: Test Agency.

Güler, N., & Penfield, R. D. (2009). A comparison of logistic regression and contingency table methods for simultaneous detection of uniform and nonuniform DIF. *Journal of Educational Measurement*, *46*, 314-329. doi: 10.1111/j.1745-3984.2009.00083.x

Haggerty, M. E., Terman, L. M., Thorndike, R. L., Whipple, G. M., & Yerkes, R. M. (1920). *National intelligence tests: Manual of directions*. New York: World Book Co.

Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory*. Newbury Park, CA: Sage.

Harris, B. (1988). Tetrachoric correlation coefficient. In L. Kotz & N. L. Johnson (Eds.), *Encyclopedia of statistical sciences* (Vol. 9, pp. 223–225). New York: Wiley.

Hedges, L. V. (1981). Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational Statistics*, *6*, 107-128. doi: 10.3102/10769986006002107

Herrnstein, R. J., & Murray, C. (1996). *The bell curve: Intelligence and class structure in American life*. New York: Free Press.

History of Nations. (2004). *History of Estonia.* Retrieved from http://www.historyofnations.net/europe/estonia.html.

Holland, P. W., & Thayer, D. T. (1988). Differential item performance and the Mantel-Haenszel procedure. In H. Wainer & H. I. Braun (Eds.), *Test validity* (pp. 129–146). Hillsdale, New Jersey: Lawrence Erlbaum.

Horn, J. L., & McArdle, J. J. (1992). A practical and theoretical guide to measurement invariance in aging research. *Experimental Aging Research*, *18*, 117-144.

Hu, L., & Bentler, P. M. (1998). Fit indicies in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods*, *3*, 424-453.

Hu, L. T., & Bentler, P. (1995). Evaluating model fit. In R. H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues, and applications* (pp. 76–99). London: Sage.

Jacobs, J. A. (1996). Gender inequality and higher education. *Annual Review of Sociology*, *22*, 153-185. doi: 10.1146/annurev.soc.22.1.153

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

Jerison, H. J. (1982). The evolution of biological intelligence. In R. J. Sternberg (Ed.), *Handbook of human intelligence* (pp. 723–791). Cambridge: Cambridge University Press.

Kanaya, T., Scullin, M. H., & Ceci, S. J. (2003). The Flynn effect and U.S. policies: The impact of rising IQ scores on American society via mental retardation diagnoses. *American Psychologist*, *58*, 778-790. doi: 10.1037/0003-066X.58.10.778

Kane, H., & Oakland, T. D. (2000). Secular declines in Spearman's g: Some evidence from the United States. *The Journal of Genetic Psychology*, *131*, 337-345. doi: 10.1080/00221320009596716

Kane, M. T. (2006). Validation. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 17–64). Westport, CT: Praeger.

Khaleefa, O., Abdelwahid, S. B., Abdulradi, F., & Lynn, R. (2008). The increase of intelligence in Sudan 1964-2006. *Personality and Individual Differences*, *45*, 412-413. doi: 10.1016/j.paid.2008.05.016

Kim, S. H., Cohen, A. S., & Park, T. H. (1995). Detection of differential item functioning in multiple groups. *Journal of Educational Measurement*, *32*, 261-276. doi: 10.1111/j.1745-3984.1995.tb00466.x

Kirk, R. E. (2007). *Statistics: An introduction.* Belmont, CA: Wadsworth.

Kolen, M. J., & Brennan, R. L. (2004). *Test equating, scaling, and linking: Methods and practices* (2nd ed.). New York: Springer-Verlag.

Li, Y., Brooks, G. P., & Johanson, G. A. (2012). Item discrimination and Type 1 error in the detection of differential item functioning. *Educational and Psychological Measurement*. Advance online publication. doi: 10.1177/0013164411432333

Light, M. L., & Chambers, W. R. (1958). A comparison of the Wechsler Adult Intelligence Scale and Wechsler-Bellevue II with mental defectives. *American Journal of Mental Deficiency*, *62*, 878-881.

Little, T. D. (1997). Mean and covariance structures (MACS) analyses of cross-cultural data: Practical and theoretical issues. *Multivariate Behavioral Research*, *32*, 53-76. doi: 10.1207/s15327906mbr3201$_$3

Lord, F. (1956). A study of speed factors in tests and academic grades. *Psychometrika*, *21*, 31-50. doi: 10.1007/bf02289085

Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Hillsdale, NJ: Lawrence Erlbaum.

Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental tests*. Reading, MA: Addison-Wesley.

Lu, Y., & Sireci, S. G. (2007). Validity issues in test speededness. *Educational Measurement: Issues and Practice*, *26*, 29-37. doi: 10.1111/j.1745-3992.2007.00106.x

Lynn, R. (1982). IQ in Japan and the United States shows a growing disparity. *Nature*, *306*, 291-292. doi: 10.1038/297222a0

Lynn, R. (1989). A nutrition theory of the secular increases in intelligence: Positive correlations between height, head size, and IQ. *British Journal of Educational Psychology*, *59*, 372-377. doi: 10.1111/j.2044-8279.1989.tb03112.x

Lynn, R. (1990). The role of nutrition in secular increases of intelligence. *Personality and Individual Differences*, *11*, 273-285. doi: 10.1016/0191-8869(90)90241-I

Lynn, R. (1998). In support of the nutrition theory. In U. Nesser (Ed.), *The rising curve: Long-term gains in IQ and related measures* (pp. 207–215). Washington, DC: American Psychological Association.

Lynn, R. (2009). What has caused the Flynn effect? Secular increases in the Development Quotients of infants. *Intelligence*, *37*, 16-27. doi: dx.doi.org/10.1016/j.intell.2008.07.008

Lynn, R., Allik, J., & Must, O. (2000). Sex differences in brain size, stature and intelligence in children and adolescents: Some evidence from Estonia. *Personality and Individual Differences*, *29*, 555-560. doi: 10.1016/S0191-8869(99)00215-9

Lynn, R., Allik, J., Pullman, H., & Laidra, K. (2002). A study of intelligence in Estonia. *Psychological Reports*, *91*, 1022-1026. doi: 10.2466/pr0.2002.91.3.1022

Lynn, R., & Hampson, S. (1986). The rise of national intelligence: Evidence from Britain, Japan, and the USA. *Personality and Individual Differences*, *7*, 23-32.

Lynn, R., & Hampson, S. E. (1989). Secular increases in reasoning and mathematical abilities in Britain, 1972-84. *School Psychology International*, *10*, 301-304. doi: 10.1177/0143034389104008

Lynn, R., & Pagliari, C. (1994). The intelligence of American children is still rising. *Journal of Biosocial Science*, *26*, 65-67. doi: 10.1017/S0021932000021064

Lynn, R., Pullman, H., & Allik, J. (2003). A new estimate of the IQ in Estonia. *Perceptual and Motor Skills*, *97*, 662-664. doi: 10.2466/PMS.97.5.662-664

Mackintosh, N. J., Wilson, B., & Boakes, R. A. (1985). Differences in mechanisms of intelligence among vertebrates. *Philosophical Transactions of the Royal Society of London B*, *308*, 53-65. doi: 10.1098/rstb.1985.0009

Magis, D., Beland, S., & Raiche, G. (2012). *difR: Collection of methods to detect dichotomous differential item functioning (DIF) in psychometrics.*

Magis, D., & Facon, B. (2012a). Angoff's delta method revisited: Improving the DIF detection under small sample. *British Journal of Mathematical and Statistical Psychology*, *65*, 302- 321. doi: 10.1111/j.2044-8317.2011.02025.x

Magis, D., & Facon, B. (2012b). Item purification does not always improve DIF detection: A counter example with angoff's delta plot. *Educational and Psychological Measurement*. Advance online publication. doi: 10.1177/0013164412451903

Magis, D., Raiche, G., Beland, S., & Gerard, P. (2011). A general framework and an R package for the detection of dichotomous differential item functioning. *Behavior Research Methods*, *42*, 847-862. doi: 10.3758/BRM.42.3.847

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. doi: 10.1016/j.intell.2004.11.005

McDonald, R. (1999). *Test theory: A unified treatment.* Mahwah, NJ: Erlbaum.

Messick, S. (1988). The once and future issues of validity: Assessing the meaning and consequences of measurement. In H. Wainer & H. I. Braun (Eds.), *Test validity* (pp. 33–46). Hillsdale, NJ: Lawrence Erlbaum.

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13–105). New York: American Council on Education/Macmillan.

Millsap, R. E. (2011). *Statistical approaches to measurement invariance.* New York: Routledge.

Millsap, R. E., & Hartog, S. B. (1988). Alpha, beta, and gamma change in evaluation research: A structural equation approach. *Journal of Applied Psychology*, *73*, 574-584. doi: 10.1037/0021-9010.73.3.574

Mingroni, M. A. (2004). The secular rise in IQ: Giving heterosis a closer look. *Intelligence*, *32*, 65-83. doi: 10.1016/S0160-2896(03)00058-8

Mingroni, M. A. (2007). Resolving the IQ paradox: Heterosis as a cause of the Flynn effect and other trends. *Psychological Review*, *114*, 806-829. doi: 10.1037/0033-295X.114.3.806

Munch, I., & Lundberg, I. (1994). Multivariate analyses of data from population A. In W. B. Elley (Ed.), *The IEA study of reading literacy* (pp. 193–222). Oxford: Pergamon Press.

Must, O., Must, A., & Raudik, V. (2003a). The Flynn effect for gains in literacy found in Estonia is not a Jensen effect. *Personality and Individual Differences*, *34*, 1287-1292. doi: 10.1016/S0191-8869(02)00115-0

Must, O., Must, A., & Raudik, V. (2003b). The secular rise in IQs: In Estonia, the Flynn effect is not a Jensen effect. *Intelligence*, *31*, 461-471. doi: 10.1016/S0160-2896(03)00013-8

Must, O., te Nijenhuis, J., Must, A., & van Vianen, A. E. M. (2009). Comparability of IQ scores over time. *Intelligence*, *37*, 25-33. doi: 10.1016/j.intell.2008.05.002

Muthén, L., & Muthén, B. O. (2010). *Mplus.* Muthén and Muthén.

Narayanan, P., & Swaminathan, H. (1996). Identification of items that show nonuniform DIF. *Psychometrika*, *20*, 257-274. doi: 10.1177/014662169602000306

Neisser, U. (1998). *The rising curve: Long-term gains in IQ and related measures.* Washington, DC: American Psychological Association.

Nettelbeck, T., & Wilson, C. (2004). The Flynn effect: Smarter not faster. *Intelligence*, *32*, 85-93. doi: 10.1016/S0160-2896(03)00060-6

Newton, J. H., & McGrew, K. S. (2010). Introduction to the special issue: Current research in Cattell-Horn-Carroll-based assessment. *Psychology in the Schools*, *47*, 621-634. doi: 10.1002/pits.20495

NFER. (1979). *Piagetian Reasoning Tasks.* Cambridge: Science Reasoning.

Orr, J. B. (1936). *Food, health, and income.* London: Macmillan.

Osterlind, S. J., & Everson, H. T. (2009). *Differential item functioning* (2nd ed.). Washington DC: Sage.

Palmer, C. E. (1935). Height and weight of children of the depressed poor. *Public Health Reports*, *59*, 33-35. doi: 10.2307/4581615

Penfield, R. D. (2001). Assessing differential item functioning among multiple groups: A comparison of three Mantel-Haenszel procedures. *Applied Measurement in Education*, *14*, 235-259. doi: 10.1207/S15324818AME1403$_$3

Penfield, R. D. (2003). Application of the Breslow-Day test of trend in odds ratio heterogeneity to the detection of nonuniform DIF. *Alberta Journal of Educational Research*, *49*, 231-243.

Pietschnig, J., Voracek, M., & Formann, A. (2010). Pervasiveness of the IQ rise: A cross-temporal meta-analysis. *PLoS ONE*, *5*, e14406. doi: 10.1371/journal.pone.0014406

Ployhart, R. E., & Vandenberg, R. J. (2010). Longitudinal research: The theory, design, and analysis of change. *Journal of Management*, *36*, 94-120. doi: 10.1177/0149206309352110

Pullman, H. (2005). *The development of intelligence and personality traits among Estonian schoolchildren.* Doctoral dissertation.

Pullman, H., Allik, J., & Lynn, R. (2004). The growth of IQ among Estonian schoolchildren from ages 7-19. *Journal of Biological Sciences*, *36*, 735-740. doi: 10.1017/S0021932003006503

R Development Core Team. (2012). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing. Retrieved from http://www.R-project.org/

Raju, N. S. (1990). Determining the significance of estimated signed and unsigned areas between two item response functions. *Applied Psychological Measurement*, *14*, 197-207. doi: 10.1177/014662169001400208

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests.* Copenhagen: Danish Institute for Educational Research.

Rasch, G. (1980). *Probabilistic models for some intelligence and attainment tests* (2nd ed.). Chicago: University of Chicago Press.

Resing, W. C. M., & Tunteler, E. (2007). Children becoming more intelligent: Can the Flynn effect be generalized to other child intelligence tests? *International Journal of Testing*, *7*, 191-208. doi: 10.1080/15305050701193546

Revelle, W. (2012). *psych: Procedures for psychological, psychometric, and personality research.*

Rizopoulos, D. (2011). *ltm: Latent trait models under IRT.*

Rodgers, J. L., & Wänström, L. (2007). Identification of a Flynn effect in the NLSY: Moving from the center to the boundaries. *Intelligence*, *35*, 187-196. doi: dx.doi.org/10.1016/j.intell.2006.06.002

Rogers, H. J., & Swaminathan, H. (1993). A comparison of logistic regression and Mantel-Haenszel procedures for detecting differential item functioning. *Applied Psychological Measurement*, *17*, 105-116. doi: 10.1177/014662169301700201

Roussos, L. A., & Stout, W. F. (1996). Simulation studies of the effects of small sample size and studied item parameters on SIBTEST and Mantel-Haenszel Type 1 error performance. *Journal of Educational Measurement*, *33*, 215-230. doi: 10.1111/j.1745-3984.1996.tb00490.x

Rupp, A. A., & Zumbo, B. D. (2006). Understanding parameter invariance in unidimensional IRT models. *Educational and Psychological Measurement*, *66*, 63-84. doi: 10.1177/0013164404273942

Ruston, J. P. (1999). Secular gains in IQ not related to the *g* factor and inbreeding depression-unlike Black-White differences: A reply to Flynn. *Personality and Individual Differences*, *26*, 381-389.

Sanborn, K. J., Truscott, S. D., Phelps, L., & McDougal, J. L. (2003). Does the Flynn effect differ by IQ level in samples of students classified as learning disabled? *Journal of Psychoeducational Assessment*, *21*, 145-159. doi: 10.1177/073428290302100203

Schooler, C. (1998). Environmental complexity and the Flynn effect. In U. Neisser (Ed.), *The rising curve: Long-term gains in IQ and related measures* (pp. 67–79). Washington, DC: American Psychological Association.

Shayer, M. (2008). Intelligence for education: As described by Piaget and measured by psychometrics. *British Journal of Educational Psychology*, *78*, 1-29. doi: 10.1348/000709907X264907

Shayer, M., Ginsburg, D., & Coe, R. (2007). Thirty years on- a large anti-Flynn effect? The Piagetian test Volume & Heaviness norms 1975-2003. *British Journal of Educational Psychology*, *77*, 25-41.

Sigman, M. D. (2000). *Determinants of intelligence: Nutrition and intelligence.* American Psychological Association.

Sigman, M. D., & Whaley, E. (1998). The role of nutrition in the development of intelligence. In U. Neisser (Ed.), *The rising curve: Long-term gains in IQ and related measures* (pp. 155–182). Washington, DC: American Psychological Association.

Simon, C. L., & Clopton, J. R. (1984). Comparison of WAIS and WAIS-R scores of mildly and moderately mentally retarded adults. *American Journal of Mental Deficiency*, *89*, 301-303.

Smith, S. (1942). Language and non-verbal test performance of racial groups in Honolulu before and after a fourteen year interval. *Journal of General Psychology*, *26*, 51-93.

Smotherman, W. P., & Robinson, S. R. (1995). Tracing developmental trajectories into the prenatal period. In J. P. Lecanuet, N. A. Krasnegor, W. P. Filer, & W. P. Smotherman (Eds.), *Fetal development: A psychobiological perspective* (pp. 295–314). New York: Erlbaum.

Spitz, H. H. (1983). Intratest and intertest reliability and stability of the WISC, WISC-R, and WAIS full scale IQ's in a mentally retarded population. *The Journal of Special Education*, *17*, 69-80. doi: 10.1177/002246698301700109

Spitz, H. H. (1986). *The raising of intelligence: A selected history of attempts to raise retarded intelligence.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Steenkamp, J., & Baumgartner. (1998). Assessing measurement invariance in cross-national consumer research. *Journal of Consumer Reseach*, *25*, 78-90. doi: 10.1086/209528

Steiger, J. H., & Lind, J. M. (1980). *Statistically based tests for the number of common factors.* Paper presented at the meeting of the Psychometric Society, Iowa City, IA.

Strenze, T. (2006). Who gets ahead in Estonia and America? A comparative analysis of mental ability and social origin as determinants of success. *Trames*, *10*, 232-254.

Sundet, J. M., Barlaug, D. G., & Torjussen, T. M. (2004). The end of the Flynn effect? A study of secular trends in mean intelligence test scores of Norwegian conscripts during half a century. *Intelligence*, *32*, 349-362. doi: 10.1016/j.intell.2004.06.004

Sundet, J. M., Borren, I., & Tambs, K. (2008). The Flynn effect is partly caused by changing fertility patterns. *Intelligence*, *36*, 183-191. doi: 10.1016/j.intell.2007.04.002

Swaminathan, H., & Rogers, H. J. (1990). Detecting differential item functioning using logistic regression procedures. *Journal of Educational Measurement*, *27*, 361-370. doi: 10.1111/j.1745-3984.1990.tb00754.x

Takahashi, E. (1966). Growth and environmental factors in Japan. *Human Biology*, *38*, 112-130.

Tanaka, J. S., & Huba, G. J. (1985). A fit index for covariance structure models under arbitrary GLS estimation. *British Journal of Mathematical and Statistical Psychology*, *38*, 197-201. doi: 10.1111/j.2044-8317.1985.tb00834.x

Teasdale, T. W. (2009). The Danish draft board's intelligence test, Børge Priens Prøve: Psychometric properties and research applications through 50 years. *Scandinavian Journal of Psychology*, *50*, 633-638. doi: 10.1111/j.1467-9450.2009.00789.x

Teasdale, T. W., & Owen, D. R. (1987). National secular trends in intelligence and education- A 20-year cross-sectional study. *Nature*, *325*, 119-121. doi: 10.1038/325119a0

Teasdale, T. W., & Owen, D. R. (1989). Continuing secular increases in intelligence and a stable prevalence of high intelligence levels. *Intelligence*, *13*, 255-262. doi: 10.1016/0160-2896(89)90021-4

Teasdale, T. W., & Owen, D. R. (2000). Forty-year secular trends in cognitive abilities. *Intelligence*, *28*, 115-120. doi: 10.1016/S0160-2896(99)00034-3

Teasdale, T. W., & Owen, D. R. (2005). A long-term rise and recent decline in intelligence test performance: The Flynn effect in reverse. *Personality and Individual Differences*, *39*, 837-843. doi: 10.1016/j.paid.2005.01.029

Thissen, D., Steinberg, L., & Wainer, H. (1988). Use of item response theory in the study of group differences in trace lines. In H. Wainer & H. I. Braun (Eds.), *Test validity* (pp. 147–169). Hillsdale, NJ: Lawrence Erlbaum Associates.

Thompson, M. S., & Green, S. B. (2006). Evaluating between-group differences in latent variable means. In G. R. Hancock & R. O. Meuller (Eds.), *Structural equation modeling: A second course* (pp. 119–169). Greenwich: Information Age.

Tork, J. (1940). *The intelligence of Estonian children*. Tartu, Estonia: Koolivara.

Truscott, S. D., & Frank, A. J. (2001). Does the Flynn effect affect IQ scores of students classified as LD? *Journal of School Psychology*, *89*, 319-334. doi: 10.1016/S0022-4405(01)00071-1

Tucker, L. R., & Lewis, C. (1973). A reliability coefficient for maximum likelihood factor analysis. *Psychometrika*, *38*, 1-10. doi: 10.1007/BF02291170

Tuddenham, R. D. (1948). Soldier intelligence in World Wars I and II. *American Psychologist*, *3*, 54-56. doi: 10.1037/h0054962

Uttaro, T., & Millsap, R. E. (1994). Factors influencing the Mantel-Haenszel procedure in the detection of differential item functioning. *Applied Psychological Measurement*, *18*, 15-25. doi: 10.1177/014662169401800102

Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*, *3*, 4-69. doi: 10.1177/109442810031002

van der Linden, W. J., & Hambleton, R. K. (1996). *Handbook of modern item response theory*. New York: Springer.

van Wieringen, J. C. (1978). Secular growth changes. In F. Falkner & J. M. Tanner (Eds.), *Human growth: Postnatal growth* (Vol. 2, pp. 445–474). New York: Plenum Press.

Wechsler, D. (1949). *Manual for the Wechsler Intelligence Scale for Children*. New York: Psychological Corp.

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

Wechsler, D. (1991). *Wechsler Intelligence Scale for Children–Third edition*. San Antonio, TX: Psychological Corporation.

Whipple, G. M. (1921). The National Intelligence Tests. *The Journal of Educational Research*, *4*, 16-31.

Whitehead, R. G., & Paul, A. A. (1988). Comparative infant nutrition in man and other animals. In K. Blaxter & I. MacDonald (Eds.), *Comparative nutrition* (pp. 199–213). London: J. Libbey.

Wicherts, J. M., Dolan, C. V., Hessen, D. J., Oosterveld, P., van Baal, G. C. M., Boomsma, D. I., & Span, M. M. (2004). Are intelligence tests measurement invariant over time? Investigating the nature of the Flynn effect. *Intelligence*, *32*, 509-537. doi: 10.1016/j.intell.2004.07.002

Wilke, M., Sohn, J.-H., Byars, A. W., & Holland, S. K. (2003). Bright spots: Correlations of gray matter volume with IQ in a normal pediatric population. *NeuroImage*, *20*, 202-215. doi: 10.1016/S1053-8119(03)00199-X

Willerman, L., & Churchill, J. A. (1967). Intelligence and birth weight in identical twins. *Child Development*, *38*, 623-629. doi: 10.2307/1127240

Williams, W. A. (1998). Are we raising smarter kids today? School and home-related influences on IQ. In U. Neisser (Ed.), *The rising curve: Long-term gains in IQ and related measures* (pp. 125–154). Washington DC: American Psychological Association.

Winick, M., Meyer, K. K., & Harris, R. C. (1975). Malnutrition and environmental enrichment by early adoption. *Science*, *190*, 1173-1175. doi: 10.1126/science.1198103

Winick, M., Ross, P., & Waterlow, J. (1970). Cellular growth of cerebrum, cerebellum, and brain stem in normal and marasmic children. *Experimental Neurology*, *26*, 393-400. doi: 10.1016/0014-4886(70)90134-2

Wirth, R. J., & Edwards, M. C. (2007). Item factor analysis: Current approaches and future directions. *Psychological Methods*, *12*, 58-79. doi: 10.1037/1082-989X.12.1.58

Woodley, M. A. (2009). Inbreeding depression and IQ in a study of 72 countries. *Intelligence*, *37*, 268-276. doi: 10.1016/j.intell.2008.10.007

Woodley, M. A. (2011). Heterosis doesn't cause the Flynn effect: A critical examination of Mingroni (2007). *Psychological Review*, *118*, 689-693. doi: 10.1037/a0024759

Yerkes, R. M. (1921). *Memoirs of the national academy of sciences: Psychological examining in the United States Army.* (Vol. 15). Government Printing Office.

Yoo, B. (2002). Cross-group comparisons: A cautionary note. *Psychology and Marketing*, *19*, 357-368. doi: 10.1002/mar.10014

Zhou, X., Zhu, J., & Weiss, L. G. (2010). Peeking inside the "black box" of the Flynn effect: Evidence from three Wechsler instruments. *Journal of Psychoeducational Assessment*, *28*, 399-411. doi: 10.1177/07342829

Zimowski, M. F. (2003). Multiple-group analyses. In M. du Toit (Ed.), *IRT from SSI: BILOG-MG, MULTILOG, PARSCALE, TESTFAT* (pp. 531–537). Lincolnwood, IL: Scientific Software International.