Fachbereich Erziehungswissenschaft und Psychologie der Freien Universität Berlin

The Impact of Educational Attainment and Grade Retention on the Development of Intelligence

Dissertation

zur Erlangung des akademischen Grades

Doktorin der Philosophie (Dr. phil.)

Doctor of Philosophy (Ph.D)

vorgelegt von

Diplom-Psychologin Daniela Sangme Schalke-Mandoux

Berlin, 2016

Erstgutachter: Professor Dr. Martin Brunner

Zweitgutachter: Professor Dr. Romain Martin

Tag der Disputation: 03.06.2016

Acknowledgements

Many people supported me during my Ph.D. time and were involved in the completion of this Ph.D. thesis. I greatly appreciate their many valuable suggestions, comments, and help. The following list is by no means complete and the support often exceeded what is explicitly mentioned here.

My primary and heartfelt thanks go to Prof. Dr. Martin Brunner, who always supported me and stimulated my thinking about the relationship between intelligence and educational attainment. It has been a great honor to work with him and to learn from him. I am especially thankful for all the support even after my decision of not pursuing a further scientific career. In addition, I want to thank Prof. Dr. Romain Martin as well as Dr. Marion Spengler for their support, suggestions, comments, and help with revising the manuscript.

I also want to thank the MAGRIP Team in 1968 and 2008 as well as all the participants, without them the entire Ph.D. thesis would have not been possible.

Special thanks go to my family and friends for their constant support.

Last but not least, this Ph.D. thesis was supported by a grant from the Luxembourgian Fonds National de la Recherche (VIVRE - Living tomorrow in Luxembourg, project no. FNR/06/09/18) and an AFR Ph.D. scholarship awarded to Daniela Schalke by the Fonds National de la Recherche, Luxembourg, PhD-09-216.

Declaration

I hereby declare that the present dissertation is my own work and that, to the best of my knowledge and belief, it contains no materials previously published or written by another person except where due acknowledgement has been made in the text. I furthermore declare that this work has not been submitted for the award of any other degree or diploma at a university or institution other than the Free University of Berlin.

Daniela Sangme Schalke-Mandoux
10th April 2016

Abstract

European countries are faced with a demographic change. People are living longer than ever before, while, at the same time, birth rates are below replacements levels. Thus, European countries will have to deal with an increasingly aging labor market as well as a growing amount of elderly people in need of care. Healthy cognitive aging is important for both of these challenges in Europe (Ball et al., 2002; European Commission, 2014). Thus, investing in healthy cognitive aging is not only important for individuals' well-being, but also for economic stability and growth. Crucially, people vary greatly in the development and maintenance of their cognitive abilities over the lifespan. Therefore, the European Commission (2014) recently called for strategies focused on the reduction of inequalities in cognitive development as well as research beginning already in childhood to understand the complex dynamics of cognitive development. The present Ph.D. project directly corresponds to this call. We examined the development of intelligence from late childhood at age 12 into middle adulthood at age 52 as well as the impact of educational attainment and grade retention in primary school on the development of intelligence. In 1968, data was collected from half of all Luxembourgian students in grade level six, comprising a total of 2824 students (49.9%). This sample is representative of the population of sixth graders in Luxembourg in 1968. In 2008, 745 of these former students were followed up and questioned on their educational pathways and attainment. In addition, the same intelligence test battery employed in 1968 was re-assessed. Thus, the present study's analyses are based on data covering 40 years of lifetime and provide a number of meaningful theoretical and practical implications.

In more detail, we examined three research questions in 3 studies, respectively. Study 1 tackled two key aspects of the development of intelligence concerning (a) stability and change in the structure of intelligence with reference to the age differentiation-dedifferentiation hypothesis (how different cognitive abilities relate to each other across age) and (b) differential stabilities (the rank ordering of persons' intelligence levels across time). To this end, we drew on two structural conceptions of intelligence: (a) the extended Gf-Gc model to study broad cognitive abilities and (b) the three-stratum model to decompose cognitive change into processes shared by all broad abilities (attributable to general cognitive ability *g*) and processes specific to a certain ability (independent of *g*). Data were obtained for 344 persons (56.4% female). The results showed that people differ more greatly over time with respect to all broad abilities except for fluid reasoning, whereas the rank ordering of persons on all broad abilities remains remarkably stable. These combined results yielded substantial gap-widening effects from age 12 to age 52, which were mainly accounted for by a substantial increase in *g* variance in combination with a high differential stability of *g*. Moreover, the increase in *g* variance reflects an increase in covariance among different broad abilities, indicating that the different constructs relate more closely to each other at age 52 compared to age 12 (i.e. age dedifferentiation). Two theoretical explanations of this change in the structure of intelligence are discussed (common cause hypothesis and investment theory).

Study 2 examined the long-term consequences of quantity and quality of formal education on the development of both fluid and crystalized cognitive abilities. Quantity of formal education was assessed by years of schooling, while quality of formal education was assessed by school track (i.e. academic vs. non-academic track). In addition, the study's design made testing the assumptions of Cattell's investment theory (1971, 1987) that fluid abilities are invested in the acquisition of crystallized abilities by taking advantage of environmental learning opportunities possible. A vital feature of the current study is that we were able to test the influence of educational attainment on two aspects of crystallized abilities, a verbal measure (word knowledge) and a factual knowledge measure (knowledge of the world), while controlling for childhood cognitive ability. Data were collected from 315 (55.9% female) participants. At the time their cognitive abilities were re-assessed, the participants in the current study had left formal education over 30 years previously. Most interestingly, we found an interaction effect of quantity and quality of formal education: Length of formal education had a long-term significant impact on both fluid and crystallized (i.e. word knowledge) ability for the non-academic track students only. Possible explanations of why the impact of formal education on cognitive abilities may be more persistent for participants in the non-academic track compared to participants in the academic track are discussed. In addition, evidence was found that confirms the assumptions made by investment theory that fluid abilities are invested into the acquisition of crystallized abilities by taking advantage of environmental learning opportunities. However, contrary to investment theory, we also found some investment of crystallized abilities in the acquisition of fluid abilities over the lifespan. In addition, we found that formal education related differently to the two different measures of crystallized ability in the present study, namely word knowledge and knowledge of the world. This may underline findings by previous research that these two abilities are empirically distinguishable facets of crystallized intelligence (Schipolowski, Wilhelm, & Schroeders, 2014). Taken together, the results of Study 2 of the present dissertation show that formal education has an important long-term impact on cognitive abilities, even over 30 years after participants had left formal education.

Study 3 tackled the long-term impact of grade retention in primary school on three key life outcomes in middle adulthood, namely educational attainment, income, and intelligence. To this end, we performed a multiple regression analysis for each key life outcome under study. Propensity score matching procedures were performed to control for 11 characteristics that are known to possibly influence grade retention and educational outcomes. In addition, we controlled for the influence of childhood intelligence, grade point average in primary school, and parental socioeconomic status, as well as educational attainment (for adult income and intelligence) in the regression analyses. From a base sample of 745 (53.3% female) participants, a different subsample was drawn for each life outcome under study. In the present study, we found that grade retention in primary school had a significant negative impact on all three key life outcomes, even over 40 years of the lifespan. On average, participants who were retained in primary school attended one year less of formal education than promoted participants, earned about €650 less per month at age 52, and scored about 7 IQ points lower in the intelligence test at age 52. Thus, contrary to the common belief that grade retention helps children with unsatisfactory academic achievement, in the present study grade retention in primary school has long lasting negative effects on a number of key life outcomes. Thus, alternatives should be considered and are discussed in the present dissertation.

The results of the present dissertation have a number of theoretical and practical implications that are discussed in the general discussion. The theoretical implications comprise propositions of lifespan developmental psychology sensu Li and Baltes (2006), Cattell's investment theory (1971, 1987), and the Matthew effect (or accumulating advantages). In addition, a number of practical implications are discussed, focusing on the pedagogical measure of grade retention as well as early interventions and aspects of the school system.

Keywords:

Intelligence, cognitive abilities, cognition, development, determinants, age differentiation-dedifferentiation hypothesis, differential stability, change and stability in cognitive abilities, investment theory, educational attainment, formal education, schooling, grade retention, longitudinal study, lifespan developmental psychology, lifespan cognitive psychology, childhood, adulthood, measurement invariance, and propensity score matching

Zusammenfassung

In vielen europäischen Ländern vollzieht sich ein demographischer Wandel. Die Menschen werden älter als je zuvor, während gleichzeitig die Geburtenraten zurückgehen. Als Folge werden viele europäische Länder mit einem stetig alternden Arbeitsmarkt sowie mit einer steigenden Anzahl an pflegebedürftigen Menschen konfrontiert. Gesundes kognitives Altern ist eine Schlüsselaufgabe, um diese Herausforderungen erfolgreich zu meistern (Ball et al., 2002; European Commission, 2014). Aufgrund der Auswirkungen auf den Arbeitsmarkt ist Forschung über gesundes kognitives Altern nicht nur wichtig für das individuelle Wohlerbefinden, sondern auch für die ökonomische Stabilität und das Wirtschaftswachstum. Menschen unterscheiden sich sehr stark darin, wie sich ihre kognitiven Fähigkeiten über die Lebensspanne entwickeln. Die Europäische Kommission (2014) forderte vor kurzem verstärkt Forschung zu betreiben mit dem Ziel kognitive Unterschiede in der Bevölkerung zu verringern. Weiterhin hat sie darauf hingewiesen, dass Forschung zum weiteren Verständnis der dynamischen Entwicklung kognitiver Fähigkeiten über die Lebensspanne bereits im Kindesalter anfangen muss. Das vorliegende Dissertationsvorhaben entspricht diesem Aufruf der Europäischen Kommission. Wir untersuchten die Entwicklung von Intelligenz von der späten Kindheit mit 12 Jahren hin zum mittleren Erwachsenenalter mit 52 Jahren. Des Weiteren untersuchten wir den Einfluss von Schulbildung sowie der Klassenwiederholung auf die kognitive Entwicklung über die Lebensspanne. 1968 nahm die Hälfte aller Sechstklässler in Luxemburg an unserer Studie teil, dies umfasste 2824 Studienteilnehmer (49,9% weiblich). Diese Stichprobe ist repräsentativ für die Population der Sechstklässler in Luxemburg in 1968. 2008 nahmen 745 der damaligen Teilnehmer erneut an unserer Studie teil und machten ausführliche Angaben zu ihrem Bildungsweg. Darüber hinaus wurde derselbe Intelligenztest wie 1968 erneut erhoben. Die Auswertungen der vorliegenden Doktorarbeit stützen sich daher auf Daten, die 40 Jahre Lebenszeit umfassen und eine Reihe aussagekräftige theoretischer und praktischer Implikationen zulassen.

Die vorliegende Arbeit besteht aus drei Studien, die jeweils eine Forschungsfrage behandeln. Studie 1 untersuchte zwei Aspekte der Intelligenzentwicklung (a) die Stabilität und Veränderung von Intelligenz im Hinblick auf die „Age Differentiation-Dedifferentiation" Hypothese (wie sich verschieden kognitive Fähigkeiten über die Lebensspanne zueinander Verhalten) und (b) differentielle Stabilitäten kognitiver Fähigkeiten über die Lebensspanne (die Rangfolge innerhalb der Population im Hinblick auf das Intelligenzniveau). Für diese Fragestellung zogen wir zwei verschiedene strukturelle Konzeptionen von Intelligenz heran. (a) Das erweiterte Modell fluider und kristalliner Intelligenz von Cattell und Horn (Cattell, 1987; J. L. Horn & Noll, 1997) wurde herangezogen, um die Entwicklung der „Broad Cognitive Abilities" zu untersuchen. (b) Die „Three-Stratum" Theorie von Carroll (1993) wurde verwendet, um kognitive Veränderungen, die allen „Broad Abilities" gemeinsam sind (die Veränderungen, die man auf den generellen Intelligenzfaktor *g* zurückführen kann), von denen zu unterscheiden, die spezifisch für gewisse Fähigkeiten sind (Veränderung unabhängig von *g*). Die Stichprobe bestand aus 344 Teilnehmern (56,4% weiblich). Die Ergebnisse zeigen, dass sich die Teilnehmer mit 52 Jahren in allen „Broad Abilities" stärker unterschieden als mit 12 Jahren. „Fluid Reasoning" bildete dabei eine Ausnahme. Allerdings blieben die differentiellen Stabilitäten gleichzeitig erstaunlich konstant über die Zeit. Diese beiden parallelen Entwicklungen (erhöhte Varianz auf den Konstrukten bei gleichzeitig hoher differentieller Stabilität über die Zeit) deuten auf einem Schereneffekt kognitiver Fähigkeiten hin. Der Unterschied zwischen den beiden Extremen der Intelligenzverteilung wird immer größer, je älter die Personen werden. Dieser Schereneffekt ist größtenteils auf die erhörte *g*-Varianz zurückzuführen, wobei auch *g* eine hohe differentielle Stabilität aufweist. Die erhöhte *g*-Varianz zeigt gleichzeitig eine erhöhte

Kovarianz zwischen den verschiedenen „Broad Abilities“ an, was heißt, dass die verschiedenen Konstrukte im Alter von 52 Jahren mehr miteinander korrelieren als im Alter von 12 Jahren (d.h. die Konstrukte sind sich im mittleren Erwachsenenalter ähnlicher als in der Kindheit). Eine solche Entwicklung deutet auf „Age Dedifferentiation“ hin. Zwei mögliche theoretische Erklärungsansätze für diese Veränderung in der Intelligenzstruktur werden diskutiert (die „Common Cause“ Hypothese und die Investmenttheorie).

Studie 2 untersucht die Langzeiteffekte von Quantität und Qualität schulischer Bildung auf die Entwicklung von fluiden und kristallinen Fähigkeiten. Die Quantität der Schulbildung wurde anhand der Dauer der schulischen Bildung (in Jahren) gemessen. Die Qualität der Schulbildung wurde anhand der Schulform im mehrgliedrigen luxemburgischen Schulsystem (nicht akaemische vs. akademische Schulform) erfasst. Zusätzlich erlaubte das Design der Studie, die Annahme von Cattell’s Investmenttheorie (1971, 1987) zu prüfen, dass fluide Fähigkeiten in den Erwerb kristalliner Fähigkeiten investiert werden. Eine weitere Besonderheit der vorliegenden Studie ist es, dass wir den Einfluss schulischer Bildung auf zwei Facetten kristalliner Intelligenz testen konnten:Verbale Fähigkeiten (Wortkenntnisse) und Sachwissen (Weltwissen). Zusätzlich wurde der Einfluss von kognitiven Fähigkeiten in der Kindheit auf die kognitiven Fähigkeiten im Erwachsenenalter kontrolliert. Die Stichprobe bestand aus 315 (55,9% weiblich) Teilnehmern. Bei der Interpretation der Ergebnisse ist zu beachten, dass zum Zeitpunkt der zweiten Intelligenztestung die Teilnehmer der vorliegenden Studie ihre schulische Bildung vor über 30 Jahren abgeschlossen hatten. Interessanterweise fanden wir einen Interaktionseffekt zwischen Quantität und Qualität schulischer Bildung: Die Länge der schulischen Bildung hatte einen signifikanten Langzeiteffekt aber nur für Schüler der nicht akademischen Schulform. Der Langzeiteffekt schulischer Bildung zeigte sich sowohl für fluide als auch für kristalline (Wortkenntnisse) Fähigkeiten. Mögliche theoretische Erklärungen werden diskutiert. Weiterhin bestätigen die Ergebnisse der vorliegenden Studie

die Annahme der Investmenttheorie, dass fluide Fähigkeiten in den Erwerb von kristallinen Fähigkeiten investiert werden. Allerdings zeigen die Ergebnisse der vorliegenden Arbeit, dass auch kristalline Fähigkeiten in den Erwerb fluider Fähigkeiten investiert wurden. Diese Befunde widersprechen den Annahmen der Investmenttheorie. Zusätzlich fanden wir, dass sich schulische Bildung unterschiedlich gegenüber den beiden Indikatoren kristalliner Intelligenz verhielt. Diese Befunde sind im Einklang mit Ergebnissen vorheriger Studien, die zeigen, dass verbale Fähigkeiten und Sachwissen zwei unterschiedliche Facetten kristalliner Intelligenz darstellen (Schipolowski, Wilhelm, & Schroeders, 2014). Die Befunde von Studie 2 verdeutlichen, dass Schulbildung einen Effekt auf die Entwicklung kognitiver Fähigkeiten hat, der sogar noch 30 Jahre, nachdem die Teilnehmer die Schule abgeschlossen hatten, zu finden war.

Studie 3 untersucht die Langzeiteffekte von Klassenwiederholung in der Grundschule auf die Schulbildung, sowie das Einkommen und die Intelligenz im Erwachsenenalter. Die Fragestellung wurde anhand von multiplen Regressionen überprüft, wobei für jedes Kriterium eine multiple Regression durchgeführt wurde. Weiterhin wurden „Propensity Score Matchings“ durchgeführt, um elf Variablen zu kontrollieren, die eine Einfluss auf die Nichtversetzung und die schulische Bildung haben könnten. Darüber hinaus wurde der Einfluss von Intelligenz in der Kindheit, von Grundschulnoten und des sozioökonomischen Status der Eltern auf die drei Kriteriumsvariablen kontrolliert. Die Basisstichprobe umfasste 745 Teilnehmer (53,3% weiblich). Aus dieser Basisstichprobe wurde für jede Fragestellung eine Substichprobe gezogen, die am besten für die Beantwortung der Forschungsfrage geeignet war. Die Befunde der vorliegenden Studie zeigen, dass die Nichtversetzung in der Grundschule einen signifikanten negativen Effekt auf die Schulbildung, das Einkommen und die Intelligenz im Erwachsenenalter hatte. Diese Effekte wurden gefunden, obwohl die Teilnehmer der Studie bereits vor über 40 Jahren in der Grundschule ein Schuljahr

wiederholen mussten. Im Durchschnitt hatten die in der Grundschule nicht versetzten Schüler am Ende ihrer Schullaufbahn ein Jahr weniger die Schule besucht (bereinigt auf die Jahre der Nichtversetzung während der gesamten Schullaufbahn), verdienten 650€ weniger im Monat im Alter von 52 Jahren und erzielten 7 IQ-Punkte weniger im Intelligenztest mit 52 Jahren als ihre vergleichbaren aber versetzten Mitschüler. Die pädagogische Maßnahme der Klassenwiederholung wird angewendet, da angenommen wird, dass sie Schülern mit mangelhaften akademischen Leistungen hilft, ihre schulischen Leistungen zu verbessern. Die Befunde der vorliegenden Studie deuten jedoch auf weitreichende negative Konsequenzen bis spät in das Erwachsenenalter hin. Die Befunde der vorliegenden Studie legen damit nahe, dass zukünftig Alternativen zur Nichtversetzung gesucht und bevorzugt werden sollten.

Die Befunde der vorliegenden Doktorarbeit haben eine Reihe von theoretischen und praktischen Implikationen. Die theoretischen Implikationen umfassen Annahmen zur Psychologie der Lebensspanne sensu Li und Baltes (2006), Cattells Investmenttheorie (1971, 1987), und dem Matthäus-Effekt (auch als kumulierte Vorteile bekannt). Des Weiteren werden einige praktische Implikationen, wie zum Beispiel die pädagogische Maßnahme der Klassenwiederholung, Frühfördermaßnahmen und Implikationen auf Aspekte des Schulsystems, diskutiert.

Schlüsselbegriffe:

Intelligenz, Kognitive Fähigkeiten, Kognition, Entwicklung, Determinanten, Age Differentiation-Dedifferentiation Hypothese, Differentielle Stabilität, Veränderung und Stabilität von kognitiven Fähhigkeiten, Intellektuelle Entwicklung, Investmenttheorie, Akademische Leistungen, Formelle Bildung, Schulbildung, Sitzenbleiben, Längsschnittsstudie, Lifespan Cognitive Psychology, Psychologie der Lebensspanne, Kindheit, Erwachsenenalter, Messinvarianz und Propensity Score Matching

Publications from the Present Dissertation:

Schalke, D., Brunner, M., Geiser, C., Preckel, F., Keller, U., Spengler, M., & Martin, R. (2012). Stability and change in intelligence from age 12 to age 52: Results from the Luxembourg MAGRIP study. *Developmental Psychology*. http://doi.org/10.1037/a0030623

Table of Contents

Table of Contents

Tables

Figures

Chapter 1 – Theoretical Background

1.1 Introduction: The Challenge of Healthy Cognitive Aging

Most countries in Europe are facing a demographic change. People are living longer than ever before, while, at the same time, birth rates are below replacements levels. In the coming decades, more and more of the elderly will need care, when they cannot continue to live independently. However, crucial changes in the age structure of the labor market can also be expected. Cognitive aging is important for being able to live independently in old age (Ball et al., 2002) and also for job performance (European Commission, 2014). Thus, healthy cognitive aging is a key challenge of demographic change in Europe (European Commission, 2014). Importantly, investing in healthy cognitive aging is not only important for individuals' well-being, but also for economic stability and growth. People vary greatly in the development and maintenance of their cognitive abilities over the lifespan. Therefore, the European Commission (2014, p. 42) asserts "*there is a need for strategies focusing on the reduction of inequalities in developing cognitive skills.*" In addition, the European Commission (2014) calls for approaches that include all ages of the population and research that starts in young childhood. This is in line with the theoretical assumption of lifespan developmental psychology that the impact of culture (e.g., cognitive training and interventions, formal education) decreases with increasing age (Li & Baltes, 2006). Thus, interventions that start late in life will be less effective than interventions that already start in childhood or early adulthood (Heckman, 2000, 2006, 2008). Formal education is proposed to be one of the most important socialization typical influences on cognitive development (Li & Baltes, 2006). Also, this theoretical proposition is supported by many empirical research results: Educational attainment plays an important role in cognitive development (Ceci, 1991, 1999; Mazzonna & Peracchi, 2012; Neisser et al., 1996).

The current dissertation directly corresponds to the call from the European Commission. In order to be able to develop strategies to reduce inequalities in cognitive development, the processes that take place from childhood into adulthood must be better understood. Therefore, the aim of the current dissertation is to investigate the impact of educational attainment on the development of intelligence. We were able to draw on longitudinal data over a 40 year time span from late childhood into middle adulthood. The current dissertation addresses three research questions: (a) how intelligence develops over the lifespan, (b) how the development of intelligence depends on schooling in general, and (c) how the development of intelligence is influenced by a key structural characteristic of the Luxembourgian educational system: grade retention in primary school.

1.2 Definition of Intelligence

A conceptual definition of intelligence is provided, for example, by Gottfredson (1997a, p. 13) *"Intelligence is a very general mental capability that, among other things, involves the ability to reason, plan, solve problems, think abstractly, comprehend complex ideas, learn quickly and learn from experience"*. In the current Ph.D. thesis, the terms intelligence and cognitive ability or abilities are applied interchangeably. Most of the current psychological research is based on the psychometric approach of intelligence. According to scientists in the psychometric tradition, intelligence can be well measured by tests and individual differences observed on these tests can be well captured by so-called structural models (L. S. Gottfredson & Saklofske, 2009). Within the psychometric approach, an important distinction must be made between the statistical structure of intelligence and the theoretical interpretation of this structure (Kan, Kievit, Dolan, & van der Maas, 2011). While there is consensus that intelligence factors are positively intercorrelated and hierarchically structured, researchers disagree on the theoretical interpretation of this structure. The most evident difference between theories is whether they assume a general factor of intelligence, namely *g*, at the apex of the hierarchy, such as Carroll's three-stratum theory (1993), or not, such as two component theories. A prominent two component theory is the extended Cattell-Horn theory of fluid and crystallized intelligence (extended Gf-Gc theory; Cattell, 1987; J. L. Horn & Noll, 1997).

Importantly, *g* theories and two-component theories serve different research purposes. Theories including *g* dominate studies that investigate the predictive powers of cognitive capacities, where *g* has been empirically proven to predict a number of key life outcomes such as educational achievement (Strenze, 2007), occupational success (Schmidt & Hunter, 2004), and longevity (L. S. Gottfredson & Deary, 2004; Whalley & Deary, 2001). However, *g* theories have rarely been used in developmental research, where theoretical frameworks of two component theories such as the extended Gf-Gc theory dominate (see also Lindenberger, 2001, on two-component theories).

The Gf-Gc theory distinguishes crystallized abilities, *Gc*, and fluid abilities, *Gf*, (Cattell, 1987; Nisbett et al., 2012). A conceptual definition is given by Nisbett and colleagues (2012, p. 131): Gc is *"the individual's store of knowledge about the nature of the world and learned operations such as arithmetic ones which can be drawn on in solving problems"*. Moreover, Gc comprises two distinct aspects: verbal abilities or language (e.g. word knowledge) and pure knowledge (e.g. knowledge of the world, Schipolowski et al., 2014). Gf on the other hand is *"the ability to solve novel problems that depend relatively little on stored knowledge as well as the ability to learn"* (Nisbett et al., 2012, p. 132). Typical indicators of Gf include inductive and deductive reasoning (McGrew, 2009). Importantly, the distinction between these two cognitive abilities is also strongly supported by research on the development of intelligence, showing differential trajectories of the two abilities over the lifespan (Lindenberger, 2001; Li et al., 2004; McArdle, Ferrer-Caja, Hamagami, &

Woodcock, 2002), as well as the high differential stability of both abilities over several decades of an individual's lifetime (Schalke et al., 2012).

1.3 Development of Intelligence

When describing the development of intelligence, several aspects of development are important: (1) development of the construct, (2) development of the individuals, and (3) impacts on this development. The first aspect, the development of the construct intelligence over the lifespan, addresses whether the construct is the same in childhood, adulthood, and old age. In the psychometric tradition, a change in the construct is indicated by a change in the factor structure of intelligence. This question is stated in more detail by the age differentiation-dedifferentiation hypothesis. The hypothesis postulates three developmental stages (Baltes, Cornelius, Spiro, Nesselroade, & Willis, 1980). The first stage of *differentiation* occurs in early childhood, where different broad abilities are proposed to become increasingly independent of each other with increasing age (Deary et al., 1996; Garrett, 1946; Reinert, 1970). The following period of adulthood is theoretically characterized by a fair degree of *stability* of the intelligence structure (Baltes et al., 1980). The third stage of *dedifferentiation* is again characterized by increasing dependencies among different broad abilities as people reach old age (Balinsky, 1941; Baltes et al., 1980). Chapter 2 of the current dissertation describes the age differentiation-dedifferentiation hypothesis in more detail.

The second aspect is the development of individuals over the lifespan in regard to their cognitive abilities. Here, three different aspects are of importance: (2.1) the differential stability of intelligence (i.e. the rank order of individuals), (2.2) the trajectory of the level of mean performance over the life span, and (2.3) the change in between-person variation or variance. Differential stabilities describe whether the rank order of individual persons remains stable across time. This addresses the question, whether the smarter children, will also go on to be the smarter adults? To answer this question, longitudinal data is required. The existing findings suggest that *g* shows high differential stability during different developmental stages over the lifespan. An extensive review of studies, which examined the differential stability of *g* can be found in Conley (1984) or Deary and colleagues (2000). However, the differential stabilities of lower level cognitive abilities (i.e. broad abilities such as Gf and Gc) are less well understood. Chapter 2 addresses this question in more detail. Second, consensus exists that mean performance in all broad abilities rise over childhood years until early adulthood. Then, in adulthood, fluid abilities, show linear mean decline with accelerated decline in old age. In contrast, mean performance in crystallized abilities remains stable or even increases with aging and shows only some decline in very old age (Cattell, 1987; Li et al., 2004; McArdle et al., 2002; Schaie, 2005; Tucker-Drob, 2009; Tucker-Drob & Salthouse, 2008). However, cross-sectional findings estimate an earlier onset of this decline in mean performance than longitudinal studies. Third, these mean changes of cognitive performance over the lifespan should theoretically be accompanied by changes of between person variation around the mean (Li & Baltes, 2006). The increase in variance is proposed to be greater for crystallized than for fluid abilities, as they are more sensitive to environmental influences. Thus, as people age, individual differences, particularly in crystallized abilities, should become continuously greater as person-specific environmental influences accumulate.

Chapter 2 of the current dissertation addresses the first two aspects of intelligence development described, namely the development of the construct intelligence over the lifespan and the development of individuals in regard to their intelligence over the lifespan. Chapter 3 and 4 will tackle important impact factors on the development of intelligence across the lifespan. Next, we will discuss formal education in school as it is one of the most important impact factors on cognitive development over the lifespan (Li & Baltes, 2006; Neisser et al., 1996).

1.4 Formal Education and Intelligence Development

Formal education in Luxembourg in the 1960 and 1970, when the participants of the MAGRIP study went to school comprised six years of primary school education for all Luxembourgian school children. After primary school, students were allocated to different schools depending on their academic achievement mostly in terms of their grades. These different secondary schools can be grouped into two types of schools: non-academic and academic schools. Non-academic schools are those schools that prepare students for an apprenticeship, while academic schools are those schools that prepare for college and university. While non-academic schools in Luxembourg had a duration between two and five years depending on the type of non-academic education, academic schools had a duration of seven years. Thereafter, students of both tracks had additional possibilities for further education. The non-academic track students had the option to specialize in a certain craftsmanship (Meisterausbildung) or even continue their education and attain a degree that allows entry to an advanced technical college (Fachhochschule). The academic track students could attend college and university.

The impact of formal education on intelligence development is part of several theoretical frameworks (Baltes, Staudinger, & Lindenberger, 1999; Ceci, 1999; Glaser, 1976; Li et al., 2004; Li & Baltes, 2006; Neisser et al., 1996). As the present Ph.D. study covers individual data from age 12 to 52, we will take on a lifespan point of view. Thus, we will focus on lifespan developmental psychology to understand lifespan cognition and the impact of formal education (Baltes et al., 1999; Li & Baltes, 2006). The lifespan approach focuses on the plasticity of different individual abilities, such as cognitive abilities. It states that any given developmental outcome is only one of many possible outcomes, which is affected by the reciprocal interplay of the individual's neurocognitive processes and the developmental context (Baltes et al., 1999; Li & Baltes, 2006). Brains can be conceived as open, dynamic information processors that adapt to, as well as affect, the individual's life circumstances and experiences reciprocally (Li & Baltes, 2006). Thus, cognitive abilities and formal education should affect each other in a reciprocal manner. More specifically, lifespan cognitive psychology postulates three kinds of influences that differentialy affect cognitive development over the lifespan: (a) species typical neurobiological and cultural evolutionary processes, (b) normative socialization-typical environmental influences (e.g. formal education), and (c) non-normative idiosyncratic person-specific experiences that result from self-selection into different (professional) environments (Li & Baltes, 2006). Further, Li and Baltes (2006) explain that the early stages of life are characterized by more normative processes and focused on basic competences as the development of very young children depends mostly on their species typical neurobiological processes. This does not mean that there are no inter-individual differences between small children or that early investments in child development are of little importance as the contrary has been found (compare Heckman, 2000, 2006, 2008). It means that the general developmental outcomes of early developmental stages are more or less the alike for all children (more normative) compared to later stages in life. All the children will learn to crawl, to walk, to speak, or to eat by themselves when they reach a certain age. Some will learn this faster than others, but the developmental "goal" is more or less the same for all children for most of the development that takes place in early childhood. Especially compared to the heterogeneity in development later in life, early developmental stages can be described as fairly normative. When children are schooled, the normative socialization-typical influence of schooling gains importance and affects cognitive development in childhood and adolescence. Formal education in school may be the most important socialization-typical influence throughout the lifespan in Luxembourg, as children spent several years in school. However, after individuals leave formal education, non-normative person-specific experiences resulting from self-selection into different (occupational) environments accumulate (Li et al., 2004; Li & Baltes, 2006). Thus, the later stages in life are characterized by a much greater diversity in developmental processes, so that the influence of formal education on cognitive abilities may be overlain by several other processes. However, cumulative advantages are also possible (DiPrete & Eirich, 2006). A better education may lead to a cognitively more challenging job and this may in turn foster cognitive abilities. Crucially, individualization in later life may not affect all cognitive abilities. Gf and Gc are thought to be differently affected by the environment and by biological influences. Gf is more strongly affected by species-typical neurobiological processes and biological aging effects and less by the environment or culture. On the contrary, Gc is to a large degree affected by the environment and much less by aging effects (Baltes et al., 1999; Cattell, 1987; Li et al., 2004; Li & Baltes, 2006). Thus, Gf is assumed to be less culture sensitive than Gc and subsequently formal education may affect Gf to a lesser degree than Gc. However, the relative plasticity of Gf and Gc has not yet been sufficiently studied (Li & Baltes, 2006).

Crucially, the influence of biology and culture may vary at different stages in life. Three principles have been proposed to describe the relationship between biology and culture over the lifespan (Baltes et al., 1999; Li & Baltes, 2006). First, the influence of biology-based plasticity is highest from birth to maturity and decreases thereafter. Second, the need for culture increases over the lifespan. Thus, the required cultural conditions for reaching any certain state increase as people age. Third, the efficacy of culture decreases with aging. Hence, more and more resources are needed to achieve the same level of cognitive performance. Importantly, formal education in schools affects individuals at a very sensitive point in their lives, when the influence of biological plasticity as well as the efficacy of culture are still high. Later environmental influences or interventions on cognitive functioning may be far less efficient. In addition, previous research has shown that small advantages cumulate over the lifespan and result in big advantages later in life (DiPrete & Eirich, 2006). DiPrete and Eirich (2006) discuss that the educational system with its transitions from one class to another may itself have a character that installs cumulative advantages. In addition, it has also been debated whether ability tracking as in the Luxembourgian school system may produce cumulative advantages over time or not. Thus, it is very likely that formal education has a robust and lasting effect on cognitive abilities throughout the lifespan. Chapter 3 examines the question how formal education impacts the development of Gf and Gc from late childhood into middle adulthood.

1.5 The Impact of Grade Retention on Key Life Outcomes

In many European school systems grade retention is a key pedagogical measure that is taken to help students with difficulties achieving competences demanded for a certain grade-level in school (European Commission, 2011). These students must repeat a grade level with the idea that an additional year of maturity and exposure to the curriculum will improve the students' academic achievement and social success in future grade levels. Grade retention is applied in many European countries, because it is believed that grade retention is a very beneficial pedagogical measure for the student (European Commission, 2011). In almost all school systems, poor grades/marks are the main reason for a student to repeat a year, although other criteria may play a role. The decision-making process is mainly influenced by the teacher's opinion of the student, while parental opinion plays a less important role (European Commission, 2011).

However, the general idea that grade retention is beneficial for the retained student is not well supported by previous research. Only a few research results report positive findings for grade retention on future academic performance (D. C. Gottfredson, Fink, & Graham, 1994; Hughes, Chen, Thoemmes, & Kwok, 2010) or children's perceived school competence

(Reynolds, 1992). There are some studies that report short-term positive effects, which, however, diminish over one or two years (Jimerson, 1999, 2001; Jimerson, Carlson, Rotert, Egeland, & Sroufe, 1997; Wu, West, & Hughes, 2010). Crucially, the vast majority of research findings report either no effect of grade retention on academic performance (Chen, Liu, Zhang, Shi, & Rozelle, 2010; Im, Hughes, Kwok, Puckett, & Cerda, 2013; Phelps, Dowdell, Rizzo, Ehrlich, & Wilczenski, 1992; Pierson & Connell, 1992) or even negative effects on a number of variables, such as future academic performance, parent's expectations, academic self-efficacy, children's psychological functioning, future school career, and importantly, dropping out of high school (Goos, Van Damme, Onghena, Petry, & de Bilde, 2013; Hughes, Kwok, & Im, 2013; Jimerson, Anderson, & Whipple, 2002; Reynolds, 1992). In addition, grade retention has also been identified to have a negative effect with an effect size of -.16 in the famous Hattie-Study (Hattie, 2009) on visible learning. Moreover, grade retention had a negative effect on academic achievement in several domains and fearing of getting retained did not motivate the students.

Long-term negative consequences of grade retention in primary school can be expected on several key life outcomes such as income and cognitive abilities as individuals interact in a reciprocal manner with their environment (Li & Baltes, 2006). First, previous research has shown that grade retention does not sustainably improve the retained students' performance (Hattie, 2009; Holmes & Matthews, 1984; Jimerson, 2001). Therefore, the retained students do not catch up and, thus, retained students are more likely to attain the non-academic compared to the academic track in Luxembourg. In addition, grade retention has often been found to impact negatively on a number of academically related variables such as student's attitude towards school (Holmes & Matthews, 1984) and academic self-efficacy (Hughes et al., 2013). This in combination with the poor academic performance will lead to lower academic attainment and increased drop-out rates for retained students who then leave school without a diploma (Alexander, Entwisle, & Dauber, 2003; Alexander, Entwisle, & Horsey, 1997). After formal education in school ends, person-specific influences gain a larger impact on development (Li & Baltes, 2006). Thus, individuals develop in the direction of their interests and professions, but also in the direction in which they think they are competent. Hence, due to their academic failures, retained students may believe that cognitively challenging activities and occupations will not suite them well. This, in combination with leaving school early, may cause retained students to take on occupations that are of a more physical nature and less cognitively challenging. These jobs are usually less well paid. In addition, job complexity and cognitive challenges on the job have an impact on cognitive development (Schooler, Mulatu, & Oates, 1999). Thus, grade retention may negatively impact critical key life outcomes such as educational attainment, adult income, and adult intelligence. Chapter 4 of the present dissertation investigates these relationships in more detail.

1.6 The Present Dissertation

Healthy cognitive aging has become a major challenge for most European countries as their population grows older than ever before. However, there is great variability between people in the development and maintenance of their cognitive abilities over the lifespan. Recently, the European Commission (2014) has called for strategies that focus on reducing inequalities in the development of cognitive skills and that investigations pertaining to these strategies must start early in life, even in childhood. The present Ph.D. thesis directly corresponds to this call and investigates the development of intelligence from late childhood into middle adulthood as well as one of the most important impact factors for cognitive development, namely educational attainment. The current Ph.D. thesis is based on data from the longitudinal MAGRIP.

1.6.1 The MAGRIP Project

Data for the present Ph.D. thesis stems from the longitudinal MAGRIP project. Data was collected in two waves: 1968/69 and 2008/09. The first wave of the MAGRIP project started in 1968 and was conducted under the supervision of the principle investigators Gaston Schaber, Paul Dickes, and Marcel Bamberg at the Institut Pédagogique in Walferdange, Luxembourg (Bamberg, Dickes, & Schaber, 1977). The first wave of the MAGRIP study was designed to examine children's transition from primary into secondary school in the tracked Luxembourgian school system (Dickes, 2011). In 1968/69, data was collected from half of all Luxembourgian students in grade level 6, comprising a total of 2824 students (M = 11.9 years; SD = 0.6 years; 50.1% male). The data included an intelligence test battery (the Leistungsprüfsystem - L-P-S, W. Horn, 1962, 1983), data on students' educational attainment, a questionnaire on students' behavior, and information on family background. A multi-stage sampling procedure was applied to realize two (overlapping) representative samples. First, half of all Luxembourgian school classes at grade 6 were selected randomly. All students of these classes participated. This sample was representative of sixth-graders in Luxembourg. Second, a representative age-based sample was drawn. For this purpose, all students who were enrolled in school in the school year 1963/64 were identified in the selected schools and participated. This included students that attended classes spanning from grade 4 to 6 (students in lower grades had repeated one or more classes). Of the 2824 students, 84% were in Grade 6 of primary school, 11% in Grade 5, and 5% in Grade 4. This sample is representative of students aging 12 years and attending primary school in Luxembourg.

In 2008/09, a large follow-up study (MAGRIP-R) was conducted. The former students now aged around 52 years. This follow-up was funded by the Fonds National de la Recherche Luxembourg (FNR/VIVRE/06/09/18) and led and designed under the supervision of Principal Investigators Prof. Dr. Martin Brunner and Prof. Dr. Romain Martin in close cooperation with

Frederic Berger (CEPS/INSTEAD). As the project's Research Coordinator, Dipl-Psych. Daniela Schalke was responsible for daily project management and database construction.

Data collection took place from November 2008 to August 2009. In a first step, the current addresses of the original participants were identified using the database of Luxembourg's social security agency (permission was granted by the Luxemburgian data protection committee "Commission Nationale Pour la Protection des Données"). Addresses could be found for 2377 of the former participants. 166 had died, and addresses were unknown for 281 participants. Subsequently, a stratified sample of 1632 individuals was drawn and contacted, to ask whether they wanted to participate in the follow-up study. Stratification was achieved in respect to region in Luxembourg and gender. These people were contacted randomly. 745 took part in the study, 300 could not be contacted and 587 did not want to take part in the study. Data was collected in three stages: Stage 1 consisted of a household study. 745 participants (53.3% female) were visited at home by trained assessors and interviewed on their life history in regard to their educational and professional history. In addition, data was collected on health and well-being. On average, the interview and questionnaires took 90 minutes. This stage of data collection took place from November 2008 to January 2009. Stage 2 consisted of group testing for the assessment of cognitive abilities. 247 participants that had been part of the household study came to the University of Luxembourg on 4 weekends in March 2009. The test took 90 minutes and the participants were invited for lunch afterwards to thank them for their participation. Stage 3 was conducted to collect data on the cognitive abilities of those participants who had taken part in the household study, but were not able to come to the group testing sessions. Thus, 131 people were visited at home by trained assessors and received a small monetary incentive for participating in the study. These tests took place between April and August 2009. See Figurers

9 and 10 in Appendix D for an overview that summarizes these steps for the samples in Study1 and 2.

For the analyses in Chapter 2, Chapter 3, and Chapter 4, subsamples were drawn from the base sample described above. The subsamples were selected so that the research question under study could best be addressed. Thus, samples sizes may vary for each of the studies in the present Ph.D. thesis. The subsamples are described in detail in the respective chapters.

1.6.2 Study 1: Development of Intelligence

The first study examined the development of intelligence over the lifespan (see Chapter 2) by focusing on two key aspects: (a) stability and change in the structure of intelligence with reference to the age differentiation-dedifferentiation hypothesis (how different cognitive abilities relate to each other across age) and (b) differential stabilities (the rank order of persons' intelligence levels across time). To this end, we drew on two structural conceptions of intelligence: (a) the extended Gf-Gc model (Cattell, 1987; J. L. Horn & Noll, 1997), to study broad cognitive abilities and (b) the three-stratum model (Carroll, 1993), to decompose cognitive change into processes that are shared by all broad abilities (attributable to general cognitive ability *g*) and processes specific to a certain ability (independent of *g*). For these analyses, we drew on the age-based sample in the MAGRIP data. Thus, only the 12-year-old students were included to control for aging effects on the development of cognitive abilities. This rendered a sample of 344 (56.4% female) participants for the analyses. Intelligence at ages 12 and 52 was assessed by nine subtests taken from a standardized and well validated German intelligence test battery, named the "Leistungsprüfsystem" (L-P-S; i.e. achievement test battery, W. Horn, 1962, 1983). Four broad abilities were assessed by the test: Fluid reasoning (Gf), Comprehension knowledge (Gc), Visual processing (Gv), and Processing speed (Gs).

Drawing on two different models of intelligence, Study 1 makes several important contributions to the empirical body of research: (a) The study covers 40 years from age 12 to age 52 and thereby covers an age range for which little empirical knowledge exists, middle adulthood. Longitudinal research on child development has seldom examined intelligence development beyond early adulthood and most adult developmental research focuses on old age. (b) Previous research on the development of the structure of intelligence has produced mixed results, probably as a result of several methodological difficulties. Thus, in contrast to most previous research, we studied these processes in a longitudinal study over a very long time period (i.e. 40 years) with a homogeneous age sample. Hence, we did not confound for maturation effects on cognitive abilities as the sample is homogeneous in respect to age and the participants should have had more or less the same level of maturation at both times of measurement. In addition, we did not confound for history effects on cognitive abilities as the participants were all of the same cohort (for example the Flynn effect, Flynn, 1987). The study design with the two intelligence models allowed the decomposition of change specific to certain broad abilities and change shared by all abilities. (c) Most previous research on the differential stability of intelligence has been conducted on the manifest level and therefore did not control for changes in the operational definition of the construct or for measurement error. In addition, most previous research did not examine the differential stability of specific abilities after the influence of *g* has been accounted for. The results of the analyses are described in Chapter 2 of the present dissertation.

1.6.3 Study 2: The Impact of Educational Attainment

The second study investigated the long-term impact of quantity and quality of formal education on the development of cognitive abilities over the lifespan (see Chapter 3). Quantity of formal education was assessed by years of schooling after primary school and quality of formal education was assessed by school track (i.e. non-academic vs. academic track). We focused on the development of Gf and two separate aspects of Gc, a more verbal measure (i.e.

word knowledge) and a factual knowledge measure (knowledge of the world, Schipolowski et al., 2014). The differentiation between Gf and Gc is important, as they are supposed to be differently affected by biological and environmental influences and may thus relate differently to formal education (Li & Baltes, 2006). Crucially, the study design permitted the examination of interactions between the impact of quantity and quality of formal education on Gf and Gc. We performed separate analyses for each dependent variable at age 52: fluid reasoning (Gf), word knowledge (Gc; verbal ability), and knowledge of the world (Gc; factual knowledge). Years of formal education were included in the model as a mediating variable, while we controlled for childhood fluid reasoning (Gf) and word knowledge (Gc; verbal ability). The effect of quality of formal education (i.e. school track) was examined by specifying multiple-group models, where we estimated the effects of the non-academic and the academic track simultaneously. For these analyses, we drew a sample from the MAGRIP data, which included only those students who were 12 years old and in Grade 6. In doing so, we controlled for the differential effects of aging as well as possible effects of quantity of schooling in primary school. This yielded a sample of 315 (55.9% female) students.

The analyses in Study 2 are singular, as they examine the impact of schooling on cognitive abilities even 30 years after participants had left formal education. To our knowledge, no other study has examined comparable long-term effects of formal education on cognitive abilities. In addition, we simultaneously investigate the effects of quantity and quality of formal education as well as possible interactions. Another vital feature of the present study is that we examine two distinct aspects of crystallized ability in middle adulthood: (a) word knowledge (WK), a more verbal measure, and (b) knowledge of the world (KW), a factual knowledge measure (Schipolowski et al., 2014). The results of Study 2 can be found in Chapter 3.

1.6.4 Study 3: The Impact of Grade Retention in Primary School

The third study examined the long-term impact of grade retention in primary school on three key life outcomes, namely educational attainment, adult income, and adult intelligence (see Chapter 4). Grade retention is a very frequent pedagogical method in Luxembourg. The data from the household sample in 1968 show that every sixth participant was retained at least once in primary school. In 2011, the European Commission reported that about 20% of all Luxembourgian students are currently retained at least once in their school careers (European Commission, 2011). Grade retention is supposed to be beneficial for the students, but there are controversial findings concerning this positive effect (European Commission, 2011; Holmes & Matthews, 1984; Jimerson, 2001). In the present Ph.D. thesis, we examined the effect of grade retention on educational attainment (i.e. years of formal education), adult male income, and adult cognitive abilities (i.e. general intelligence *g*). We were able to control for several important child, school, and parent characteristics that have been shown to affect grade retention and cognitive development. To this end, we first applied propensity score matching to identify suited comparison groups of retained and non-retained students and second, included control variables in the multiple regression analyses. We had to draw on different base samples from the MAGRIP Study for each key life outcome. Therefore, a separate propensity score matching procedure and regression analysis was performed for each key life outcome under study. After accounting for missing data and performing the propensity score matching procedure, the samples for the regression analysis contained the following samples sizes: First, the analyses for educational attainment were based on 472 (51.9% female) successfully matched students, of which 97 (49.5% female) were retained and 375 (52.5% female) were promoted. Second, the analyses of adult income were based on a sample of 191 (0% female) successfully matched students, of which 46 (0% female) were retained and 145 (0% female) were promoted. Third, the analyses of general cognitive ability *g* were based on 187 (52.4% female) successfully matched students, of which 37 (45.9% female) were retained and 150 (54.0% female) were promoted. The results are reported in Chapter 4.

The analyses in Study 3 are of great importance for the existing body of empirical research, as studies examining the long-term impact of grade retention are scarce. Crucially, to our knowledge, no study exists that has looked at effects of grade retention in primary school 40 years after the retention decision had been made. Grade retention is a very frequent and expensive measure, as still today 20% of all Luxembourgian students are retained for at least one school year during primary school (European Commission, 2011). Thus, teachers, teaching facilities, and teaching materials need to be provided for 20% of all Luxembourgian students for an additional year. The most important features of Study 3 are: (a) The 40-year span from late childhood to middle adulthood and (b) the examination of three important key life-outcomes (i.e. educational attainment, adult income, and adult intelligence). Previous longitudinal studies on grade retention have rarely looked at length of schooling or intelligence development beyond school age. In addition, to our knowledge, no study exists that addressed the impact of grade retention on adult income. (c) We were able to control for several important key school, child, family, and parental characteristics related to grade retention and educational attainment.

Chapter 2 – Study 1: Stability and Change in Intelligence from Age 12 to Age 52

2.1 Theoretical Background

Stability and change in intelligence across the lifespan are crucial topics in human development because intelligence is of great importance for facing challenges at school, at work, and in every-day life (L. S. Gottfredson, 1997b). To profoundly understand the developmental dynamics of cognitive aging, it is essential to study longitudinal data that extend from childhood to adulthood where the same individuals take the same cognitive measures two or more times (Schaie & Hofer, 2001). There are a number of longitudinal studies that have tackled the developmental dynamics of cognitive abilities from early adulthood to old age such as the Seattle Longitudinal Study (Schaie, 2005), in very old individuals such as the Berlin Aging Study (Baltes & Lindenberger, 1997), and from late childhood into old age such as the Scottish Mental Survey (Deary et al., 2000; Deary, Whiteman, Starr, Whalley, & Fox, 2004; see Schaie & Hofer, 2001 for a review of other longitudinal studies). However, little is known about the developmental dynamics of cognitive abilities from late childhood to middle adulthood. Thus, the present longitudinal study contributes to the existing body of research by investigating the change and stability of intelligence from late childhood (age 12) to middle adulthood (age 52). More specifically, we tackled two key aspects of lifespan intellectual development over a 40-year time period concerning (a) changes in the structure of intelligence embedded into the framework of the age differentiation-dedifferentiation hypothesis, and (b) differential stabilities, that is, the rank ordering of persons' intelligence levels across age.

2.1.1 Conceptualization and Structure of Intelligence

Intelligence can be conceptually defined as "[…] a very general mental capability that, among other things, involves the ability to reason, plan, solve problems, think abstractly, comprehend complex ideas, learn quickly and learn from experience" (L. S. Gottfredson, 1997a, p. 13). Most current psychological research is based on the psychometric approach (Neisser et al., 1996), which states that intelligence is well measured by tests, and individual differences on these tests are well represented by structural models (L. S. Gottfredson & Saklofske, 2009). These structural models (in terms of confirmatory or exploratory factor analytic models) are of central importance because they provide the starting point for relating intelligence to other theoretical concepts and for studying cognitive change (Edwards & Bagozzi, 2000).

An important distinction has to be made between the statistical structure of intelligence and the theoretical interpretation of this structure (Kan et al., 2011). Statistically, the common factors in factor models of intelligence capture the shared variance of the observed test scores and a theoretical framework is needed in order to interpret these factors. Whereas it is widely agreed that intelligence is hierarchically structured with constructs varying in their levels of generality, theories of intelligence differ in their conceptions of how broadly these constructs are defined and how many hierarchical levels are needed (Carroll, 1993; Cattell, 1987). Nevertheless, the similarities of different theories are so apparent that McGrew (2009) recently proposed the CHC model (Cattell-Horn-Carroll model) of intelligence, and thereby synthesized the two most prominent theories in the field: (a) the extended Cattell-Horn theory of fluid and crystallized intelligence (Cattell, 1987; J. L. Horn & Noll, 1997), and (b) Carroll's three-stratum theory (1993). The CHC model (McGrew, 2009) specifies a large number of primary abilities at the first level of the hierarchy. On the second level, primary abilities that rely on the same cognitive demands are structured into a system of 10 broad abilities. These broad abilities (for a description of the abilities that we examined in the present study, see Table 1) have been reproduced in several studies, and their discriminant validity has been shown (Carroll, 1993; J. L. Horn & McArdle, 2007; J. L. Horn & Noll, 1997). At the apex of the hierarchy in the CHC model is a general factor of intelligence, namely *g*, which accounts for the positive intercorrelations of the broad ability factors.

Table 1

Definitions of Abilities and Descriptions of Corresponding Measures as Applied in the Present Study

Broad ability	Measure	Description
Fluid reasoning (Gf) "describes the use of deliberate and controlled mental operations to solve novel problems that cannot be performed automatically. [..] Inductive and deductive reasoning are generally considered the hallmark indicators of Gf."	Concept Formation (*Gf_1*)	Identify, categorize, and determine rules from a complete stimulus set of patterns.
	Number and Letter Series (*Gf_2*)	Identify, categorize, and determine rules from a complete stimulus set of numbers and letters.
Comprehension-knowledge (Gc) "is typically described as a person's breadth and depth of acquired knowledge of the language, information, and concepts of a specific culture, and/or the application of this knowledge."	Vocabulary (*Gc_1*)	Identify the spelling error of a given noun.
	Word Identification (*Gc_2*)	Identify a word out of a random composition of letters.
Visual processing (Gv) "is the ability to generate, store, retrieve, and transform visual images and sensations."	Mental Figure Folding (*Gv_1*)	Identify the same position of a marker point on the layout and the folded object.
	Spatial Relations (*Gv_2*)	Identify the number of all hidden and unhidden surfaces of an object.
Processing speed (Gs) "is the ability to automatically and fluently perform relatively easy or over-learned elementary cognitive tasks, especially when high mental efficiency (i.e., attention and focused concentration) is required."	Perception Speed (*Gs_1*)	Quickly count all target objects and circle each eighth target object.
	Accuracy (*Gs_2*)	Quickly and accurately compare two rows that should be identical and find the error in the right row.

Note. Definitions are adopted from McGrew (2009, p. 5 & 6).

Even though the CHC model offers an integrating taxonomy for the similarities of the two underlying models, the extended Gf-Gc and the three-stratum model differ crucially with regard to the nature of *g*: Carroll (1993) interprets *g* as a unique cognitive ability, whereas Horn argues strongly against the existence of *g* (compare J. L. Horn & McArdle, 2007; J. L. Horn & Noll, 1997). Interestingly, theories that accept or do not accept the existence of *g* have been used for different research purposes. *g* theories dominate studies investigating the predictive powers of cognitive capacities where *g* has been empirically demonstrated to predict a number of key life outcomes such as educational achievement (Strenze, 2007), occupational success (Schmidt & Hunter, 2004), and longevity (L. S. Gottfredson & Deary, 2004). However, *g* theories have rarely been used in developmental research (Ackerman & Lohman, 2003), as "the description of a cognitive system with only a single *g* factor is an overly simplistic view of the more complex sequential dynamics" (McArdle et al., 2002, p. 134). Thus, in developmental research two-component theories such as the extended Gf-Gc theory have been prevailing (Lindenberger, 2001). These theories focus on the interplay and differences between fluid and crystallized abilities, but not on *g*. However, we think that a comprehensive study of age-related changes in the structure of intelligence should examine both broad abilities *and g*. Hence, in the current study, we scrutinized the change and stability of intelligence by capitalizing on (a) the extended Gf-Gc model (Figure 1a) and (b) the three-stratum model (Figure 1b). Importantly, these two theories differ not only in their structural conceptualization of intelligence, but they may also highlight different aspects of change. Specifically, a first-order model like the extended Gf-Gc model allows examination of change in broad abilities and their intercorrelations. *g* captures these intercorrelations in a higher order model like the three-stratum model, and the different abilities statistically represent residual factors where the influence of *g* is partialled out (Brunner, Nagy, & Wilhelm, 2012). These residual factors capture only what is specific to each ability and are referred to as $Gf_{specific}$, $Gc_{specific}$, $Gv_{specific}$, and $Gs_{specific}$ in the model. Hence, a higher order model allows separating change specific to each ability from change shared by all abilities, captured by *g*.

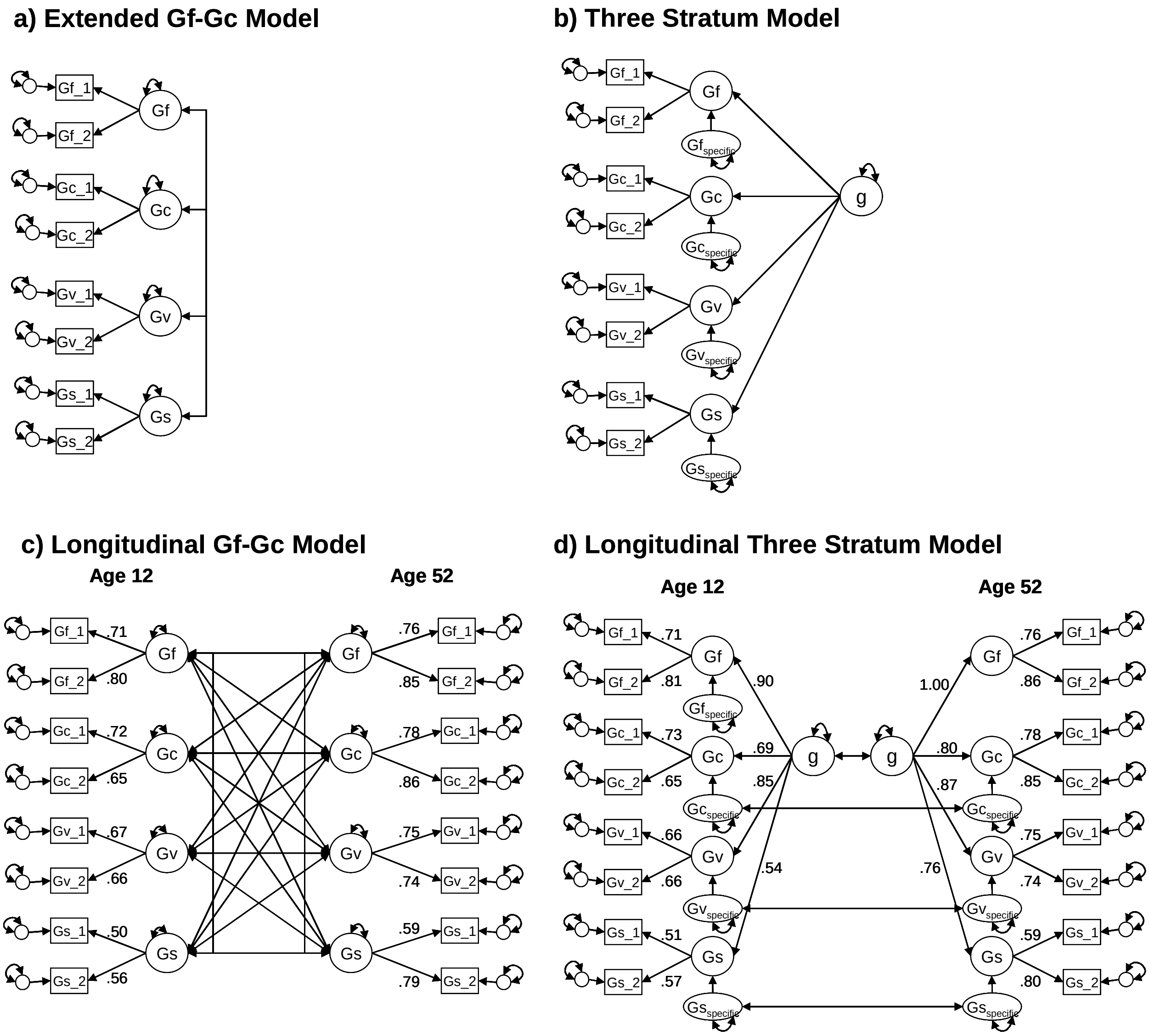

Figure 1

Alternative structural conceptualizations of intelligence: (a) first-order factor model, representing the extended Gf-Gc model, (b) higher order factor model, representing the three-stratum model, (c) longitudinal extension of the extended Gf-Gc model, and (d) longitudinal extension of the three-stratum model. Models C and D show the standardized factor loadings as obtained from Model T1.6 for the extended Gf-Gc model and Model C.3 for the three-stratum model. Gf = fluid reasoning, Gc = comprehension-knowledge, Gv = visual processing, Gs = processing speed, *g* = general cognitive ability. In Figures 1b and 1d, the suffix *specific* indicates specific abilities from which the influence of *g* was partialled out. Squares represent manifest test scores, circles represent latent variables; one-headed asymmetrical arrows represent directional regression coefficients (factor loadings), whereas two-headed symmetrical arrows represent variances or covariances. Correlated uniqueness terms of the manifest indicators Gc_1, Gv_2, Gs_1, and Gs_2 in the longitudinally extended models (see text) are not shown to ensure clarity of presentation.

In the current study, we focus on the developmental dynamics of four broad abilities: fluid reasoning (*Gf*), visual processing (*Gv*), processing speed (*Gs*), and comprehension-knowledge (*Gc*). Gf and Gc resemble the two opposing ends of abilities in two-component theories of intelligence, namely fluid and crystallized abilities (Li et al., 2004). Moreover, Gv has shown incremental validity in predicting educational and vocational attainment (Shea, Lubinski, & Benbow, 2001), and Gs has been shown to play an important role in the development of cognitive abilities (Salthouse, 1996).

2.1.2 The Distinction between Fluid and Crystallized Abilities

According to lifespan cognitive psychology proposed by Li and Baltes (2006), three kinds of influences and their interactions affect cognitive development: (a) biological processes, (b) normative environmental processes (e.g., formal education), and (c) non-normative person-specific experiences that result from self-selection into different environments. However, some broad abilities may be more sensitive to the environment than others. For example, the extended Gf-Gc theory allocates broad abilities on a continuum between two poles (Cattell, 1987; Li et al., 2004): fluid abilities (e.g. Gf, Gv, Gs) that are more strongly based on biological processes, and crystallized abilities (e.g. Gc) that are to a larger extent influenced by the environment. As people age, environmental influences, especially person-specific experiences, accumulate and should result in increased individual differences between persons. However, this increase in variance should be more pronounced for crystallized than for fluid abilities. Moreover, because fluid and crystallized abilities are predicted to be influenced differently by biological processes (e.g. aging) and the environment, crystallized abilities decline less and later in life compared to fluid abilities (McArdle et al., 2002; Schaie, 2005; Tucker-Drob, 2009).

2.1.3 The Age Differentiation-Dedifferentiation Hypothesis

One of the most comprehensive hypotheses regarding the development of intelligence is the age differentiation-dedifferentiation hypothesis that postulates three developmental stages across the lifespan (Baltes et al., 1980). The first stage of *differentiation* occurs during maturation, but especially in early childhood when different broad abilities are proposed to become increasingly independent of each other with increasing age (Deary et al., 1996). This effect is statistically represented by a decline in intercorrelations among broad abilities in the extended Gf-Gc model (Deary et al., 2004) or a progressively decreasing role of the influence of *g* in the three-stratum model (Escorial, Juan-Espinosa, García, Rebollo, & Colom, 2003). The following time of adulthood is described as a stage of *stability* in the structure of intelligence (Baltes et al., 1980). The third stage of *dedifferentiation* is characterized by again increasing dependencies among different broad abilities as people reach old age (Baltes et al., 1980). This effect is statistically represented by increases in intercorrelations among broad abilities in the extended Gf-Gc model (Deary et al., 2004) or an increasing influence of *g* in the three-stratum model (Escorial et al., 2003).

Theoretical accounts of age differentiation-dedifferentiation. Theoretical accounts refer to the differential impact of biological and environmental influences on fluid and crystallized abilities as well as their interdependencies. Specifically, Cattell’s investment theory (1987) postulates that fluid abilities are invested into the acquisition of crystallized abilities by taking advantage of environmental learning opportunities. When the environment becomes more heterogeneous as life unfolds, so do crystallized but not fluid abilities because crystallized abilities are more strongly impacted by the environment. This in turn leads to a differentiation of fluid and crystallized abilities. Lifespan developmental psychology has built upon these ideas and proposed comparable mechanisms for dedifferentiation in old age (Baltes & Lindenberger, 1997; Li & Baltes, 2006). At this stage, biological influences regain in importance by restricting cognitive performance in fluid abilities (Lindenberger & von Oertzen, 2006). Comparable to investment theory, declines in fluid abilities limit the acquisition or expression of crystallized abilities, and hence, the two broad categories of cognitive functioning grow closer together again (i.e. common cause hypothesis; Baltes & Lindenberger, 1997). However, some empirical findings point to qualitatively different processes that operate during maturation and senescence (Hommel, Li, & Li, 2004) so that they cannot simply be interpreted as the reverse of each other (Li & Baltes, 2006; Li et al., 2004).

Empirical results on age differentiation-dedifferentiation. The empirical results for the age-dependent differentiation-dedifferentiation are mixed (see Tucker-Drob & Salthouse, 2008 and ; Zelinski & Lewis, 2003 for a good overview of additional studies). For instance, support was found in cross-sectional comparisons by Baltes and Lindenberger (1997), Deary and colleagues (2004), Hayslip and Sterns (1979), Li and colleagues (2004), as well as longitudinal studies by Ghisletta and colleagues (Ghisletta & De Ribaupierre, 2005; Ghisletta & Lindenberger, 2003). By contrast, no support was found in cross-sectional studies by Escorial and colleagues (2003), Molenaar, Dolan, Wicherts, and van der Maas (2010), Tucker-Drob (2009), and Tucker-Drob and Salthouse (2008), or in longitudinal studies by Zelinski and Lewis (2003), and Schaie, Maitland, Willis, and Intrieri (1998). This leads to the conclusion that still, little is known about the age level at which differentiation and dedifferentiation actually occurs or whether the effects exist at all.

Problems in the study of the age differentiation-dedifferentiation hypothesis. To some extent, these mixed findings may be attributed to a number of methodological challenges. First, studies have applied different ability measures and/or have used samples that differed in their composition and age, which may render the findings somewhat incomparable (Lindenberger & von Oertzen, 2006). Second, the operationalization of the age variable in the analyses poses a problematic question (Molenaar et al., 2010). Combining different age levels into one age group and dividing the sample into respectively younger and older age groups, categorizes a continuous variable and is problematic because little is known about the age level at which differentiation and dedifferentiation occurs. Third, the effect can be conceptualized in a number of different ways. Some studies have contrasted the proportion of variance accounted for by the first unrotated principal component (Li et al., 2004) or the (mean) subtest correlations among two or more age groups (Deary et al., 2004). However, these approaches reveal little as to where in the model an increase (decrease) in correlations among different broad abilities originates and thus preclude a better understanding of the effect. Others have tested the factor structure of different age groups by casting constraints on factor covariances, variances, and/or loadings (Zelinski & Lewis, 2003). This approach is much more specific, but it does not solve the problem of categorizing the age variable. Only a few studies have analyzed the effect within a structural equation modeling approach by casting explicit age constraints on the parameters of the model (Tucker-Drob & Salthouse, 2008; Tucker-Drob, 2009) or by using age-moderated factor analysis (Molenaar et al., 2010).

2.1.4 Differential Stabilities

Age-dependent differentiation-dedifferentiation concerns the stability and change in the structure of intelligence; that is, whether and to what extent ability constructs operate similarly across time. A second key aspect regarding the developmental dynamics of cognitive abilities concerns differential stabilities; that is, whether the rank ordering of individuals remains stable across time. Statistically, this is represented by the autocorrelation of cognitive abilities across time, which requires longitudinal data. The existing findings suggest that *g* shows high differential stability across the lifespan. An extensive review of studies that have examined the differential stability of *g* can be found in Conley (1984) or Deary and colleagues (2000). For example, Deary and colleagues (2000, 2004) report differential stability estimates across almost the entire lifespan from age 11 to age 77 as well as age 11 to age 80 with correlation coefficients (not corrected for measurement error) of $r = .63$ and $r = .66$, respectively. Hertzog and Schaie (1986) analyzed differential stability by means of a *g*-factor spanning an age range of 20 to 74 years at the first test session, over a time span of 14 years.

They found differential stability estimates for *g* that were corrected for measurement error of $r = .92$ for the whole age sample as well as comparable correlations when they divided their sample into three age groups (young: $r = .93$; middle age: $r = .96$; old age: $r = .89$).

Only a few studies have addressed differential stabilities of different broad abilities over long time periods. The results of some key studies indicate comparable differential stabilities for broad abilities as were found for *g* (for a summary see Table 2). However, some studies have found higher differential stabilities for crystallized than for fluid abilities (Eichorn, Hunt, & Honzik, 1981; Kangas & Bradway, 1971; Nisbet, 1957; Owens, 1966; Pushkar Gold et al., 1995; Schwartzman, Gold, Andres, Arbuckle, & Chaikelson, 1987), but others have not (Larsen, Hartmann, & Nyborg, 2008; Schaie & Strother, 1968; Tuddenham, Blumenkrantz, & Wilkin, 1968). Interestingly, Larsen and colleagues (2008) found a vast decrease in differential stabilities of both verbal and arithmetic reasoning from $r = .82$ and .79 to $r = .44$ and .36, respectively, after the influence of *g* had been partialled out. This indicates that a large proportion of the differential stabilities of broad abilities (as represented by first-order factors) may be attributed to the differential stability of *g*. Hence, the rank ordering of specific abilities may be subjected to change to a larger degree than *g*. However, this conclusion is tentative because, to our knowledge, the study by Larsen and colleagues (2008) was the only study that took the stability of *g* into account when studying the differential stabilities of (specific) abilities

Table 2

Summary of Previous Key Studies on the Differential Stability (r) of Cognitive Abilities

Study	Mean age in years: Initial	Mean age in years: Follow-up	*N*	Cognitive ability	Correlation	Corrected for m.e.	*g* partialled	Measure
Eichorn et al. (1981)	17- 18	36 - 48	250	Verbal	.84 (men)	No	No	Stanford-Binet or Wechlser
					.81 (women)		No	Bellevue (initial) and
				Performance	.69 (men)		No	Wechsler Adult Intelligence
					.63 (women)		No	Scale (follow-up)
Kangas & Bradway (1971)	30	42	48	Verbal	.70	No	No	Wechsler Adult Intelligence
				Performance	.57		No	Scale
Larsen (2008)	20	38	4321-4385	Verbal reasoning	.82	No	No	Army Classification Battery
			(only men)	Arithmetic reasoning	.79		No	
				Verbal reasoning	.44	No	Yes	
				Arithmetic reasoning	.36		Yes	
Nisbet (1957)	22	47	141	Vocabulary	.48	No	No	Simplex Group Test
				Verbal	.44		No	
				Number	.39		No	
Owens (1966)	19	61	96	Verbal	.52 - .60	No	No	Army Alpha
			(only men)	Reasoning	.41 - .54		No	
Pushkar Gold et al. (1995)	25	65	316	Verbal abilities	.93	Yes	No	Revised Examination "M"
			(only men)	Nonverbal abilities	.64		No	
Schaie & Strother (1968)	20-70	5 year intervals	302	Verbal meaning	.88	No	No	Primary Mental Abilities
				Reasoning	.93		No	
				Space	.75		No	
Schwartzman et al. (1987)	25	65	260	Verbal abilities	.82	No	No	Revised Examination "M"
			(only men)	Nonverbal abilities	.54		No	
				Mechanical abilities	.66		No	
Tuddenham et al. (1968)	30	43	164	Reading and vocabulary	.69	No	No	Army General Classification
			(only men)	Arithmetic reasoning	.74		No	Test
				Pattern analysis (visual processing)	.64		No	

Note. m.e. = measurement error

However, previous findings on differential stabilities (Table 2) should be interpreted with some caution for two reasons. First, there are only a few longitudinal studies that have used latent variables that are free of measurement error. Hence, the reported results may underestimate the true differential stabilities because the stabilities reported for manifest test scores are attenuated by measurement error. Second, the differential stabilities of broad abilities (e.g., in a first-order model) may be overestimated because they do not separate the stability of *g* from the stabilities of broad abilities. Consequently, the differential stabilities of specific abilities as conceptualized in terms of a higher order model may be somewhat lower.

2.1.5 Methodological Requirements—Measurement Invariance

According to Little (1997), two types of measurement invariance (MI) can be distinguished: Type 1 MI concerns properties of the measurement scale (i.e., the measurement part of a model) across time, and Type 2 MI concerns latent variances, covariances, and means (i.e., the structural part of a model) across time. Type 1 invariance of measurement properties is needed in order to make meaningful comparisons of any latent construct in the intelligence models described above across time (age) by separating true changes in latent abilities from changes in operational definitions of the constructs. Thus, we first have to ensure that the measured (sub)tests relate to the latent common factors in the same way at all times of measurement (Meredith & Horn, 2001). More specifically, Type 1 MI concerns four different properties of the measurement scale (Cheung & Rensvold, 2002). First, *configural invariance* requires that the pattern of zero and nonzero loadings of observed indicators on the common factors remain the same across time. Second, *metric invariance* requires invariant factor loadings across time (i.e., the magnitudes of the unstandardized factor loadings have to be equal at all measurement occasions) and allows for the application of meaningful analyses of correlations and variances across time (Lubke, Dolan, Kelderman, & Mellenbergh, 2003). Third, *error invariance* requires the residual variances of the observed indicators (unique indicator variance and measurement error variance) to be invariant across time to ensure that the indicators are measured with the same amount of precision. A lack of error invariance may complicate the meaningful interpretation of latent variances, covariances, and means even when other invariance constraints are tenable (DeShon, 2004). Fourth, *scalar invariance* requires time-invariant intercepts and is needed for a meaningful comparison of means. Horn, McArdle, and Mason (1983) have questioned whether even metric invariance can realistically be expected in complex data sets used in developmental studies. However, some studies have shown that cognitive measures can demonstrate metric invariance across several age groups (for an overview see Zelinski & Lewis, 2003).

Type 1 MI is important and necessary but only a prerequisite for studying so-called Type 2 differences in latent variances, covariances, and means (Cheung & Rensvold, 2002; T. D. Little, 1997). Crucially, the Type 2 differences represent the substantive research interest in the present study because changes across time in the covariances and variances of the latent broad abilities directly tackle age differentiation-dedifferentiation. Remember that the hypothesis postulates changes in intercorrelations of different broad abilities across life stages. In a first-order model, such as the extended Gf-Gc model, changes in correlations among broad abilities can be caused by changes in covariances and/or changes in variances because a correlation between two broad abilities is computed by dividing their covariance by the product of their standard deviations. In a higher order model, such as the three-stratum model, a change in the intelligence structure is captured by changes in the variance of specific abilities and *g*, as well as second-order factor loadings of the different ability constructs on *g*.

2.2 The Present Study

The present study tackles two key aspects of the developmental dynamics of cognitive abilities concerning (a) stability and change in the structure of intelligence with reference to the age differentiation-dedifferentiation hypothesis and (b) differential stabilities across a 40-year time period from late childhood (at age 12) into middle adulthood (at age 52). A vital feature of the present study is that we examined these developmental dynamics by means of two alternative structural conceptualizations of intelligence: (a) the extended Cattell-Horn Gf-Gc model (Cattell, 1987; J. L. Horn & Noll, 1997) and (b) Carroll's three-stratum model (1993). Most previous developmental studies have conceptualized intelligence by applying two-component models, such as the extended Gf-Gc model, whereas psychometric research has been dominated by theoretical models that include *g*, such as Carroll's three-stratum model. Crucially, each model highlights different aspects of the data that are not visible from the vantage point of the other model. In particular, the extended Gf-Gc model emphasizes change in broad abilities as a whole, whereas the three-stratum model divides this change into change that is specific to each ability and change shared by all abilities and thus captured by *g*.

Drawing on these alternative conceptualizations of intelligence, the present longitudinal study makes several important contributions to the empirical body of research on the developmental dynamics of cognitive abilities. (a) It spans 40 years from late childhood to middle adulthood. Previous longitudinal studies on child development have rarely looked at intelligence development beyond early adulthood, and most of the developmental research on adults focuses on old age but not on middle adulthood. Hence, the present study provides vital information on cognitive development for an age range for which little empirical knowledge exists. (b) Previous results on the age differentiation-dedifferentiation hypothesis were mixed, and still little is known about the differentiation and dedifferentiation of the structure of cognitive abilities from late childhood to middle adulthood. Crucially, and in contrast to most previous research, we studied these processes across 40 years of people's lifetimes by means of a longitudinal sample that is highly homogeneous with respect to age. Moreover, as most previous developmental research on this hypothesis was embedded in the extended Gf-Gc model, it is not clear whether changes in the structure of intelligence can be attributed to a common core in terms of *g* or whether these changes are limited to specific abilities. (c) Most previous research on the differential stability of intelligence was conducted on the manifest level and therefore did not control for changes in the operational definition of the construct or for measurement error. Moreover, previous research mostly drew on the extended Gf-Gc model. Thus, little is known about the differential stability of specific abilities after the influence of *g* has been accounted for. Taken together, the current study provides a more detailed picture of the developmental dynamics of cognitive abilities for the time span from late childhood (at age 12) into middle adulthood (at age 52) by disentangling change that is attributed to specific abilities from change that is attributable to *g*.

2.3 Method

2.3.1 Participants and Procedure

The current longitudinal study (entitled MAGRIP) covers a time span of 40 years and encompasses two points of measurement: 1968 and 2008. In 1968, a multistage sampling procedure was applied to create two (overlapping) representative samples. First, half of all Grade 6 school classes in Luxembourg were selected randomly. All students from these classes participated. This sample is representative of sixth graders in Luxembourg. Second, a representative age-based sample was drawn that included all students in the selected schools who were enrolled in school in the school year 1963-1964. These were students who attended classes spanning from Grades 4 to 6 (students in lower grades had repeated one or more classes). To control for differential effects of age on cognitive development, we drew from this age-based sample, which included 2,450 children[1] (50.0% female) who were about 12

[1] One student was excluded because of severe outlying values on one of the intelligence subtests.

years old (*M* = 11.7 years; *SD* = 3.8 months) at the time of testing. All children completed a comprehensive intelligence test, the "Leistungsprüfsystem" (i.e. achievement test battery, W. Horn, 1962, 1983), which was administered by trained university students in a group setting.

In 2008, a sample that was stratified by region of residence in 1968 and gender of 344 (56.4% female) of these former students retook the same intelligence test at about 52 years of age. About two thirds of the retested age based sample (n = 227) took this test in a group setting; the remaining participants were visited at home to take the test individually. All tests were administered by trained assessors and the test taking procedure strictly followed the standardization of the test manual. Estimates of selective attrition of the retested age based sample show that (relative to the age base sample in 1968), the people who participated at both waves of measurement were slightly positively selected with respect to mean childhood g (Cohen's d = 0.34), parental socioeconomic status in childhood (d = 0.08), and grade point average (i.e., the mean grades computed across the last four trimesters prior to data collection in 1968; d = 0.28). Additional information on sample selectivity of the retested age based sample is depicted in Table 3 (for a detailed overview of the data collection stages and attrition see Figure 9 in Annex D).

Table 3

Estimates of Sample Selectivity

	Data collected in 1968				Effect size in *SD* units
	Total base sample *N* = 2,450		Longitudinal sample *n* = 344		
	M	*SD*	*M*	*SD*	Cohen's *d*
pSES	39.84	13.67	40.94	12.81	0.08
GPA	45.02	8.50	47.41	7.46	0.28
Childhood intelligence					
Gc	100.00	15.00	103.27	13.77	0.22
Gs	100.00	15.00	102.57	13.98	0.17
Gv	100.00	15.00	104.25	14.64	0.28
Gf	100.00	15.00	105.32	13.36	0.35
g	100.00	15.00	105.11	13.20	0.34

Note. Effect sizes indicate the selectivity of the longitudinal sample as used in the present dissertation: Positive effect sizes indicate that the value of a certain childhood characteristic was larger in the longitudinal sample compared to the total base sample. pSES = parental socioeconomic status measured on the ISEI scale (Ganzeboom, De Graaf, Treiman, & De Leeuw, 1992; Ganzeboom & Treiman, 1996); GPA = grade point average; *Gc* = comprehension knowledge; *Gf* = fluid reasoning; *Gv* = visual processing; *Gs* = processing speed; *g* = general cognitive ability.

2.3.2 Measures

Intelligence at ages 12 and 52 was assessed by nine subtests taken from a standardized and well validated German intelligence test battery, named the “Leistungsprüfsystem” (L-P-S; i.e. achievement test battery, W. Horn, 1962, 1983). Gf, Gv, and Gs were each assessed with two subtests. Gc was captured by three subtests. Each subtest contained 40 items and had to be completed within strict time constraints that were specified in the test manual. Because two of the three subtests of Gc contained the same kinds of items, we merged the scores on these two subtests into a single composite score to avoid having variance specific to this kind of subtest reflected in the factor Gc. Hence, every broad ability factor was assessed by two subtests, which are described in Table 1. Split-half reliabilities of single subtests as reported in the L-P-S test manual range between r_{tt} = .89 for the subtest Gs_1 and r_{tt} = .97 for subtest Gc_2 (Sturm, Willmes, & Horn, 1993), as well as split-half reliabilities for scales range between r_{tt} = .90 for Gf and r_{tt} = .99 for Gs (W. Horn, 1983). Strum and Büssing (1982) report a correlation of .94 between the L-P-S total score and the total score on the German version of the Wechsler Adult Intelligence Scale (WAIS; Tewes, 1991, note that Annex A contains detailed information on the reliability and validity of the L-P-S). In 1968, the children were randomly administered one of two parallel test forms of the L-P-S. Because the means and variances of subtests differed slightly across test forms, we used a linear-conversion rule (Kolan & Brennan, 1995) to equate the test scores. To this end, we standardized the subscales separately for each test form to an IQ metric with a mean of 100 and a *SD* of 15 for the base sample. In 2008, the participants were given the exact same test form and items that they had completed in 1968. To allow meaningful comparisons across time, subtest scores obtained for the second wave of measurement in 2008 were equated by using the same conversion rules as applied in 1968 (i.e., the standardization of measures in 2008 was based on means and *SD*s obtained from the entire age based sample in 1968).

2.3.3 Statistical Analysis

Strategy of analyses. Longitudinal confirmatory factor analysis was used to test the age differentiation-dedifferentiation hypothesis as well as differential stabilities in both the extended Gf-Gc and the three-stratum model. Some of the subtest scores were approximately but not strictly normally distributed. We therefore used maximum likelihood estimation with robust standard errors (MLR) as implemented in the Mplus program (Mplus 6; Muthén &

Muthén, 1998, 1998-2010). We conducted our main analyses in consecutive steps. In a first step, we tested for MI of the psychometric properties of the subtest scores (Type 1 MI) across age 12 and age 52. Because the three-stratum model rests on the extended Gf-Gc model, they share the same measurement model. Thus, the test for Type 1 MI applied to both models. To study Type 1 MI, we tested configural invariance first and metric invariance second. We then proceeded by testing the equality of error variances because these residual variances could also contain reliable unique sources of variance, and changes in the residual part of the model might complicate the substantial interpretation of factor variances and covariances, which were the main focus of the present study. We tested for scalar invariance last as this level of Type 1 MI was least important for our hypotheses. To assess model fit we applied nested-model comparisons and consulted several fit indices that are recommended in the literature (see Annex for details). In the second step of our analyses, we tested for Type 2 MI of latent variances, covariances, and factor loadings across time, which tackles the age differentiation-dedifferentiation hypothesis. Third, we assessed the differential stabilities of broad abilities, specific abilities, and *g*.

Handling correlated residual terms. A vexing problem of research on cognitive development is that an observed subtest score may not only capture the target ability construct(s) but also some unique ability that is specific to a certain subtest. The latter is represented by the subtests' residual terms in factor models (Brunner et al., 2012). Preliminary analyses showed that for some of the subtests (i.e., Gc_1_{12} with Gc_1_{52}, Gv_2_{12} with Gv_2_{52}, Gs_1_{12} with Gs_1_{52}, and Gs_2_{12} with Gs_2_{52}) the residual terms were significantly correlated across the two measurement occasions. As recommended in the literature for longitudinal studies (Cole & Maxwell, 2003), we therefore allowed the residual terms of these subtests to correlate across time in all models that we investigated.

Handling missing data. Missing values were not a severe problem in our data. For the 344 participants in the longitudinal sample, data were missing on one variable (Gv_1) for 11 participants on the other variables for only one or two participants. Full information maximum likelihood (FIML) estimation was used to handle missing data (R. J. A. Little & Rubin, 2002).

2.4 Results

2.4.1 Descriptive Statistics

The manifest scores of all indicators showed substantial mean increases across time with large effect sizes (ranging from d = 0.52 to d = 2.71; see Table 4). Furthermore, the correlations between the measures at age 12 and 52 ranged between r = .40 and .62, suggesting moderate to high differential stabilities of the observed subtest scores (see Table 13 in Annex C for a full correlation matrix of all measures applied and information on reliabilities of subtest scores). Moreover, to measure changes in variance across time, we computed variance ratios by dividing the variance of a subtest score at age 52 by the variance of the same subtest score at age 12: a value of 1 indicates no change in variance, values greater and smaller than 1 indicate an increase or respectively decrease in variance at age 52. The variance ratios of subtest scores indicated that the variances for measures of Gc and Gs increased more than the variances for measures of Gf and Gv.

Table 4

Mean Change, Differential Stability, and Change in Variability as Obtained for Observed Subtest Scores in the Longitudinal Sample

Measure	Age 12		Age 52		Age 12 vs. age 52		
	M	*SD*	*M*	*SD*	*ES*	*r*	Variance ratios
Comprehension-knowledge							
Gc_1	102.80	13.04	161.98	24.14	2.71	0.62	3.43
Gc_2	103.11	14.68	133.63	21.13	1.63	0.47	2.07
Fluid reasoning							
Gf_1	104.67	14.18	117.42	15.80	0.85	0.48	1.24
Gf_2	104.86	13.02	121.76	13.83	1.26	0.54	1.13
Visual processing							
Gv_1	103.51	15.34	112.52	18.91	0.52	0.40	1.52
Gv_2	103.55	14.12	116.32	14.52	0.89	0.57	1.06
Processing speed							
Gs_1	101.70	14.79	121.19	23.55	0.96	0.40	2.53
Gs_2	102.22	12.99	121.95	17.40	1.27	0.41	1.79

Note. $N = 344$, full information maximum likelihood estimates for missing data. ES = effect size for mean differences of correlated measures across time computed according to Dunlap, Cortina, Vaslow, and Burke (1996). r = differential stability. Variance ratios = variance at age 52 / variance at age 12; values larger than 1 indicate larger variability at age 52 compared to age 12. Gc = comprehension-knowledge; Gf = fluid reasoning; Gv = visual processing; Gs = processing speed; _1 and _2 refer to manifest variables 1 and 2 that measure the respective broad ability.

2.4.2 Invariance of Psychometric Properties of Subtest Scores across Time

To study the Type 1 MI of subtest scores, we examined a series of increasingly constrained models. The key results of these analyses can be summarized as follows (see Annex B for a detailed description of these analyses. Model fit indices are depicted in Table 5). A partial scalar invariant measurement model (i.e., T1.6) where subtests demonstrated complete metric invariance of factor loadings, partial invariance of the residual terms (the residual variances of the subtest scores Gc_1 and Gs_1 were not invariant across time), and partial invariance of the intercepts (the intercepts of the subtests Gc_1 and Gv_1 were not invariant across time) provided a good fit to the data. The standardized factor loadings (λ) obtained for this model (see Figure 1c) show that each factor representing a broad ability was well defined with values ranging between $\lambda = .50$ (for the loading of Gs_1 on Gs at age 12) and $\lambda = .86$ (for the loading of Gc_2 on Gc at age 52). As noted above, the residual terms of some subtests were significantly correlated across the two measurement occasions, involving Gc_1with $r = .37$, Gv_2 with $r = .27$, Gs_1 with $r = .21$, and Gs_2 with $r = .26$. Note that these correlated uniqueness terms remained approximately the same when we tested the three-stratum model (see below). Model T1.6 also provides some insights into changes in latent means of broad abilities across time (a question that was, however, not central to the present manuscript). We observed substantial and statistically significant increases in mean changes from age 12 to age 52 representing very large effect sizes[2]: $d_{Gf} = 1.34$, $d_{Gs} = 1.48$, $d_{Gv} = 1.16$, and $d_{Gc} = 1.58$. Note that the mean changes observed for Gc and Gv should be interpreted with caution as partial scalar invariance implies that mean changes in the (observed) subtest scores represent not only changes in the latent means of the corresponding broad abilities, but also mean change attributable to subtest-specific abilities (see Millsap & Olivera-Aguilar, 2012). To conclude, our results concerning Type 1 MI indicate that the operational definition of the four broad abilities is fundamentally the same at age 12 and age 52 and allows meaningful comparisons of the latent covariances and variances in order to test the age differentiation-dedifferentiation hypothesis based on the extended Gf-Gc model and the three-stratum model, respectively.

[2] Effect sizes were computed according to Dunlap, Cortina, Vaslow, and Burke (1996) for mean differences of correlated measures across time.

Table 5

Evaluation of Model Fit to Study the Psychometric Properties of the Cognitive Measures (Type 1 Measurement Invariance) and the Age-Differentiation-Dedifferentiation Hypothesis (Type 2 Measurement Invariance) for the Extended Gf-Gc and the Three-Stratum Models

Model Constraint	χ2	df	CFI	RMSEA	SRMR	AIC	Differences (Δ)						
							Compare	χ2	df	CFI	RMSEA	SRMR	AIC
Type 1 Measurement Invariance (T1)													
T1.1 configural invariance	87.79	72	.996	.025	.029	43830.60	—	—	—	—	—	—	—
T1.2 metric invariance	107.03	76	.993	.034	.064	43842.24	T1.2 vs. T1.1	18.51	4	–.003	.009	.035	11.63
T1.3 error invariances	--	--	--	--	--	--	--	--	--	--	--	--	--
T1.4 partial error invariance	120.30	82	.991	.037	.074	43844.01	T1.4 vs. T1.2	13.09	6	–.002	.003	.010	1.78
T1.5 scalar invariance	234.94	86	.967	.071	.142	43949.86	T1.5 vs. T1.4	135.95	4	–.025	.034	.068	105.85
T1.6 partial scalar invariance	139.44	84	.988	.044	.086	43859.20	T1.6 vs. T1.4	19.83	2	–.004	.007	.012	15.18
Type 2 Measurement Invariance: Extended Cattell-Horn Gf-Gc Model (CH)													
CH.1 all covariances equal	255.83	90	.963	.073	.218	43964.11	CH.1 vs.T1.6	117.89	6	–.025	.029	.132	104.91
CH.2 all variances equal	292.85	88	.954	.082	.279	44007.42	CH.2 vs. T1.6	137.19	4	–.034	.038	.193	148.22
CH.3 variances of Gf equal	147.99	85	.986	.046	.101	43866.19	CH.3 vs. T1.6	7.17	1	–.002	.002	.015	7.00

(Table 5 to be continued)

Table 5. (continued)

Model Constraint	χ2	df	CFI	RMSEA	SRMR	AIC	Differences (Δ)						
							Compare	χ2	df	CFI	RMSEA	SRMR	AIC
Type 2 Measurement Invariance: Carroll's Three Stratum Model (C)													
C.1 configural invariance[3]	165.09	101	.986	.043	.089	43851.66	C.1 vs. T1.6	25.70	17	–.002	–.001	.003	–7.54
C.2 second order metric invariance	220.52	104	.974	.057	.139	43900.98	C.2 vs. C.1	61.04	3	–.012	.014	.050	49.32
C.3 second order partial metric invariance	165.10	102	.986	.042	.089	43849.69	C.3 vs. C.1	0.01	1	.000	–.001	.000	–1.97
C.4 all specific variances equal	205.65	106	.978	.052	.114	43882.66	C.4 vs. C.3	40.54	4	–.008	.010	.025	32.97
C.5 specific variances of Gv & Gs equal	175.27	104	.984	.045	.097	43856.15	C.5 vs. C.3	9.84	2	–.002	.003	.008	6.46
C.6 equal variances of *g*	201.54	105	.978	.052	.140	43880.79	C.6 vs. C.3	35.41	3	–.007	.010	.051	31.10

Note. Model CH1 and C1 are nested within Model T1.6. Except for Model CH2 (nested within Model T1.6), each model is nested within the previous one. df = degrees of freedom; CFI = comparative fit index; RMSEA = root mean square error of approximation; SRMR = standardized root mean square residual; AIC = Akaike's Information Criterion. For computing the χ^2 difference test as derived from the MLR estimator we calculated scaled χ^2 difference values using the procedure described in Satorra and Bentler (1999)

[3] Preliminary analyses indicated that parameter estimates for a fully configural invariance specification of the three-stratum model were not admissible. To overcome this problem, we constrained the variance of $Gf_{specific}$ at age 52 to zero

2.4.3 Testing the Age Differentiation-Dedifferentiation Hypothesis

The extended Cattell-Horn Gf-Gc model. Model T1.6 reflects the structural propositions of the extended Gf-Gc model, and therefore served as the baseline model for testing the age differentiation-dedifferentiation hypothesis within this theoretical framework. In the extended Gf-Gc model, age differentiation or dedifferentiation is captured by changes in the intercorrelations among broad abilities. Table 6 shows that except for the correlation of Gv with Gc, the correlations between broad abilities increased from age 12 to age 52. To assess the overall effect of dedifferentiation, we computed mean correlations of broad abilities at age 12 and age 52, respectively. Mean correlations were computed by transforming the correlations among abilities into Fisher's *z*-values, averaging the *z*-values, and retransforming the average *z*-value into a correlation coefficient (Cohen, Cohen, West, & Aiken, 2003). This yielded a mean correlation of $\bar{r}_{12}$ = .57 with a 95% confidence interval of [.48; .67] and $\bar{r}_{52}$ = .75 with a 95% confidence interval of [.71; .79] at age 12 and age 52, respectively.

Table 6

Correlations among Broad Abilities as Obtained for the Extended Gf-Gc Model

	At age 12				At age 52			
	Gc	Gs	Gv	Gf	Gc	Gs	Gv	Gf
At age 12								
Gc	—							
Gs	.44	—						
Gv	.67	.43	—					
Gf	.59	.49	.74	—				
At age 52								
Gc	.81	.38	.59	.60	—			
Gs	.40	.72	.50	.57	.67	—		
Gv	.50	.22	.87	.71	.68	.56	—	
Gf	.57	.42	.76	.82	.79	.77	.90	—

Note. These correlation coefficients are based on the extended Gf-Gc model with metric invariance, partial residual invariance, and partial scalar invariance (Model T1.6 in Table 5). These correlations are identical to those obtained using the extended Gf-Gc model with metric invariance and partial residual invariance

To study the source of the increased intercorrelations, we examined the age-group-specific variances and covariances of broad abilities. Our results showed that the increased correlations resulted from increased covariances for all broad abilities (see Figure 2a) as well as increased variances for all broad abilities (see Figure 2b) with the largest variance increases for Gc, followed by Gs and Gv. The variance of Gf did not change much across time. These conclusions were corroborated by statistical tests (see Table 5) in which we imposed equality constraints on covariances (Model CH.1) and variances (Models CH.2 and CH.3) across time and compared the resulting models with the baseline Model T1.6. To conclude, in the extended Gf-Gc model, a significant increase in the mean correlation could be observed from age 12 to age 52, which is indicative of age dedifferentiation. Further, the age dedifferentiation effect was the product of both an increase in covariances and variances of broad abilities.

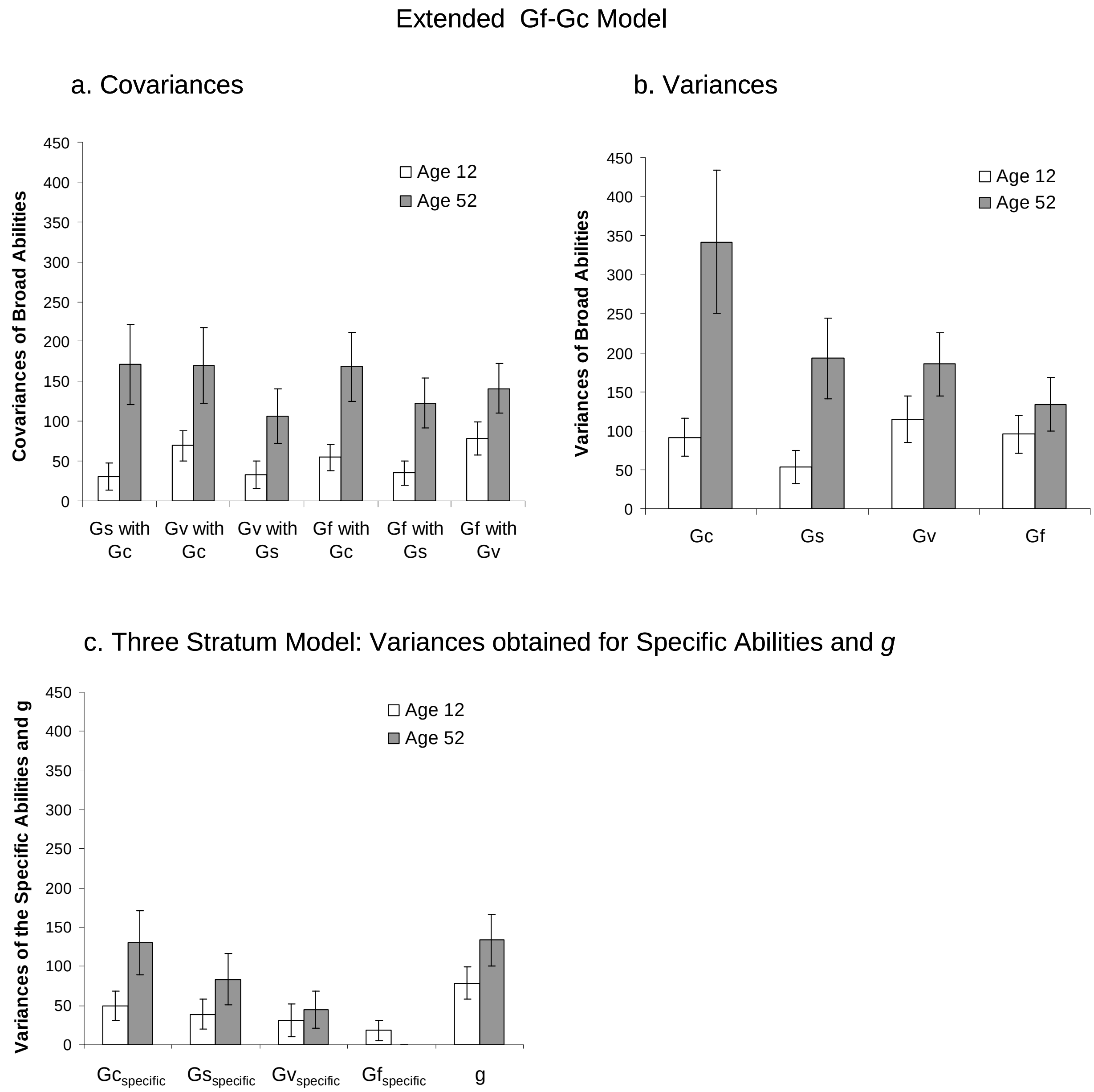

Figure 2

Dedifferentiation of cognitive abilities across time as observed in the extended Gf-Gc model and the three-stratum model: (a) covariances of broad abilities in the extended Gf-Gc model (Model T1.6), (b) variances of the broad abilities in the extended Gf-Gc model (Model T1.6), and (c) variances of specific abilities and *g* in the three-stratum model (Model C.3). Error bars represent the 95% confidence intervals. Gf = fluid reasoning, Gc = comprehension-knowledge, Gv = visual processing, Gs = processing speed, *g* = general cognitive ability.

Carroll's three-stratum model. In the three-stratum model, age dedifferentiation can be caused by (a) increases in the second-order factor loadings of the broad abilities on *g*, (b) decreases in the variances specific to Gf, Gc, Gv, or Gs, and (c) an increase in *g* variance. To study these sources of dedifferentiation, we first needed to test the structural propositions of the model. To this end, we drew on the measurement model T1.6 and introduced a higher order factor representing *g* at age 12 and age 52, respectively. To examine differential stabilities in the framework of the three-stratum theory (described in the next section), we specified correlations between matching specific ability factors and *g* across time, respectively. Preliminary results indicated that the loading of Gf on *g* at age 52 was estimated to be greater than one and thus not admissible. To overcome this problem, we constrained the variance of $Gf_{specific52}$ to zero (Model C.1). This model fit the data well and not considerably worse than Model T1.6 (see Rindskopf & Rose, 1988, who provide the rationale that the higher order factor model is nested within the corresponding first-order factor model).

To identify the various sources of age dedifferentiation, we drew on Model C.1 and imposed several equality constraints across time (see Annex B for a detailed description of these results). In sum, our results showed that the increase in intercorrelations observed in the extended Gf-Gc model is the result of several age-specific changes: (a) increases in the factor loadings of Gc and Gs on *g*, (b) a decrease in the variance specific to Gf with Gf even becoming indistinguishable from *g* at age 52, and (c) by a substantial increase in the variance of *g* over time. However, at the same time, the variance specific to Gc increased, which is indicative of differentiation.

2.4.4 Differential Stabilities

Figure 3 shows the differential stabilities (i.e., the correlations of corresponding factors across age) of the broad abilities in the extended Gf-Gc model as well as the specific abilities and *g* in the three-stratum model. These differential stabilities span 40 years of the participants' lifetimes from late childhood to middle adulthood. Model parameters were taken from Model T.1.6 for the extended Gf-Gc model and Model C.3 for the three-stratum model. The values ranged from r = .72 to r = .87 in the extended Gf-Gc model and from r = .75 to r = .91 in the three-stratum model. Thus, the results of both intelligence models show very high differential stabilities for all broad abilities, for all specific abilities, and for *g*. This further shows that the differential stabilities of specific abilities, after the influence of *g* has been partialled out, remained high and comparable to the differential stabilities of broad abilities when the influence of *g* had not been controlled for. Finally, no differences between fluid and crystallized abilities were observed for differential stabilities. Hence, both fluid and crystallized abilities as well as *g* were found to be highly stable personal traits from late childhood to late adulthood with only minor shifts in the rank ordering of persons.

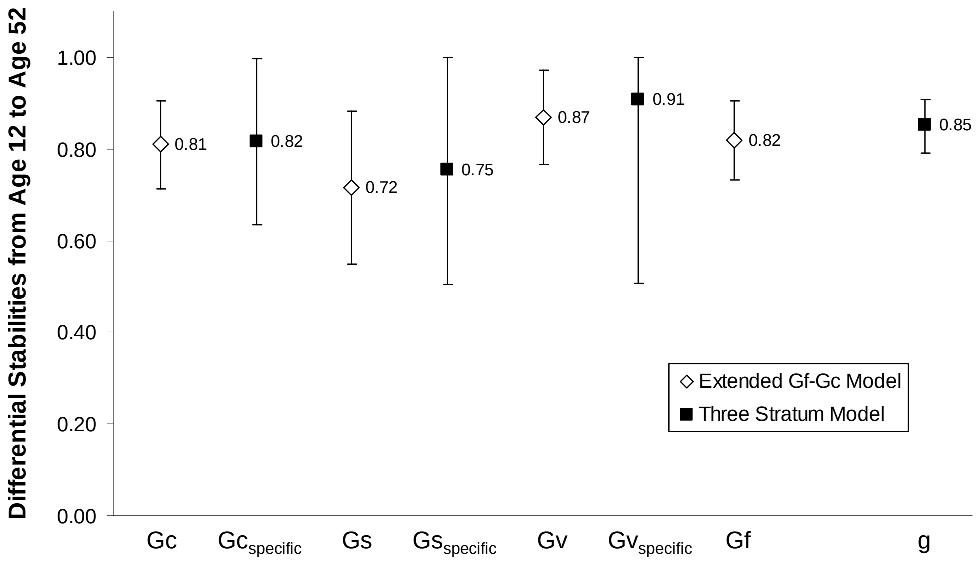

Figure 3

Differential stabilities as correlations of corresponding broad abilities in the extended Gf-Gc model (Model T1.6) and specific abilities as well as *g* in the three-stratum model (Model C.3) from age 12 to age 52. Gf, Gv, Gs, and Gc refer to the broad abilities in the extended Gf-Gc model. $Gf_{specific}$, $Gv_{specific}$, $Gs_{specific}$, and $Gc_{specific}$ refer to specific abilities in the three-stratum model and represent the correlations of the broad ability factors after the influence of *g* has been partialled out. The bars represent the 95 % confidence intervals. Gf = fluid reasoning, Gc = comprehension-knowledge, Gv = visual processing, Gs = processing speed, *g* = general cognitive ability

2.5 Discussion

2.5.1 Cognitive Change in Alternative Structural Conceptualizations of Intelligence

The present study examined two key aspects of the developmental dynamics of cognitive abilities across the lifespan: (a) stability and change in the structure of intelligence with reference to the age differentiation-dedifferentiation hypothesis and (b) differential stabilities from late childhood (age 12) into middle adulthood (age 52). To this end, we took advantage of two alternative structural conceptualizations of intelligence. The extended Cattell-Horn Gf-Gc model (Cattell, 1987; J. L. Horn & Noll, 1997) has dominated previous developmental research and examines broad abilities. Carroll's (1993) three-stratum model is strongly grounded in psychometrically oriented intelligence research and highlights aspects of the data that are not visible when using the extended Gf-Gc model. Specifically, the three-stratum model disentangles developmental processes that are attributable to what is specific to a certain ability (independent of *g*) from those processes that are shared by all abilities and that are therefore attributable to *g*.

2.5.2 Age Differentiation-Dedifferentiation

According to the age differentiation-dedifferentiation hypothesis, the structure of intelligence is expected to differentiate during childhood until late adolescence, keep a fairly stable structure during adulthood, and dedifferentiate in old age. However, irrespective of the structural model applied, the results of the present study seem not to fit well into the theoretically expected pattern, as we found age dedifferentiation from age 12 to age 52. Specifically, in the extended Gf-Gc model, we saw that all covariances between broad abilities increased significantly and substantially from age 12 to age 52. In contrast, the variance increases were substantial for Gc and Gs only, much smaller for Gv, and did not reach significance for Gf. However, the increases in the covariances among and the variances of the broad abilities in the extended Gf-Gc model were largely accounted for by variance increases in *g* in the three-stratum model, because the variances of specific abilities (except for Gc) did not increase significantly across time. Gc was an exception, demonstrating a large increase in the variance that was specific to Gc even after the influence of *g* was taken into account. This points to the conclusion that changes in the variance of Gc are influenced by a source other than only *g* (to be explained below). Crucially, the two pure markers of fluid (Gf) and crystallized (Gc) abilities exhibited complementary patterns. This was especially visible in the three-stratum model, because the specific variance of Gc increased significantly, whereas the specific variance of Gf decreased to zero and hence became indistinguishable from *g* at age 52.

2.5.3 Differential Stabilities

Our results showed that persons' rank ordering across time concerning (a) their broad abilities in the extended Gf-Gc model and (b) their specific abilities and *g* in Carroll's three-stratum model remained largely stable. This suggests that, across a time span of 40 years, individuals may keep their relative standing with reference to the population in all broad abilities, all specific abilities, and *g*. Thus, in contrast to the study by Larsen et. al. (2008), the differential stabilities of specific abilities remained high even though the influence of *g* had been accounted for. In addition, no differences in differential stabilities were indicated for fluid and crystallized abilities or for *g*. Thus, in line with other studies (Conley, 1984; Deary et al., 2000, 2004), our results suggest that the various aspects of intelligence and general intelligence comprise a highly differentially stable construct across age. Importantly, the results obtained for the three-stratum model also show that when individual differences in *g* are held constant, specific strengths and weaknesses in the cognitive profile (as reflected by the specific abilities) are highly stable. Thus, these results point to the conclusion that it is not only the level of an ability profile (as indicated by g) that remains stable across time, but also the pattern of the cognitive profile with regard to an individual's configuration of specific abilities.

2.5.4 Combined Effect of Age Differentiation-Dedifferentiation and Differential Stabilities

In the extended Cattell-Horn Gf-Gc model, we saw that the variances of broad abilities (except Gf) increased, which suggests that people differ more with respect to their broad abilities at age 52 than at age 12. At the same time, all differential stabilities of broad abilities remained high, which shows that individuals keep their relative standing in the population. Hence, initial differences between people on Gc, Gs, and Gv appear to become increasingly larger as life unfolds, and the gap between the two ends of the ability distribution widens across the lifespan. This effect (in combination with the observed increases in latent means and means of the manifest subtest scores shown in Table 4) can be described in the words of Ceci and Paperierno (2005, p. 149) as, “the ‘have-nots’ gain but the ‘haves’ gain even more”. This effect is also known as the Matthew effect or cumulated advantages (DiPrete & Eirich, 2006). In the three-stratum model, we saw that (a) the main reason why people differ more greatly at age 52 is captured by an increased *g* variance, though (b) the differential stability of *g* also remains extremely high. Thus, initial differences in *g* become amplified and increasingly important as life unfolds. Moreover, this gap-widening effect of *g* seems to account for large parts of the age dedifferentiation effect, which we observed as increases in the covariances of the broad abilities in the extended Gf-Gc model.

2.5.5 Explanations of Age Dedifferentiation in the Current Study

How can we explain the current finding of age dedifferentiation from age 12 to age 52? Several processes may have acted in combination to produce these results. First, the results of the current study are in line with propositions made by lifespan cognitive psychology by Li and Baltes (2006) that the increasingly heterogeneous environment across the lifespan leads to greater increases in the variance of crystallized than of fluid abilities, because crystallized abilities are more sensitive to the environment. In the extended Gf-Gc model, the variance of Gc increased substantively, whereas the variances of Gv and Gf showed smaller gains. Intuitively, Gs might be expected to be an exception to the proposed pattern, since processing speed is generally considered to be a biologically determined and fluid ability. However, as processing speed mainly involves the ability to concentrate and to focus, it is presumably also affected by environmental opportunities to train these abilities, which can explain the large increase in variance.

Second, according to Ceci and Papierno (2005), gap widening occurs because (a) more gifted people may profit more from environmental opportunities by learning faster (see also Kan et al., 2011), and (b) more gifted people may take better advantage of environmental opportunities (e.g. by seeking environments that are cognitively more challenging and thus more profitable for their cognitive development). This may result in an interaction of the environment with the initial ability level because people actively select or are placed into environments that match their abilities (for similar explanations, see also Dickens & Flynn, 2001; Scarr & McCartney, 1983; van der Maas et al., 2006).

Third, it seems that the observed process of age dedifferentiation is not explained well by the common cause hypothesis (Baltes & Lindenberger, 1997). According to the common cause hypothesis, decreases in fluid abilities limit the acquisition of crystallized abilities, and as a result, the two kinds of abilities become more similar. This explanation does not fit well with the current results for a number of reasons: (a) Longitudinal studies do not show declines in cognitive abilities until age 50 (Tucker-Drob & Salthouse, 2011). Likewise, we did not observe a decrease in mean levels of fluid abilities from age 12 to age 52 in the current study. On the contrary, the latent means of Gf and Gs point to a substantial increase from age 12 to age 52. (b) The age dedifferentiation in the current study is caused by initial differences between people that become more pronounced. Thus, the current effect originates because the “have-nots” gain but the “haves” gain even more, and a gap widens between the two ends of the distribution. In other words, the effect seems to be caused by unequal gains in cognitive functions between people and not by losses in cognitive functions. Taken together, our results suggest that the common cause hypothesis might be more appropriate for explaining ability dedifferentiation in older age groups.

Fourth, the current pattern of results may be partly explained by several propositions of the investment theory. Specifically, investment theory proposes that fluid abilities are invested into the acquisition of crystallized abilities by taking advantage of environmental learning opportunities. Kvist and Gustafsson (2008) have further argued that if this proposition holds true, Gf and *g* should be the same entity because Gf is postulated to be involved in all kinds of learning (see also Kan et al., 2011). This is exactly what we found in the current study: Gf and *g* became indistinguishable at age 52. Further, according to investment theory, age differentiation occurs because the environment becomes increasingly heterogeneous as life unfolds, which affects crystallized abilities to a greater extent than fluid abilities. The described mechanisms are used to explain differentiation of crystallized and fluid abilities. Our results partly supported this prediction, as we found a significant change in $Gc_{specific}$ (which is indicative of age differentiation) that may resemble the strong influence of environmental learning opportunities on crystallized abilities. However, the substantial increase in the variance of Gc (in the extended Gf-Gc model) was also to a large degree accounted for by variance increases in *g* (in the three-stratum model), which implies dedifferentiation of broad abilities with age and not differentiation as proposed by the investment theory.

2.6 Strengths and Limitations of the Current Study

In the current study, we examined cognitive development across 40 years of participants' lifespans in a longitudinal sample that was highly homogeneous with respect to age. For this reason, we did not have to arbitrarily divide our sample into two age groups to study the age differentiation-dedifferentiation hypothesis, as has been done in most previous research. Moreover, the longitudinal data base made it possible to analyze both the age differentiation-dedifferentiation hypothesis as well as differential stabilities in the same study. Further, our research design also allowed us to effectively address one major validity threat that longitudinal designs usually suffer from: The time span of 40 years in between the two measurement occasions rendered retest effects almost impossible (Tucker-Drob & Salthouse, 2011).

Despite these strengths, our study was subject to several limitations which should be born in mind when interpreting the present findings and addressed in future research. First, we could not directly tackle one key problem of longitudinally designs – selective attrition (Tucker-Drob & Salthouse, 2011). Notably, the present longitudinal sample reflected important characteristics of the representative base sample of 12 year old students fairly well, as it was only slightly positively selected in terms of several childhood characteristics including cognitive abilities, parental socioeconomic status, grade point average, gender, and migration background (see Tables 3 and also Figure 9 in Annex D. Nevertheless, we cannot rule out that the statistical estimates of individuals' cognitive development are distorted because of the selective attrition of study participants. For example, aging related deficits may be underestimated (Tucker-Drob & Salthouse, 2011) because participants with lower cognitive abilities were more likely to drop-out of the current study. This is a typical problem of most longitudinal studies (Tucker-Drob & Salthouse, 2011) and can be a result of several reasons, like (a) an association of lower cognitive abilities and death or illness (Deary, 2009), or (b) disinterest of lower functioning participants as they might have less confidence in their cognitive abilities or in the test.

Second, data were available only for two points of measurement. With only two points of measurement, it is impossible to provide a comprehensive picture of the course of cognitive development (e.g., growth-curve modeling of individuals' cognitive development). Thus, we do not know whether the mean performance of our individuals was already declining after it peaked in late adolescence, as often found in cross-sectional studies, or was not yet in decline, as often estimated by longitudinal studies (Tucker-Drob & Salthouse, 2011). Moreover, we cannot rule out the possibility that the age dedifferentiation effect in the current study was a result of initial differentiation until late adolescence followed by dedifferentiation as proposed by the theory. According to the findings by Tucker-Drob (2009), even the reverse pattern would be possible. Thus, having more measurement occasions at important developmental stages such as in early childhood or late adolescence would have been very valuable to better portray individuals' cognitive development.

Third, Type 1 invariance of measurement properties is needed in order to make meaningful comparisons of latent ability constructs across time. In the present paper, we based our conclusions on a measurement model (i.e., T1.6) where subtests demonstrated complete metric invariance of factor loadings, partial invariance of the intercepts, and partial invariance of the residual terms. Particularly, the residual variances of two out of eight subtests were higher at age 52 than at age 12. This may reflect an increase of variance attributable to (a) random measurement error and/or (b) subtest-specific abilities. The latter would indicate another potential source of differentiation of abilities. The available data, however, are insufficient to examine this possibility, as a separate analysis of subtest-specific abilities would require two parallel subtests at each point of measurement, which are not available in the present data set. Clearly, given this level of measurement invariance, changes in mean levels and variances of subtests across time need to be interpreted with great caution, as these changes may reflect changes in target ability constructs (specified as factors in the extended Gf-Gc model or the three-stratum model), as well as changes in subtest-specific abilities or random measurement error. Moreover, it has been debated whether partial invariance of residual terms may complicate the interpretation of factor variances and covariances (DeShon, 2004) or not (T. D. Little, 1997). Here we take the stance that it is reasonable to compare factor variances and covariances (which were central to our research goals) across time even when only metric invariance of factor loadings holds (see for example Widaman & Reise, 1997). Note that estimates of factor variances and covariances may be severely biased when residual terms are specified to be invariant though they are in fact not (as found in the present study). To obtain precise estimates of cognitive development given our data, we therefore followed Little's (T. D. Little, 1997, p. 55, Footnote 1) advice and based our conclusion on a measurement model with partial invariance of residual terms rather than forcing the residual variances to be invariant.

Fourth, we had only two observed indicators as measures of each broad ability factor, which constitutes the lower limit for assessing latent factors in structural equation models. To be able to measure cognitive change, we had to use the same subtests that were given in 1968. Notably, these subtests represent widely-used indicators of the broad abilities under investigation. However, when ability factors are measured using only two subtests, the factors may not represent the full conceptual scope of the abilities in question (e.g., the measurement of Gc would have profited from including a curriculum-based test of students' knowledge). Further, subtest scores were not perfectly reliable (see Table 13 in Annex C), which in turn affects the precision (in terms of standard errors) of statistical parameters reflecting age differentiation-dedifferentiation or differential stabilities. In sum, it is an open question whether the present results on cognitive development are tied to the specific subtests applied or whether the present results may also reflect cognitive change when ability factors are measured using a broader set of subtests. Future research will therefore benefit from administering a broader set of subtests to overcome this limitation and to yield more precise estimates of cognitive development.

2.7 Conclusion

The present study examined age differentiation-dedifferentiation and differential stabilities of cognitive abilities in the theoretical framework of (a) the extended Gf-Gc model for studying broad abilities and (b) the three-stratum model for decomposing cognitive change into those processes that are attributable to a certain specific ability (which is independent of *g*) and those that are shared by all broad abilities (which are thus attributable to *g*). The present results suggest that people differ more greatly with respect to broad abilities (except for Gf) as life unfolds and that the rank ordering of persons on all broad abilities remains remarkably stable across time. The combined results of these developmental processes points to considerable gap-widening effects from age 12 to age 52 that can be mainly accounted for by a substantial increase in *g* variance in combination with the high differential stability of *g*. The described gap-widening also led to substantial age dedifferentiation effects. The pattern of results in the current study seems to be well aligned with the predictions of the investment theory and lifespan cognitive psychology, that fluid and crystallized abilities are differentially affected by learning environments and that fluid abilities are invested into the acquisition of crystallized abilities. However, the proposition of the investment theory that these processes lead to age differentiation could only partially be supported, as we found that these processes mainly lead to age dedifferentiation.

Chapter 3 – Study 2: Long-term Consequences of the Quantity and Quality of Formal Education for the Development of Intelligence

3.1 Theoretical Background

In our increasingly complex environment, cognitive abilities have become one of the most important determinants for facing the challenges of everyday life in our modern society (L. S. Gottfredson, 1997b, 2002; Lubinski, 2004). Trajectories of cognitive change may be more or less successful as people grow older (Tucker-Drob & Salthouse, 2011) and differences in these trajectories become especially important as people reach old age. In this phase of life, successful cognitive aging is critical for mastering everyday activities and living independently, as well as for general health (Tucker-Drob, 2011). However, the later stages in life cannot be adequately addressed without understanding what happens earlier in life. It has been shown that early interventions are the most effective (Heckman, 2006, 2008). Thus, in order to develop interventions that promote successful cognitive aging, it is imperative to gain empirical knowledge about determinants that affect cognitive development from childhood into adulthood and old age. For instance, many theoretical frameworks propose learning opportunities provided by formal education in school to be such a critical determinant (Adey, Csapo, Demetriou, Hautamäki, & Shayer, 2007; Cattell, 1987; Ceci, 1991; Neisser et al., 1996; Snow, 1996). However, the dynamics between cognitive abilities and educational attainment are complex and the long-term impact of educational attainment on the development of adult cognitive abilities, above and beyond the impact of childhood cognitive abilities, has seldom been researched (Ferrer & McArdle, 2004; Gustafsson & Undheim, 1992). Therefore, the present study aims to investigate the long-term consequences of different educational pathways on the development of cognitive abilities by simultaneously taking into account aspects of quantity and quality of formal education in school (i.e. schooling).

3.1.1 Cognitive Abilities, Plasticity, and Formal Education

Intelligence can be conceptually defined as the *"ability to understand complex ideas, to adapt effectively to the environment, to learn from experience, to engage in various forms of reasoning, to overcome obstacles by taking thought"* (Neisser et al., 1996, p. 77). In the study of intelligence, most current psychological research is based on the psychometric approach (Neisser et al., 1996). This approach states that intelligence can be measured well by tests and individual differences observed on these tests can be represented well by so-called structural models (Gottfredson & Saklofske, 2009). In developmental research, two-component theories such as the extended Gf-Gc theory (J. L. Horn & Noll, 1997) have prevailed (see Lindenberger, 2001, on two-component theories). The Gf-Gc theory distinguishes two poles of broad cognitive abilities: crystallized intelligence (Gc) and fluid intelligence (Gf; Cattell, 1987; Nisbett et al., 2012). Gc can be defined as *"the individual's store of knowledge about the nature of the world and learned operations such as arithmetic ones which can be drawn on in solving problems"* (Nisbett et al., 2012, p. 131). Moreover, Gc comprises several aspects represented by typical indicators for Gc, which include knowledge of the world, word knowledge or lexical knowledge, reading comprehension, and general (verbal) information, among others (Adey et al., 2007; McGrew, 2009; Nisbett et al., 2012). However, it should be noted that there has been some debate on the interpretation of Gc (Kan et al., 2011; Schipolowski et al., 2014). While researchers agree that Gc represents the influence of acculturation and learning opportunities, theories differ with respect to the role of verbal ability. While Cattell (1971, 1987) defined Gc as knowledge in all kinds of content domains, Carroll (1993) emphasized language as defining Gc. Schipolowski and colleagues (2014) showed that both factual knowledge and verbal ability reflect empirically distinct constructs and facets of Gc. It is therefore reasonable to distinguish verbal ability from factual knowledge, as we will do in the current study.

Gf on the other hand is *"the ability to solve novel problems that depend relatively little on stored knowledge as well as the ability to learn"* (Nisbett et al., 2012, p. 132). Gf comprises mental operations such as *"drawing inferences, concept formation, classification, generating and testing hypothesis, identifying relations, comprehending implications, problem solving, extrapolating, and transforming information."* (McGrew, 2009, p. 5). Typical indicators of Gf include inductive and deductive reasoning (McGrew, 2009). Importantly, the distinction between these two cognitive abilities is also strongly supported by research on the development of intelligence showing differential trajectories of the two abilities over the lifespan (Lindenberger, 2001; Li et al., 2004; McArdle et al., 2002), as well as high differential stability of both abilities over several decades of an individual's lifetime (Schalke et al., 2012). Thus, the differentiation of Gc and Gf is of some importance as they are affected by different influences and have been shown to develop differently over the lifespan.

Lifespan cognitive psychology (Baltes et al., 1999; Li & Baltes, 2006) provides the following explanation as to why Gc and Gf vary in their developmental trajectories over the lifespan. The theory proposes that development is a reciprocal process between the individual and its developmental context. Further, the theory postulates three kinds of influences on cognitive development: (a) species typical neurobiological and cultural evolutionary processes, (b) socialization-typical environmental influences (e.g. formal education), and (c) idiosyncratic non-normative person-specific experiences as a result of self-selection into different (professional) environments (for a more detailed discussion see Li & Baltes, 2006). In very young age, development is much more normative than later in life. Very young children mainly depend on their neurobiological processes, but at the time they enter school, the shared environment largely affects these processes. Formal education in schools is probably the most important socialization-typical environmental influence on cognitive development in Luxembourg and affects all school children in a comparable way. However, as people age, all kinds of environmental influences, especially non-normative person-specific experiences, accumulate (Li et al., 2004; Li & Baltes, 2006). Importantly, after individuals leave formal education, these person-specific (professional) influences will become very important to the development of cognitive abilities. Thus, after formal education ends, a process of individualization starts and people become increasingly heterogeneous with respect to their cognitive abilities. However, the sensitivity for biological and cultural influences varies greatly over the lifespan. Culture comprises several environmental influences in the lifespan developmental framework such as medicine and education. While the need for culture increases with aging, the efficacy of culture and biological plasticity decrease with aging. Thus, formal education may play a crucial role as it affects individuals, when their biological plasticity is still high and when they are still very sensitive to cultural influences. Later cultural experiences may also be less efficient. In addition, it is often argued that Gf and Gc may be differently affected by the environment and by biological influences. While Gf is more strongly based on neurobiological processes and biological aging effects, Gc is to a larger extend influenced by the environment, including formal education (Baltes et al., 1999; Cattell, 1987; Li et al., 2004; Li & Baltes, 2006). Thus, Gf is supposed to be less culture sensitive than Gc and therefore also less affected by environmental influences such as formal education and later individualization processes. However, the relative plasticity of Gf and Gc has not yet been sufficiently studied. We therefore focus on indicators of both Gf and Gc in the current study and investigate the influence of formal education on Gc and Gf separately.

3.1.2 The Reciprocal Relationship between Cognitive Abilities and Formal Education

In many theoretical frameworks, formal education in school is a critical determinant of cognitive development from childhood into adulthood and old age (Baltes et al., 1999; Cattell, 1987; Ceci, 1991; Li et al., 2004; Li & Baltes, 2006; Neisser et al., 1996; Snow, 1996). However, the relationship between cognitive abilities and formal education is complex and has also been proposed to be reciprocal (Brody, 1997; Li & Baltes, 2006; Neisser et al., 1996; Nisbett et al., 2012). This means that cognitive abilities affect and predict formal education, but formal education also affects cognitive abilities. Figure 4 shows a schematic representation of the various possible relationships between cognitive abilities and formal education for the statistical analyses in the current study. Three cognitive abilities are included in the current study and will be analyzed separately. These are Gf and two aspects of Gc, namely word knowledge (WK), a verbal ability measure, and knowledge of the world (KW), a factual knowledge measure (see also Schipolowski et al., 2014).

Four different types of relationships can be distinguished in the schematic model. The main focus of the current study is how formal education or schooling affects the development of cognitive abilities. Crucially, two paths of influence of education on cognitive abilities must be distinguished: (a) quantity (see Paths 4 in Figure 4) and (b) quality of formal education in school (i.e. the distinction between the non-academic and academic track in Figure 4, Carroll, 1989; Haertel, Walberg, & Weinstein, 1983). Quantity of formal education is represented in the current study and in Figure 4 by years of formal education (YoFE). The length of schooling is important, as more time spent in formal education offers students more opportunities to enhance their knowledge and skills. Thus, quantity of schooling has an important impact on all cognitive abilities (Carroll, 1989). Quality of schooling or formal education is also crucial for acquiring a deeper and more sophisticated knowledge base, as well as other cognitive skills (Glaser, 1982). Quality of instruction involves several aspects, one of which is the tailoring of instruction to fit student characteristics such as prior knowledge or ability level (Becker, Lüdtke, Trautwein, Köller, & Baumert, 2012; Haertel et al., 1983). This also means that the same subject matter can be taught at different levels of complexity. Hence, the level of complexity taught (e.g. richness and complexity of a curriculum) is another crucial aspect of instruction (Becker et al., 2012) and one of the underlying rationales of school tracking. Thus, in the current study, we use the distinction between the non-academic and academic track as an indicator of quality of instruction.

Second, we can distinguish the influence of cognitive abilities on formal education. Historically, intelligence and education are closely connected: When Simon and Binet (1905, 1916) devised their first intelligence test, this was with the intention of measuring children's potential to succeed in school (Brody, 2000). IQ tests have since shown good predictive value for academic success, as noted by Jensen (1998, p. 277): *"If there is any unquestioned fact in applied psychometrics, it is that IQ tests have a high degree of predictive validity for many educational criteria, such as [...] school and college grades, [...] number of years of schooling, probability of entering college, and after entering, probability of receiving a bachelor's degree"*. This predictive value of intelligence for educational attainment is depicted by Paths 1 and 2 in Figure 4.

Third, certain cross-lagged effects of different cognitive abilities on each other could be expected across time, for example $Gf_{\_12}$ on $WK_{\_52}$ (i.e. Paths 3 and 5 in Figure 4, depending on the dependent variable at age 52). An effect of $Gf_{\_12}$ on $WK_{\_52}$ or $KW_{\_52}$ is proposed by Cattell's (1987) investment theory. The theory postulates that fluid abilities are invested into the acquisition of crystallized abilities when individuals take advantage of environmental learning opportunities such as formal education. This would be shown in (a) an indirect effect of $Gf_{\_12}$ on $WK_{\_52}$/$KW_{\_12}$ over years of formal education (YoFE) and (b) possibly also by a direct effect of $Gf_{\_12}$ on $WK_{\_52}$/$KW_{\_12}$ for the learning opportunities that have not been assessed in the current study. On the contrary, investment theory does not expect an effect of Gc on Gf over time. However, evidence exists that prior knowledge facilitates information processing and further acquisition of new knowledge in various stages of knowledge acquisition (Renkl, 1996). More specifically, prior knowledge helps (a) to focus attention on the really important aspects for solving a problem, (b) to encode information by relating new information to existing knowledge, (c) to form larger, meaningful composites of new pieces of information (i.e. chunks), (d) to efficiently store new information in long term memory by linking it to previous knowledge, and importantly (e) to free working memory capacity resources that can be used for further cognitive processing. Thus, prior knowledge may not only facilitate the acquisition of new knowledge, but also improve processing speed and fluid reasoning by freeing cognitive resources. Previous research has shown that expertise may also (partially) compensate for age-related losses in fluid abilities (Baltes et al., 1999). However, there is only limited evidence that crystallized abilities may affect fluid abilities (Baltes et al., 1999). In addition, the described relationships are probably only applicable to problems of a certain complexity and in order for prior knowledge to facilitate tasks that measure Gf, the knowledge must be relevant to the task. Thus, it is feasible to assume effects of Gc on Gf, but only for certain constellations. In the current study, the measure of Gc in childhood is most probably not task-relevant for the measure of Gf in adulthood and therefore no influence of Gc on Gf across time is expected in the current study.

Fourth, as shown by ample research findings, both Gf and Gc are highly stable differential personal characteristics across time (Larsen et al., 2008; Pushkar Gold et al., 1995; for an overview of other studies see Schalke et al., 2012). Thus, cognitive abilities also have strong autoregressive effects over time for the constructs assessed both at age 12 and age 52 (i.e. Paths 3 and 5, respectively for Gf and WK as dependent variables at age 52, in Figure 4). It is therefore crucial to control for the impact of prior cognitive abilities in any study of interventions that foster cognitive abilities over time. If these controls are neglected, it is likely that the effect of the intervention (i.e. formal education in the current study) will be overestimated.

In summary, lifespan developmental psychology proposes a reciprocal relationship between cognitive abilities and formal education (Li & Baltes, 2006). Thus, cognitive abilities do not only influence the length and quality of formal education, importantly, length and quality of formal education also affect cognitive abilities (Adey et al., 2007; Becker et al., 2012; Brody, 1997; Ceci, 1991; Neisser et al., 1996; Nisbett et al., 2012). The former relationship has been well studied and documented (see for example Jensen, 1998). Thus, in the following section, we will cite studies that focus only on the later relationship: the impact of formal education on cognitive abilities.

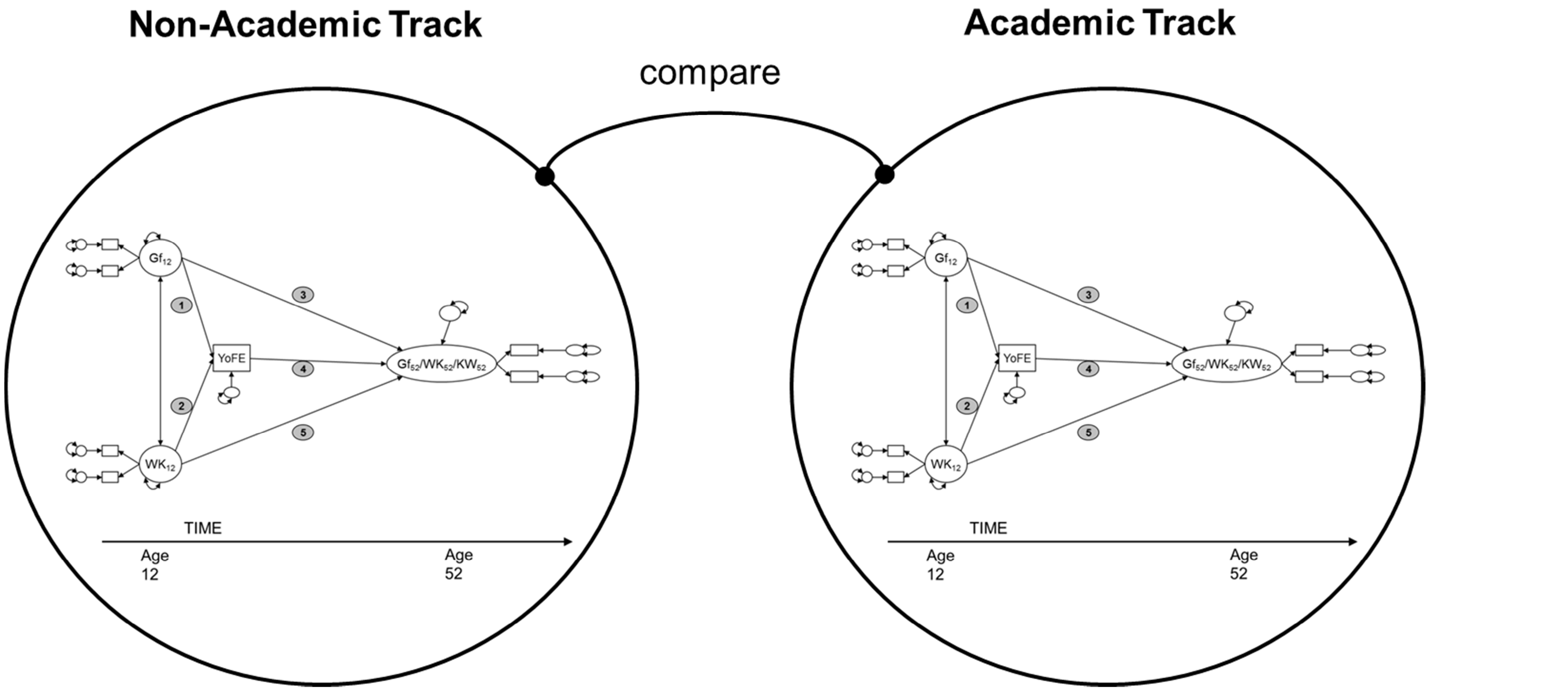

Figure 4

Schematic representation of the multiple group design and the various possible relationships between cognitive abilities and formal education for the three cognitive abilities included in the current study: (a) Fluid reasoning, (b) word knowledge, and (c) knowledge of the world. Gf = fluid reasoning, WK = word knowledge, KW = knowledge of the world, YoFE = years of formal education. The effect of school track (i.e. quality) is represented by the multiple group design, meaning that these models are estimated simultaneously for both the non-academic and academic tracked students. Squares represent manifest test scores, circles represent latent variables; one-headed asymmetrical arrows represent directional regression coefficients, whereas two-headed symmetrical arrows represent variances or covariances. Correlated uniqueness terms of the manifest indicators WK_1, WK_2, Gf_1, and Gf_2 (see text) are not shown to ensure clarity of presentation.

3.1.3 Evidence from Studies that Focused on Length of Education

Ceci (1991; Ceci & Williams, 1997) has gathered a large body of empirical evidence from natural experiments (e.g. summer vacation, effects of intermittent school attendance, delayed onset of schooling, or early termination of schooling) that all support the proposition that longer schooling enhances cognitive abilities. A difficult problem in the study of cognitive development and length of schooling is the confounding effect of chronological age on cognitive development (Ackerman & Lohman, 2003). Children who have received one more year of formal education are usually also one year older. Thus, in the studies reported by Ceci (1991; Ceci & Williams, 1997), it remains ambiguous, whether the effect depends on the year of extra education or maturation processes of natural cognitive development. However, some studies have separated the effect of schooling on intelligence from effects of chronological age and maturation by taking advantage of the fact that children in a given school grade vary in chronological age by about one year depending on the cut-off birthdate for school entrance (see Cahan & Cohen, 1989; Stelzl, Merz, Ehlers, & Remer, 1995 for a more detailed description of the procedures). Cahan and Cohen (1989) report effect sizes for one year of schooling independent of age on cognitive development that range between d = .11 and d = .50, depending on the cognitive subtest under study. Importantly, in their study, the effects of schooling were higher than those of age for all verbal (6) and numerical (1) subtests. For the 5 figural tests reported in their study, the effect was higher for schooling than aging for 2 subtests (including the matrices subtest), lower for schooling than for aging for 2 subtests and equal for 1 subtest. Stelzl and colleagues (1995) undertook a study of 10-year-old children with a similar design and found that effect sizes of one year of primary school attendance on cognitive abilities ranged between d = .24 and d = .80, depending on the cognitive subtest. After attenuation for measurement error, the effects on fluid intelligence did not differ significantly from the effects on crystallized intelligence.

3.1.4 Evidence from Studies that Focused on Quality of Education

Whereas Ceci (1991) reports little support for the effect of quality of schooling on cognitive development, Nisbett and colleagues (2012) give an overview on different studies showing that cognitive performance is impacted by various aspects of quality of education such as teachers' experience, quality of teaching, as well as specifically designed interventions.

Other studies have used school tracking (i.e. explicit ability grouping) to study the impact of quality of schooling, as represented by the level of complexity taught (e.g. richness and complexity of a curriculum), on cognitive abilities. In a recent longitudinal study, Becker and colleagues (2012) took advantage of the tracked German school system in order to investigate the impact of quality of education on fluid intelligence. In Germany, children are placed into different school tracks (i.e. school levels) after primary school. These school tracks differ in the richness and level of educational curriculum and can thus be taken as a good indicator of quality of schooling. The longitudinal study spanned a 4-year time interval from the beginning of grade 7 to the end of grade 10. The sample contained 1,038 participants and the study controlled for several covariates (i.e. prior Gf and Gc, school grades, sex, parental socio-economic status). The authors found that students in the academic track gained about 25% more on fluid intelligence over the 4-year time span than did students in the non-academic track. Similar results were found by Härnquist (1968), by Shavit and Featherman (1988), as well as by Husén and Tuijnman (1991), who also report higher cognitive gains for the higher educational tracks in their study, even when prior cognitive ability was controlled for.

A study with a related design was undertaken in Sweden by Clifford und Gustafsson (2008). However, contrary to the study by Becker and colleagues (2012), the tracks in this study differed only in terms of specialization and not so much in terms of ability level.

Clifford und Gustafsson (2008) simultaneously researched the effects of age and length of schooling. The participants of the study were allocated into 7 different tracks, 2 tracks were vocational and 5 were academic with a different field of specialization. They had about 48,000 male participants that performed a military enlistment test on Gf, Gc, and General Visualization. The authors found effects for age and length of schooling, but the effects of age were generally lower than those of schooling. No general effect of track was reported. However, differential effects were found for the academic specialization of the track. For instance, scores on the indicator of Gc were influenced by length of schooling to a larger degree for students in the Social Science track than for students in the other tracks. However, this study did not control for prior cognitive abilities. This means that the effects of schooling for students in the different tracks cannot clearly be attributed to the academic specialization of the track, since students' allocation to a track could have been influenced by their prior cognitive abilities.

In summary, there is vast and consistent body of empirical evidence showing that both quantity and quality of schooling impact the development of cognitive abilities. However, the effects of quantity and quality are empirically difficult to distinguish, as in most educational systems longer educational pathways are commonly associated with increased quality or higher level of education (e.g. attending a university vs. immediately starting a profession). To our knowledge, only few researchers have examined the effects of both quantity and quality of formal education in a single study. Importantly, the few studies that exist show that effects of schooling are found for both Gc and Gf and also persist even when prior cognitive ability level is controlled for. However, to our knowledge no study has examined the long-term effects of formal education over several decades after the participants have left school.

3.2 The Present Study

The current study examines the complex dynamics of formal education and intelligence over a lifespan of 40 years. Crucially, it investigates the long-term consequences of formal education on both fluid and crystallized cognitive ability up to 30 years after participants left formal education. In addition, the analyses control for initial cognitive ability in childhood. Another vital feature of the present study is the examination of two distinct aspects of crystallized ability in middle adulthood: (a) word knowledge (WK), a more verbal measure, and (b) knowledge of the world (KW), a factual knowledge measure (Schipolowski et al., 2014). Four different types of relationships between formal education and cognitive abilities are tested in the current study, as depicted in Figure 4.

Moreover, we simultaneously test for effects of quantity (i.e. length) and quality of formal education by taking advantage of the tracked school system in Luxembourg. After primary school, Luxembourgian students are placed in formally different types of secondary schools. These tracks either prepare for an apprenticeship (i.e. non-academic track) or for university (i.e. academic track, OECD, 1999). The two tracks differ in terms of level or quality of schooling (i.e. their curriculum), but also in length or quantity of schooling, as the average length in the non-academic track is shorter than in the academic track. Placement into one of these tracks is mainly based on students' academic achievement in primary school. Thus, students in the current study most probably differed with respect to their initial ability before they were placed into one of the two tracks. In the present study, we examine the relationships depicted in Figure 4 for both the academic and non-academic track by simultaneously estimating the represented model for both groups.

More specifically, we hypothesize the following for the four different types of relationships as depicted in Figure 4. In line with lifespan developmental psychology, we expect that schooling and cognitive abilities reciprocally relate to each other (Li & Baltes, 2006). Schools are a socialization-typical environmental influence where all children from a certain society may train their cognitive abilities. The more challenging this context and the longer a child profits from it, the better the influence should be on the child's cognitive abilities. Thus, first, we hypothesize that quantity of formal education measured in years of formal education (see Path 4 in Figure 4) positively affects Gf and Gc in middle adulthood. Moreover, quality of formal education as measured by school track also has a positive effect on both Gf and Gc. As crystallized abilities are supposed to be more sensitive to environmental influences such as formal education, it is feasible to expect a larger effect of formal education on Gc than on Gf. In addition, due to the design of the present study, we are able to explore any interactions between quantity and quality of schooling (i.e., whether the strength of the relationship between years of education and cognitive outcomes depends on type of school track).

Second, since the conception of intelligence testing, cognitive abilities have shown high predictive power for educational outcomes (Jensen, 1998). Thus, we expect that both Gf and Gc in childhood influence length of education (i.e. Paths 1 and 2 in Figure 4). Third, as proposed by Cattell's investment theory (Cattell, 1987) fluid abilities are invested into the acquisition of crystallized abilities over the lifespan. Thus, we expect an indirect effect of Gf on WK and KW over years of formal education. In addition, to map learning opportunities other than formal education in the present study, we hypothesize that Gf may have a direct influence on the development of both aspects of Gc (i.e. WK and KW) over time (Path 3, respective for the dependent variable). However, we do not expect to find an effect of Gc on Gf in the current study. Although previous research has shown that crystallized abilities such as previous knowledge may affect fluid abilities (Renkl, 1996), this only applies to complex problems and Gc measures relevant to the task in Gf. This is not the case in the current study and thus, no effects of WK on Gf over time are expected. Fourth, both Gf and WK will have high autoregressive effects over time (see Paths 3 and 5 respectively, for the same construct over time) because of their high differential stabilities (Schalke et al., 2012), meaning that the rank order in the population does not change much with age and Gf and Gc are stable personal traits. Thus, the more intelligent children are also the more intelligent adults. This also means that a large share of the variance in the dependent variable will be explained by the differential stabilities. We cannot estimate differential stabilities for KW, as this cognitive ability was only measured in adulthood in the current study. However, as both WK and KW assess crystallized abilities, we assume that WK will have high predictive power for KW (Path 5, depending on dependent variable).

In summary, the present longitudinal study contributes to the empirical body of research on the impact of formal education on cognitive development in several ways: (a) It spans 40 years from late childhood to middle adulthood, which makes it possible to study the long-term impact of formal education on cognitive abilities. (b) In contrast to most previous results, we separate different aspects of crystallized ability and simultaneously control for prior cognitive ability. (c) Most previous research has either looked at the impact of quantity or quality of education. In the current study, we simultaneously examine both quantity of formal education measured in years of formal education and quality of education in terms of school tracking. (d) The design of the current study makes it possible to largely control for aging effects because of the homogeneous composition of the sample with respect to age.

3.3 Methods

3.3.1 Participants and Procedure:

The current longitudinal study (MAGRIP) encompasses two points of measurement - 1968 and 2008, covering a time span of 40 years. In 1968, a multistage sampling procedure was applied to all elementary schools in Luxembourg. First, half of all school classes at grade

6 were selected randomly and all students of these classes participated. Second, a representative age-based sample was drawn that included all students enrolled in the selected schools in the school year 1963-1964. These were students who attended classes spanning from Grades 4 to 6 (students in lower grades had repeated one or more classes). To control for differential effects of age on cognitive development as well as quantity of schooling, we drew from a sample that included only those students who were in grade 6 and of approximately the same age. This age and grade-based sample included 2,007 children[4] (51.0% female) who were about 12 years old (*M* = 11.7 years; *SD* = 3.8 months) at the time of testing. All children completed a comprehensive intelligence test, the "Leistungsprüfsystem" (i.e. achievement test battery, W. Horn, 1962, 1983), which was administered by trained university students in a group setting.

In 2008, 315[5] (55.9% female) of these former students, now aged around 52 years, were visited at home and interviewed on their educational history. In a second stage, the participants retook the same intelligence test that they had completed 40 years earlier. About two thirds of these persons (*n* = 212) took this test in a group setting; the remaining participants were visited at home to take the test individually. All tests were administered by trained assessors and the test-taking procedure strictly followed the standardization protocol of the test manual. Information on sample selectivity of the retested age-based and grade-based samples is depicted in Figure 10 in Annex D. Figure 10 also shows a detailed overview of the data collection stages and attrition at each stage of data collection.

---

[4] Four students were excluded because of severe multivariate outlying values in some of the intelligence subtests.

[5] For the 316 participants in the longitudinal sample, data was missing for one person on the school type variable. This person had to be excluded from the structural models, leaving a longitudinal sample of 315 participants for the current study.

3.3.2 Measures

Longitudinal Measures of Gf and WK

At age 12 and 52, all participants took a standardized intelligence test battery, named the "Leistungsprüfsystem" (L-P-S; i.e. achievement test battery W. Horn, 1962, 1983). Five of the subtests in this test battery were taken to assess fluid and crystallized ability. Each of these subtests contained 40 items and had to be completed within strict time constraints that were specified in the test manual. In this test battery, fluid intelligence (Gf) was assessed by two measures of fluid reasoning, (a) concept formation (Gf_1) and (b) number and letter series (Gf_2). In more detail, the participants had to identify, categorize, and determine rules from a complete stimulus set of (a) patterns (Gf_1) and (b) numbers and letters (Gf_2). Crystallized intelligence (Gc) was assessed by three verbal measures of word knowledge (WK). As two of the three tests contained the same kind of items, we merged the scores on these two subtests into a single composite score (i.e. WK_1). In the first word knowledge measure, participants identified the spelling error in a given noun (WK_1), and in the second measure they named a word that could be formed out of a random composition of letters (WK_2).

In 1968, the children were randomly administered one of two parallel test forms of the "Leistungsprüfsystem". Because the means and variances of subtests differed slightly across test forms, we used a linear-conversion rule (Kolan & Brennan, 1995) to equate test scores. To this end, we standardized the subscales separately for each test form to an IQ metric with a mean of 100 and a *SD* of 15 for the base sample. In 2008, participants were given the exact same test form and items that they had completed in 1968. To allow meaningful comparisons across time, subtest scores obtained from the second wave of measurement in 2008 were equated using the same conversion rules applied in 1968 (i.e. the standardization of measures in 2008 is based on means and *SDs* as obtained for the base sample in 1968).

Adulthood Measure: Knowledge of the World

In 2008, in addition to the measures of fluid reasoning and word knowledge described above, a comprehensive declarative knowledge test was applied to assess another aspect of Gc. To this end, we used a shortened version of the knowledge test, which serves as a measure of Gc in the standardized intelligence test battery named the “Intelligenz-Struktur-Test” (IST-2000R; Liepmann, Beauducel, Brocke, & Amthauer, 2001). The questions in this test were presented in a verbal (kv), a numerical (kn), and a figural (kf) format. Each format was assessed with one question on history and geography, art and culture, economy, science, mathematics, and general everyday knowledge. Thus, knowledge of the world (KW) was assessed with 18 items in total and can be broken down into 6 items for each of the three representation formats.

Formal Educational Attainment

Trained assessors interviewed the participants on their educational history and recorded the information in the form of an educational curriculum vitae (CV), where start and end year of each school visited was reported. From this educational CV, we calculated educational attainment in terms of type of school visited (i.e. school tracks as an indicator of quality of schooling) and length of schooling. These school types were coded according to the International Standard Classification of Education (ISCED; OECD, 1999; UNESCO, 2006) and applied to the Luxembourgian context as proposed by the OECD (OECD, 1999). We then grouped the different types of schools into non-academic and academic schools. Non-academic schools were defined as those schools that prepare students for an apprenticeship as well as vocational training itself. In the sample, this comprised the ISCED codes 2C, 2B, 3C, 3B, and 4B. Academic schools were defined as those preparing for college and university as well as college or university attendance itself. In the sample, this comprised the ISCED codes 3A, 5B and 5A. Participants were allocated to either the non-academic group (n = 184) or the academic group (*n* = 131), depending on which school type they had attended for most of their educational career. This procedure was chosen in order to take into account that some students changed from one track to another. We decided that the school type students had attended for most of their educational career was best suited as an indicator of quality of schooling.

Quantity of formal education was assessed by years of formal education after primary school. In addition, number of retained school years was assessed by self-report of the participants and these repeated school years were not considered in computing overall length of education. The overall mean of length of schooling was 6 years and 9 months (*SD* = 3 years and 5 months). Participants in the non-academic track had a mean length of schooling of 5 years and 1 month (*SD* = 2 years and 1 month) and participants in the academic track had a mean of 9 years and 1 month (*SD* = 3 years and 6 months). This shows that participants in the non-academic track left formal education on average at the age of 17 years, while participants in the academic track left formal education on average at the age of 21 years.

3.2.3 Statistical Analysis

Structural equation modeling was applied to the data using Mplus version 6.0 (Mplus 6; Muthén & Muthén, 1998-2010). This allowed us to control for measurement error of the observed subtest scores. Some of the subtest scores were approximately but not strictly normally distributed. We therefore used maximum likelihood estimation with robust standard errors (MLR) as implemented in Mplus.

Strategy of analyses. The analyses took place in 2 steps. First, we estimated the model depicted in Figure 4 for the entire sample. Thus, contrary to Figure 4, we did not separate the sample into school tracks. This allowed the investigation of the main effect for quantity of education. In addition, the impact of quantity of education was estimated with a bigger sample size and the variable years of formal education also has a larger range in the entire sample compared to the range for each sample in the multiple group models. We examined the 5 indicated effects in Figure 4 by fitting a separate model for each dependent variable at age 52 (i.e. Gf_{52}, WK_{52}, KW_{52}). Moreover, as recommended in the literature for longitudinal studies (Cole & Maxwell, 2003) , we introduced correlated uniqueness terms for the repeated measures in the models. Thus, we allowed residual variances of the same indicators to correlate across time to account for unique aspects of a certain measure.

Second, for the simultaneous investigation of the impact of quantity and quality of formal education, we specified the models as multiple-group models with type of school track as grouping variable. In doing so, we could study (a) the effect of quantity of education (in terms of years of education) on cognitive outcomes for each track, (b) the main effect of quality of education (in terms of type of school track), and (c) the interaction between quantity and quality of education (i.e., whether the strength of the relation between years of education and cognitive outcomes depends on type of school track). To this end, it was necessary to ensure that the measurement models for the two groups were comparable for the constructs under investigation, allowing differences between the two groups to be attributed to the different tracks. Thus, we first tested for measurement invariance between the two groups. We applied a stepwise approach as recommended in the literature (Byrne & Stewart, 2006; T. D. Little, 1997; Lubke et al., 2003; Meredith, 1993). We first tested for *configural invariance*, which requires that the pattern of zero and nonzero loadings of observed indicators on the common factors remain the same between the groups. Second, *metric invariance* was established, which requires invariant factor loadings between groups (i.e., the magnitudes of the unstandardized factor loadings have to be equal) and allows for the application of meaningful analyses of correlations and variances between the two groups (Lubke et al., 2003). Third, *scalar invariance* requires group-invariant intercepts and is needed for a meaningful comparison of means between the groups (Cheung & Rensvold, 2002). Only after scalar invariance is achieved between the two groups, can all differences between the two groups (i.e. non-academic and academic track) be interpreted substantially. Measurement invariance for corresponding measures across time is not required for cross-lagged models and thus also not for the models we specified in the current study. To summarize, first, we estimated the model depicted in Figure 4 for the entire sample without splitting the sample by academic track. Second, we established measurement invariance between the non-academic and academic group. Third, we simultaneously evaluated the impact of length of formal education for both types of school track. Fourth, we tested for differences in the multi group models between the non-academic and academic group.

Assessing model fit. Goodness of model fit was assessed by the χ^2 goodness-of-fit test as well as by several descriptive measures of fit that are recommended in the literature: the χ^2 statistic in relation to degrees of freedom, the Root Mean Square Error of Approximation (RMSEA, Steiger, 1990), the Standardized Root Mean Square Residual (SRMR, Bentler, 1995)(*SRMR*, Bentler, 1995), Akaike's Information Criterion (AIC, Akaike, 1974), and the Comparative Fit Index (CFI, Bentler, 1990). RMSEA values smaller than .05 show an acceptable fit (Hu & Bentler, 1998). SRMR values smaller than .05 indicate good fit, and values smaller than .10 point to an adequate model fit (Schermelleh-Engel, Moosbrugger, & Müller, 2003). The model with the smallest AIC value is considered to be the best fitting model (Schermelleh-Engel et al., 2003). CFI values greater than .95 suggest good model fit (Hu & Bentler, 1998, Schermelleh-Engel et al., 2003). The CFI shows the relative fit of a given model compared to a null model. An appropriate null model must be nested within the model of interest, and Widaman and Thompson (2003) have argued that an appropriate null model must be able to account for MI restrictions. Thus, the appropriate null model for the current study is a model with zero covariation among the observed variables, as well as group invariant intercepts (i.e. Model 0B in Widaman & Thompson, 2003).

Assessing measurement invariance and differences between groups. Model fit differences are most commonly assessed using the χ^2 difference test (Bollen, 1989). To compute the χ^2 difference test as derived from the MLR estimator in the current study, we calculated scaled χ^2 difference values using the procedure described in Satorra and Bentler (1999). Moreover, Cheung and Rensvold (2002) have demonstrated that the change in descriptive fit indices (i.e. the CFI) is an equally good or even better-suited model comparison criterion compared to the classical χ^2 difference test, because change in the CFI is not affected by sample size. Further, Cheung and Rensvold (2002) have shown that a change in the CFI smaller than -.01 indicates adequate fit of the model with additional MI constraints. In some cases, measurement invariance holds for some but not for all indicators; this is called partial measurement invariance (Byrne, Shavelson, & Muthén, 1989). Byrne et al. (1989, p. 458) suggested that a sufficient degree of partial invariance for meaningful comparisons of different models is established, if multiple indicators for a construct exist and at least one measure is invariant.

Handling Missing Data. Missing data was not a severe problem in our analyses. Data was missing for only one to two participants, depending on the variable. Full information maximum likelihood (FIML) estimation was therefore used to handle this missing data (Arbuckle, 1996; R. J. A. Little & Rubin, 2002; Wothke, 2000).

3.4 Results

3.4.1 Descriptive Statistics

The descriptive statistics depicted in Table 7 show differences between the non-academic (non-AT, n= 184) and academic track (AT, n = 131) in Gf and Gc in both childhood and adulthood. The academic track students scored higher than the non-academic track students on Gf, WK, and KW. Moreover, the mean difference between the tracks increased from age 12 to age 52. This effect was especially pronounced for our longitudinal measure of Gc (i.e. WK), with a mean difference between the two tracks of 10.77 IQ points at age 12 and 16.02 IQ points at age 52. The mean difference between tracks for Gf increased from 8.20 IQ points at age 12 to 9.08 IQ points at age 52. However, the increased difference between the tracks across age was not accompanied by an increase in effect size (for Gf: ES_{12} = .62 and ES_{52} = .61; for WK: ES_{12} = .84 and ES_{52} = .76), because the standard deviations for both groups also increased across age. Thus, in terms of effect size, the initial difference between both groups remained comparable from age 12 to age 52. Further, Table 7 also shows a mean increase across time in participants' test performance in Gf as well as in WK *within* each track. This increase in test performance across age showed a large effect size for Gf with $ES_{\text{non-AT}}$ = 1.18 and ES_{AT} = 1.39. The effect was even larger for WK with $ES_{\text{non-AT}}$ = 2.84 and ES_{AT} = 3.79. This shows that, on average, participants scored higher on Gf and WK at age 52 than they did at age 12. Moreover, this increase in test performance was more pronounced in the academic track group than the non-academic track group. However, we also see that the standard deviations increased over time, especially in the non-academic track. This effect shows that the non-academic track group became more heterogeneous over time than the academic track group. Importantly, this also indicates that initial differences between people became more pronounced for non-academic track students than for academic track students.

We also find differences between the non-academic and academic track in the test assessing knowledge of the world (KW), which was only assessed at age 52 (see Table 7). On average, the participants in the academic track had about one more correct answer per presentation format than the participants in the non-academic track. This difference can also be described as medium (kv & kf) or large (kn) in terms of effect size.

Table 7

Means, Standard Deviations, and Differences in Means Between School Tracks as Obtained for Observed Subtest Scores and Over Time for Longitudinal Subtest Scores

	non-Academic Track		Academic Track		Difference in mean	
	mean	**SD**	**mean**	**SD**	**across groups**	**ES**
Fluid Reasoning (Gf)						
Age 12	100.90	13.39	109.10	12.14	8.20	0.62
Age 52	118.21	14.90	127.29	12.36	9.08	0.61
Difference in mean over time (ES)	17.31 (1.18)		18.19 (1.39)			
Word Knowledge (WK)						
Age 12	97.49	13.38	108.26	11.71	10.77	0.84
Age 52	150.56	22.15	166.57	17.50	16.02	0.76
Difference in mean over time (ES)	53.07 (2.84)		58.31 (3.79)			
Knowledge of the World Subtests at Age 52						
kn	2.31	1.10	3.53	1.27	1.22	1.04
kv	3.31	1.17	4.05	1.22	0.74	0.62
kf	3.11	1.22	3.75	1.15	0.64	0.54

Note. N = 315, $n_{\text{non-academic track}}$ = 184, $n_{\text{academic track}}$ = 131. *ES* = effect size for mean differences Cohen's *d* with pooled variance. Gf = fluid reasoning. WK = word knowledge. kn = numerical subtest. kv = verbal subtest. kf = figural subtest

3.4.2 Measurement Invariance Between Groups

To study the invariance of subtest scores between the non-academic and academic group, we examined a series of increasingly constrained models. A report of the model fit indices is summarized in Table 8. The key results of these analyses can be summarized as follows. We could establish full configural, full metric, and full scalar invariance for the models with Gf and WK as the dependent variable at age 52. For KW as the dependent variable at age 52, we could establish full configural and full metric invariance. Full scalar invariance did not hold, so we set the intercepts of the numerical knowledge subtest free across school tracks in order to obtain partial scalar invariance. In summary, in the current study, the full and partial scalar invariant models for the non-academic and academic track groups indicate that the two groups do not differ much in terms of their measurement properties for all three dependent variables at age 52. Hence, all differences between the non-academic track and academic track in the remaining analyses can be interpreted as differences between the two tracks. For the model estimates of the entire sample, where we did not examine the tracks separately, no measurement invariance is necessary.

Table 8

Evaluation of Model Fit to Study Measurement Invariance and the Impact of Different Variables of Cognitive Abilities for the Non-Academic and Academic School Tracks

Model	Base	Additional Equality Constraint	$\chi2$	df	p	CFI	RMSEA	SRMR	AIC	Differences (Δ)							
										compare	$\chi2$	df	p of $\Delta\chi2$	CFI	RMSEA	SRMR	AIC
Group-Model: Fluid Reasoning (Gf) at age 52 as Dependent Variable																	
Gf1		configural invariance	12.37	14	0.58	1.000	0.000	0.022	16279.92								
Gf2	Gf1	metric invariance	20.38	17	0.26	0.995	0.036	0.089	16281.84	Gf2 vs. Gf1	7.13	3	0.068	-0.005	0.036	0.067	1.92
Gf3	Gf2	scalar invariance	22.07	20	0.34	0.997	0.026	0.093	16277.44	Gf3 vs. Gf2	1.71	3	0.635	0.002	-0.010	0.004	-4.39
Gf4	Gf3	Gf12 & Gc12 on YoFE	24.07	22	0.34	0.997	0.024	0.098	16275.20	Gf4 vs. Gf3	1.99	2	0.370	0.000	-0.002	0.005	-2.24
Gf5	Gf4	Gf12 & Gc12 on Gf52	24.60	24	0.43	0.999	0.013	0.095	16272.01	Gf5 vs. Gf4	0.74	2	0.691	0.002	-0.011	-0.003	-3.19
Gf6	Gf5	YoFE on Gf52	30.84	25	0.19	0.991	0.039	0.099	16276.12	Gf6 vs. Gf5	5.36	1	0.021	-0.008	0.026	0.004	4.11
Gf7	Gf5	intercept of Gf52	26.73	25	0.37	0.997	0.021	0.096	16272.19	Gf7 vs. Gf5	1.96	1	0.162	-0.002	0.008	0.001	0.18
										Gf7 vs. Gf3	4.71	5	0.452	0.000	-0.005	0.003	-5.25
Group-Model: Word Knowledge (WK) at age 52 as Dependent Variable																	
WK1		configural invariance	17.09	14	0.25	0.995	0.037	0.026	16773.37								
WK2	WK1	metric invariance	23.56	17	0.13	0.990	0.050	0.053	16773.47	WK2 vs. WK1	6.36	3	0.095	-0.005	0.013	0.027	0.10
WK3	WK2	scalar invariance	28.04	20	0.11	0.988	0.051	0.060	16772.26	WK3 vs. WK2	4.44	3	0.218	-0.002	0.001	0.007	-1.21
WK4	WK3	Gf12 & Gc12 on YoFE	30.36	22	0.11	0.987	0.049	0.066	16770.39	WK4 vs. WK3	2.30	2	0.317	0.000	-0.002	0.006	-1.88
WK5	WK4	Gf12 & Gc12 on WK52	30.83	24	0.16	0.990	0.043	0.067	16767.05	WK5 vs. WK4	0.65	2	0.723	0.002	-0.006	0.001	-3.33
WK6	WK5	YoFE on WK52	34.91	25	0.09	0.985	0.050	0.067	16768.73	WK6 vs. WK5	4.58	1	0.032	-0.005	0.007	0.000	1.67
WK7	WK5	intercept of WK52	33.51	25	0.12	0.987	0.046	0.068	16767.57	WK7 vs. WK5	2.71	1	0.100	-0.003	0.003	0.001	0.52
										WK7 vs. WK3	5.49	5	0.359	-0.001	-0.005	0.008	-4.69
Group-Model: Knowledge of the World (KW) at age 52 as Dependent Variable																	
KW1		configural invariance	29.92	30	0.47	1.000	0.000	0.036	14265.39								
KW2	KW1	metric invariance	37.32	34	0.32	0.995	0.025	0.053	14264.34	KW2 vs. KW1	7.72	4	0.102	-0.005	0.025	0.017	-1.05
KW3	KW2	scalar invariance	55.02	38	0.04	0.973	0.053	0.065	14274.38	KW3 vs. KW2	16.47	4	0.002	-0.022	0.028	0.012	10.04
KW4	KW2	partial scalar invariance	41.11	37	0.30	0.993	0.027	0.058	14262.51	KW4 vs. KW2	3.73	3	0.292	-0.001	0.002	0.005	-1.83
KW5	KW4	Gf12 & Gc12 on YoFE	43.07	39	0.30	0.993	0.026	0.062	14260.41	KW5 vs. KW4	1.96	2	0.375	0.000	-0.001	0.004	-2.11
KW6	KW5	Gf12 & Gc12 on KW52	46.59	41	0.25	0.991	0.029	0.065	14259.60	KW6 vs. KW5	3.73	2	0.155	-0.002	0.003	0.003	-0.81
KW7	KW6	YoFE on KW52	46.82	42	0.28	0.992	0.027	0.065	14257.62	KW7 vs. KW6	0.05	1	0.823	0.001	-0.002	0.000	-1.98
KW8	KW4	intercept of KW52	49.43	43	0.23	0.990	0.031	0.068	14257.94	KW8 vs. KW7	3.01	1	0.083	-0.003	0.004	0.003	0.32
										KW8 vs. KW4	8.52	6	0.359	-0.004	0.004	0.010	-4.57

Note. Constraints that proved feasible were kept in the next model test, so that each model is nested within the previous one. The Gf = fluid reasoning; WK = word knowledge; KW = knowledge of the world; df = degrees of freedom; CFI = comparative fit index; RMSEA = root mean square error of approximation; SRMR = standardized root mean square residual; AIC = Akaike's Information Criterion. For computing the χ^2 difference test as derived from the MLR estimator we calculated scaled χ^2 difference values using the procedure described in Satorra and Bentler (1999). In model KW4, the intercepts of the numerical knowledge subtest were set free across school tracks to obtain partial scalar invariance

3.4.3 Model Testing Results

In a first step, we examined the main effect of quantity of formal education. We therefore estimated the model in Figure 4 for the entire sample. Thus, contrary as to the representation in Figure 4, we did not split the sample into the different tracks in this first step of the analyses. We estimated a separate model for each dependent variable at age 52 (i.e. Gf, WK, and KW). Model fit for Gf as dependent variable at age 52 was excellent (χ^2 = 4.99, *df* = 7, *p* = .66, CFI = 1.00, RMSEA = .000, SRMR = .014). Model fit for WK as dependent variable at age 52 was good (χ^2 = 14.18, *df* = 7, *p* = .05, CFI = .99, RMSEA = .057, SRMR = .022). Model fit for KW as dependent variable at age 52 was also excellent (χ^2 = 11.59, *df* = 15, *p* = .70, CFI = 1.00, RMSEA = .000, SRMR = .022).

In a second step, we examined the effect of quality of formal education in addition to the effect of quantity of formal education. Therefore, we simultaneously estimated the model shown in Figure 4 for the non-academic and the academic tracks. Figure 5 shows all estimated model parameters for Gf (Figure 5a), WK (Figure 5b), and KW (Figure 5c) as dependent variable, respectively. The path coefficients in each figure show the unstandardized model parameters and, in parenthesis, the standardized model parameter for all participants, the non-academic, and the academic track, respectively. The model estimates for the group models are taken from the scalar invariant group models for Gf (i.e. Model Gf3) and WK (i.e. Model WK3), and the partial scalar invariant model for KW (i.e. Model KW4, see Table 8).

Main Effect of Quantity of Formal Education on Cognitive Abilities (i.e. Paths 4 in Figure 4 for the Entire Sample without a Split by Track)

As shown in Figure 5a for Gf as the dependent variable at age 52, years of formal education had a small significant effect on Gf at age 52 with an standardized path coefficient of .12 (p < .05). Figure 5b shows the results for WK as the dependent variable at age 52.

Years of formal education had a standardized path coefficient of .07 (ns) on WK_{52}. This small effect fails to reach significance for the entire sample. Figure 5c shows the results for KW_{52} as a dependent variable. Years of formal education had a very large and significant effect on knowledge of the world at age 52 with a standardized path coefficient of .56 ($p < .05$). However, for KW at age 52, no autoregressive effects over time could be controlled for as we did not measure KW in childhood. Thus, the influence of years of education could be overestimated.

Main Effect of Quality of Schooling as Assessed by the Type of School Track (see Group Model Comparison in Figure 4)

In a second step, we simultaneously estimated the group models. To test for significant differences in model parameters between the two tracks, we set model parameters equal between the non-academic and academic track in a stepwise procedure. The model fits for these tests are shown in Table 8. The impact of school track on cognitive abilities at age 52 can be seen in the difference of model intercepts between tracks. The model intercept for the non-academic group was fixed at zero and the intercept for the academic group was freely estimated. Any difference in the intercept indicates a difference in cognitive abilities at age 52 that can be accounted for by school track. For Gf as the dependent variable at age 52, we found an unstandardized difference in the intercept of 5.63 IQ points on Gf_{52} with a 95% confidence interval of [-1.84; 13.10] (taken from model Gf3). For WK as the dependent variable at age 52, we found an unstandardized difference in the intercept of 9.90 IQ points on WK_{52} with a 95% confidence interval of [-2.14; 21.93] (taken from model WK3). Importantly, for the models of Gf and WK at age 52, we were able to establish full configural, full metric, and full scalar invariance so that the intercept differences can clearly be attributed to the different tracks. Thus, in the models for both Gf and WK at age 52, we observe a higher increase in these two cognitive abilities for the academic track participants than for the non-academic track participants. However, this difference fails to reach significance as indicated by the 95% confidence intervals and also shown in Table 8 (see test for Models Gf7 and WK7).

For KW as the dependent variable at age 52, we found an unstandardized difference in the intercept of -0.15 correct answers in the questionnaire on knowledge of the world with a 95% confidence interval of [-0.59; 0.29] (taken from model KW4). This rather small negative difference would indicate that the non-academic track had a more positive effect on participants' knowledge of the world than the academic track education. However, as indicated by the 95% confidence interval and shown in Table 8 (see test for Model KW8), the differences in intercepts between the groups was not significant. In addition, for KW at age 52 as the dependent variable, we could establish full configural and full metric invariance, but only partial scalar invariance. This shows that the subtest kn is affected by academic track. But the lack of full scalar invariance also makes the interpretation of the difference in the intercept less clear. Taken together, we did not find a significant long-term main effect of academic track for the dependent cognitive abilities at age 52 in the current study.

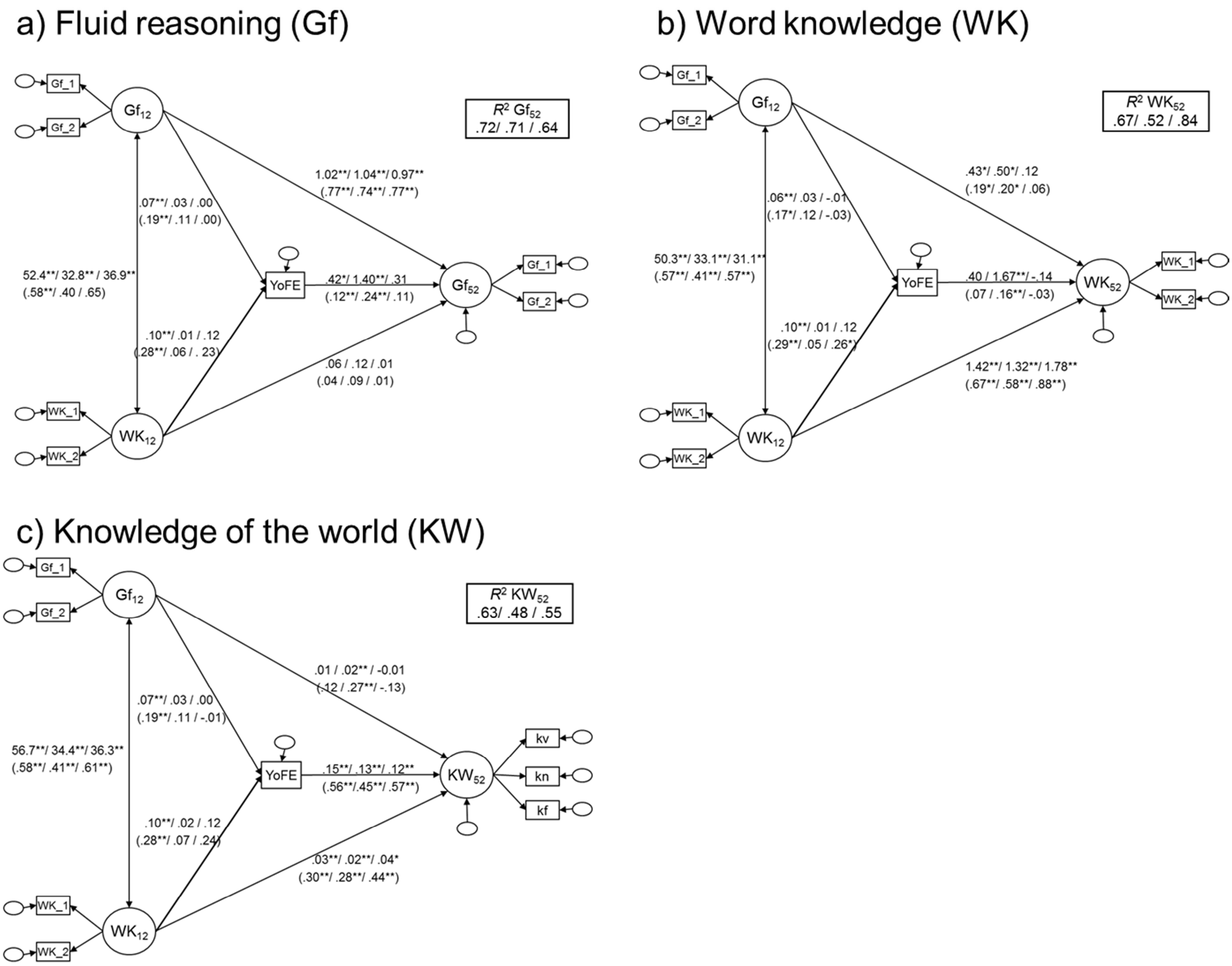

Figure 5

Unstandardized path coefficients, covariances, and R^2 for all participants, the non-academic track, and the academic track, respectively. Numbers in parentheses show standardized values for all participants, the non-academic track, and the academic track, respectively. These models are based on the full scalar and partial scalar invariant group models for the three dependent variables in the current study: (a) fluid reasoning, (b) word knowledge, and (c) knowledge of the world. Gf = fluid reasoning, WK = word knowledge, KW = knowledge of the world, YoFE = years of formal education, _1 and _2 refer to manifest variables 1 and 2 that measure the respective cognitive ability. Asterisks represent significance level. Squares represent manifest test scores, circles represent latent variables; one-headed asymmetrical arrows represent directional regression coefficients, whereas two-headed symmetrical arrows represent variances or covariances. Correlated uniqueness terms of the manifest indicators WK_1, WK_2, Gf_1, and Gf_2 (see text) are not shown to ensure clarity of presentation.
** $p < .05$; * $p < .10$.

Interaction Effect of Quantity and Quality of Formal Education on Cognitive Abilities by Track (i.e. Paths 4 in Figure 4)

While simultaneously estimating the group models, we were able to test for interactions between quantity and quality of formal education. Thus, to test for significant differences in model parameters between the two tracks, we set model parameters equal between the non-academic and academic track in a stepwise procedure. The model fits for these tests are shown in Table 8. For the two longitudinal dependent variables measured at age 12 and age 52 (i.e. Gf in Figure 5a and WK in Figure 5b), we found that years of formal education had a significant effect on Gf and WK in middle adulthood for the participants who visited the non-academic track only. These effects were of small to medium effect size ($.24_{\text{non-AT}}$, $p < .05$, for Gf52 as the dependent variable and $.16_{\text{non-AT}}$, $p < .05$, for WK52 as the dependent variable; Cohen, 1988, 1992). In contrast, for participants who visited the academic track, Gf and WK at middle adulthood were not significantly affected by years of formal education. This difference between the two tracks also proved to be significant, as shown by the test of Models Gf6 and WK6 in Table 8. Thus, in the current study, years of formal education had a long lasting positive effect on the cognitive development of students attending the non-academic track, but not of those students attending the academic track. In addition, no such interaction effect was found for KW as dependent variable at age 52. Years of formal education had a significant and large effect on KW at age 52 for both students in the non-academic and the academic track ($.45_{\text{non-AT}}$ and $.57_{\text{AT}}$, see Figure 5c). The size of the effects does not differ significantly between the two tracks (see test for Model KW7 in Table 8). Unfortunately, we could not account for the influence of KW at age 12. Thus, it is possible that the effect of length of education may be overestimated, although we were able to account for the influence of other cognitive abilities in childhood, i.e. WK and Gf at age 12.

To summarize, we found an interaction between length of formal education and school track for Gf and WK as dependent variables. Length of formal education had a significant small to medium sized effect on both Gf and WK at age 52 only for those participants that had visited the non-academic track. For KW at age 52, length of formal education had a large significant effect for both tracks. This effect may however be an overestimation due to the lack of measuring autoregressive effects of KW over time.

Cattell's Investment Theory (i.e. Paths 3 and 5 respective of the Dependent Variable in Figure 4)

Cattell's investment theory (Cattell, 1971, 1987) predicts that Gf is invested into the acquisition of Gc by taking advantage of environmental learning opportunities. Thus, the predictions of investment theory can either be seen by an indirect effect of Gf_{12} on WK_{52} and KW_{52} via years of formal education to map learning opportunities in school or a direct effect of Gf_{12} on WK_{52} and KW_{52} to map learning opportunities not assessed in the current study. We drew on the analyses for the entire sample to examine the propositions by Cattell's investment theory (i.e. the same analyses as for the main effect of length of education) as the investment theory does not make any propositions on quality of education. In the present study, no indirect effect of Gf_{12} on WK_{52} via years of formal education was found. The indirect effect is the product of the two direct effects. For the Model with WK at age 52 as the dependent variable, this would be represented by the product of path 1 and path 4 in Figure 4. The effect of years of formal education on WK_{52} is not significantly different from zero; therefore, the product would not be significantly different from zero. However, we found a marginally significant (significant on an alpha level of 10%) direct effect of Gf_{12} on WK_{52} (see Figure 5b) for the entire sample (.19, $p < .10$). In summary, we did not find an indirect effect of Gf in childhood on WK in adulthood via formal education. However, we found a marginally significant direct effect of Gf in childhood on WK in adulthood that may map learning opportunities other than those provided by formal education.

Cattell´s investment theory also applies to KW_{52} as the dependent variable. For the entire sample, we found an indirect effect of .11 for Gf_{12} via years of formal education on KW in adulthood. This finding is in line with Cattell’s investment theory. In addition, we also found an indirect effect of WK_{12} on KW_{52} via years of formal education of .16. Thus, in terms of Cattell’s reasoning, word knowledge was also invested in the acquisition of knowledge of the world during learning opportunities in school. In addition, we did not find a significant direct effect of Gf_{12} on KW_{52} for the entire sample. Taken together, for KW learning opportunities provided by formal education were of importance. Both Gf and WK in childhood were invested into the acquisition of KW in adulthood during learning opportunities provided by formal education. However, no direct effect of Gf_{12} was found on KW in adulthood, while WK_{12} did have a direct effect on KW in adulthood.

In addition, Figure 5a shows that WK_{12} did not have any direct cross-lagged effects on Gf_{52}. However, for the entire sample, we found a very small indirect effect via years of formal education for both Gf_{12} (.02) and WK_{12} (.03).

To summarize, when looking at the entire sample, Gf in childhood did have an effect on both WK and KW in adulthood as proposed by the investment theory. However, WK and KW differed concerning the mediating effect of formal education. While formal education did not play a role for WK_{52} in the current study, the effect of Gf_{12} on KW_{52} was mediated via formal education. However, we also found that WK_{12} was invested into the acquisition of both Gf_{52} and KW_{52} by taking advantage of learning opportunities provided by formal education.

Impact of Cognitive Abilities on Years of Formal Education (i.e. Paths 1 and 2 in Figure 4)

As shown in Figure 5, for the entire sample, both Gf and WK at age 12 had a significant effect on years of education, which was almost identical for all dependent cognitive abilities in the present study. Varying by the dependent variable under study, the standardized effect of Gf_{12} on years of education was of small to medium effect size and ranged between .17 ($p < .05$, for WK_{52}) and .19 ($p < .05$, for Gf_{52} and KW_{52}). The effects of WK_{12} on years of education were of medium effect size and ranged between .28 ($p < .05$, for Gf_{52} and KW_{52}) and .29 ($p < .05$, for WK_{52}). Thus, as expected, we found that cognitive abilities impact the length of formal education in the entire sample.

A little surprisingly, when we estimated the models simultaneously by school track in the multi-group model, the effect of Gf and WK at age 12 did not reach significance (as indicated by the model test of Model Gf4, WK4, and KW5 in Table 8). Thus, *within* each track, Gf_{12} and WK_{12} did not affect the length of formal education in the current study. This finding may be a result of the reduced variance in $Gf_{12,}$ WK_{12}, as well as length of schooling, when analyzing the non-academic and academic group, separately. However, explanations of this finding are discussed in more detail in the discussion.

Differential Stabilities (i.e. Paths 3 and 5 respective of the Dependent Variable in Figure 4)

Differential stabilities are shown in the autoregressive effects over time in the models that we have estimated. As expected for the two longitudinally assessed cognitive measures Gf and WK, we found large significant autoregressive effects even over the 40 years of lifespan that are captured in the current study. Figure 5a shows the autoregressive effect for Gf. For the entire sample, the standardized autoregressive effect of Gf amounts to .77 ($p < .05$). For the group model, the effect amounts to .74 ($p < .05$) for the non-academic track and .77 ($p < .05$) for the academic track. Figure 5b shows the autoregressive effects for WK. For the entire sample, the standardized autoregressive effect of WK across 40 years of lifetime amounts to .67 ($p < .05$). For the group models, the effect amounts to .58 ($p < .05$) for the non-academic track and .88 ($p < .05$) for the academic track. The difference in the path coefficients between the tracks does not reach significance, as shown in Table 8 (see test for

Model WK5). Importantly, the high autoregressive effects show that a large part of the variance in Gf_{52} and WK_{52} is explained by the autoregressive effect across time.

For KW as the dependent variable at age 52 (see Figure 5c), a childhood measure is lacking in the current study, thus we could not estimate the autoregressive effects of this cognitive ability. However, it is shown that WK at age 12 (i.e. another aspect of Gc) has a medium-sized effect on KW_{52} for the entire sample (.30, $p < .05$) as well as the non-academic track (.28, $p < .05$), and an even larger impact on the academic group (.44, $p < .05$). However, the difference between the two tracks fails to reach significance (see Table 8, test for Model KW6).

3.5 Discussion

The present study examined the complex dynamics of formal education and cognitive abilities over the lifetime. We were especially interested in the impact of quantity of formal education (as measured by years of formal education) and quality of formal education (as measured by school track) on fluid intelligence (Gf), and two aspects of crystallized intelligence (Gc), namely word knowledge (WK) and knowledge of the world (KW). We were able to control for prior cognitive ability in late childhood on Gf and WK, and the study design also enabled us to examine possible interactions between quantity and quality of formal education. Crucially, the present study spanned 40 years from late childhood at age 12 into middle adulthood at age 52. On average, participants who had visited the non-academic track left formal education at the age of 17 years and participants in the academic track at the age of 21 years. For the discussion it is important to bear in mind that when our participants' cognitive abilities were re-assessed in middle adulthood, the participants had left formal education over three decades ago. Thus, compared to many other previous studies, we investigated the long-lasting impacts of formal education on cognitive abilities.

The statistical analyses consisted of two steps. First, we analyzed the entire sample without splitting the sample by academic track. Thus, we did not simultaneously account for quantity and quality of education, but examined only the impact of quantity of formal education on the three dependent variables under study. These analyses had some statistical advantages: (a) By not splitting the sample into two groups, we had a bigger sample size and thus more statistical power. (b) The distinction into the two school tracks is based on students' academic performance in primary school. Academic performance is connected to cognitive ability. Thus, when divided by school track, a restriction of range on cognitive abilities in childhood can be expected within each school track (see Table 7). Due to the high differential stabilities, a restriction in range in adult cognitive abilities can also be expected (see Table 7). (c) In Luxembourg, the non-academic secondary school track is shorter than the academic secondary school track. In addition, academic track students have the opportunity to visit all kinds of career tracks including university and thus again prolong their formal education, while non-academic track students do not have as many options to pursue their educational career. Thus, splitting the sample by school track also results in a restriction of range in the variable years of formal education within each track (see section on Measures in the current study). In summary, the analysis of the entire sample without splitting it by school track has the advantage of higher statistical power and no restrictions of range can be expected. In addition, the analyses on the entire sample are better suited to estimate the main effect of quantity of education and also to investigate the investment theory.

Second, we estimated the multi group models by academic track to simultaneously account for the impact of quantity and quality of formal education. These analyses enable investigating the main effect of quality of formal education in terms of school track as well as any possible interactions between quantity and quality of formal education.

3.5.1 Impact of Quantity of Formal Education on Cognitive Abilities

The main effect of quantity of formal education is shown in the analyses of the entire sample. For the models of Gf and WK at age 52, we were able to control for autoregressive effects from childhood into adulthood. We observed a small effect of years of formal education for both Gf_{52} and WK_{52}. However, the effect only reached significance for Gf_{52} and failed to reach significance for WK_{52}. Although this effect is rather small for the entire sample, we see later in the interaction that the effect is significant for both Gf_{52} and WK_{52} and of small to medium effect size but only for the non-academic group. However, the finding that Gf is stronger impacted by formal education than Gc (WK) is a bit surprising. Two component theories of intelligence postulate that crystallized abilities are more sensitive to changes in the environment than fluid abilities. The finding will be discussed in more detail in the interaction section.

For knowledge of the world (KW) at age 52 as the dependent variable, we found that years of formal education had a very large and significant effect on KW_{52}. However, we were able to control for Gf_{12} and WK_{12} in childhood, but we could not control for autoregressive effects of KW over the lifespan. It could be for this reason that we see a different pattern of results for KW. However, another reason for the different outcomes for KW is that KW might be differently affected by formal education than WK. The knowledge test was constructed so that it does not favor very specific knowledge domains (Beauducel, Liepmann, Felfe, & Nettelnstroth, 2007). Thus, the test covers a broad area of different subjects and included the following subject areas: history and geography, art and culture, economy, science, mathematics, and general everyday knowledge. It is possible that much of the knowledge covered by the test is learned in formal education. This could explain why KW, as assessed in the current study, may be differently affected by years of formal education than WK.

3.5.2 Impact of Quality of Formal Education as Assessed by the Type of School Track

We did not find a significant main effect of quality of formal education as assessed by school track in the current study. However, for Gf_{52} and WK_{52}, the results in the current study pointed to a more positive effect on cognitive abilities for the academic track, as also found in previous research (Becker et al., 2012; Husén & Tuijnman, 1991; Shavit & Featherman, 1988). In the descriptive analysis, Gf increased from childhood into adulthood with an effect size of Cohen's d = 1.18 for the non-academic group and an effect size of Cohen's d = 1.39 for the academic group. Thus, the manifest increase in Gf across age differs for the two groups by about .20 of a standard deviation. However, the descriptive statistics also showed no increase in effect size in the *between* group difference from age 12 to age 52. Thus, even though the academic track had a higher increase in Gf from age 12 to age 52, the difference between the groups remained stable in terms of effect size. This pattern of effects can probably be explained by the increased gap between the two poles of the ability distribution across age as shown in Study 1. The difference in Gf in the intercept in the latent group model also pointed to a more positive effect for the academic group with a difference between the non-academic and the academic group of 5.63 IQ points (taken from model Gf3). However, the difference between the tracks was not significant in the latent group model.

For WK the pattern was the same as for Gf but with larger effect sizes. In the descriptive analysis, WK increased from childhood into adulthood with an effect size of Cohen's d = 2.84 for the non-academic group and an effect size of Cohen's d = 3.79 for the academic group. Thus, the two groups differed by almost one standard deviation. However, the difference *between* groups was the same at age 12 and age 52 in terms of effect size. The difference in the intercept in the latent group model for WK also pointed to a more positive effect for the academic track with a difference between the non-academic and the academic track of 9.90 IQ points (taken from model WK3). However, the difference between the tracks was not significant. Thus, the effect for Gf may at least be small and the effect for WK may even be somewhat larger, possibly a medium effect size. However, these effects did not reach significance in the current study. One explanation for not finding a significant effect for school track (although the direction of our findings is in line with previous research) could be that the statistical power in the current study was too low for the effects to reach significance. We would have needed a larger sample size for the group models under study in order to find an effect of school track that reaches significance.

Another explanation is that in Luxembourg students are allocated to the non-academic and the academic track mainly on the basis of their academic performance in primary school. As academic performance is connected to cognitive abilities (Jensen, 1998), the tracking decision is also affected by a students' cognitive abilities. Thus, students in the non-academic track have a lower average on Gf and WK than students that attend the academic track (see Table 7). It is possible that formal education in the two tracks is tailored to the academic and cognitive needs of the attending students. This is also one of the main rationales for sorting students into different tracks. Thus, formal education in the non-academic track is adapted to best promote the academic and cognitive abilities of the students with the lower average on cognitive abilities. The same argument also accounts for the cognitively better scoring students in the academic track. Thus, both groups' gaining in cognitive abilities (see Table 7) may be due to the fact that individual students attend the school track best suited to them. Thus, it would not be possible to make a general statement that the academic track is better for everyone, but that increase in cognitive abilities is more a function of fit between students' abilities and school track. In this case, we would not find a main effect for school track, as found in the current study, even though the quality of formal education is very important to students' cognitive development. The finding in the descriptive statistics of the manifest variables that the difference *between* the groups did not increase with increasing age in terms of effect size also points into this direction. This explanation is opposed by other studies that have found a more positive effect for the higher academic track (see Becker et al., 2012; Husén & Tuijnman, 1991; Shavit & Featherman, 1988). However, the contrast to other findings may be grounded in a number of methodological differences, for example, country specific school systems, length of study, or taking into account quantity of schooling in different tracks (i.e. number of hours per day and years of schooling per track). However, much more research is necessary to investigate this possibility.

3.5.3 Interaction between Quantity and Quality of Formal Education

In the multi group models, we found an interaction between school track and the influence of years of formal education on WK_{52} and Gf_{52}, the two measures were we could control for autoregressive effects over the lifespan. Years of education had a significant effect on WK_{52} and Gf_{52} only for the non-academic track students, but not for the academic track students. The effects were of small to medium effect size and unexpectedly the effect for WK_{52} was a bit smaller than the effect for Gf_{52}. On the contrary, the cognitive abilities in Gf_{52} and WK_{52} of the students in the academic track were only impacted by the autoregressive effect over time of the cognitive ability under study.

Thus, years of formal education was significant only for the non-academic but not for the academic track for the analyses on Gf_{52} and WK_{52}. When interpreting these findings, we must keep in mind that over 30 years have passed between participants leaving formal education and the re-assessment of cognitive abilities. Lifespan developmental psychology may provide an explanation for the interaction. It is postulated that person-specific (professional) influence factors gain in importance after individuals leave formal education. Thus, after formal education ends people develop in the direction of their professions and interests. Therefore, the professional and leisure environment becomes increasingly important for the development of cognitive abilities. Schooler and colleagues (1999) have for example shown that job complexity affects cognitive abilities. In addition, individuals themselves are responsible for the level of cognitive challenge of their leisure activities. It is very feasible to assume that the former students of the academic and the non-academic differ greatly in respect to both professional and leisure environment in later life. An academic track diploma enables to obtain more intellectual occupations. Thus, academic track students are more likely to take on occupations that are of an intellectual nature. Many of the academic track students will also have interests that are of a more cultural and intellectual nature. Thus, the later professional and leisure environment of academic track students is very likely to be cognitively challenging, which in turn will train their cognitive abilities after formal education has ended. Importantly, the content and intensity of this “cognitive training” will be very individualized. Thus, even though a positive effect of formal education on cognitive abilities may have most probably existed for the academic track at the end of formal education. This effect may have been compensated or overlain by the professional and leisure environment in the 30 years after leaving formal education.

On the contrary, the non-academic track prepares students to take on occupations of a more physical and less intellectual nature. In addition, the non-academic track students are less likely to have interests of a cultural or intellectual nature. Thus, the professional and leisure environment of former non-academic track students will be less cognitively challenging than formal education. Therefore, the effect of formal education on the cognitive abilities of non-academic track students may persist even up until age 52. The effect may be very robust as formal education takes place at a time in life, when the students are still very sensitive to environmental influences (Li & Baltes, 2006).

Crucially, these findings show that formal education may indeed be an important instrument when it comes to closing gaps in cognitive abilities. The cognitive abilities of participants in the non-academic track profited from formal education over 30 years after leaving school. Thus, it is especially this subgroup, probably the lower performing subgroup in the population, which can profit from structured interventions or formal education. It is also possible that this subgroup may be much more dependent on external help and programs to keep them cognitively challenged. Future research on the development of cognitive abilities that takes into account the interactions between quantity of education and quality of education or level of cognitive abilities is needed. Future research is also needed to determine whether formal education also has an impact on the academic track students and when this influence is no longer detectable as in the current study.

In addition, we found in the present study that Gf was more sensitive to length of formal education than WK, a more crystallized variable. This finding is contrary to assumptions of lifespan cognitive psychology (Li & Baltes, 2006). It is postulated that crystallized and fluid abilities, are differentially affected by the environment and biological factors. Crystallized abilities are more sensitive to the environment, while fluid abilities are more affected by biological factors such as aging. However, the plasticity of different aspects of cognition is not sufficiently studied. The present study's findings show that Gf was more affected by quantity of formal education than WK. It is possible that we found a larger effect on Gf than on WK, because word knowledge may be more strongly affected by influences other than only formal education. Word knowledge is an ability that is not only trained in school, but, for example, whenever a person opens the paper or the internet and starts reading. Thus, it is possible that for this specific measure of Gc, person-specific influences have become more important than formal education in the 30 years between leaving formal education and data collection. On the contrary, Gf could be better trained in school. Crucially, the current findings show that the plasticity of fluid abilities to environmental factors may be largely underestimated at the moment. This point is important as it enables very many possibilities to support and maintain cognitive functioning into old age for a society, which is constantly growing older.

3.5.4 Cattell's Investment Theory

Cattell's investment theory (1987) postulates that fluid abilities are invested into the acquisition of crystallized abilities when individuals take advantage of environmental learning opportunities such as formal education. Such an effect could be shown in (a) an indirect effect of Gf_{12} on WK_{52}/KW_{52} via years of formal education and (b) a direct effect of Gf_{12} on WK_{52}/KW_{52}. The direct effect is expected as it may map learning opportunities other than those provided by formal education. For WK_{52} as the dependent variable, we did not find that Gf in childhood was invested into the acquisition of WK during formal education. However, learning opportunities other than those provided by formal education may have played a role as we did find a significant direct effect of Gf_{12} on WK_{52} of approximately small to medium effect size. This explanation is very feasible in the case of word knowledge as this ability can be trained whenever a person is faced with verbal material, for example in television, the newspaper or magazines. A different pattern was found for KW, the other facet of crystallized abilities in the present study. For KW_{52} as the dependent variable, we found significant indirect effect for both Gf_{12} and WK_{12} on KW_{52} via years of formal education, while no direct effect of Gf_{12} on KW_{52} was found. Thus, in terms of Cattell´s investment theory, both Gf_{12} and WK_{12} were invested into the acquisition of knowledge of the world at age 52. This investment mainly took place during formal education, as we do not find a significant direct influence of Gf_{12}.

The reason for the different pattern of results for WK_{52} compared to KW_{52}, may be twofold: (a) the lack of being able to measure autoregressive effects for KW may bias the results for KW so that the effect of formal education may be overestimated. (b) Knowledge of the world may be differently affected by formal education than word knowledge. Former research has shown that factual knowledge and verbal ability are two distinct aspects of Gc (Schipolowski et al., 2014), thus they could also differently relate to formal education. Factual knowledge may depend more on formal education as school is an important place for the acquisition of general knowledge. Verbal ability on the contrary is surely also trained it school, but also easily trained in everyday life when faced with verbal material. This explanation applies especially to the measures of verbal ability and factual knowledge in the present study. The measure of factual knowledge (KW) related very closely to knowledge obtained in formal education and the ability for verbal measure (WK) is very likely to be trained in everyday life.

Further, as expected by investment theory, we did not find a direct impact of crystallized abilities on fluid abilities across the lifetime (i.e. WK_{12} on $Gf_{52)}$. However, we did find an indirect effect of WK in childhood on Gf in adulthood via formal education. Thus, word knowledge may have been invested into the acquisition of Gf in adulthood via learning opportunities provided by formal education. It is feasible to expect certain positive effects of previous knowledge on fluid ability (Renkl, 1996). But the knowledge needs to show some task-relevance for the task of Gf (Li & Baltes, 2006). We did not expect this for the current study. However, as we found a small indirect effect of WK_{12} on Gf in adulthood, WK_{12} may have had some task-relevance for the measure of Gf. In addition, this finding may also point to a reciprocal relationship between fluid and crystallized abilities across the lifespan.

To summarize, the predictions of investment theory that fluid abilities are invested into the acquisition of crystallized abilities are supported by the data in the present study. Depending on the crystallized ability under study, the investment of fluid ability into crystallized ability took place during formal education or must have probably taken place during other learning opportunities. However, the effects were not of large effect size and partially only marginally significant. One explanation could be that the time interval of the current study may have been too long to test the postulations of investment theory in detail. Investment theory describes the process of how fluid abilities affect the acquisition of crystallized abilities. Thus, more points of measurement with shorter intervals would be necessary to test the assumptions of investment theory in more detail and in order to find larger effect sizes. In the present study, we had a time interval of 40 years between measurements. However, despite this long time interval we still found effects to support the proposition that fluid abilities are invested into the acquisition of crystallized abilities by taking advantage of learning opportunities. Thus, these effects can be considered as extremely robust.

3.5.5 Impact of Cognitive Abilities on Quantity of Formal Education

Intelligence tests measure the potential of children to succeed in school (Brody, 1997). Thus, in line with ample previous research (Jensen, 1998), both Gf and WK in childhood had a significant small to medium sized impact on years of formal education for the entire sample in the current study. However, this effect was not found in the separate analyses for the non-academic and academic tracks. This finding may be a result of the restriction of range of childhood cognitive abilities as well as the length of formal education within each track as described earlier in the discussion. However, it could also show an indirect effect of intelligence in childhood on the tracking decision. The more intelligent students are allocated to the academic track, while the not so intelligent students are allocated to the non-academic track. Thus, when analyzing the models by track, the effect of intelligence on educational attainment may already be indirectly included in the tracking variable.

3.5.6 Differential Stabilities

Differential stabilities are indicated in autoregressive effects of the same variables in childhood and adulthood across time. However, differential stabilities are usually the correlation of the same variables across time. In contrast, the autoregressive effects in the current study show the differential stabilities in the context of the remaining variables in the structural equation model under examination. We found high autoregressive effects for the cognitive abilities that were assessed both at age 12 and at age 52. These high autoregressive effects for Gf and WK were found for both the non-academic and academic track, showing that the constructs have high differential stability across time even, when other important impact factors are controlled for. Similar effects have been found in previous research (Deary et al., 2000). In addition, even though we did not have an equivalent longitudinal measure for KW in childhood, WK – the other aspect of Gc assessed in childhood – had a medium to large effect on KW in adulthood. This finding supports the close relationship between the two aspects of Gc as assessed in the current study.

3.6 Strength and Limitations

In the current study, we examined the long-term consequences of formal education on cognitive development across 40 years of participants' lifespans in a longitudinal sample in which most of the participants had left formal education over 30 years ago. The fact that we found an impact of formal education over such a long time span shows the robustness of the effect and great importance of formal education on cognitive development. Further, our research design allowed us to effectively address one major validity threat that longitudinal designs usually suffer from: The time span of 40 years between the two measurement occasions rendered retest effects almost impossible (Tucker-Drob & Salthouse, 2011). This made it possible to effectively control for childhood cognitive abilities (i.e. autoregressive effects) for almost all of the cognitive abilities under study. In addition, our study design enabled us to address two important indicators of formal education: quantity and quality (i.e. school track). Importantly, in contrast to most previous research, we were able to study possible interactions between these two indicators.

Despite these strengths, our study was subject to several limitations, which should be borne in mind when interpreting the present findings and addressed in future research. First, our only re-test measure of cognitive abilities was taken over 30 years after most participants had left formal education. One more measure taken shortly after the participants had left formal education would have been very helpful for two reasons (a) with only two points of measurement, it is impossible to provide a comprehensive picture of the course of cognitive development (e.g., growth-curve modeling of individuals' cognitive development), and (b) we could have assessed the impact of formal education much better for both tracks. This would have enabled us to see whether length of formal education also affected the academic track immediately after leaving formal education.

Second, we could not directly tackle one key problems of longitudinal designs – selective attrition (Tucker-Drob & Salthouse, 2011). Notably, the present longitudinal sample reflected important characteristics of the representative base sample of 12 year old students fairly well, as it was only slightly positively selected in terms of several childhood characteristics including cognitive abilities, parental socioeconomic status, grade point average, gender, and migration background (see Figure 10 in the Annex D). Nevertheless, we cannot rule out that the statistical estimates of individuals' cognitive development are distorted because of the selective attrition of study participants. For example, aging related deficits may be underestimated (Tucker-Drob & Salthouse, 2011) because participants with lower cognitive abilities were more likely to drop-out of the current study. This is a typical problem of most longitudinal studies (Tucker-Drob & Salthouse, 2011) and can be a result of several reasons, such as (a) an association of lower cognitive abilities and death or illness (Deary, 2010), or (b) disinterest of lower-functioning participants as they might have less confidence in their cognitive abilities or in the test.

Third, we lacked a longitudinal measure for KW. Thus, the different patterns of results for KW and WK could not be clearly interpreted. Nevertheless, the results of the current study show the importance of testing different aspects of Gc, as they may also differently relate to other important constructs such as formal education.

Fourth, it is possible that a larger sample size would have been required to find significant differences on the main effect of quality of formal education as assessed by academic track. In the current study, we had an overall sample size of $N = 315$, of which 184 students were in the non-academic track and 131 students were in the academic track. Given the latent multi group models that were specified, it could be possible that the difference in main effect between the non-academic and the academic track would have become significant, if we had had more statistical power (i.e. a larger sample size).

3.7 Conclusion

The present study examined the complex relationships between cognitive abilities and formal education over a time span of 40 years from late childhood into middle adulthood. More specifically, we were interested in the long-term consequences of educational pathways in terms of quantity of formal education (i.e. years of formal education) and quality of formal education (i.e. non-academic track vs. academic track) for the development of cognitive abilities for over 30 years after participants left school. We found a significant small to medium sized effect of quantity of formal education on Gf and WK in adulthood for the non-academic track but no effect for the academic track. Lifespan developmental psychology (Li & Baltes, 2006) provides an explanation. After individuals leave formal education, their cognitive abilities are influenced more and more by person-specific professional influences and socialization typical influences such as formal education loose importance. The non-academic track prepares students to take up occupations that are of a more physical or manual nature, whereas the academic track prepares for occupations of a more intellectual career. Hence, after formal education, the participants in the non-academic track may have had fewer opportunities than participants in the academic track to train their cognitive skills in everyday life. Therefore, the impact of formal education on cognitive abilities may be more persistent for participants in the non-academic track compared to participants in the academic track. In addition, length of formal education had a large significant effect on knowledge of the world in the current study. This effect was found for the entire sample as well as for the non-academic and academic group. Importantly, the current study shows that formal education has a very robust and enduring effect on cognitive abilities over most of a person's lifespan. In addition, we have found that the two aspects of Gc, word knowledge, a verbal measure, and knowledge of the world, a factual knowledge measure, related differently to formal education. This may underline the finding that they are distinct aspects of Gc (Schipolowski et al., 2014).

Chapter 4 – Study 3: Long-Term Consequences of Grade Retention on Educational Attainment, Adult Income, and Adult Intelligence

4.1 Theoretical Background

The European Commission has identified grade retention as a very critical issue in the wider struggle against school failure and early school leaving, which are two high priority problems on the European policy agenda. Strategies to combat school failure are at the center of discussion and the European Commission is focusing on grade retention and its impact on children with difficulties at school (European Commission, 2011, p. 3). In all educational systems, students' achievement and progress in school is assessed throughout the year and several measures are taken to help students who do not perform well. Grade retention is one of these pedagogical measures. It is the practice of making students with unsatisfactory academic achievement repeat the current grade level, before they can proceed to the next grade in school.

Grade retention is a very frequent pedagogical measure at schools in several European countries. According to the European Commission (2008), in some countries grade retention concerns up to 25% of students at least once in their school career, while in other countries hardly any students repeat a year. In Luxembourg, grade retention rates at the primary education level are over 20%, meaning that every fifth student has fallen behind by the end of primary school (European Commission, 2011, p. 35).

Grade retention is meant to help students with difficulties at school and improve their academic performance. Grade retention is also a very expensive measure. In Luxembourg's case, a 20% repetition rate in primary school means that 20% of students are kept in school one additional year during their primary school career. Thus, additional teachers, classrooms, furniture and so on must be provided and financed for an additional year for every fifth student. However, the reported outcomes of grade retention most often stand in sharp contrast to the idea with which this pedagogic measure is applied. According to a report by the European Commission (2011), some repeaters catch up, while the vast majority does not. Moreover, retained students often show lower long-term achievements than weak students who have been promoted (European Commission, 2008). Crucially, in most research studies, grade retention has been found to have either no or even negative short- to mid-term outcomes (Goos et al., 2013; Holmes & Matthews, 1984; Hughes et al., 2013; Jimerson, 2001; Jimerson, Anderson, et al., 2002; Jimerson, Ferguson, Whipple, Anderson, & Dalton, 2002; Reynolds, 1992). In addition, it has been shown that the retention decision is biased by socio-economic background, resulting in children with a less advantaged socio-economic background to have higher repetition rates (European Commission, 2008). For these reasons and the great number of affected students, it is of crucial importance to study the long-term consequences of grade retention.

4.1.1 Grade Retention

Students who fail to attain the competencies required to achieve a certain grade level have to repeat that grade level with the idea that an additional year of maturity and exposure to the curriculum will improve the students' academic achievement and social success in future grade levels. In many countries, there is a general consensus that grade retention is a very beneficial measure for students (European Commission, 2011). In almost all school systems, poor grades/marks are the main reason for a student to repeat a year, although other criteria may play a role. The decision-making process is mainly influenced by the teacher's opinion of the student, while parental opinion has a lesser influence (European Commission, 2011). Contrary to the general idea that grade retention is supposed to be beneficial for the students, research findings are very controversial and the majority of findings report either no or even negative effects of grade retention on a number of variables related to educational attainment and psychological functioning (Holmes & Matthews, 1984; Jimerson, 2001).

Reasons for Grade Retention

What are the characteristics of students who are retained compared to those students who are promoted? Surely, students are retained when they fail to achieve certain competences that are expected of their grade level, which is often reflected in poor grades. Thus, achievement, in terms of grades and intelligence, plays an important role in the retention decision (McCoy & Reynolds, 1999; Sandoval, 1984). However, several research results show that differences in grades and intelligence may not be significant between retained students and those students who are at risk of being retained but are promoted (Jimerson et al., 1997; Niklason, 1984; Sandoval, 1984). Crucially, social aspects may also play a role.

Several parties such as the teacher, the school, and the parents are passively or actively involved and play an important role in the decision of whether a child will be retained or not. Most often, other criteria than only achievement variables are considered in the process and decisions are made on an individual basis (Jimerson et al., 1997). Socioemotional adjustment is also a variable, which plays a role in the retention decision (Holmes & Matthews, 1984; Jimerson, 2001; Jimerson et al., 1997). Moreover, it has been repeatedly found that boys are more often retained than girls (Byrd & Weitzman, 1994; Holmes & Matthews, 1984; Jimerson, 2001; Jimerson et al., 1997; McCoy & Reynolds, 1999). Poverty and low socioeconomic status (SES) have also been identified as risk factors for being retained (Byrd & Weitzman, 1994; Holmes & Matthews, 1984; Jimerson, 2001).

Interestingly, parental characteristics have been found to play an important role in whether a child will be retained or not. In a study by Byrd and Weitzmann (1994) based on data from almost 10,000 children aged 7-17 who participated in the Child Health Supplement to the 1988 National Health Interview Survey in the USA, single parenthood and low maternal education were identified as risk factors for grade retention. On the other hand, high maternal education and both biological parents living at home were found to play a protective role. Another large-scale study by McCoy and Reynolds (1999) included 1,164 low-income minority children from the Chicago Longitudinal Study. They found that parental participation in school prevented children from being retained. However, parental education, free lunch eligibility (as an indicator for poverty), number of years in CPC participation (i.e. an early intervention program), special education placement, and teacher ratings of classroom adjustment did not have any incremental predictive power for retention in their analyses. These results might partially be a result of the special sample of the study, which focused on low-income minority children. In another study, Jimerson and colleagues (1997) found that promoted children (comparison group) had parents with higher IQs and those parents were more involved in the school than those of retained children.

In summary, grades, intelligence, gender, parental SES, socioemotional adjustment, living situation, parental involvement in school, parental intelligence, and parental education are predictors of grade retention.

Impact on School Careers

Only a few research results report positive findings of grade retention on future academic performance (D. C. Gottfredson et al., 1994; Hughes et al., 2010) or children's perceived school competence (Reynolds, 1992). Some studies report short-term positive effects that diminish over one or two years (Jimerson, 1999, 2001; Jimerson et al., 1997; Wu et al., 2010). In contrast, the great majority of research findings reports either no effect of grade retention on academic performance (Chen et al., 2010; Im et al., 2013; Phelps et al., 1992; Pierson & Connell, 1992) or even an negative effect of grade retention on a number of variables such as future academic performance (Goos et al., 2013; Hughes et al., 2013;

Reynolds, 1992), parent's expectations (Hughes et al., 2013), academic self-efficacy (Hughes et al., 2013), children's psychological functioning (Goos et al., 2013), future school career (Goos et al., 2013), and importantly, dropping out of high school (Jimerson, Anderson, et al., 2002). In addition, several negative consequences of grade retention on students' school careers have been reported in the meta-analyses by Holmes and Matthews (1984) as well as by Jimerson (2001). Holmes & Matthews (1984) report that grade retention has a negative effect on personal adjustment (*ES* = -.27), self-concept (*ES* = -.19), attitude towards school (*ES* = -.16), and school attendance (*ES* = .-12). They conclude that retained students favor school less than promoted students. In the more recent meta-analysis by Jimerson (2001), negative effects of grade retention were also found on self-concept (*ES* = -.04) and school attendance (*ES* = -.65), but with different effect sizes compared to Holmes and Matthews (1984). Moreover, Jimerson (2001) emphasized the consideration of long-term consequences of grade retention. Some of the studies in the meta-analysis reported short-term advantages for retained students during the year following retention. However, these advantages diminished or even reversed in many longitudinal studies covering a longer time span (Jimerson et al., 1997).

This pattern of effects has also been found in more recent studies. Im and colleagues (2013) found effects of grade retention in primary school on students' reading and math achievement, teacher-rated engagement, and student-reported school belonging in middle school. The authors used propensity score matching to control for 67 covariates, the matched sample included almost 400 children. No positive longitudinal effect of grade retention was found in their study. Both retained and continuously promoted students had comparable outcome measures in the year prior to and the year after transition to middle school.

Wu and colleagues (Wu, West, & Hughes, 2008) conducted a longitudinal study that spanned 4 years with a sample of 784 students. After matching retained with comparable promoted students on the basis of propensity score matching, their sample contained 196 students. They performed growth curve analyses to estimate the effect of grade retention. Compared to national grade norms, retained students experienced an increase in the year that they repeated, but a decrease in growth rate after that year. Thus, the students only benefited from grade retention in the short-term, but not in the longer term. The authors concluded that the short-term benefit for the retained students vanishes when they face new and unfamiliar material. In addition, the students might experience a sequence of failure (the year before repetition), success (the repeated year) and failure again (when they face new material), and that this chain of events might have a crucial negative impact on students' academic-related self-beliefs.

Goos and colleagues (2013) computed a longitudinal study over 6 primary school years in Belgium that covered 122 schools and more than 3,600 students (298 retained students). They found that retained students in first grade were less likely to retain another school grade. However, compared to equally at risk, but promoted first graders, they had a 12.5% higher probability to move to a special education primary school, were 21,0% more likely to move to another primary school, and 35,4% less likely to move to the A track (the, higher school track versus B track) in the first year of secondary school. Thus, in a tracked school system such as in Belgium (or also in Luxembourg), retained students are more likely to attend the lower track than their equally at risk, but promoted, peers.

In addition, research findings show that retained students are also more likely to drop out of school than their low achieving but promoted peers (Alexander et al., 2003, 1997; Jimerson, Ferguson, et al., 2002; Rumberger, 1995). In addition, previously retained students are also less likely to obtain a high school diploma, even when they have achievements comparable to those of continuously promoted students in grades 8 and 9 (Alexander et al., 2003). Jimerson and colleagues (2002) reviewed 17 studies that examined the relationship between grade retention and school dropout. They concluded that "*grade retention is one of the most powerful predictors of dropout status*" (Jimerson, Anderson, et al., 2002, p. 441), a finding that is shared by Rumberger's work (1995). The link between grade retention and school dropout has often been explained by a process of disengagement from school (Alexander et al., 2003; Holmes & Matthews, 1984; Pagani, Tremblay, Vitaro, Boulerice, & McDuff, 2001). Retained children might like school less than their promoted peers or link school to more negative events than their continuously promoted but low achieving peers.

To summarize, grade retention in primary school has been shown to have a lasting negative effect on most students' future school careers, even though many findings show no negative effects on students' achievement. Retained students are more likely to attend a lower (and shorter) track in a tracked school system or drop out of school. Thus, by reaching young adulthood, most retained students have received less years of schooling at a lower level than their comparable but continuously promoted peers.

Long-Term Consequences: Educational Attainment, Income, and Intelligence

Lifespan developmental psychology (Baltes, 1987; Li & Baltes, 2006) may provide a theoretical framework to explain possible negative long-term consequences of grade retention in primary school. Grade retention takes place at a point in life, when individuals are still very sensitive to environmental interventions. Thus, grade retention may have long lasting impacts as later environmental influences are less effective (Baltes, 1987; Li & Baltes, 2006). Further, it is proposed that individuals interact with their environment in a reciprocal manner. Therefore, the three key life outcomes in the present study are closely connected. As described above, most retained students receive less education than their promoted but equally at risk peers. This can affect a number of other life outcomes over several decades of the lifetime. Two prominent factors are income and intelligence. In Study 3, we focus on the relationship between grade retention and general intelligence (g) as, to our knowledge, no hypothesis exist why grade retention should be differently affect by fluid and crystallized abilities.

Schooling, intelligence, and income are closely connected. Schooling and intelligence affect economic life outcomes. It is a well-documented fact that longer school attendance results in a higher income (Ceci & Williams, 1997). However, the possible reasons for this are more debated. Some researchers (L. S. Gottfredson, 1997b, 1998, 2002; Schmidt & Hunter, 2004) argue that the influence of schooling on income is mainly determined by intelligence. More intelligent people stay in school longer and will also perform better in the job, which leads to a higher income. Ceci and Williams (1997) propose that there are two ways how schooling may influence income: (a) An indirect effect as higher schooling is an indicator of higher levels of intelligence and (b) a more direct effect as many jobs require a minimum entry school qualification. Thus, the higher the qualification level, the higher the income.

The effect of intelligence on schooling is well documented. In fact, the first IQ tests were designed and invented with the objective of predicting school performance (Brody, 2000) and they have served their purpose well. Thus, the predominant amount of research on the relation between intelligence and education focuses on intelligence being the predictor and education being the criterion. It therefore comes as no surprise that the predictive value of intelligence on academic success is well established (Jensen, 1998). The effect of intelligence on job performance and income is equally well established. Intelligence has been found to be the single best predictor of job performance in several studies and meta-analyses (L. S. Gottfredson, 1997b, 2002; Hunter & Hunter, 1984; Schmidt & Hunter, 2004).

However, even though there is wide consensus among scholars in the field that more intelligent students stay in school longer, many researchers underestimate the reciprocal relationship between schooling and intelligence (Brody, 1997; Li & Baltes, 2006). Thus, staying in school longer may also positively affect the student's intelligence level (Ceci &

Williams, 1997). Therefore, the influence of intelligence goes beyond the decision to remain in school, as schooling may also increase intelligence (Brody, 1997; Ceci & Williams, 1997; Neisser et al., 1996). There is vast evidence that this is the case and that schooling does indeed affect intelligence (Ceci, 1991, 1999; Neisser et al., 1996). The positive effect of length of schooling on cognitive abilities was also found in Study 2 of the current Ph.D. thesis for non-academic track students even over 30 years after participants had left formal education. Thus, less schooling negatively impacts a person's intelligence level over several years. Grade retention has been found to be connected to school drop-out and to negatively affect a number of other academically relevant variables. Hence, retained students are more likely to leave school early. Subsequently, it is possible that retained children's intelligence is negatively affected by the retention decision and that this effect accumulates over several decades after having been retained.

In summary, retained children drop out of school more often and are less likely to attain a higher track. Thus, retained school children receive less schooling. In the long-term, this can have a large impact on their educational attainment, their economic success and thus income, but also on their level of intelligence. Hence, the retention decision in primary school can have a lasting impact on key life outcomes in adulthood.

4.1.2 Selection Bias and Propensity Score Matching

A great problem for most research tackling grade retention is the methodological limitation of lacking a randomized experimental design (Thoemmes & Kim, 2011; Wu et al., 2008). For a causal attribution of an outcome to a treatment (e.g. retention), it is required that the assignment to the treatment and control groups be uncorrelated to the outcome (Morgan & Winship, 2007). This is usually ensured by means of randomization. If this is not the case, changes in the outcome may be attributed to either the treatment or pre-existing group differences. However, in retention research, it is neither practically nor ethically possible to randomly assign students to the retention group ("treatment") or the promotion group ("control"). The assignment to these two groups is made on the basis of several characteristics including the students' grades and thus also the students' level of intelligence. This could result in a potential selection bias, so that only the academically and cognitively weakest students experience grade retention. It is possible that the retained and promoted groups differ too greatly to be compared.

Previous research on grade retention has used different approaches to address the problem of biased selection: (a) a promoted comparison group was identified, which was comparable to the retained group (Jimerson et al., 1997; Pierson & Connell, 1992; Reynolds, 1992; Rust & Wallace, 1993). Most commonly, students in the same class are identified, who score below a certain achievement or intelligence measure in a certain year, but are promoted to the next class level in the following year. However, in this approach, it is possible that the retained and the promoted groups may not be completely equivalent on the achievement variables. Another disadvantage of this approach is that potential differences on other important variables (e.g. parental socioeconomic status) between the retained and promoted groups may exist, which may affect interpretation of the effect.

(b) Regression-based models, which include the analysis of covariance or multiple regression are used (D. C. Gottfredson et al., 1994; Meisels & Liaw, 1993). This approach makes several assumptions that may not be appropriate to study grade retention effects (Cohen et al., 2003): First, only a limited number of covariates can be included in the regression model due to power considerations (for a more elaborated explanation see Cohen et al., 2003, p. 185ff). This is a problem for research in the social sciences, where the outcomes are very complex and more factors may be involved than can be included in a regression-based model. Second, the type of relationship between predictor and criterion must be known in advance and correctly specified. Most of the models assume a linear relationship, but a misspecification can have a very negative effect on model estimation. Third, the retained and promoted groups may differ greatly on their covariates at pretest. Reliable predictions can only be drawn when the two groups overlap on the covariates and any extrapolation of the effect of retention beyond the region where the two groups overlap is risky (Becker et al., 2012).

c) A more recent and increasingly popular approach to disentangle confounding factors and control for selection bias in retention research is propensity score matching (Becker et al., 2012; Chen et al., 2010; Goos et al., 2013; Thoemmes, 2012; Thoemmes & Kim, 2011; Wu et al., 2008, 2010). In comparison to regression-based models, it is possible to include many covariates, which, in prior research and theory, have been shown to be related to treatment selection and the outcome variable. The approach allows the balancing of these variables between the treatment and control groups. Propensity score matching is comprised of two steps: In a first step, a single propensity score for each participant is estimated using logistic regression. In the present study, this is the probability that the participant will be retained. In a second step, this probability is used to create comparable groups of retained and promoted students via a matching algorithm. The downside to this approach is that it reduces sample size.

Propensity score matching procedure consists of five different steps, as described by Thoemmes (Thoemmes, 2012; Thoemmes & Kim, 2011): First, a set of baseline covariates is selected on the basis of previous research and theory. It is important that these baseline variables are not affected by the retention decision and should therefore be assessed before the students are either retained or promoted. It is also important to be as comprehensive as possible, when selecting appropriate variables. Second, on the basis of the selected covariates, a single propensity score (i.e. the probability that the student will be retained) is calculated. The score is most commonly calculated by means of logistic regression, where group assignment is the outcome variable and the covariates are the predictors (Thoemmes, 2012). Thus, the propensity score conveys the likelihood of a person being retained a year in school on the basis of the selected covariates, which were measured before the retention decision was made. Third, retained and promoted students with similar estimated propensity scores are matched. The matching process equates the retained and promoted students on the propensity score and thus, also on the covariates used to estimate the score. Several different matching algorithms can be used. The most common is 1:1 nearest neighbor matching (Thoemmes & Kim, 2011). In this algorithm, one retained students is matched to one promoted student with the most similar propensity score. However, when sample sizes of the retained and promoted group differ greatly (as in the current study) one-to-many matching can be applied (Thoemmes, 2012). Matching one treated participant to a number of controls has been shown to produce even better gains in bias reduction (Ming & Rosenbaum, 2000; Thoemmes, 2012). However, not more than 5 matches to a single unit are necessary (Ming & Rosenbaum, 2000). In order to control matching quality, a caliper value is defined, which is reported in units of standard deviation of the logit of the estimated propensity score. The caliper distance is the maximum allowable difference between the propensity scores of the matched students. There is no recommended value for the caliper distance (Bacher, 2002), although some researchers provide rules of thumb (Rosenbaum & Rubin, 1985). On the one hand, the lower the caliper is set, the fewer matches will be found. On the other hand, the higher the caliper is set, the larger the difference between matched students. Thus, a caliper value must be deliberately chosen and re-adjusted if needed, taking the model adequacy tests in step 4 into account (Bacher, 2002). Fourth, model adequacy is tested. In order to check whether a balance of the covariates has been achieved by the matching procedure, several statistics from the retained and promoted groups are compared before and after matching. The main indicator is the standardized mean difference of covariates, which should be close to 0 after matching

(Thoemmes, 2012). Finally, the retention effect is estimated via multiple regression in the matched subsample.

4.2 The Present Study

The present study examines the important question of the long-term effects of grade retention in primary school up into middle adulthood. We will address three key life outcomes (a) educational attainment in terms of years of education, (b) adult male income, and (c) intelligence in middle adulthood. Most previous research on the long-term effects of grade retention covers only a couple of years. Even longitudinal studies, which cover the time until participants leave school, are scarce. Thus, to our knowledge, the effect of grade retention in primary school on income and intelligence in middle adulthood has rarely been studied, if ever. However, long-term effects of grade retention can be expected, as retained students are more likely to leave school earlier than their peers, who were also at risk of being retained, but were promoted. In turn, leaving school early affects job qualification, income, and also intelligence level.

In addition, another important feature of the present study is that we were able to perform a propensity score matching procedure to control for important child, family, parent, and school characteristics that have been shown to affect the retention decision. Furthermore, in the multiple regression analyses, we also controlled for childhood intelligence and grade point average in the most important school subjects in primary school. Due to the study design, we lacked a measure of intelligence and grades before students were retained. Therefore, we could not include these two important covariates in the propensity score matching procedure, as we expected an effect of grade retention on exactly the same variables: intelligence and educational attainment.

In summary, the present longitudinal study makes several important contributions to the empirical body of research on the long-term effects of grade retention in primary school. (a) It spans 40 years from late childhood to middle adulthood and (b) examines three key life outcomes (i.e. educational attainment, adult male income, and adult intelligence). Previous longitudinal studies on grade retention have rarely looked at the length of formal education or intelligence development beyond school age. To our knowledge, no study exists that addressed the impact of grade retention on adult income. (c) We were able to control for key child, family, parent, and school characteristics, as well as childhood intelligence and grade point average in the analyses.

4.3 Methods

4.3.1 Participants and Procedure:

The current longitudinal study (MAGRIP) encompasses two points of measurement - 1968 and 2008, covering a time span of 40 years. In 1968, a representative sample of 2,824 students (49.9% female) was drawn from all elementary schools in Luxembourg, who were approximately 12 years old (*M* = 11.9 years; *SD* = 7.2 months) at the time of testing. Detailed information was collected on the parents' working situation, the students' living situation, grades, school characteristics, as well as other student characteristics. In addition, all children completed a comprehensive intelligence test, the "Leistungsprüfsystem" (i.e. achievement test battery, Horn, 1962, 1983), which was administered by trained university students in a group setting.

In 2008, 745 (53.3% female) of these former students, now aged approximately 52 years, were visited at home and interviewed on their educational history as well as their income (i.e. household sample). For the analyses concerning the influence of grade retention on educational attainment in terms of years of education, we drew on the entire household sample. For analyses of adult income, we drew on the male participants (*n* = 348) of the household sample only. Most of the women in this cohort had stopped working when they married, or did not achieve their career potential if they continued working. Thus, including them in the analyses could have biased the results.

In a second stage, 378 participants (51.8% female) from the household sample retook the same intelligence test that they had completed 40 years earlier (i.e. intelligence test sample). About two thirds of these persons (n = 247) took this test in a group setting; the remaining participants were visited at home to take the test individually. All tests were administered by trained assessors and the test-taking procedure strictly followed the standardization of the test manual. We drew on this intelligence test sample to study the influence of grade retention on intelligence in middle adulthood. More details on the sample sizes of the three propensity score subsamples for the three criterion variables are included in the missing data section.

4.3.2 Measures

In the following, we will describe our measures of grade retention and the three key life outcomes under study: (a) educational attainment in terms of years of formal education, (b) income, and (c) intelligence in middle adulthood. We will then describe the covariates that were controlled for in the current study, namely intelligence in childhood, grade point average, parental socioeconomic status, and the covariates included in the propensity score matching analyses.

Grade Retention in Primary School

In the current study, grade retention was assessed by the participants via self-report[6]. For this purpose, participants were asked whether they were retained in primary school or not during the home interviews in 2008.

Educational Attainment in Terms of Years of Education

In 2008, trained assessors interviewed the participants on their educational and occupational history and recorded the information in the form of an educational curriculum vitae (CV), where start and end year of each school visited were reported. From the educational CV, we calculated educational attainment in terms of length of formal education. Length of formal education was assessed by years of formal education after primary school. In addition, the number of retained school years in primary and secondary school were not considered for the calculation of years of education.

Income at Age 52

Monthly income was measured with a 15 point scale that clustered several net monthly income categories. The lowest possible category was named “no own income” the highest category was “€10,000 and more”. The categories were chosen in reference to Luxembourgian monthly income and were designed in cooperation with experts in the field, namely the research institute CEPS/INSTEAD in Luxembourg, which was also responsible for the Luxembourgian survey on “European Union-Statistics on Income and Living

[6] Note. In 1968, teachers were also asked to report the “retention” career of the students in the study. However, the coding system in 1968 was rather complicated so that the variable was coded in several different ways and the data was not clearly interpretable for a large proportion of study participants. “Cleaning” the data would have involved a great deal of subjective interpretation and speculation. In addition, the data from 1968 did not take into account participants that repeated grade level 6 after the MAGRIP study was conducted in 1968. We therefore decided to rely on the self-reported data in 2008. The bias in self-report is clearly attributable and not subject to the interpretation of the person “cleaning” or interpreting the data from 1968.

Conditions / Panel Socio-Economique Liewen zu Lëtzebuerg". For the statistical analyses, we took the mean per category with a maximum value of €10,000 net per month. The detailed categories can be found in Table 14 in Annex C.

The analyses of own income were only conducted for male participants as most of the female participants had stopped working or had not achieved their career potential. This is a typical pattern in Luxembourg for women in the cohort of the current study.

Intelligence in Late Childhood and Middle Adulthood

Intelligence at ages 12 and 52 was assessed by nine subtests taken from a standardized and well validated German intelligence test battery, named the "Leistungsprüfsystem" (L-P-S; i.e. achievement test battery, W. Horn, 1962, 1983). The L-P-S includes 9 subtests on four different broad ability factors of intelligence such as fluid reasoning, comprehension-knowledge, visual processing, and processing speed (Schalke et al., 2012). Each subtest contained 40 items and had to be completed within strict time constraints, which were specified in the test manual. For the analyses in the current study on grade retention, we used the L-P-S sum-score of the 9 assessed subtests to assess general intelligence *g*. A comprehensive description of the L-P-S, that includes reliability and validity estimates, can be found in Annex A. In 1968, the children were randomly administered to one of two parallel test forms of the L-P-S. Because the means and variances of subtests differed slightly across test forms, we used a linear-conversion rule (Kolan & Brennan, 1995) to equate the test scores. To this end, we standardized the sum-score of the L-P-S separately for each test form to an IQ metric with a mean of 100 and an *SD* of 15 for the base sample. In 2008, the participants were given the exact same test form and items that they had completed in 1968. To allow meaningful comparisons across time, the sum-scores obtained for the second wave of measurement in 2008 were equated by using the same conversion rules as applied in 1968 (i.e., the standardization of measures in 2008 was based on *means* and *SD*s obtained from the entire sample in 1968).

Grade Point Average in Primary School

In 1968, grades in primary school were assessed in mathematics, oral and written German, and oral and written French. Grades were assessed for 4 trimesters. The trimesters included the three trimesters of the school year prior to assessment of the study as well as the first trimester of the year of data assessment. Grades in Luxembourg were assessed on a scale from 1 to 60, were 60 was the highest achievable grade and 1 the lowest possible grade. For the current study, we calculated the mean for all assessed grades by weighting the mathematics, the German, and French grades 1/3 respectively. Thus, this grade total included the mathematics (weight 1/3), oral and written German (weight 1/3), and oral and written French (weight 1/3) grades from all 4 assessed trimesters. The grade total was then included in the current study's multiple regression analyses.

Parental Socioeconomic Status in Primary School

In 1968, the school children reported their fathers' and mothers' occupations. These occupations were then used to assess parental socioeconomic status (SES). To this end, the occupations were coded using the International Standard Classification of Occupations (ISCO-68; International Labour Organization (ILO), 1968) and then transformed into the International Socio-Economic Index of Occupational Status (ISEI; Ganzeboom et al., 1992) according to the transformation tables provided by Ganzeboom and colleagues (1992). The ISEI was developed by Ganzeboom, De Graaf, and Treiman (1992; Ganzeboom & Treiman, 1996) and has also been used as a standard indicator for parental SES in the PISA study (Baumert et al., 2001). The ISEI is a linear continuously scaled variable that is based on the education and income of a certain occupational group (Ganzeboom & Treiman, 2003). To determine a single indicator for a child's parental SES, we took the higher ISEI of the mother's and father's occupation.

Covariates for the propensity score matching procedure

The pattern of data collection did not allow grades or intelligence in childhood to be controlled for in the propensity score analyses, because we expected an impact of grade retention on these very same variables. Thus, in order to control for grades and intelligence these variables must be assessed before the student is retained. In the current study, both grades and intelligence were collected after most of the children had already been retained in primary school. However, we tried to make as much use of the rich dataset of the MAGRIP Study as possible and applied a procedure similar to Reynolds (1992). For propensity score matching, we identified those variables, which were collected after the student had been retained, but could not be affected by grade retention and are very stable (e.g. gender, number of people in household, parental SES). As described above, in addition to grades and intelligence, many child, family, and parental characteristics have been shown to affect the retention decision. Thus, the retained and promoted students were matched based on the following covariates:

Several studies have shown that boys are much more likely to be retained than girls (Byrd & Weitzman, 1994; Holmes & Matthews, 1984; Jimerson, 2001; Jimerson et al., 1997; McCoy & Reynolds, 1999). Therefore, (1) gender was included in the covariates. We also included (2) school gender configuration (i.e. only boys, only girls, mixed), as it is possible that this might influence the finding that boys are more likely to be retained than girls. In addition, poverty and low socioeconomic status (SES) have been identified to be predictors of grade retention (Byrd & Weitzman, 1994; Holmes & Matthews, 1984; Jimerson, 2001). Thus we included the following variables as they are indicators of SES (3) total number of children in household, (4) highest parental SES in terms of ISEI of parental occupation, (5) parents working, (6) total number of people living in household, and (7) nationality. Single parenthood has also been shown to influence whether a child is retained or not (Byrd & Weitzman, 1994), so we included the (8) family situation (i.e. parents married, divorced, widowed, etc). As found by McCoy and Reynolds (1999), parental participation prevents children from being retained. We included the variable (9) person learning with participant, as indicator of family support and participation. Furthermore, we included two child characteristics, namely (10) birth rank and (11) month participant born. Birth rank was included as first-born children are likely to get much more attention from their parents. Month participant was born was chosen because school enrollment in Luxembourg is connected to a fixed reference birth date. Thus depending on the month born, students can be either young or old for their grade. It is possible that children "young" for their grade are more likely to be retained as the retention decision is also often justified with the developmental status of a child.

To summarize, we were able to control for 11 important child, family, parental, and school characteristics in the study. However, due to the lack of another measurement point before primary school (i.e. before the retention decision), we could not include intelligence, and grades as covariates in the propensity score matching procedure, as we expected an effect of grade retention on exactly these two variables. We therefore included these two variables as covariates in the multiple regression analyses.

4.3.3 Propensity Score Matching Procedure and Data Analyses Approach

In order to control for a possible selection bias on family and child characteristics between the retained and promoted children, we applied propensity score matching. The propensity score can be defined as the probability of a student to be retained in elementary school based on the predictive power of a number of covariates on retention. In the current study, we were able to control for a number of family, parental, child, and school related variables that have been found to correlate with grade retention in previous research.

We had to draw on a different sample for each criterion variable (i.e. educational attainment, adult income, and adult intelligence). Thus, the propensity score matching procedure was performed for each criterion variable in 5 steps as described above (Thoemmes, 2012; Thoemmes & Kim, 2011). The propensity score was calculated in SPSS 23 using the SPSS propensity score matching (PS) add-in version 2.0 provided and programmed by Thoemmes and based on the underlying R packages Matchit, Rltools, and cem (Hansen, 2004; Hansen & Bowers, 2008; Ho, Imai, King, & Stuart, 2007a, 2007b; Thoemmes, 2011). We used logistic regression analysis to calculate the propensity score. Retained and promoted students with similar estimated propensity scores were matched with a nearest neighbor algorithm. As recommended in the literature (Ming & Rosenbaum, 2000; Thoemmes & Kim, 2011), a one-to-many matching procedure was applied. For each criterion variable, the ratio of promoted to retained students in the unmatched samples ranged from 1:5 (income) to 1:9 (adult intelligence). In the literature, it is recommended that a 1:5 ratio is sufficient (Ming & Rosenbaum, 2000). Thus, we chose a ratio of 1:5 for all analyses in the current study. The maximal allowed caliper distance, with which we started was set at $c = .2$, a value often chosen and recommended in the literature (Austin, 2011). A caliper value of .2 indicates that the retained and promoted matched groups do not differ more than .2 standard deviations on the logit of the propensity score. After the matching procedure was finished, model adequacy was tested to verify that a balance of the covariates had been achieved by the matching procedure. To this end, we compared the standardized mean difference of covariates (Cohen's *d*) between groups before and after matching. For this test, values close to 0 are desirable (Thoemmes, 2012) with a recommended cut-off value of .30 (Im et al., 2013). Should the standardized mean difference of covariates have differed greater than the recommended cut-off value, the caliper distance would have been adjusted and a new matching procedure would have been performed (Bacher, 2002).

After propensity score matching had successfully been accomplished, we estimated the retention effect by applying multiple regression models to the data for each matched subsample (i.e. for each criterion variable). We included grade point average in primary school, childhood intelligence at age 12, highest parental SES, and the propensity score in the multiple regression analyses. Childhood intelligence, school grades, and parental SES are three of the most prominent impact factors on educational attainment, adult income, and adult intelligence (Bradley & Corwyn, 2002; Ceci & Williams, 1997; Schmidt & Hunter, 2004). Therefore, we decided to control for these important influences in the regression analyses. Parental SES was already controlled for in the propensity score matching, so that parental SES was equally distributed between the promoted and retained group. However, parental SES can still have an impact on the three criterion variables under study (i.e. educational attainment, adult income, and adult intelligence) over the lifespan. Thus, including parental SES in propensity score matching controls for its influence on being retained or not, while including parental SES in the multiple regression analyses controls for its influence on educational attainment, adult income, and adult intelligence across the lifespan. As recommended in the literature, the propensity score was included in the regression analyses for a better adjustment of the model (Rubin & Thomas, 2000). In addition, previous research has shown that much of the influence of grade retention on key life outcomes could be mediated by educational attainment. Therefore, we included years of education as a predictor in the analyses for adult income and adult intelligence. The calculations were performed using SPSS 23[7].

[7] For some of the criterion variables, the models did not converge in MPlus, probably a result of the too small sample size of some of the matched subsamples. Therefore, all analyses were taken out in SPSS.

In addition to the standard indices reported in multiple regression models, we used several different types of correlations to describe the complex relationships between multiple predictors and criteria, which are described by Cohen and colleagues (2003). The product moment regression coefficient (*r*) is the standard measure of the linear relationship between two variables. The squared regression coefficient (r^2) represents the proportion of variance in the criterion that can be explained by one predictor in a multiple regression (not the unique proportion of variance). The semi-partial correlation coefficient (sr) is the correlation between the criterion and the portion of one predictor that is uncorrelated with the remaining predictors in a multiple regression. The squared semi-partial correlation coefficient (sr^2) is the unique variance of a single predictor in a multiple regression on total criterion variance. Therefore, sr^2 equals the unique contribution of one predictor to the R^2 of the multiple regression in the context of the remaining predictors. The partial correlation coefficient (*pr*) is the correlation between the portion of the criterion that is uncorrelated with the remaining predictors and the portion of one predictor that is uncorrelated with the remaining predictors in a multiple regression. The squared semi-partial correlation coefficient (pr^2) is the unique variance of a single predictor in a multiple regression on that part of the criterion variance that is not accounted for by the remaining predictors in a multiple regression. Therefore *pr* will always be larger than sr as it represents the same unique contribution of one predictor on the criterion, but in proportion to a smaller part of the criterion variance (i.e. the part of the criterion variance that cannot be accounted for by the remaining predictors).

4.3.4 Missing Data

We applied listwise deletion for missing data. Thus, participants with missing data on at least one of the covariates in the propensity score analyses were excluded from the analyses. In addition, participants with missing values on at least one of the predictors in the multiple regression analyses were also excluded.

Educational Attainment. The household sample from which we drew comprised 745 (53.3% female) students. 20 participants had missing values on one of the covariates for the propensity score matching (3 parental SES, 12 family situation, and 5 nationality) and were excluded from the analyses. Of the remaining 725 students, 485 students were successfully matched, of which 101 (49.5% female) were retained and 384 (52.6% female) were promoted. For the regression analyses, data was missing for 11 participants on intelligence in childhood and for 2 participants on grade point average. These 13 participants were excluded from the multiple regression analyses. Thus, a sample of n = 472 (51.9% female) constituted the basis for the multiple regression analyses, of which 97 (49.5% female) were retained and 375 (52.5% female) were promoted.

Adult Income. The income sample from which we drew comprised 348 (0% female) students. 10 participants had missing values on one of the covariates for the propensity score matching (1 parental SES, 5 family situation, and 4 nationality) and had to be excluded from the analyses. Of the remaining 338 students, 213 students were successfully matched, of which 49 (0% female) were retained and 164 (0% female) were promoted. Of the 213 successfully matched participants, data was missing for 5 participants on intelligence in childhood and for 17 participants on income. These 22 participants were excluded from the multiple regression analyses. Thus, a sample of n = 191 (0% female) constituted the basis for the multiple regression analyses, of which 46 (0% female) were retained and 145 (0% female) were promoted.

Adult Intelligence. The intelligence test sample from which we drew comprised 378 (51.8% female) students. 10 participants had missing values on one of the covariates for the propensity score matching (1 parental SES, 6 family situation, and 3 nationality) and were excluded from the analyses. Of the remaining 368 students, 203 students were successfully matched, of which 41 (46.3% female) were retained and 162 (53.7% female) were promoted.

For the multiple regression analyses, 5 participants had missing data on intelligence in childhood and 12 participants had missing data on intelligence in middle adulthood. 16 participants were excluded from the analyses as one participant had missing values on both variables. Thus, a sample of 187 (52.4% female) participants constituted the basis for the multiple regression analyses, of which 37 (45.9% female) were retained and 150 (54.0% female) were promoted.

4.4 Results

4.4.1 Propensity Score Matching and Descriptive Statistics

In the following, we report the results of the propensity score matching procedure as well as the descriptive statistics for the matched sample. This is done separately for each criterion variable, namely (a) educational attainment, (b) adult male income, and (c) intelligence in middle adulthood. As described in the participants and procedure section of the current study, we drew on a different subsample for the analyses of each criterion variable.

Educational Attainment

As recommended in the literature (Ming & Rosenbaum, 2000; Thoemmes & Kim, 2011), a one-to-many matching procedure was applied with a ratio of 1:5. We started with a maximal allowed caliper distance of $c = .2$. 485 students were successfully matched, of which 101 (49.5% female) were retained and 384 (52.6% female) were promoted. The propensity scores in the sample ranged from .02 to .70 ($M = .19$, $SD = .12$). Standardized mean differences between the unmatched and matched sample were calculated and all mean differences were below the recommended cut-off value of .30 (Im et al., 2013). See Figure 1 for histograms with overlaid kernel density estimates of standardized differences before and after matching. The pair of histograms shows the standardized differences of all terms (covariates, quadratic term, and interactions) before and after matching. For a good comparative view of the magnitude of differences before and after matching, the two histograms use the same scale. The histograms in Figure 1 show that after matching the standardized differences are centered on zero and therefore no systematic differences remain. Thus, the matching procedure with a caliper value of .2 was successful. Further details comparing the unmatched and matched samples are included in Table 15 and Table 16 in Annex C.

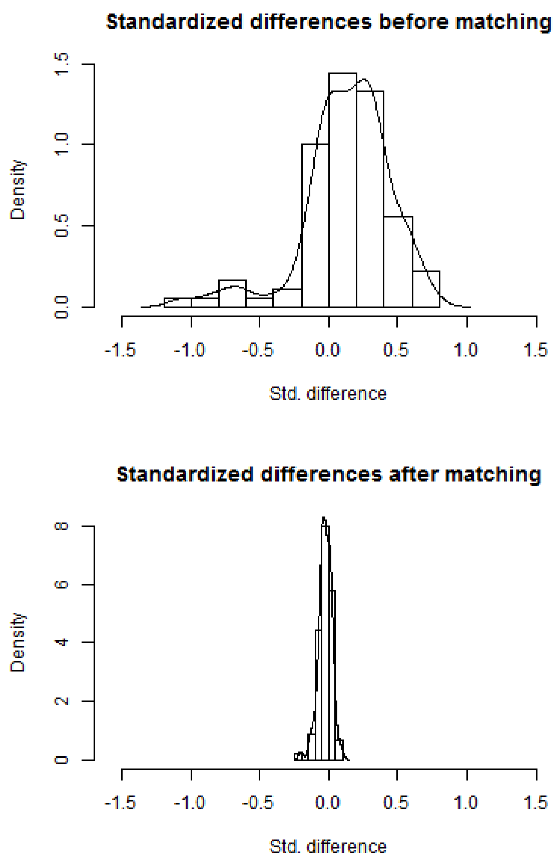

Figure 6

Histograms with overlaid kernel density estimates of standardized differences of all terms (covariates, quadratic term, and interactions) before and after matching for the sample with educational attainment as the criterion variable.

Table 9 shows descriptive statistics for the variables included in the multiple regression analyses for the promoted and retained students before and after matching. The predictors in the regression analysis are highest parental SES, grade point average, and intelligence in childhood. The criterion is educational attainment in terms of years of formal education. The medium to large group differences in the unmatched sample were reduced in the multiple regression analyses for some of the variables. The difference between groups on parental SES became much smaller (Cohen's *d* = .18) and fell below the recommended cut-off value (Im et al., 2013). However, in the regression analyses, the promoted and retained students still differed greatly on grade point average (Cohen's *d* = 1.12) and intelligence in childhood (Cohen's *d* = .84). These two variables could not be included as covariates in the propensity score matching procedure, as we expected an impact of grade retention on exactly these two variables and we lacked a measurement before the retention decision was made. To best possibly control for the effect of these significant group differences in grade point average and intelligence in childhood, we included them in the regression analyses. In addition, after matching the not retained students visited formal education on average 2.5 years longer than the retained students (Cohen's *d* = .82).

Table 9

Descriptive Statistics for the Variables in the Multiple Regression Analyses of Years of Education of the Promoted and Retained Students Before and After Matching

	Before Matching							After Matching						
	Not retained			Retained			Effect size	Not retained			Retained			Effect size
Unmatched Variables	mean	*SD*	*n*	mean	*SD*	*n*	Cohen's *d*	mean	*SD*	*n*	mean	*SD*	*n*	Cohen's *d*
Highest parental SES	41.2	12.9	618	33.7	8.6	107	0.61	35.5	9.7	384	33.8	8.7	101	0.18
Grade point average (0-60 points)	47.1	7.6	617	37.3	7.5	106	1.29	45.9	7.9	383	37.2	7.5	100	1.12
Childhood general intelligence	104.8	13.7	609	92.8	14.0	104	0.87	104.1	13.6	376	92.6	13.9	98	0.84
Years of education	6.3	3.4	618	3.2	2.2	107	0.97	5.8	3.3	384	3.3	2.1	101	0.82

Note. SES = socioeconomic status

Adult Male Income

We matched using a 1:5 ratio and a maximal allowed caliper distance of *c* = .2. 213 students were successfully matched, of which 49 (0% female) were retained and 164 (0% female) were promoted. The propensity scores in the sample ranged from .01 to .86 (M = .21, SD = .13). The matching procedure was successful. All standardized mean differences were below the recommended cut-off value of .30 (Im et al., 2013). Figure 2 shows histograms with overlaid kernel density estimates of standardized differences before and after matching. We see that after matching the standardized differences are centered on zero and thus no systematic differences remain. Further details comparing the unmatched and matched samples are included Table 17 and Table 18 in Annex C.

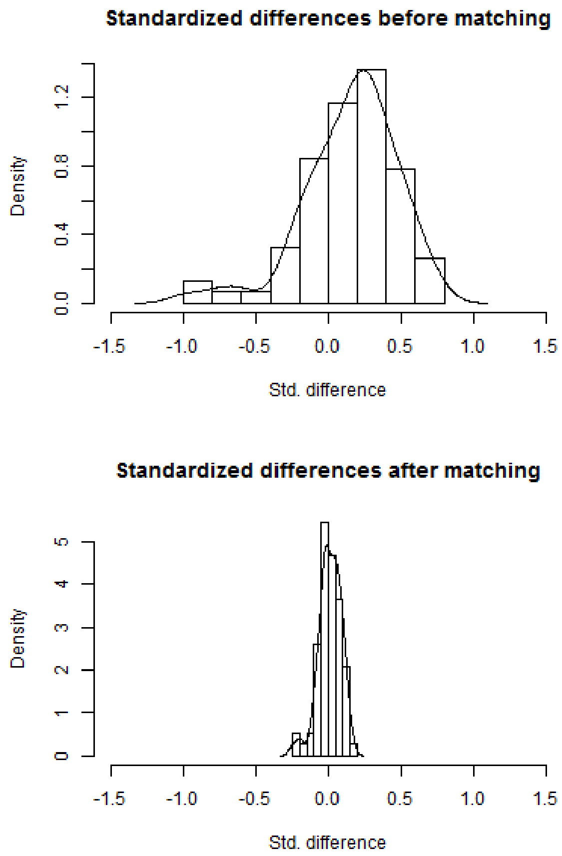

Figure 7

Histograms with overlaid kernel density estimates of standardized differences of all terms (covariates, quadratic term, and interactions) before and after matching for the sample with adult income as the criterion variable.

Descriptive statistics for the variables included in the multiple regression analyses on income for the promoted and retained students before and after matching are shown in Table 10. The pattern is comparable to the pattern seen for years of education. Before the matching procedure, promoted and retained students differed greatly on all predictors. After matching, the promoted and retained groups differed much less on highest parental SES (Cohen's *d* = .18). The group differences on grade point average (Cohen's *d* = 1.02), childhood intelligence (Cohen's *d* = .94), and years of education after primary school (Cohen's *d* = .90) remained large, as they were not included in the propensity score matching procedure. In addition, after matching promoted and retained students, the group difference on income amounted to almost €1,500 (Cohen's *d* = .83).

Table 10

Descriptive Statistics for the Variables in the Multiple Regression Analyses of Adult Income of Promoted and Retained Students Before and After Matching

	Before Matching							After Matching						
	Not retained			Retained			Effect size	Not retained			Retained			Effect size
Unmatched Variables	mean	*SD*	*n*	mean	*SD*	*n*	Cohen's *d*	mean	*SD*	*n*	mean	*SD*	*n*	Cohen's *d*
Highest parental SES	40.8	12.7	285	34.3	8.1	53	0.53	35.8	10.2	164	34.0	8.2	49	0.18
Grade point average (0-60)	45.9	8.0	285	37.3	7.8	53	1.08	45.3	8.0	164	37.2	7.9	49	1.02
Childhood general intelligence	105.8	13.9	280	92.7	15.5	51	0.92	106.6	14.2	161	93.0	15.7	47	0.94
Years of education	7.1	3.7	285	3.5	2.4	53	1.00	6.2	2.9	164	3.7	2.4	49	0.90
Adult income in Euro	4716	2251	260	3091	1150	52	0.77	4602	2070	148	3031	1158	48	0.83

Note. SES = socioeconomic status

Intelligence in Middle Adulthood

We matched using a ratio of 1:5 and a starting caliper distance of *c* = .2. 203 students were successfully matched, of which 41 (46.3% female) were retained and 162 (53.7% female) were promoted. The propensity scores in the sample ranged from .03 to .60 (*M* = .16, *SD* = .10). All standardized mean differences were below the recommended cut-off value of .30 (Im et al., 2013). Thus, the matching procedure was successful. Figure 3 shows histograms with overlaid kernel density estimates of standardized differences before and after matching. We see that after matching the standardized differences are centered on zero and thus no systematic differences remain. Further details comparing the unmatched and matched samples are included Table 19 and Table 20 in Annex C.

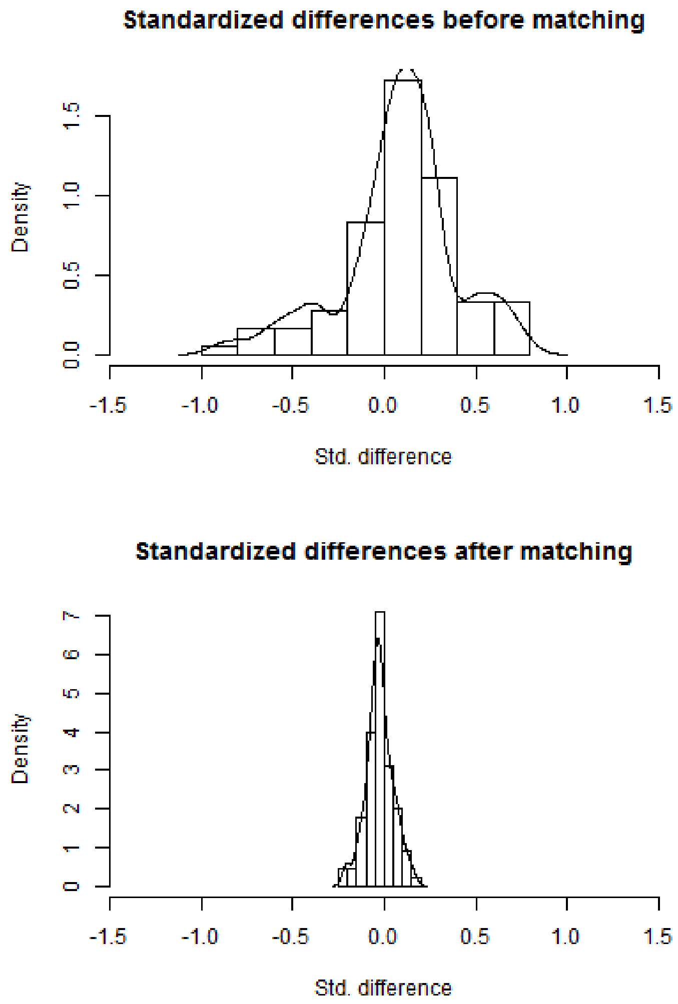

Figure 8

Histograms with overlaid kernel density estimates of standardized differences of all terms (covariates, quadratic term, and interactions) before and after matching for the sample with adult intelligence as the criterion variable.

Descriptive statistics for the predictor and criterion variables in the multiple regression analyses on intelligence in adulthood are included in Table 11. Differences between the promoted and retained students are shown before and after matching for the variables included in the multiple regression analyses. Before matching, the two groups differed greatly on all predictors as well as on the criterion variable adulthood intelligence. The group difference for highest parental SES was reduced after matching (Cohen's d = .12). The group differences for the remaining variables were not affected by the matching procedure, as it was not possible to control for grade point average, intelligence in childhood, and years of education in the propensity score matching procedure. However, we controlled for these variables in the regression analyses. In addition, after matching, the promoted students had scored on average 23 IQ points more in the intelligence test at age 52 compared to the retained students in Study 3 (Cohen's d = 1.16).

Table 11

Descriptive Statistics for the Variables in the Multiple Regression Analyses of Adult Intelligence of the Promoted and Retained Students Before and After Matching

	Before Matching							After Matching						
	Not retained			Retained			Effect size	Not retained			Retained			Effect size
Unmatched Variables	mean	*SD*	*n*	mean	*SD*	*n*	Cohen's *d*	mean	*SD*	*n*	mean	*SD*	*n*	Cohen's *d*
Highest parental SES	41.3	13.1	327	35.6	9.2	41	0.45	36.8	11.0	162	35.6	9.2	41	0.12
Grade point average (0-60 points)	48.1	6.9	327	37.6	6.9	41	1.52	47.4	6.8	162	37.6	6.9	41	1.43
Childhood general intelligence	106.7	12.7	321	93.3	14.7	40	1.03	106.5	12.2	158	93.3	14.7	40	1.04
Years of education	6.6	3.4	327	3.1	2.2	41	1.04	6.2	3.4	162	3.1	2.2	41	0.96
Adult general intelligence	142.2	19.3	314	117.9	21.4	38	1.25	141.2	19.8	153	117.9	21.4	38	1.16

Note. SES = socioeconomic status

4.4.2 Effect of Grade Retention in Primary School on Educational Attainment

A multiple regression was conducted to examine the effect of grade retention in primary school on educational attainment in terms of years of education. In addition, we simultaneously controlled for intelligence in childhood, highest parental SES in childhood, and grade point average of German, French, and mathematics grades in primary school. For a better adjustment of the regression model, we also included the propensity score. A sample of $n = 472$ constituted the basis for the multiple regression. All effects were estimated simultaneously.

Grade retention in primary school had a significant negative effect on years of education ($B = -.94$, Beta = -.12, $t = -2.65$, $p < .01$). The unstandardized B coefficient of grade retention shows that retained children left school almost 1 year earlier than their promoted peers. Crucially, the negative effect of grade retention is found even when the effects of the three (probably) most important impact factors on years of education, namely intelligence in childhood, highest parental SES, and grade point average, were controlled for. Table 12a summarizes the results of the regression analysis. The squared product moment correlation of grade retention with years of education ($r^2 = .10$) shows that 10% of the variance in years of education can be explained by grade retention, when not accounting for the influence of the other predictors in the study. Grade retention accounts uniquely for only 1% for the variance in years of education as shown in the squared semi-partial correlation coefficient ($sr^2 = .01$). This equals an effect size of Cohen f^2 for partial coefficients of 0.014 (Jacob Cohen et al., 2003, p. 94), and thus slightly fails the convention of a small effect. However, a significant effect that explains 1% incremental variance over a time span of 40 years when so many very important factors are controlled for is a meaningful result. Especially, since both childhood intelligence ($sr^2 = .02$) and parental SES in childhood ($sr^2 = .02$) also explain only 2% incremental variance in years of education. Thus, the incremental effect of grade retention on years of education can compare to the effects of childhood intelligence and parental SES in childhood.

In addition, a significant amount of the variance in years of education is explained in the multiple regression analysis ($F(5, 466) = 39.56$, $p < .01$, $R^2 = .30$, $R^2_{\text{Adjusted}} = .29$). An R^2 value of .30 is a large effect ($f^2 = .43$) according to Cohen´s convention of effect sizes (J.

Cohen, 1988, 1992). Moreover, the analysis shows that intelligence in childhood (*B* = .04, Beta =.16, $t = 3.52$, $p < .01$), grade point average (*B* = .12, Beta =.31, $t = 6.45$, $p < .01$), and highest parental SES (*B* = .05, Beta =.15, $t = 3.38$, $p < .01$) all significantly predict years of education. Grade point average is the most important predictor and explains 23% of variance in years of education (when not accounting for the other predictors) as shown in the squared correlation coefficient ($r^2 = .23$). It also accounts uniquely for 6% as shown in the squared semi-partial correlation coefficient ($sr^2 = .06$, Cohen's $f^2 = .09$), a small to medium effect. Thus, even the unique effect of the most important predictor in the present regression analysis can be considered rather small.

Importantly, R^2 of the regression analysis has a value of .30 and is rather large. Thus, approximately 30% of the variance in years of education is explained by the current regression analysis. However, all squared semi-partial correlation coefficients (sr^2) and squared partial correlation coefficients (pr^2) are rather small. Especially the small squared semi-partial correlations show that none of the included variables explains a large unique part of the total variance in years of education. Hence, most of the explained variance in years of education, as indicated in the large R^2 of the regression analyses, is explained by two or more of the included variables jointly. In conclusion, parental SES, grade point average, intelligence in childhood, and grade retention are all very closely interrelated and related all closely to years of education.

4.4.3 Effect of Grade Retention in Primary School on Adult Income

The effect of grade retention in primary school on adult male income was examined via a multiple regression. We simultaneously controlled for intelligence in childhood, highest parental SES in childhood, grade point average in childhood, years of education. For a better adjustment of the regression model, we included the propensity score. A sample of $n = 191$ constituted the basis for the multiple regression. All effects were estimated simultaneously.

We found that grade retention in primary school had a marginally significant negative impact on adult male income at age 52 (B = -650.04, Beta = -.14, t = -1.89, $p < .10$). The unstandardized B coefficient of grade retention shows that retained children earned approximately €650 less per month at age 52 than their promoted peers, even when the effects of intelligence in childhood, highest parental SES, and grade point average were controlled for. Grade retention in primary school explains 11% of variance in income as shown by the r^2 of grade retention in primary school (not accounting for the effect of the other predictors). In addition, 1% of the variance in income is uniquely explained by grade retention as indicated by the squared semi-partial correlation coefficient (f^2 = .014). Moreover, grade retention in primary school significantly impacts adult income in the current multiple regression analysis, while childhood intelligence (B = 16.53, Beta =.13, t = 1.65, ns), grade point average (B = 29.90, Beta =.13, t = 1.52, ns), and childhood parental SES (B = 10.39, Beta =.05, t = .72, ns) do not have a significant impact on adult income. Crucially, the most important predictor in the regression analysis is years of education (B = 146.96, Beta = .22, t = 2.84, $p < .01$). Every additional year in formal education after primary school earns the participant about 150 additional Euros per month in the current regression analysis. However, even though years of education explains 15% of variance in income as shown by r^2 of years of education in the regression analysis, years of education only explains 3% of the variance in income uniquely, as indicated by the squared semi-partial correlation coefficient. The squared semi-partial correlation has an effect size of Cohen's f^2 = .04 and is by convention a small effect (Cohen, 1992; Cohen et al., 2003).

In addition, a significant amount of the variance in adult income is explained by the five predictors ($F(6, 184) = 9.66$, $p < .01$, $R^2 = .24$, $R^2_{\text{Adjusted}} = .22$). Thus, 24% of the variance in income can be explained by the five predictors in the current study. This is a large effect with an effect size of Cohen´s f^2 of 0.32 (Cohen, 1992). Table 12b summarizes the results.

However, even though a large part of the variance in adult male income in the current study can be explained by the five predictors jointly, none of the predictors explains a large or medium proportion of the variance in income uniquely. This is indicated by the very low squared semi-correlations of all the predictors in the present multiple regression. Thus, similar to the results for years of education, most of the explained variance in income is shared by two or more of the predictors. This, in turn, points to large interactions between the predictors from late childhood into middle adulthood in the current study.

4.4.4 Effect of Grade Retention in Primary School on Intelligence in Middle Adulthood

A multiple regression was conducted to examine the impact of grade retention in primary school on intelligence in middle adulthood when simultaneously controlling for intelligence in childhood, highest parental SES in childhood, grade point average in primary school, and years of formal education. A sample of 187 participants constituted the basis for the multiple regression analysis. All effects were estimated simultaneously.

Grade retention in primary school had a significant negative effect on adult intelligence at age 52 (B = -6.78, Beta = -.12, t = -2.04, p < .05) even after controlling for probably the most important other impact factors on intelligence development over the lifespan. The unstandardized regression coefficient of grade retention shows that students retained in primary school scored approximately 7 IQ points lower at age 52 than promoted students. Thus, two students with comparable scores on intelligence at age 12, comparable highest parental SES, comparable grade point averages in primary school, and comparable years of education after primary school will statistically differ by 7 IQ points at age 52, depending on whether they had been retained over 40 years ago in primary school or not. In addition, in the current regression analysis, grade retention in primary school is the third most important predictor of intelligence at age 52 after intelligence at age 12 and years of formal education. However, even though grade retention in primary school can explain 19 % of the variance in intelligence at age 52 (r^2 = .19, when not accounting for the other predictors), only 1 % of the variance is explained uniquely by grade retention (sr^2 = .01) in the present regression analysis. Table 12c summarizes the results.

Childhood intelligence at age 12 is by far the most important predictor of adult intelligence at age 52 (B = 0.88, Beta =.54, t = 8.85, p < .01). Childhood intelligence accounts for 48% of variance in adult intelligence (r^2 = .48) in the current regression analysis and 20% is accounted for uniquely (sr^2 = .20) in the present regression analysis. Years of education after primary school also has a significant impact on adult intelligence at age 52 (B = .98, Beta =.15, t = 2.47, p < .05). However, the unique variance share in the criterion is only 2 % (sr^2 = .02). The influences of childhood parental SES and grade point average in primary school did not reach significance in the current regression analysis.

Moreover, it was found that a large significant amount of the variance in intelligence in middle adulthood is explained by the predictors in the current study ($F(6, 180) = 36.04$, $p < .01$, $R^2 = .55$, $R^2_{\text{Adjusted}} = .53$). The predictors in the current regression analysis explained 55% of the variance in adult intelligence at age 52. This is a very large effect (Cohen's f^2 = 1.22). As described above, 20% of the explained variance is uniquely accounted for by childhood intelligence in the present regression analysis. Apart from this large unique share, the remaining explained variance share is not largely accounted for uniquely by any other variable. This is indicated by the other predictors' very low squared semi-correlations. Thus, similar to the results in the other multiple regression analyses of the present study, half of the explained variance in adult intelligence is shared by two or more of the predictors. This may, in turn, point to large interactions between the predictors from late childhood into middle adulthood in the current study.

Table 12

a) Regression Analysis Summary and Correlations of Childhood Variables Predicting Years of Education

Predictor variable	B	95% CI	β	t	r	r^2	sr	sr^2	pr	pr^2	p
Parental socioeconomic status (ISEI range 16-90)	0.05	[0.02, 0.08]	0.15	3.38	0.25	0.06	0.13	0.02	0.15	0.02	.001
Grade point average (0-60 points)	0.12	[0.08, 0.16]	0.31	6.45	0.48	0.23	0.25	0.06	0.29	0.08	.000
Intelligence age 12 (mean 100, SD 15)	0.04	[0.02, 0.06]	0.16	3.52	0.37	0.14	0.14	0.02	0.16	0.03	.000
Grade retention in primary school (0 = no; 1 = yes)	-0.94	[-1.64, -0.24]	-0.12	-2.65	-0.32	0.10	-0.10	0.01	-0.12	0.01	.008
Propensity score	-1.98	[-4.53, 0.57]	-0.07	-1.53	-0.24	0.06	-0.06	0.00	-0.07	0.00	.127

Note. R^2 = .30 (n = 472, p < .01). *CI* = confidence interval for *B*. *r* = correlation of predictor with criterion. *sr* = semi-partial correlation coefficient. *pr* = partial correlation coefficient. r^2 = proportion of criterion variance accounted for by predictor (Cohen et al., 2003, p. 85). sr^2 = unique variance of predictor on total criterion variance (Cohen et al., 2003, p. 84ff). pr^2 = unique variance of predictor on criterion variance not accounted for by other predictors (Cohen et al., 2003, p. 85).

b) Regression Analysis Summary and Correlations of Childhood Variables Predicting Adulthood Income

Predictor variable	B	95% CI	β	t	r	r^2	sr	sr^2	pr	pr^2	p
Parental socioeconomic status (ISEI range 16-90)	10.39	[-18.19, 38.98]	0.05	0.72	0.18	0.03	0.05	0.00	0.05	0.00	0.474
Grade point average (0-60 points)	29.90	[9.03, 68.82]	0.13	1.52	0.38	0.15	0.10	0.01	0.11	0.01	0.131
Intelligence age 12 (mean 100, SD 15)	16.53	[-3.29, 36.35]	0.13	1.65	0.33	0.11	0.11	0.01	0.12	0.01	0.102
Years of education	146.96	[44.95, 248.98]	0.22	2.84	0.39	0.15	0.18	0.03	0.21	0.04	0.005
Grade retention in primary school (0 = no; 1 = yes)	-650.04	[-1329.53, 29.45]	-0.14	-1.89	-0.34	0.11	-0.12	0.01	-0.14	0.02	0.061
Propensity score	-428.61	[-2595.95, 1738.74]	-0.03	-0.39	-0.18	0.03	-0.03	0.00	-0.03	0.00	0.697

Note. R^2 = .24 (n = 191, p < .01). *CI* = confidence interval for *B*. *r* = correlation of predictor with criterion. *sr* = semi-partial correlation coefficient. *pr* = partial correlation coefficient. r^2 = proportion of criterion variance accounted for by predictor (Cohen et al., 2003, p. 85). sr^2 = unique variance of predictor on total criterion variance (Cohen et al., 2003, p. 84ff). pr^2 = unique variance of predictor on criterion variance not accounted for by other predictors (Cohen et al., 2003, p. 85).

c) Regression Analysis Summary and Correlations of Childhood Variables Predicting Adulthood Intelligence

Predictor variable	B	95% CI	β	t	r	r^2	sr	sr^2	pr	pr^2	p
Parental socioeconomic status (ISEI range 16-90)	0.05	[-0.18, 0.27]	0.02	0.41	0.09	0.01	0.02	0.00	0.03	0.00	.679
Grade point average (0-60 points)	0.22	[-0.16, 0.60]	0.08	1.15	0.52	0.27	0.06	0.00	0.09	0.01	.254
Intelligence age 12 (mean 100, SD 15)	0.88	[0.69, 1.08]	0.54	8.85	0.69	0.48	0.44	0.20	0.55	0.30	.000
Years of education	0.98	[0.20, 1.77]	0.15	2.47	0.46	0.21	0.12	0.02	0.18	0.03	.014
Grade retention in primary school (0 = no; 1 = yes)	-6.78	[-13.33, -0.23]	-0.12	-2.04	-0.43	0.19	-0.10	0.01	-0.15	0.02	.042
Propensity score	-6.20	[-32.21, 19.81]	-0.03	-0.47	-0.16	0.03	-0.02	0.00	-0.04	0.00	.639

Note. R^2 = .55 (n = 187, p < .01). *CI* = confidence interval for *B*. *r* = correlation of predictor with criterion. *sr* = semi-partial correlation coefficient. *pr* = partial correlation coefficient. r^2 = proportion of criterion variance accounted for by predictor (Cohen et al., 2003, p. 85). sr^2 = unique variance of predictor on total criterion variance (Cohen et al., 2003, p. 84ff). pr^2 = unique variance of predictor on criterion variance not accounted for by other predictors (Cohen et al., 2003, p. 85).

4.5 Discussion

According to data from the European Commission, every 5th student in Luxembourg repeats at least one year in primary school and thus falls behind his/her peers (European Commission, 2011). In the household study data in 1968, every sixth participant of the MAGRIP study had fallen behind by the end of primary school. Considering the retested sample is slightly positively biased compared to the Luxembourgian population in 1968, it is very reasonable to assume that the percentage of students affected by grade retention has remained fairly stable since 1968. Grade retention is commonly applied, because the measure is widely believed to have a positive impact on student's academic career (European Commission, 2011). Contrary to this belief, in the current study, grade retention in primary school had a significant negative impact on all three key life outcomes, even over 40 years of the lifespan and even when intelligence in childhood, grade point average in primary school, parental SES, educational attainment, and 11 further characteristics that are known to possibly influence grade retention and educational outcomes were controlled for. On average, participants who were retained in primary school attended one year less of formal education than promoted participants, earned about €650 less per month at age 52, and scored about 7 IQ points lower in the intelligence test at age 52.

First, we performed propensity score matching to find comparable groups of retained and not retained subsamples by controlling for 11 school, child, and family characteristics. Second, we performed a separate regression analysis for the three criterion variables: educational attainment, adult male income, and adult intelligence, respectively. Educational attainment in terms years of formal education, was predicted by childhood intelligence, parental SES, grade point average, and grade retention in primary school. For the analyses on adult income and adult intelligence, we added years of formal education as an additional predictor. The R^2 of each of the performed regression analyses on years of formal education, income, and adult intelligence was large, the predictors in the current study explained 30%, 24%, and 55% of the criterion variance, respectively. The examination of partial correlations showed that most of the explained variance in the criterion was shared by several predictors. Hence, the unique contribution of each predictor was rather small (with the exception of intelligence in childhood for the prediction of intelligence in adulthood). Thus, most of the variance in years of education, adult income, and adult intelligence was jointly explained by two or more of the included predictors. This may indicate reciprocal interactions between parental SES, grade point average, childhood intelligence, years of formal education, and grade retention over the lifespan.

In the multiple regression analyses, grade retention accounted for 10%, 11%, and 19% of the variance in years of formal education, adult income, and adult intelligence, respectively, when not accounting for the influence of the other predictors in the regression. In addition, grade retention in primary school uniquely accounted for 1% of the variance in each criterion. The unique contribution of grade retention may not be very large, but the effect of grade retention was significant for each of the criterion variables and we simultaneously controlled for other extremely important impact factors. Crucially, we found this effect 40 years after the retention decision had been taken. Thus, grade retention in primary school has very robust long-term negative consequences on several important key life outcomes.

How can a single life-event such as grade retention have such a lasting impact that is still detectable even over 40 years after the retention decision was made? The explanation could lie within a chain of reactions, which take place after the retention decision is made. We draw in lifespan developmental psychology (Baltes, 1987; Li & Baltes, 2006) to explain these negative long-term consequences of grade retention in primary school. Children interact reciprocally with their environment. It has been shown that grade retention may lead to a more negative attitude towards school (Holmes & Matthews, 1984) and lower academic self-efficacy (Hughes et al., 2013). Thus, retained students favor school less, become more likely to drop out of school (Jimerson, Anderson, et al., 2002; Jimerson, Ferguson, et al., 2002) and in general leave school early. In the current study, we found exactly this pattern: grade retention in primary school had a negative impact on length of formal education. It was predicted that retained students left school one year earlier than non-retained students, even when we simultaneously controlled for childhood intelligence, grade point average in primary school, and parental SES.

Leaving school early with less academic qualifications is associated with taking on occupations that are of a more physical nature and less cognitively challenging. These occupations are also usually less well paid. The results of Study 3 also confirm this relationship. Grade retention in primary school had a significant negative impact on adult income, while years of formal education had a significant positive impact on adult income. Childhood intelligence, grade point average in primary school, and parental SES did not affect adult income significantly in the regression analysis in Study 3. The fact that childhood intelligence did not have a significant impact on adult income is probably grounded in what Ceci and Williams (1997) describe as the indirect effect of schooling on income: A higher level of schooling is also an indicator of higher levels of intelligence. Thus, in the regression analyses, length of education better predicts adult income than childhood intelligence, as it may also be an indirect indicator of intelligence level. As a result, childhood intelligence has no incremental share in the variance of adult income in the present regression anylsis. On the contrary, it is remarkable that grade retention in primary school had an incremental share in the variance in adult income at age 52 above years of education. This shows that grade retention is not only an indicator of lower educational attainment. Thus, other mechanisms must also play a role; one possibility is that formerly retained students still have a lower level of self-efficacy, leading them to not ask for promotions as often as their non-retained colleagues.

Person-specific influences gain a larger impact on cognitive development after formal education in school ends and individuals will develop more and more in the direction of their self-perceived abilities and interests (Li & Baltes, 2006). Retained students may often receive feedback from their environment that training in general and cognitively challenging activities are not what suits them well. Thus, they might not be as eager to attend additional job training and strive for additional qualifications after formal education. Also, in their free time, they may become less likely to visit museums or engage in other cultural events. As described in the previous paragraph, retained students may also have less cognitively challenging jobs. All these developments may impact on their intelligence over the lifespan. The results of our current study reflect this explanation. Retained students had 7 predicted IQ points less at age 52 than their promoted peers. Interestingly, grade retention in primary school had again a significant incremental influence above childhood intelligence, years of formal education, grade point average in primary school, and parental SES. The effect of grade retention in primary school that is incremental and independent of the other predictors in the study may point to the impact of free time activities or occupational complexity on cognitive development over the lifespan.

In conclusion, the results of the current study suggest that the negative effects of grade retention accumulate over the lifetime in a reciprocal manner, resulting in a detectable negative effect even over 40 years after the retention decision was made. In the present study, we found a significant negative impact of grade retention in primary school on all three key life outcomes: years of formal education, adult income, and intelligence in middle adulthood. This is especially dramatic, when we remember that grade retention is actually supposed to be a measure to support a child’s development and school performance.

The present research findings suggest that alternative measures to grade retention should be considered and these results are in line with conclusions drawn by the European Commission (2011, p. 60): "*The existence of a culture of grade retention is the reason why the practice is used more often in certain countries. In these countries, the idea that repeating a year is beneficial for pupils learning remains prevalent. […] Changes in regulations on grade retention are not enough to modify this belief; it should be supplanted by an alternative approach to managing children´s learning difficulties. The challenge lies more in questioning certain assumptions and beliefs rather than regulatory change.*" The belief that grade retention is beneficial has been sustained in Luxembourg until today. In our data collected in 1968, every sixth student was retained at least once during his or her primary school career. According to recent data by the European Commission, every fifth Luxembourgian student continues to fall behind by the end of primary school (European Commission, 2011, p. 35). This is the case, although grade retention is a very costly measure, as 20% of all Luxembourgian students stay in primary school one year longer resulting in increased costs for teachers, facilities, and teaching material. Alternatives such as formative assessment combined with short, intensive interventions or individual lessons with support staff seem to be very promising alternatives in some countries (European Commission, 2008). A number of alternatives to grade retention exist and are summarized, for example, by Owings and Kaplan (2001) or McDonald and Bean (1992).

4.6 Strength and Limitations

The dataset spans 40 years from late childhood to middle adulthood and comprises very many variables on achievement as well as on child, family, school, and parental characteristics. Nevertheless, it is neither feasible nor ethically acceptable to study the effects of grade retention in an experimental design. Therefore, a possible selection bias for the retained student sample cannot be excluded. Retained children may have differed on variables that affected years of formal education, adult income, and adult intelligence that could not be controlled for in the current study. We tried to minimize this selection bias by applying propensity score matching, given the data that was available from 1968. Unfortunately, the data in 1968 was collected at the end of primary school, and thus only after most of the children had already been retained. Therefore, we could not use variables such as childhood intelligence, mathematics, German, or French grades to estimate the propensity score as we expected an impact of grade retention on the very same variables. However, the impact of these variables on our criterion variables was controlled for in the multiple regression analyses. Further, we were able to control for other important child and family characteristics that have been found to be correlated with grade retention. These covariates on child, family, and school characteristics that we included in the propensity score analyses were collected after the retention decision, but these characteristics are supposedly unaffected by the retention decision and at the same time very stable (e.g. gender, parental SES). This strategy has also been applied in other studies (Reynolds, 1992).We were able to achieve a good balance between the retained and promoted participants across the 11 covariates that had been assessed. In addition, we included childhood intelligence, grade point average in primary school, parental SES, and length of education as covariates in the multiple regression analyses where we investigated the effect of grade retention on or three key life outcomes under study. Thus, we combined two well established methods to control for selection bias in grade retention research.

Another limitation of Study 3 is that due to the reduced sample sizes after the propensity score matching procedure, we were not able to apply latent models. Due to the low sample size for some of the sub-samples, some of the models did not converge in MPlus. This precluded controlling for measurement error and controlling for measurement invariance of the latent constructs between the groups.

4.7 Conclusion

We found that grade retention has negative long-term effects on important key life outcomes, namely educational attainment, income at age 52, and adult intelligence at age 52. In the current study, we were able to control for very many variables that are associated with being retained and that have been found to affect the key life outcomes under study. However, it is neither feasible nor ethically acceptable to study effects of grade retention in an experimental design. So we cannot rule out that the retained students may differ on characteristics not assessed in the current study that are related to the examined key life outcomes. However, the negative effects of grade retention in primary school in the present study persisted even over 40 years after the retention decision had been made. For a pedagogical measure that is supposed to help the student and is also relatively expensive, these kinds of results could be sufficient reason for contemplation and reconsideration. School systems in many European countries have renounced the application of grade retention in recent years and promising alternatives do exist (European Commission, 2011). Thus, it is less a question of what can be changed and how, but more a question of changing the persisting belief that grade retention is beneficial, even though most of the existing research findings report the contrary.

Chapter 5 – General Discussion

5.1 Summary of Main Outcomes

People differ greatly with respect to developing cognitive abilities and the European Commission has identified reducing these inequalities as one of the major challenges in order to face demographic change in most European countries (European Commission, 2014). The present dissertation directly corresponds to this call. The aim of the present Ph.D. thesis was to provide a better understanding of the development of intelligence as well as the long-term impact of educational attainment on cognitive abilities. This knowledge may enable future research and practice to reduce the existing inequalities in cognitive aging. The present dissertation is based on the longitudinal MAGRIP study with first data collection in 1968 and a follow-up from November 2008 to August 2009 at the University of Luxembourg. The main outcomes as well as the theoretical and practical implications of our findings are described in the following.

5.1.1 Study 1: Development of Intelligence

The first study described in Chapter 2 tackled the question of how intelligence developed over a 40-year time span from age 12 to age 52. In more detail, we examined two key aspects: (a) stability and change in the structure of intelligence with reference to the age differentiation-dedifferentiation hypothesis and (b) differential stabilities from late childhood (age 12) into middle adulthood (age 52). To this end, we drew on two different conceptualizations of intelligence, the extended Gf-Gc Model of intelligence (Cattell, 1987; J. L. Horn & Noll, 1997), which focuses on broad abilities and the three-stratum model (Carroll, 1993), which incorporates a general intelligence factor *g*. By examining development in both models, we were able to separate change that is shared by all broad abilities (change attributable to *g*) from change specific to a certain broad ability (change independent of *g*).

Age Differentiation-Dedifferentiation Hypothesis. Stability and change in the structure of intelligence is captured by the age differentiation-dedifferentiation hypothesis. It is postulated that the structure of intelligence differentiates during childhood until late adolescence, stays fairly stable during adulthood, and dedifferentiates again in old age. However, the results of Study 1 do not fit well into the pattern expected by this theory, as we found age dedifferentiation instead of differentiation from age 12 to age 52. In more detail, in the extended Gf-Gc model, we saw a large increase in the covariances between all broad abilities, but no increase in the variance of all broad abilities. Gc and Gs showed a substantially increased variance, Gv had a much smaller increase in variance and the variance of Gf did not change significantly. When we examined the same findings in the three-stratum model, we found that the increases in covariances and the variances of the broad abilities (except Gc) could be accounted for by variance increases in *g*. Gc was an exception, additionally demonstrating a large increase in the variance that was specific to Gc, even after the influence of *g* was taken into account. This points to the conclusion that changes in Gc's variance are influenced by a source other than only *g*, for example environmental influences. Crucially, the two pure markers of fluid (Gf) and crystallized (Gc) abilities exhibited complementary patterns. This was especially visible in the three-stratum model, because the variance specific to Gc increased significantly, whereas the variance specific to Gf decreased to zero and hence became indistinguishable from *g* at age 52.

Differential Stabilities. Our results showed that persons' rank ordering across time concerning (a) their broad abilities in the extended Gf-Gc model as well as (b) their specific abilities and *g* in the three-stratum model remained largely stable from age 12 to age 52. This means that in the current study, individuals kept their relative standing with reference to the population in all broad abilities, all specific abilities, and *g*, across a time span of 40 years. Thus, in line with previous research (Conley, 1984; Deary et al., 2000, 2004), we found that the various aspects of intelligence and *g* are highly differentially stable constructs. Importantly, the results obtained for the three-stratum model also show that when individual differences in *g* are held constant, specific strengths and weaknesses in the cognitive profile (as reflected by the specific abilities) are also highly stable. Thus, it may not only be the level of an ability profile (as indicated by g; Lubinski, 2004) that remains stable across time, but also the pattern of the cognitive profile with regard to an individual's configuration of specific abilities.

Combined Effect of Age Differentiation-Dedifferentiation and Differential Stabilities. The results of the extended Cattell-Horn Gf-Gc model indicate that people differ more greatly as they grow older (shown in increased variances). In this respect, the only exception is Gf. This finding can be well explained by the predictions of lifespan cognitive psychology that crystallized abilities are more strongly affected by person-specific environmental influences than fluid abilities. Thus, people may differ more greatly in crystallized abilities as life unfolds, while the variability in fluid abilities may not change as much. However, we also found that all constructs showed high differential stability (i.e. rank order). Thus, individuals keep their relative standing in the population. The findings of the increased variance combined with the constant rank order indicates that initial differences between people on Gc, Gs, and Gv become greater and greater as people grow older. Hence, the gap between the two ends of the ability distribution widens across the lifespan. This effect (in combination with the observed increases in latent means and means of the manifest subtest scores) can be described in the words of Ceci and Paperierno (2005, p. 1) as, "*the 'have-nots' gain but the 'haves' gain even more*". In the three-stratum model, we also saw this gap widening effect in *g*. Thus, initial differences in *g* become amplified and increasingly important as life unfolds.

5.1.2 Study 2: The Impact of Educational Attainment

The aim of Study 2, described in Chapter 3, was to examine whether educational attainment in terms of quantity (i.e. years of formal education) and quality (i.e. type of school track) affects the development of cognitive abilities over the lifespan. As fluid (Gf) and crystallized (Gc) cognitive abilities are predicted to be differently affected by biological and environmental influences (Li & Baltes, 2006), we examined the impact of formal education separately for Gf and two separate aspects of Gc, namely word knowledge (WK, verbal measure) and knowledge of the world (KW, factual knowledge measure; Schipolowski et al., 2014). This also allowed us to test central assumptions made by Cattell's investment theory (Cattell, 1971, 1987). In addition, a crucial aspect of Study 2 is that we were able to control for childhood cognitive ability and that we examined the long-term impact of formal education after participants had left formal education over 30 years previously. To our knowledge, no other study has examined the long-term impacts of formal education on cognitive abilities over such a long time period.

Main Effect of Quantity of Formal Education on Cognitive Abilities. Years of formal education had a small effect on Gf at age 52 but not on Gc (i.e. word knowledge) at age 52 for the entire sample. Thus, length of formal education had a positive effect on fluid abilities but not on crystallized abilities for the entire sample. A different pattern was found for the other facet of Gc in the present study, knowledge of the world: Years of formal education had a very large effect on KW at age 52 in the entire sample. Thus, the two facets of crystallized ability in the present study related differently to formal education. This may provide further evidence that verbal ability and factual knowledge are two distinct aspects of Gc that can be empirically distinguished (Schipolowski et al., 2014). However, we could not control for autoregressive effects over time for KW, which hinders a clear attribution of the different pattern of effects

Main Effect of Quality of Formal Education. In the current study, we did not find a significant main effect of quality of formal education as assessed by school track. This is in contrast to previous research findings (see Becker et al., 2012; Husén & Tuijnman, 1991; Shavit & Featherman, 1988). The results of the current study pointed into the direction of a more positive effect on cognitive abilities for the academic track compared to the non-academic track. But the difference between the tracks did not reach significance. Several explanations are possible for this effect. One is that a main effect of school track existed in the current study, but the sample size was too small to find a significant effect. However, in addition to the non-significant difference between the latent means between tracks, we also saw in the manifest variables that the difference *between* the tracks did not increase with increasing age in terms of effect size. Thus even though the academic track students gained more on Gf and WK over time, the difference between groups remained stable in terms of effect size as the overall variance increased. Thus, maybe the lack of a main effect of quality of education is not only a statistical power problem. The lack of a main effect of school track may also indicate that formal education in the two tracks was tailored to the academic and cognitive needs of the attending students. Thus, it would not be possible to say that the academic track is better for everyone in general, but more a function of fit between student's abilities and school track. In addition, the contrast to other findings may be grounded in a number of methodological differences between the studies, for example country specific school systems, length of education, or quantity of schooling in different tracks (i.e. number of hours per day and years of schooling per track).

Interaction Between Quantity and Quality of Formal Education. In the current Ph.D. thesis, we found an interaction between quantity and quality of formal education for

WK and Gf at age 52. Years of formal education had a significant positive effect on WK_{52} and Gf_{52} for the non-academic track students, but not for the academic track students. The effect was of small to medium effect size for WK_{52} and Gf_{52}, even after the participants had left formal education over 30 years previously. On the contrary, we did not find an interaction between quantity and quality of formal education for KW at age 52. Years of formal education had a large impact on KW at age 52 in the entire sample, as well as in the non-academic and the academic track. However, KW was assessed at age 52 only. Thus, the difference between WK and KW may be grounded in the different nature of the two abilities or the lack of the longitudinal control.

Cattell's Investment Theory. Cattell's investment theory (1987) postulates that fluid abilities are invested into the acquisition of crystallized abilities by taking advantage of environmental learning opportunities such as formal education. Such an effect could be shown in (a) an indirect effect of Gf_{12} on WK_{52} and KW_{52}, respectively, over years of formal education and (b) a direct effect of Gf_{12} on WK_{52} and KW_{52}, respectively. The direct effect may map other learning opportunities than formal education. For word knowledge as dependent variable at age 52, we did not find an indirect effect, but we found direct effect of Gf_{12} on WK_{52}. Thus, it is possible that childhood abilities in Gf were invested in the acquisition of WK_{52}, but the learning opportunities were not provided by formal education. This explanation may be feasible in the case of word knowledge, as this ability can be trained whenever a person is faced with verbal material. Thus, the learning opportunities provided by formal education may only play a minor role in the acquisition of WK especially after formal education. For knowledge of the world as a dependent variable at age 52, we found an indirect effect for both Gf_{12} and WK_{12} on KW_{52} through years of formal education. However, no direct effect of Gf_{12} on KW_{52} was found. Thus, in terms of Cattell´s investment theory, it is feasible to assume that both Gf_{12} and WK_{12} were invested in the acquisition of knowledge of the world at age 52 by taking advantage of learning opportunities provided by formal education. In addition, opposing propositions by Cattell's investment theory (Cattell, 1971, 1987), we also found an indirect effect of WK at age 12 on Gf at age 52 via years of formal education. This may indicate that WK, a crystallized ability, was invested in the acquisition of Gf via formal education.

5.1.3 Study 3: The Impact of Grade Retention in Primary School

Study 3 tackled the question whether grade retention in primary school affects educational attainment, adult income, and adult intelligence. Grade retention is a very frequently applied pedagogical measure in Luxembourg. In our sample from 1968, every sixth students was retained at least once in primary school. According to more recent data from the European Commission every fifth student continues to be retained in primary school in Luxembourg today (European Commission, 2011).

The methods applied in Study 3 varied from Study 2 in two ways: (a) the applied intelligence model and (b) the statistical approach. This was done for the following reasons. (a) In Study 2, we examined the impact of formal education on fluid and crystallized abilities separately as we wanted to test several hypotheses that predicted a different pattern of development for fluid and crystallized abilities (Cattell, 1987; Li & Baltes, 2006), remembering, that two-component theories of intelligence are frequently applied in developmental psychology (Lindenberger, 2001). However, in the study on grade retention (i.e. Study 3), previous research has made no distinction in between fluid and crystallized abilities. Thus, no hypotheses existed that predicted a differential pattern of the impact of grade retention on fluid and crystallized abilities. Therefore, we chose to measure intelligence in terms of *g*. This is also in accordance with the tradition of other applications of *g* theories, where *g* is applied to relate intelligence to other measures (Ackerman & Lohman, 2003), as was done in Study 3.

(b) The statistical data analyses approach for Study 2 and Study 3 was chosen to best address the research question under study and best control for the most critical confounding factors. In Study 2, we applied latent multi group models to study the effect of formal education on the development of Gc and Gf across age by controlling for childhood Gc and Gf. Thus, we were able to compare latent constructs, where we could control for measurement error. In addition, we were able to establish scalar group invariance between the non-academic and the academic group to ensure that the measurement model of the two groups did not differ. We also controlled for the most important impact factor on adult intelligence, which is childhood intelligence (Deary et al., 2000). In addition, we simultaneously researched the influence of the probably second most important impact factor on adult intelligence, formal education. In Study 3, we applied a propensity score matching procedure and multiple regression models to study the effect of grade retention on years of education, adult income, and adult intelligence. Propensity score matching is applied, to address the problem of not being able to apply randomization in the study design. Thus, the selection into the retained and the promoted group may be biased by other variables known to influence the retention decision as well as the dependent variable under study. In the case of grade retention and the dependent variables under study in Study 3 ample research exists that has identified such confounding variables. Thus, controlling for these confounding variables such as parental SES, grades, childhood intelligence, years of education, as well as other child, family, and school characteristics, was the main priority in Study 3. However, applying propensity score matching has the downside that it reduces the sample size. Therefore, due to the low sample sizes in Study 3, we were not able to apply latent models and control for measurement error. In addition, we could also not control for measurement invariance between the groups, which we would have done in an ideal world. On the contrary, in Study 2, we would have ideally applied propensity score matching to control for confounding variables on the tracking decision that also affect cognitive abilities. However, the sample sizes were already rather small for the applied latent models. Thus, both in Study 2 and Study 3, we could have controlled for a number of additional influences but the accompanying disadvantages seemed larger compared to the benefits for the research question under study. Hence we chose the statistical approach that controlled best for the most important confounding factors for the research question under study.

The Impact of Grade Retention. Grade retention is applied because it is believed to be beneficial to the student's academic career. However, contrary to this believe, grade retention in primary school had a significant negative impact on all three key life outcomes in the current study, even over 40 years of the life time and even when intelligence in childhood, grade point average, parental SES, educational attainment, and 11 further characteristics that are known to possibly influence educational outcomes are controlled for. On average, children that were retained in primary school attended one year less of formal education than promoted children, earned about 650€ less per month at age 52, and scored approximately 7 IQ points lower on the intelligence test at age 52. The non-incremental share of criterion variance of grade retention in the regression model was (1) 10% of the variance in educational attainment, (2) 11% in the variance of male income at age 52 and (3) 19% of the variance in intelligence at age 52. The incremental significant contribution of grade retention in the applied regression models was 1% for each of the criterion variables under study. Thus, the incremental share might not be large, but still impressive when we consider that the retention decision took place over 40 years ago and we simultaneously controlled for many of the probably most important other predictors of the criterion under study. Moreover, the examination of partial correlations showed that most of the explained variance in the criterion was shared by several predictors. Hence, the unique contribution of each predictor was rather small (with the exception of the intelligence in childhood, intelligence in adulthood prediction). Thus, most of the variance in years of education, adult income, and adult intelligence was jointly explained by two or more of the included predictors. This indicates strong interactions among parental SES, grade point average, childhood intelligence, years of education, and grade retention over the life time. It also showed that the incremental share of grade retention in the current regression analyses was comparable to the other important predictors in the regression.

5.2 Theoretical Implications

The empirical results of each of the studies in the present dissertation suggest a number of theoretical and practical implications. In the following, we will first discuss the theoretical implications for the underlying theoretical frameworks of the current Ph.D. thesis. The next section will then focus on practical implications and deductions of the results of the present dissertation.

5.2.1 Lifespan Developmental Psychology sensu Li and Baltes

In the current dissertation, we drew on a number of theoretical assumptions by Lifespan Developmental Psychology (Baltes, 1987) or Lifespan Cognitive Psychology sensu Li and Baltes (2006). First, lifespan cognitive psychology distinguishes between fluid and crystallized abilities. While crystallized abilities are more sensitive to environmental influences, fluid abilities are more affected by biological processes. Thus, as life unfolds and person-specific environmental influences accumulate, it is postulated that people will become increasingly different with regard to crystallized but not so much with regard to fluid abilities. The results of the current dissertation can mostly confirm these postulations. In Study 1, we did find a different pattern of development for more crystallized and more fluid abilities. The variance in crystallized abilities increased from age 12 to age 52, while the variance of fluid abilities stayed largely stable across age. This pattern was found for broad abilities as well as for specific abilities and *g*. Thus, as life unfolds and environmental person-specific influences accumulate, people become more different with regard to crystallized but not so much with regard to fluid abilities. It would be possible to conclude from this finding that crystallized abilities are more sensitive to environmental influences than fluid abilities and that the plasticity of fluid abilities to environmental influences is rather limited. However, in Study 2 we even found that length of formal education had a larger impact on fluid than on crystallized abilities. This effect was especially pronounced for the non-academic track participants. These findings are contrary to the proposition that fluid abilities may be less sensitive to environmental influences (Li & Baltes, 2006). The results of Study 2 show that the plasticity of fluid abilities to environmental influences such as formal education may be largely underestimated in the current literature. This knowledge is important as it enables many possibilities to support and maintain cognitive functioning in terms of crystallized *and* fluid abilities into old age for a society that is constantly growing older.

Taken together, the results of the current dissertation underline the proposition by lifespan developmental psychology that crystallized and fluid abilities are distinct abilities and that people grow more apart on crystallized than on fluid abilities as life unfolds and person-specific influences accumulate. Until now, little is known about the relative plasticity across different aspects of cognition (Li & Baltes, 2006). However, the results of the present dissertation show that crystallized abilities may be more sensitive to the environment, but the plasticity of fluid abilities to environmental influences must not be underestimated. In the current dissertation, we found that formal education had a significant and very persistent impact on fluid abilities from childhood into middle adulthood.

The second theoretical proposition that we tested was the age differentiation-dedifferentiation hypothesis (Baltes et al., 1980) on the stability and structure of intelligence over the lifespan. The age differentiation-dedifferentiation hypothesis proposes three stages.

The first stage of differentiation occurs during maturation, when different broad abilities are proposed to become increasingly independent of each other with increasing age (Deary et al., 1996). This effect is statistically represented by a decline in intercorrelations among broad abilities in the extended Gf-Gc model (Deary et al., 2004) or a progressively decreasing role of the influence of *g* in the three-stratum model (Escorial et al., 2003). The following period of adulthood is described as a stage of stability in the structure of intelligence (Baltes et al., 1980). Thus, statistically no changes in the magnitude of intercorrelations among broad ability factors or the influence of *g* would be expected. The third stage of dedifferentiation is characterized by again increasing dependencies among different broad abilities as people reach old age (Baltes et al., 1980). This effect is statistically represented by increases in intercorrelations among broad abilities in the extended Gf-Gc model (Deary et al., 2004) or an increasing influence of *g* in the three-stratum model (Escorial et al., 2003).

While we observed that the participants grew increasingly different in terms of almost all their cognitive abilities (except Gf) from age 12 to age 52 in the current study, all cognitive abilities under study dedifferentiated at the same time. Thus, we saw that covariances (and intercorrelations) among all broad abilities increased in the extended Gf-Gc model and we also found that these increases could be explained by *g* in the three-stratum model. Such a pattern describes age dedifferentiation and indicates that the different cognitive abilities become more and more similar. However, according to the hypothesis age dedifferentiation is not supposed to occur before old age. Thus, our results do not fit well with the assumptions described above.

The theoretical assumptions of the age dedifferentiation-dedifferentiation hypothesis are intuitively very comprehensible. However, the statistical interpretation may have been too simplistic to clearly and unambiguously prove the postulations made by the hypothesis. Increasing correlations among different broad abilities may indicate changes in the construct in the Gf-Gc model of intelligence. However, correlations are a product of two parameters, covariances and variances. Thus, an increase in intercorrelations among factors can be caused by (a) an increase in the covariances given the variances stay stable, (b) a decrease in the variances given the covariances stay stable, or (c) a increase in both but so that the covariances increase proportionally more than the variances, (d) a decrease in both so that the variances decrease proportionally more than the variances. Importantly, each of the above described mechanisms may point to a different explanation of why the intercorrelations among different intelligence factors increases. In the current study, we found mechanism "c", both covariances and variances increased. However, almost all the increases in the variances of the broad abilities under study were shared by two or more broad abilities. Thus, almost the entire increase in variances was also an increase in covariances. This was also shown in the three-stratum model as the increase in variance in broad abilities could almost completely be explained by *g* alone (except for Gc). Thus, the smarter participant's became smarter in *all* broad abilities, while the not so smart participants did also become smarter on *all* broad abilities but much less compared to the smart. This is the effect that we saw for the gap widening effect described by Ceci and Papierno (2005) as "*The 'have-nots' gain but the 'haves' gain even more*". Hence, the dedifferentiation among different factors of intelligence in the present study cannot be clearly attributed to a change in factor structure, but is much more a result of the gap widening effect in the population.

Several other explanations not related to a change in the construct may also account for higher correlations among broad abilities. For example, intelligence and educational attainment are two constructs that are highly related. Some studies report correlation coefficients as high as .81 (Deary, Strand, Smith, & Fernandes, 2007). However, most scholars in the field would not go as far as to assume that intelligence and educational achievement are the same construct. Most researchers would probably state that the two constructs capture different aspects of two things that are admittedly closely related to each other. As described in more detail in Study 2 of the current dissertation, education and intelligence may be reciprocally related to each other over the lifespan (Li & Baltes, 2006). A reciprocal relationship over time will most probably also result in higher variances and covariances on the reciprocally related constructs as people age. But this does not imply that the two constructs become the same.

A second example from Study 1: in the three-stratum model we saw that both individual differences in *g* and the specific abilities were highly stable across age. Thus, it may not only be the level of an ability profile (as indicated by *g*) that remains stable across time, but also the pattern of the cognitive profile with regard to an individual's configuration of specific abilities (Lubinski, 2004). Hence, we saw that specific profiles or talents are stable characteristics across the lifespan. These talents can even be seen in small children; some learn to talk more quickly while others love to do puzzles. Thus, when the factor structure would collapse in old age, these different profiles of specific abilities would also collapse in old age, which is not so intuitively conceivable. Instead it is much more feasible that only the level of the profile as indicated by *g* may drop in old age, which, in turn, would affect performance in all the specific abilities. However, the particular strengths and weaknesses of each person would persist, just on a lower general level. Yet, as all specific abilities are affected by the decrease in *g*, differences between different specific abilities could become smaller and this could make it harder to identify the stable patterns of specific abilities even though the factor structure would theoretically persist into old age.

Taken together, the results of the current dissertation cannot support the age-differentiation-dedifferentiation hypothesis, as we found statistical dedifferentiation from late childhood to middle adulthood. A possible alternative interpretation may be to assume stable ability profiles over the lifespan with a changing influence of *g* as an indicator of the level of the ability profile, instead of changes in the factor structure of intelligence. This interpretation would also be supported by the high differential stabilities of all broad abilities, all specific abilities, and *g*. In addition, the increased intercorrelation among different broad abilities found in the current study could also be well explained by reciprocal relationships over the lifespan. In addition, investment theory provides another possible mechanism of how intercorrelations of different broad abilities increase across the lifespan and will be described in the next section.

5.2.2 Investment Theory

Investment theory (Cattell, 1967, 1971, 1987) proposes that fluid abilities are invested in the acquisition of crystallized abilities by taking advantage of environmental learning opportunities. Kvist and Gustafsson (2008) have further argued that if this proposition holds true, Gf and *g* should be the same entity, because Gf is postulated to be involved in all kinds of learning (see also Kan et al., 2011). The assumption by Kvist and Gustafsson (2008) is supported by the present study's finding that Gf became indistinguishable from *g* at age 52. In addition, if Gf really is involved in all kinds of learning, this will also result in an increased interrelation between Gf and other broad abilities as people age. In turn, as increases in all broad abilities have Gf's involvement in common, the interrelations among all broad cognitive abilities should also rise. These mechanisms described by investment theory (Cattell, 1971, 1987) can also explain the shared increase in both variances and covariances of broad abilities that can be attributed to increases in *g* variance in Study 1 of the current dissertation. Thus, investment theory may provide an explanation of the dedifferentiation among cognitive ability constructs across age accompanied by widening gaps in the cognitive ability distribution in the population.

Contrary to the above described mechanisms of investment theory that can explain dedifferentiation among abilities, investment theory has been drawn on in previous research to explain age differentiation of ability constructs (Tucker-Drob, 2009; Tucker-Drob & Salthouse, 2008). The reason for doing so was that investment theory proposes that different people's environments become increasingly heterogeneous as life unfolds. Crystallized abilities are more sensitive to the environment than fluid abilities. Therefore, crystallized abilities may be more affected by the environment and thus crystallized and fluid abilities were proposed to grow increasingly apart as life unfolds. However, as described in the previous section, this line of reasoning explains the increase in inter-individual variability (i.e. variance) in crystallized ability well. Yet, it may explain why *people* become more different, it does not necessarily explain, why the *intelligence constructs* should become increasingly different as life unfolds. On the contrary, if fluid ability is involved in all kinds of learning as interpreted by Kvist and Gustafson (2008), the mechanisms of investment theory would lead to an increase in interrelation among different broad abilities as described above. Thus, it would be the people that differ more greatly with increasing age, but not necessarily the factor structure. This is in line with what we found in Study 1 of the current dissertation. Thus, the proposition that investment theory explains age differentiation (Tucker-Drob, 2009) cannot be supported by the pattern of results found in the current dissertation.

In Study 2 of the current dissertation, we were able to examine whether fluid abilities were invested in the acquisition of crystallized abilities by taking advantage of learning opportunities provided by formal education. Interestingly, we found a different pattern of results for the two crystallized dependent abilities under study, namely word knowledge and knowledge of the world. We found that fluid reasoning in childhood influenced word knowledge in adulthood. However, we did not find evidence that learning opportunities provided by formal education played a role in this process. Thus, fluid reasoning may have been invested in the acquisition of word knowledge by taking advantage of learning opportunities other than formal education. This may be very feasible as word knowledge is likely trained in everyday life, when opening the paper, switching on the television, or reading on the Internet. In addition, the participants in the current study had left formal education 30 years before word knowledge ability was reassessed, thus the influence of formal education on word knowledge may have been overlaid by other learning opportunities since participants left school. A different pattern of results was found for knowledge of the world. We found that fluid reasoning was invested in the acquisition of knowledge of the world during learning opportunities provided by formal education. However, other learning opportunities may have not played a significant role as we did not find a direct effect of Gf in childhood on knowledge of the world in adulthood. Thus, as proposed by investment theory, we did find that fluid reasoning was involved in the acquisition of word knowledge and knowledge of the world. However, learning opportunities provided by formal education only played a role for knowledge of the world, but not so much for word knowledge. Two explanations may account for this finding: (a) a more methodological explanation as we could not control for childhood knowledge of the world in the current study, which might have led to an overestimation of the effect of formal education. (b) Word knowledge as a more verbal measure and knowledge of the world as factual knowledge measure may be differently affected by formal education and learning opportunities outside of formal education. Similarly, Schipolowski and colleagues (2014) have found that verbal ability and factual knowledge are distinct abilities that may capture two different aspects of crystallized ability.

In addition, contrary to the predictions made by investment theory, we also found that word knowledge in childhood was invested in the acquisition of knowledge of the world and fluid reasoning in adulthood by taking advantage of learning opportunities provided by formal education. This finding was especially interesting as we did not expect that word knowledge had any task-relevance for the measure of fluid reasoning. A possible explanation of this finding could be that the broad ability word knowledge is to some degree affected by *g*, as shown in the three-stratum model in Study 1. Thus, it could be the *g* component in word knowledge that is invested in the acquisition of knowledge of the world and fluid reasoning during formal education. However, as we did control for the investment of childhood fluid reasoning that is very closely related to *g* in the current study (see results of Study 1), the investment of word knowledge into knowledge of the world and fluid reasoning may exceed *g* component in word knowledge. In any case, these mechanisms show that fluid and crystallized abilities as well as formal education are reciprocally related to each other over the lifespan, which, in turn, most probably leads to dedifferentiation of different broad ability constructs or respectively a more pronounced influence of *g*. Importantly, the investment of crystallized abilities, especially in fluid abilities, has seldom been studied in previous research and may be largely underestimated.

To summarize, the predictions of investment theory that fluid abilities are invested in the acquisition of crystallized abilities are supported by the data in the present study. Depending on the crystallized ability under study, the investment of fluid ability into crystallized ability took place during formal education or outside of formal education. However, the effects were not of large effect size and partially only marginally significant. One explanation could be that the time interval of the current study may be too long to test the postulations of investment theory in detail. Investment theory describes the process of how fluid abilities affect the acquisition of crystallized abilities. Thus, more points of measurement with shorter intervals would be necessary to test the assumptions of investment theory in more detail and in order to find larger effect sizes. In the present study, we had a time interval of 40 years between measurements and despite this long time interval, we still found effects to support the postulations made by investment theory. In addition, the results of the present dissertation indicate that the investment of crystallized abilities into fluid abilities may deserve more attention than it has received in the past.

5.2.3 “*The ‘have-nots’ gain but the ‘haves’ gain even more*” or the Matthew Effect

The term “Matthew effect” is used to describe cumulated advantages as people grow older such as the sociological phenomenon when the “rich get richer and the poor get poorer”. In fact, these mechanisms are not new and seem to occur in many different life domains as life unfolds. Ceci and Papierno (2005) have called this phenomenon, “*the ‘have-nots’ gain but the ‘haves’ gain even more*”. In Study 1 of the current dissertation, we found exactly these effects in cognitive abilities from late childhood to middle adulthood. The effect is shown by the stable differential stabilities in combination with the increases in variance in almost all broad abilities (except Gf) in the extended Gf-Gc model as well as almost all specific abilities (except Gf) and *g* in the three-stratum model. Thus, initial differences in cognitive abilities in childhood became amplified in middle adulthood and inequalities in cognitive abilities become more and more apparent as life unfolds. Ceci and Papierno (2005) have pointed out that intervention programs will lead to gap widening effects, when they are made available for the entire population. The ‘haves’ and ‘have-nots’ will unequally profit from the intervention with a more positive effect for the ‘haves’. Thus, it is probably best to install intervention programs as early as possible in life to minimize increasing inequalities in cognitive abilities. Small improvements in childhood may amplify and result in large improvements in adulthood. Thus, in order to tackle the challenges of the demographic change in Europe, it is of crucial importance to start as early in life as possible. As pointed out by the European Commission (2014), reducing inequalities in cognitive functioning will be a major challenge for the future. The European population is growing older than ever before and healthy cognitive aging is not only important for well-being in old age, but also for the increasingly aging work force in Europe. Formal education in school may be a promising instrument to close some of the gaps in cognitive abilities. Formal education in school affects virtually all children in a certain country at a time in life when environmental interventions are still most effective (Li & Baltes, 2006). However, as pointed out before, a treatment that affects all the population will lead to gap widening effects because it will foster accumulating advantages in the population. Thus, if formal education affects cognitive abilities, attention has to be paid on how formal education is designed so that it will indeed reduce the gaps instead of widening them even more.

Study 2 of the current dissertation examined the impact of quantity (in terms of years of schooling) and quality (in terms of school track) of formal education on the development of cognitive abilities over the lifespan. We found that 30 years after the participants had left formal education, years of formal education was still positively and significantly related to the cognitive abilities of the students that had visited the non-academic track. However, no effect of formal education was found on the cognitive abilities of students that had visited the academic track. This finding may be explained by lifespan developmental psychology in combination with cumulating advantages. Lifespan developmental psychology proposes that person-specific professional influences gain in importance after individuals leave formal education. Thus, after formal education people develop in the direction of their interests and professions. The students in the academic track left school with a school degree that opened the door to cognitively challenging occupations. Thus, for these people, many advantages accumulated over the lifetime and led to a professional and leisure environment that may have kept them cognitively very challenged. Therefore, the influence of formal education may be overlaid or compensated by other cognitively challenging environmental person-specific influences in the 30 years after the participants of the academic track had left school. On the contrary, the later professional and leisure environment of individuals in the non-academic track may have been less cognitively challenging. The students in the non-academic track may have taken on occupations of a more physical and less cognitively challenging nature. Thus, in the non-academic track group, the influence of formal education may not be overlaid or compensated by environmental person-specific influences as these person-specific influences may have not been as cognitively challenging as formal education. Another aspect of why the effects of formal education may be so pronounced for the non-academic track is that if an upward spiral of cumulating advantages exists, it is very likely that a downward spiral of cumulating disadvantages may also exists. Thus, for the former participants in the non-academic track, formal education may have worked against these cumulating disadvantages. Crucially, these findings show that formal education may indeed be an important instrument when it comes to reducing gaps in cognitive abilities. The cognitive abilities of participants in the non-academic track profited from formal education over 30 years after leaving school. Thus, it is especially this, probably lower performing, subgroup in the population that can profit from formal education and additional structured interventions or programs. We could also say that this subgroup is much more dependent on external help and programs to keep them cognitively challenged and to possibly stop the accumulation of disadvantages. Thus, despite the fact that the 'haves' profit more from interventions, the interventions are of crucial importance especially for the 'have-nots' and special programs are needed that focus on the 'have-nots' in order to effectively reduce the cognitive gap. In addition, these interventions should start as early as possible in the lifetime, when gaps are still small and children's cognitive abilities are still very sensitive to environmental influences.

5.3 Practical Implications

In addition to the above described theoretical implications, the findings of the present Ph.D. thesis may provide some practical implications concerning the practice of grade retention. In addition, a number of additional practical implications can be deduced from the findings in the current dissertation, for example concerning early interventions, career counseling, and school tracking.

5.3.1 Grade Retention in Primary School

Grade retention is the practice that students who fail to attain the competencies required to achieve a certain grade level have to repeat their current grade level. The pedagogical measure is applied with the idea that an additional year of maturity and exposure to the curriculum will improve the students' academic achievement and social success in future grade levels. Opposed to this theoretical idea, we found negative long-term impacts of grade retention in primary school on three key life outcomes, namely educational attainment, adult income, and adult intelligence, 40 years after the students had been retained. How can a single life event have such an impact? The explanation is that grade retention is a decision, but the consequences of this decision have a great influence on the student to school relationship. Lifespan developmental psychology (Baltes, 1987; Li & Baltes, 2006) proposes that children interact reciprocally with their environment, thus the negative effects of grade retention are likely to accumulate over the lifespan. Ample previous research findings have shown that grade retention has several negative impacts on a student's academic career . For example, grade retention leads to a negative attitude towards school (Holmes & Matthews, 1984) and lower academic self-efficacy (Hughes et al., 2013). Thus, retained students favor school less and become more likely to drop out of school (Jimerson, Anderson, et al., 2002; Jimerson, Ferguson, et al., 2002). Moreover, retained students are more likely to leave school early. The results of the current study confirm this line of reasoning. The results of Study 3 showed that grade retention had a negative impact on length of formal education. It was predicted that the retained students left school one year earlier than not retained students even when simultaneously controlling for childhood intelligence, grade point average in primary school, parental SES, and 11 other variables that are possibly associated with grade retention.

Leaving school early with fewer academic qualifications is associated with taking on occupations that are more of a physical nature and less cognitively challenging. These occupations are also usually less well paid. The results of Study 3 also confirm this relationship. Grade retention in primary school had a significant negative impact on adult income, while years of formal education had a significant positive impact on adult income. Childhood intelligence, grade point average in primary school, and parental SES did not affect adult income significantly in the regression analysis in Study 3. Crucially, it is remarkable that grade retention in primary school had an incremental share in the variance in adult income at age 52 above years of education or childhood intelligence. This shows that grade retention is not only an indicator of lower educational attainment or lower childhood intelligence. Thus, other mechanisms than only lower educational attainment must play a role in why retained students in primary school have lower incomes in middle adulthood. One possibility is that formerly retained students still have a lower self-efficacy so that they may not ask for promotions as often as their non-retained colleagues.

As described above, retained students are more likely to have a less cognitively challenging job environment and probably also a less cognitively challenging leisure environment, which is in turn negatively related to cognitive functioning (Schooler & Mulatu, 2001; Schooler et al., 1999). In Study 3, we found a significant negative impact of grade retention in primary school on adult intelligence at age 52. Interestingly, again grade retention had a significant incremental influence above childhood intelligence, years of formal education, grade point average in primary school, and parental SES. Crucially, even after childhood intelligence was controlled for, grade retention and years of education still had a significant and incremental impact on adult intelligence.

Taken together, the results of Study 3 in the present dissertation tell a dramatic story of a pedagogical measure in childhood that was supposed to do good for the student, but that had negative consequences even 40 years after it had been applied. Contrary to the belief that grade retention is beneficial to the student, it has long lasting negative consequences on several key life outcomes, as shown in the present dissertation. Crucially, grade retention in primary school is a measure that is applied when children are most sensitive to environmental interventions, this may also be one reason why grade retention in primary school has such persisting negative consequences. Unfortunately, the belief that grade retention is beneficial has sustained until today in Luxembourg and grade retention is still a very frequently applied pedagogical measure. In 2011, the European Commission (2011) has made clear that grade retention is not beneficial for students and that changing existing regulations on grade retention will not be enough to solve the problem. Instead, the European Commission has called for a change in the belief that grade retention could be beneficial. Alternative approaches to help students with learning difficulties at school are needed. Alternatives such as formative assessment combined with short, intensive interventions or individual lessons with support staff seem to be very promising alternatives in some countries (European Commission, 2008). However, future research is needed to identify alternative approaches as well as to evaluate these alternative approaches with regard to long-term consequences.

5.3.2. Early Interventions

As stated by Heckman (2008) „Ability gaps between the advantaged and disadvantaged open up early in the lives of children". In the current study, we have seen that initial differences in late childhood become more and more pronounced until middle adulthood. In addition, the results of the present dissertation show that cognitive abilities and formal education relate reciprocally to each other and advantages accumulate. This is one mechanism of how initial differences become increasingly pronounced as life unfolds. Thus, the results of the present dissertation suggest that reducing cognitive differences at an early age will be most beneficial and effective in closing or minimizing large cognitive gaps later in life. This is in line with arguments by Heckman (2008), who has shown that early interventions have high benefit-cost ratios and rates of return. Early interventions, especially for children from disadvantaged families, have been empirically proven to promote educational attainment, reduce crime, foster workforce productivity and reduce teenage pregnancy (Heckman, 2008). In addition, interventions that have started at kindergarten age, have been shown to positively affect cognitive ability up until early adulthood (Heckman, 2008). Early interventions are much more promising than late interventions as they affect children's cognitive abilities at a stage in life, when their biological plasticity is still very high and they are still very sensitive to environmental influences (Baltes, 1987; Li & Baltes, 2006). Thus, although the data from the present dissertation starts in late childhood, it is very likely that gaps between advantaged and disadvantaged children already open up much earlier in life. Heckman (2008), for example, shows that ability gaps at age 12 already existed when the children entered primary school. In addition, Heckman (2008) argues that children of disadvantaged families that life in low quality nurturing environments (i.e. low parenting quality) are most dependent on external early interventions. This is similar to the finding of the present dissertation that especially the students in the non-academic track were in need of external help. Thus, it is particularly important to help the more disadvantaged children as early as possible, so that the cognitive and also non-cognitive ability gaps are not enlarged by the dynamic and reciprocal processes of skill and ability formation.

5.3.3 Career Counseling – Educational Diagnostics (Förderdiagnostik)

In the present dissertation, we found that all broad abilities, as well as *g* and the specific abilities showed very high differential stabilities from late childhood into middle adulthood. Thereby, the configuration of specific abilities can be seen as the specific pattern or profile of talent, while the level of *g* may indicate the level of the profile (Lubinski, 2004). Previous research has shown that these profiles are important to understand why certain learning and working environments are found attractive or not (Shea et al., 2001). While the profile of specific abilities is important to predict the content of the learning or working environment, the level of *g* will predict the level of complexity of the learning or working environment (L. S. Gottfredson, 2003; Lubinski, 2004). As a simple example, two students with a profile of specific abilities that points in the medical-social direction may choose jobs in the healthcare system. The student with the higher level in *g* is likely to become a medical doctor, while a lower level in *g* may lead to an apprenticeship to become a nurse. As the results of the current dissertation have shown that the inter-individual profiles as well as the level of *g* are highly stable across the lifetime, career counseling may already start very early in childhood. Today, career counseling in schools starts only very late in secondary school, mostly only shortly before students leave formal education. Thus, as school is supposed to prepare for later life and work, it would be justified and important to offer courses that make students familiar with different job families much earlier in the school career. This, in combination with targeted diagnostics to identify ability profiles and level, could facilitate career choices for many students.

Intelligence tests can help to provide optimal cognitive fostering for children by identifying slow learners and gifted children (Anastasi & Urbina, 1997; Hunt, 2011). Thus, the early and reliable identification of strengths and weaknesses in the ability profile can also be used for educational diagnostics (Förderdiagnostik). The results of the current study show that strengths and weaknesses in a certain profile can be identified very early in life. This would enable educationists, teachers, parents, and psychologists to intervene at a very early age, so that the gap between children in the same class does not become too large. In addition, knowledge about strengths in the profile can be used to balance weaknesses in the profile.

5.3.4 School System

Two implications related to the school system can be deduced from the current dissertation's findings. The first implication concerns the tracking decision in tracked school systems and whether the tracking decision should be more strongly based on cognitive abilities or not. Per se, tracked school system may bear the threat that they increase the ability and attainment gap between the lower performing and the better performing students (DiPrete & Eirich, 2006). Although, in the present dissertation, no significant evidence was found that the higher school track improved cognitive abilities, the results of the current dissertation pointed in this direction and may have been significant with more statistical power. Thus, basing the tracking decision more strongly on cognitive abilities may widen the gap even more. However, cognitive abilities are important determinants of academic success in school (Jensen, 1998), yet other non-cognitive abilities such as socioemotional skills, attention, motivation, and self-confidence are also of importance for later success (Heckman, 2008). Thus, academic achievement in terms of grades is always a result of cognitive abilities as well as non-cognitive abilities, plus time spent practicing what has been taught in school. In other words, academic achievement (i.e. grades) can be seen as a result of cognitive ability as an indicator of potential as well as the actual realization of this potential (Ceci & Williams, 1997). Therefore, grades as an indicator of the realized potential are the better indicator for future academic success and may be the better main indicator for the tracking decision. Grade point average has also been the predictor with the highest standardized regression coefficient (beta) in the analysis on years of education in Study 3.

Nevertheless, cognitive abilities can be used as a diagnostic tool before the tracking decision. One or two years before the tracking decision is to be made, cognitive abilities can provide some insight into whether students may be underperforming compared to their actual potential in terms of cognitive abilities. This could be the case for students with high ability levels but poor grades. For these students, additional measures could be taken to improve their actual performance so that it matches their abilities (e.g. enhance motivation, improve attention, raise their self-confidence, or prepare them against test anxiety). However, it is important that there is enough time in between this intervention and the tracking decision, so that actual performance may improve.

A second implication of the current dissertation's findings concerns the new educational standards (Bildungsstandards) in Germany. In 2004, the Kultusministerkonferenz (KMK) in Germany decided on new educational standards for all schools in Germany to foster newly defined competences (Kultusministerkonferenz, 2005a). For example, in primary schools, these competences focus on mathematics and German instructions (Kultusministerkonferenz, 2005b, 2005c). These newly defined competences are more closely related to fluid and crystallized intelligence, respectively. These changes in the educational system in Germany may affect the degree to which formal education in Germany will affect cognitive abilities. In the current study, we found that length of formal education had an impact on fluid and crystallized abilities. Thus, one can only assume that these new educational standards will yet enlarge the impact of formal education on cognitive abilities in the years to come. However, the implementation of these new standards may increase the ability gap, as the 'haves' may profit much more from these new standards than the 'have-nots'. Nevertheless, the opposite result is also possible, as found in the present dissertation, the non-academic track participants (i.e. probably also more of the 'have-nots') did profit from formal education for almost all their lives, while no lasting effect of formal education on the cognitive abilities of the students in the academic track was found. An explanation for this effect is that formal education offered the participants in the non-academic track a possibility to train their cognitive abilities, a possibility that they would have not had without school. Thus, it could be that training these new competences closer to fluid and crystallized abilities in schools may be especially beneficial for lower performing students. As a result the lower end of the ability distribution would be the ones who profit more, and also more enduringly, from such a change, resulting in a reduced gap. Crucially, as both scenarios are possible (a yet increased gap between the lower and the upper level of the distribution or a reduced gap between the two poles), it is of importance to monitor the changes in cognitive abilities that come along with these new educational standards.

5.4 Strengths, Limitations, and Future Avenues

The strengths of the present dissertation are grounded in the rich database that we could draw from. We were able to base our analyses on a well validated IQ test, the Leistungsprüfsystem by Horn (1962, 1983). This enabled us to test several hypotheses regarding the development of intelligence across 40 years of the lifetime. In addition, we could examine the influence of formal education and grade retention in primary school on the development of cognitive abilities. Moreover, when relevant, we were able to control for key variables associated with the development of intelligence, most notably childhood cognitive ability, educational attainment in primary school, and parental socioeconomic status in childhood, among others. Importantly, the sample of the present dissertation did not only span 40 years of the lifetime, but was homogeneous with respect to age. Crucially, the sample was representative of the Luxembourgish population of all 12 year olds and sixth graders in 1968, as half of all schools in Luxembourg participated in the study in 1968.

Our research design allowed us to effectively tackle two major validity threats that longitudinal designs usually suffer from: (a) selective attrition and (b) retest effects (Tucker-Drob & Salthouse, 2011). In the present dissertation, we were able to reliably estimate selective attrition, because our base sample was representative of Luxembourg’s population. In addition, the estimates of selective attrition showed that our longitudinal sample was only slightly positively selected. Crucially, the time span of 40 years between the two measurement occasions rendered retest effects almost impossible.

Despite all these strengths, the present dissertation was also limited in several aspects.

The most unfortunate limitation is that we could base our analyses only on two points of measurement. Measurement waves at critical stages in life would have been needed to draw more precise conclusions on the course of cognitive development and the influence of educational attainment. Such important points in the life course would have been before children entered primary school, after adolescents left formal education, and at about age 30 years. With these additional points of measurement, we would have been able to draw inferences on the path of development as well as have a better understanding of the effects of formal education on cognitive abilities. In addition, more measurement occasions with shorter time intervals would have also been needed to test the propositions of investment theory in more detail. Moreover, a measurement occasion before the participants entered school or in first grade would have been very valuable for Study 3. If childhood intelligence and grades or a teacher rating had been assessed before the retention decision was made, we could have included this information in the propensity score matching analyses. By doing so, we would have been able to control the selection bias between promoted and retained children in an even stricter way.

The IQ re-test sample was much smaller than the household sample in the MAGRIP-R study in 2008/09. This was a result of the two-step data collection process, where each step took about 90 minutes or more. Thus, on some occasions more participants would have been needed to have enough statistical power in order to find significant effects for all research questions under study. This problem is especially applicable to Study 2 and the main effect of quality of formal education. A larger number of participants would have also been very valuable in Study 3 on grade retention, because the propensity score matching procedure led to an additional reduction in sample size. With a larger initial sample, we could have also performed the analyses on grade retention on a latent level.

Moreover, we had only two observed indicators as measures of each broad ability factor, which reflects the lower limit for assessing latent factors in structural equation models. However, to be able to measure change, we had to use the same test that was applied in 1968. Nevertheless, the observed indicators were well-selected and widely-used indicators of the broad abilities under investigation. In addition, in Study 2, we lacked a longitudinal measure of knowledge of the world. This would have enabled us to directly compare word knowledge and knowledge of the world as two different aspects of crystallized abilities. The present results cannot clearly distinguish whether the difference in the two variables in respect to formal education is accounted for by the different nature of the variables or simply the lack of a longitudinal control for knowledge of the world.

Future avenues should be built on well planned longitudinal studies. Structural equation modeling offers many interesting new approaches to data analyses of longitudinal data. However, these approaches require large sample sizes as well as several waves of data collection. Therefore, longitudinal studies that start as early as kindergarten age and that cover large periods of the lifetime are needed to answer important questions on cognitive aging. The European Commission (2014) has realized that cognitive interventions for an aging society must start in childhood. If we do not succeed in tackling the gaps in cognitive functioning and improving cognitive aging, many European societies will struggle with demographic change in their countries. However, today still too little is known about many underlying mechanisms of cognitive development as well as the impact of formal education on cognitive abilities. This is why representative longitudinal studies that measure cognitive abilities as well as several important characteristics that are associated with cognitive development (e.g. educational attainment, parental SES, parental education) are needed. Ideally, these studies would start assessing the participants characteristics at age 3 or at the latest, before they enter primary school, at the end of primary school, one year after the transition to secondary school, at the end of school, and thereafter in five to ten year intervals. In addition, several aspects of cognitive abilities should be assessed to paint a clearer picture of how these different broad abilities influence each other across time as well as the interplay and importance of specific abilities and *g*. In addition, it would be very interesting to also assess other non-cognitive influences on key life outcomes such as motivation, self-esteem, goals, values, and personality.

Annex

Annex A: Comprehensive Description of the Measures of Intelligence

Childhood intelligence was assessed by the Leistungsprüfsystem (L-P-S [Performance Test System]; W. Horn, 1962, 1983). The L-P-S is a standardized, objective, and comprehensive German intelligence test based on the model of primary mental abilities formulated by Thurstone (1938). Its 14 subtests provide a measure of general intelligence (total IQ score) as well as scores for more specific intellectual facets, such as crystallized intelligence and fluid intelligence (Neubauer, Fink, & Schrausser, 2002). The scores for crystallized intelligence are based on three subtests. Two subtests consist of misspelled six-letter words; participants have to identify the appropriate words as well as the spelling errors. The other subtest consists of anagrams (Borkenau & Liebler, 1993). The scores for fluid intelligence are based on two subtests inspired by Raven's Progressive Matrices (W. Horn, 1983). For both subtests, participants have to identify the inappropriate element in a series of eight elements, the elements of the first subtests being geometric figures and those of the second subtest being letters and digits.

Split-half reliability of the overall test is .99, parallel-forms reliability is .94. Retest reliability across a time span of 32 months is .83 for the overall test score (W. Horn, 1983; Tent, 1969), .94 for the combined score for crystallized intelligence, and .78 for the combined score for fluid intelligence (W. Horn, 1983). There is ample evidence for the construct validity of the L-P-S. Specifically, the correlation of the L-P-S total score with the total score on the German version of the Wechsler Adult Intelligence Scale (WAIS)—the Hamburg Wechsler Intelligenztest für Erwachsene (HAWIE-R; Tewes, 1991)—is .94 (Sturm & Büssing, 1982). Furthermore, the correlation of the standardized L-P-S total score with the standardized total score of the Intelligenz-Struktur-Test (IST; Liepmann et al., 2001) is .72. The IST is another well-validated and widely used German intelligence test that also correlates substantially with the HAWIE-R (Tewes, 1991). In a recent meta-analysis, Hülsheger, Maier, Stumpp, and Muck (2006) compared the predictive validity of the L-P-S and five other intelligence tests widely used in German-speaking countries, including the IST and Raven's Progressive Matrices (Kratzmeier, 1979), for the outcomes of vocational education. The authors found the L-P-S to be one of the instruments with the highest criterion-related validity. Further, the total and subtest scores of the L-P-S showed high correlations with grades in various school subjects (W. Horn, 1983). For instance, the total score showed a correlation of .55 with grade point average in Grade 4 of elementary school (Tent, 1965). The crystallized intelligence score showed a correlation of .47 with German grades, and the fluid intelligence score a correlation of .80 with mathematics grades (W. Horn, 1983). Given the strong empirical evidence for its reliability and validity, the L-P-S is widely employed in various areas of psychological research, such as research on gender differences in cognitive functions (Weiss, Kemmler, Deisenhammer, Fleischhacker, & Delazer, 2003) or clinical and neuropsychology (Kuelz, Hohagen, & Voderholzer, 2004).

Annex B: Study 1- Invariance across Time of Psychometric Properties

Further Details on Assessment of Model Fit in Study 1

Applied fit indices. To test for increasing levels of MI, we used a stepwise procedure, in which increasingly more equality constraints were introduced into the models of the 12 and 52 year-olds. Goodness of model fit was assessed by the χ^2 goodness-of-fit test as well as by several descriptive measures of fit that are recommended in the literature: the χ^2 statistic in relation to degrees of freedom, the Root Mean Square Error of Approximation (*RMSEA*, Steiger, 1990), the Standardized Root Mean Square Residual (*SRMR*, Bentler, 1995), Akaike's Information Criterion (*AIC*, Akaike, 1974), and the Comparative Fit Index (*CFI*, Bentler, 1990). RMSEA values smaller than .05 show an acceptable fit (Hu & Bentler, 1999). SRMR values smaller than .05 indicate good fit, and values smaller than .10 point to an adequate model fit (Schermelleh-Engel et al., 2003). The model with the smallest AIC value is considered to be the best fitting model (Schermelleh-Engel et al., 2003). CFI values that are greater than .95 suggest good model fit (Hu & Bentler, 1999). The CFI shows the relative fit of a given model compared to an appropriate null model that accounts for MI restrictions (Widaman & Thompson, 2003). The appropriate null model for the current study is a model with zero covariation among the observed variables, time invariant measurement error, as well as time invariant intercepts (i.e., Model 0A in Widaman & Thompson, 2003).

Assessing measurement invariance. Our stepwise procedure implied that each model is nested within the previous one, and hence the model fits could be compared (Bentler & Bonett, 1980). When the overall model fit of a given model was acceptable, we made comparisons across the nested models to assess MI (T. D. Little, 1997). Model fit differences are most commonly assessed by the χ^2 difference test (Bollen, 1989). However, Cheung and Rensvold (2002) have shown that a change in the Comparative Fit Index CFI smaller than -.01

indicates adequate fit of the model with additional MI constraints. Importantly, when the χ^2 difference test and the change in the CFI disagree regarding which model should be accepted, we based our decision on the change in the CFI because of its better statistical properties (i.e., change in the CFI is not affected by sample size; Cheung & Rensvold, 2002). In some cases, MI holds for some but not for all indicators; this is called partial MI (Byrne et al., 1989). Byrne et al. (1989, p. 458) suggested that a sufficient degree of partial metric invariance for meaningful comparisons of different models is established if, in addition to the indicator whose loading is fixed to identify the scale of the latent factor, at least one additional loading is invariant across time.

Results Type 1 MI

To study the Type 1 MI of subtest scores, we examined a series of increasingly constrained models (these models are abbreviated T1). Measurement models with configural (T1.1) and metric invariance (T1.2) constraints demonstrated excellent fit (see Table 5). However, when we imposed the constraint of equal error variances across time, parameter estimation did not converge (T1.3). We therefore relaxed the equality constraints on the residual variances of two subtest scores (Gc_1 and Gs_1). The resulting model (T1.4) showed good fit to the data, and the model comparison with the metric invariant model (T1.2) indicated an acceptable deterioration in model fit. When we imposed the constraints of scalar invariance (T.5), overall fit deteriorated markedly. We therefore relaxed the constraints on the intercepts of two indicators (Gc_1 and Gv_8), yielding a model with partial scalar invariance (T1.6) that fit the data well and not considerably worse than Model T1.4. To conclude, our results indicate that configural, metric, and partial error invariance in combination with partial scalar invariance could be established for all manifest subtest scores. This level of Type 1 MI indicates that the operational definition of the four broad abilities is fundamentally the same at age 12 and age

52 and allows meaningful comparisons of the latent covariances and variances in order to test the age differentiation-dedifferentiation hypothesis based on the extended Gf-Gc model and the three-stratum model, respectively.

Further Results of Testing the Age Differentiation-Dedifferentiation Hypothesis

Here we present some additional statistical results of our study of the Age Differentiation-Dedifferentiation Hypothesis when taking the perspective of Carroll's three-stratum model. To identify the various sources of age dedifferentiation in this model, we drew on Model C.1 (where the variance of $Gf_{specific52}$ was fixed to zero) and imposed several equality constraints across time (see Table 5). First, in Model C.2, we constrained the unstandardized factor loadings to be equal across time (i.e., reflecting second-order metric invariance). This model did not show acceptable overall model fit. The reason was that the unstandardized loadings of Gc on *g* and Gs on *g* were much higher at age 52 than at age 12. When we relaxed the equality constraints of these second-order factor loadings (i.e., reflecting partial metric invariance; Model C.3), model fit was acceptable and not considerably worse than the configurally invariant model (Model C.1). Second, in Model C.4, we constrained the variances of specific abilities to be equal across time. This model did not show acceptable overall model fit. The reason was that the variance of $\mathrm{Gc}_{\mathrm{specific}}$ increased whereas the variance of $\mathrm{Gf}_{\mathrm{specific}}$ decreased with age (see Figure 2c). When the equality constraints on the variances of these specific abilities were relaxed (Model C.5), model fit was acceptable and not considerably worse than that of Model C.3. Third, in Model C.6, we imposed equality constraints on the variance of *g* across time. This model did not fit the data well because the variance of the *g* factor increased significantly from age 12 to age 52 (see Figure 2c). Taken together, these results indicate that the increase in intercorrelations observed in the extended Gf-Gc model is the result of several age-specific changes: (a) increases in the factor loadings of Gc and Gs on *g*, (b) a decrease in the variance specific to Gf, with Gf even becoming indistinguishable from *g* at age 52, and (c) by a substantial increase in the variance of *g* over time. However, at the same time, the variance specific to Gc increased, which is indicative of differentiation.

Annex C: Tables

Table 13

Intercorrelations, Descriptive Statistics, and Model-Based Reliability Estimates of Manifest Measures as Applied in the Current Study

	Age 12								Age 52							
	Gc_1	Gc_2	Gf_1	Gf_2	Gv_1	Gv_2	Gs_1	Gs_2	Gc_1	Gc_2	Gf_1	Gf_2	Gv_1	Gv_2	Gs_1	Gs_2
At age 12																
Gc_1	–															
Gc_2	.51	–														
Gf_1	.30	.29	–													
Gf_2	.32	.34	.54	–												
Gv_1	.31	.30	.33	.36	–											
Gv_2	.30	.37	.41	.37	.42	–										
Gs_1	.20	.25	.21	.29	.18	.21	–									
Gs_2	*.09*	.17	*.07*	.19	*.09*	*.12*	.28	–								
At age 52																
Gc_1	.62	.47	.31	.42	.33	.33	.24	*.12*	–							
Gc_2	.46	.47	.32	.40	.28	.31	.28	*.05*	.64	–						
Gf_1	.30	.31	.48	.52	.41	.45	.21	*.08*	.49	.49	–					
Gf_2	.38	.35	.49	.54	.39	.40	.26	.17	.57	.56	.70	–				
Gv_1	.27	.27	.43	.43	.40	.47	*.16*	*.01*	.43	.45	.60	.53	–			
Gv_2	.24	.30	.43	.35	.40	.57	*.12*	*.08*	.38	.39	.59	.53	.59	–		
Gs_1	.19	.17	.19	.22	.15	.18	.40	.29	.29	.34	.31	.36	.23	.19	–	
Gs_2	.22	.22	.33	.38	.26	.26	.29	.41	.45	.41	.46	.55	.36	.34	.49	–
M	102.8	103.1	104.7	104.9	103.5	103.5	101.7	102.2	162.0	133.6	117.4	121.8	112.5	116.3	121.2	121.9
SD	13.0	14.7	14.2	13.0	15.3	14.1	14.8	13.0	24.1	21.1	15.8	13.8	18.9	14.5	23.5	17.4
RTT	.52	.42	.50	.64	.44	.43	.25	.32	.60	.73	.58	.72	.56	.55	.35	.62

Note. All entries are based on full information maximum likelihood estimates for missing data. Gc = comprehension knowledge; Gf = fluid reasoning; Gv = visual processing; Gs = processing speed; _1 and _2 refer to manifest variables 1 and 2 that measure the respective broad ability; RTT = lower-bound model-based estimates of subtest score reliabilities obtained from Model T1.6 (see Bollen, 1989); these estimates take into account the reliable variance due to the first-order factors and (in case of Gc_1, Gv_2, Gs_1, and Gs_2) correlated residual terms.

Table 14

Income Categories and Means per Category adapted to Luxembourgish Income Levels as used in the Study 3 of the present dissertation.

Income categories as assessed in the study	Mean per category used for the data analyses
No own income	0 €
less than 150 €	75 €
150 € to 299 €	225 €
300 € to 499 €	375 €
500 € to 999 €	750 €
1,000 € to 1,499 €	1,250 €
1,500 € to 1,999 €	1,750 €
2,000 € to 2,499 €	2,250 €
2,500 € to 2,999 €	2,750 €
3,000 € to 3,999 €	3,500 €
4,000 € to 4,999 €	4,500 €
5,000 € to 6,249 €	5,625 €
6,250 € to 7,499 €	6,875 €
7,500 € to 9,999 €	8,750 €
10,000 € or more	10,000 €

Table 15

Descriptive Statistics of the Interval Scaled Variables in the Propensity Score Matching for Years of Education by Promoted and Retained Students Before and After Matching in Study 3.

	Before Matching					After Matching				
	Not retained (*n* = 618)		Retained (*n* = 107)		Effect size	Not retained (*n* = 384)		Retained (*n* = 101)		Effect size
Matched Variables	mean	*SD*	mean	*SD*	Cohen´s *d*	mean	*SD*	mean	*SD*	Cohen´s *d*
Total number of children in family	2.8	1.5	3.6	1.9	-0.52	3.1	1.6	3.4	1.8	-0.19
Total number of people living in household	5.0	1.6	5.6	1.9	-0.36	5.3	1.7	5.5	1.9	-0.14
Highest parental SES	41.2	12.9	33.7	8.6	0.61	35.5	9.7	33.8	8.7	0.18
Participant: Birth rank	1.9	1.2	2.6	1.6	-0.59	2.1	1.3	2.4	1.3	-0.21

Note. SES = socioeconomic status

Table 16

Descriptive Statistics of the Not Interval Scaled Variables in the Propensity Score Matching for Years of Education by Promoted and Retained Students Before and After Matching in Study 3.

	Before Matching					After Matching				
	Not retained (*n* = 618)		Retained (*n* = 107)		Effect size	Not retained (*n* = 384)		Retained (*n* = 101)		Effect size
Matched Variables	*n*	%	*n*	%	Cohen´s *h*	*n*	%	*n*	%	Cohen´s *h*
Gender										
male	285	46.1	53	49.5	-0.07	182	47.4	51	50.5	-0.16
female	333	53.9	54	50.5	0.07	202	52.6	50	49.5	-0.04
Class gender configuration										
only boys	131	21.2	31	29.0	-0.18	103	26.8	30	29.7	-0.12
only girls	164	26.5	28	26.2	0.01	98	25.5	27	26.7	-0.09
mixed	323	52.3	48	44.9	0.15	183	47.7	44	43.6	-0.01
Parents working situation										
both parents working	116	18.8	23	21.5	-0.07	80	20.8	22	21.8	-0.07
only father working	478	77.3	78	72.9	0.10	289	75.3	75	74.3	-0.14
only mother working	8	1.3	0	0.0	0.23	4	1.0	0	0.0	0.20
parents unemployed	16	2.6	6	5.6	-0.16	11	2.9	4	4.0	-0.08
Person learning with participant										
nobody	182	29.4	28	26.2	0.07	124	32.3	27	26.7	0.06
father	79	12.8	17	15.9	-0.09	38	9.9	17	16.8	-0.24
mother	232	37.5	29	27.1	0.22	140	36.5	28	27.7	0.12
siblings	28	4.5	17	15.9	-0.39	20	5.2	14	13.9	-0.33
somebody else / teacher	9	1.5	4	3.7	-0.15	8	2.1	3	3.0	-0.07
somebody else / privat	27	4.4	5	4.7	-0.01	18	4.7	5	5.0	-0.03
father and mother	36	5.8	3	2.8	0.15	16	4.2	3	3.0	0.05
mother and somebody else	8	1.3	2	1.9	-0.05	6	1.6	2	2.0	-0.04
father, mother, and siblings	17	2.8	2	1.9	0.06	14	3.6	2	2.0	0.09
Nationality										
Luxembourgish	534	86.4	86	80.4	0.16	320	83.3	80	79.2	-0.10
Italian	40	6.5	12	11.2	-0.17	29	7.6	12	11.9	-0.18
Portuguese	1	0.2	0	0.0	0.08	1	0.3	0	0.0	0.10
German	7	1.1	2	1.9	-0.06	5	1.3	2	2.0	-0.07
Spanish	4	0.6	0	0.0	0.16	4	1.0	0	0.0	0.20
Dutch	7	1.1	3	2.8	-0.12	4	1.0	3	3.0	-0.16
Belgian	9	1.5	0	0.0	0.24	9	2.3	0	0.0	0.30
French	10	1.6	3	2.8	-0.08	8	2.1	3	3.0	-0.07
none / other	6	1.0	1	0.9	0.00	4	1.0	1	1.0	0.00
Month participant was born										
January	40	6.5	6	5.6	0.04	27	7.0	6	5.9	0.02
February	32	5.2	7	6.5	-0.06	22	5.7	7	6.9	-0.07
March	58	9.4	6	5.6	0.14	31	8.1	6	5.9	0.06
April	41	6.6	11	10.3	-0.13	22	5.7	10	9.9	-0.19
May	54	8.7	10	9.3	-0.02	35	9.1	9	8.9	-0.02
June	50	8.1	10	9.3	-0.05	25	6.5	9	8.9	-0.12
July	46	7.4	3	2.8	0.22	32	8.3	3	3.0	0.22
August	37	6.0	10	9.3	-0.13	22	5.7	9	8.9	-0.15
September	48	7.8	13	12.1	-0.15	28	7.3	12	11.9	-0.19
October	51	8.3	12	11.2	-0.10	34	8.9	12	11.9	-0.13
November	79	12.8	6	5.6	0.25	50	13.0	6	5.9	0.22
December	82	13.3	13	12.1	0.03	56	14.6	12	11.9	0.04
Family situation										
parents married	585	94.7	89	83.2	0.39	356	92.7	87	86.1	-0.07
parents divorced	4	0.6	3	2.8	-0.18	2	0.5	3	3.0	-0.21
parents seperated	6	1.0	6	5.6	-0.28	6	1.6	6	5.9	-0.26
parents remarried	6	1.0	4	3.7	-0.19	5	1.3	3	3.0	-0.13
foster family	1	0.2	1	0.9	-0.11	1	0.3	0	0.0	0.10
widowed	16	2.6	4	3.7	-0.07	14	3.6	2	2.0	0.09

Table 17

Descriptive Statistics of the Interval Scaled Variables in the Propensity Score Matching for Male Adult Income by Promoted and Retained Students Before and After Matching in Study 3.

	Before Matching					After Matching				
	Not retained (*n* = 285)		Retained (*n* = 53)		Effect size	Not retained (*n* = 164)		Retained (*n* = 49)		Effect size
Matched Variables	**mean**	***SD***	**mean**	***SD***	**Cohen´s *d***	**mean**	***SD***	**mean**	***SD***	**Cohen´s *d***
Total number of children in family	2.7	1.4	3.3	1.7	-0.43	2.9	1.5	3.1	1.5	-0.17
Total number of people living in household	4.9	1.6	5.4	1.6	-0.31	5.0	1.7	5.2	1.5	-0.09
Highest parental SES	40.8	12.7	34.3	8.1	0.53	35.8	10.2	34.0	8.2	0.18
Participant: Birth rank	1.8	1.0	2.6	1.5	-0.75	2.1	1.1	2.4	1.3	-0.25

Note. SES = socioeconomic status

Table 18

Descriptive Statistics of the Not Interval Scaled Variables in the Propensity Score Matching for Adult Male Income by Promoted and Retained Students Before and After Matching in Study 3.

	Before Matching					After Matching				
	Not retained (*n* = 285)		Retained (*n* = 53)		Effect size	Not retained (*n* = 164)		Retained (*n* = 49)		Effect size
Matched Variables	***n***	**%**	***n***	**%**	**Cohen´s *h***	***n***	**%**	***n***	**%**	**Cohen´s *h***
Gender										
male	285	100.0	53	100.0	0.00	164	100.0	49	100.0	0.00
female	0	0.0	0	0.0	0.00	0	0.0	0	0.0	0.00
Class gender configuration										
only boys	131	46.0	31	58.5	-0.25	93	56.7	28	57.1	0.04
only girls	0	0.0	0	0.0	0.00	0	0.0	0	0.0	0.00
mixed	154	54.0	22	41.5	0.26	71	43.3	21	42.9	0.05
Parents working situation										
both parents working	54	18.9	12	22.6	-0.09	34	20.7	10	20.4	0.03
only father working	224	78.6	40	75.5	0.08	125	76.2	38	77.6	0.05
only mother working	2	0.7	0	0.0	0.17	2	1.2	0	0.0	0.23
parents unemployed	5	1.8	1	1.9	-0.01	3	1.8	1	2.0	-0.01
Person learning with participant										
nobody	62	21.8	12	22.6	-0.02	39	23.8	11	22.4	0.06
father	38	13.3	8	15.1	-0.05	20	12.2	8	16.3	-0.10
mother	119	41.8	15	28.3	0.29	70	42.7	15	30.6	0.29
siblings	14	4.9	9	17.0	-0.40	14	8.5	8	16.3	-0.23
somebody else / teacher	6	2.1	3	5.7	-0.19	4	2.4	2	4.1	-0.09
somebody else / privat	13	4.6	2	3.8	0.04	8	4.9	2	4.1	0.05
father and mother	20	7.0	2	3.8	0.15	4	2.4	2	4.1	-0.09
mother and somebody else	3	1.1	2	3.8	-0.19	0	0.0	1	2.0	-0.29
father, mother, and siblings	10	3.5	0	0.0	0.38	5	3.0	0	0.0	0.36
Nationality										
Luxembourgish	250	87.7	39	73.6	0.37	138	84.1	36	73.5	0.39
Italian	15	5.3	7	13.2	-0.28	11	6.7	6	12.2	-0.18
Portuguese	1	0.4	0	0.0	0.12	1	0.6	0	0.0	0.16
German	3	1.1	1	1.9	-0.07	3	1.8	1	2.0	-0.01
Spanish	3	1.1	0	0.0	0.21	3	1.8	0	0.0	0.28
Dutch	4	1.4	3	5.7	-0.24	0	0.0	3	6.1	-0.51
Belgian	1	0.4	0	0.0	0.12	1	0.6	0	0.0	0.16
French	7	2.5	2	3.8	-0.08	6	3.7	2	4.1	-0.01
none / other	1	0.4	1	1.9	-0.16	1	0.6	1	2.0	-0.13
Month participant was born										
January	16	5.6	3	5.7	0.00	8	4.9	3	6.1	-0.05
February	17	6.0	5	9.4	-0.13	10	6.1	5	10.2	-0.14
March	34	11.9	1	1.9	0.43	18	11.0	1	2.0	0.41
April	19	6.7	8	15.1	-0.27	15	9.1	7	14.3	-0.15
May	27	9.5	3	5.7	0.15	13	7.9	3	6.1	0.08
June	21	7.4	6	11.3	-0.14	12	7.3	4	8.2	-0.02
July	23	8.1	1	1.9	0.30	14	8.5	0	0.0	0.61
August	16	5.6	5	9.4	-0.15	11	6.7	5	10.2	-0.12
September	25	8.8	6	11.3	-0.08	12	7.3	6	12.2	-0.16
October	20	7.0	6	11.3	-0.15	14	8.5	6	12.2	-0.11
November	36	12.6	4	7.5	0.17	18	11.0	4	8.2	0.11
December	31	10.9	5	9.4	0.05	19	11.6	5	10.2	0.06
Family situation										
parents married	273	95.8	47	88.7	0.29	153	93.3	44	89.8	0.42
parents divorced	1	0.4	1	1.9	-0.16	1	0.6	1	2.0	-0.13
parents seperated	3	1.1	2	3.8	-0.19	3	1.8	2	4.1	-0.13
parents remarried	4	1.4	2	3.8	-0.15	3	1.8	1	2.0	-0.01
foster family	0	0.0	0	0.0	0.00	0	0.0	0	0.0	0.00
widowed	4	1.4	1	1.9	-0.04	4	2.4	1	2.0	0.03

Table 19

Descriptive Statistics of the Interval Scaled Variables in the Propensity Score Matching for Adult Intelligence by Promoted and Retained Students Before and After Matching in Study 3.

	Before Matching					After Matching				
	Not retained (*n* = 327)		Retained (*n* = 41)		Effect size	Not retained (*n* = 162)		Retained (*n* = 41)		Effect size
Matched Variables	mean	*SD*	mean	*SD*	Cohen´s *d*	mean	*SD*	mean	*SD*	Cohen´s *d*
Total number of children in family	2.9	1.5	3.3	1.6	-0.29	3.3	1.6	3.3	1.6	-0.03
Total number of people living in household	5.1	1.6	5.5	1.7	-0.29	5.5	1.7	5.5	1.7	-0.02
Highest parental SES	41.3	13.1	35.6	9.2	0.45	36.8	11.0	35.6	9.2	0.11
Participant: Birth rank	1.9	1.2	2.3	1.2	-0.28	2.2	1.4	2.3	1.2	-0.07

Note. SES = socioeconomic status

Table 20

Descriptive Statistics of the Not Interval Scaled Variables in the Propensity Score Matching for Adult Intelligence by Promoted and Retained Students Before and After Matching in Study 3.

	Before Matching					After Matching				
	Not retained (*n* = 327)		Retained (*n* = 41)		Effect size	Not retained (*n* = 162)		Retained (*n* = 41)		Effect size
Matched Variables	*n*	%	*n*	%	Cohen´s *h*	*n*	%	*n*	%	Cohen´s *h*
Gender										
male	143	43.7	22	53.7	-0.22	75	46.3	22	53.7	-0.36
female	184	56.3	19	46.3	0.19	87	53.7	19	46.3	-0.08
Class gender configuration										
only boys	60	18.3	14	34.1	-0.38	46	28.4	14	34.1	-0.27
only girls	91	27.8	12	29.3	-0.04	42	25.9	12	29.3	-0.21
mixed	176	53.8	15	36.6	0.34	74	45.7	15	36.6	-0.01
Parents working situation										
both parents working	57	17.4	10	24.4	-0.18	44	27.2	10	24.4	-0.07
only father working	255	78.0	31	75.6	0.03	112	69.1	31	75.6	-0.48
only mother working	6	1.8	0	0.0	0.27	4	2.5	0	0.0	0.29
parents unemployed	9	2.8	0	0.0	0.33	2	1.2	0	0.0	0.20
Person learning with participant										
nobody	112	34.3	11	26.8	0.15	53	32.7	11	26.8	-0.02
father	43	13.1	6	14.6	-0.05	12	7.4	6	14.6	-0.30
mother	108	33.0	15	36.6	-0.09	54	33.3	15	36.6	-0.23
siblings	18	5.5	2	4.9	0.02	13	8.0	2	4.9	0.07
somebody else / teacher	4	1.2	1	2.4	-0.09	2	1.2	1	2.4	-0.12
somebody else / privat	12	3.7	3	7.3	-0.17	9	5.6	3	7.3	-0.13
father and mother	20	6.1	1	2.4	0.18	11	6.8	1	2.4	0.16
mother and somebody else	5	1.5	1	2.4	-0.07	3	1.9	1	2.4	-0.07
father, mother, and siblings	5	1.5	1	2.4	-0.07	5	3.1	1	2.4	0.00
Nationality										
Luxembourgish	284	86.9	32	78.0	0.21	135	83.3	32	78.0	-0.29
Italian	21	6.4	6	14.6	-0.28	13	8.0	6	14.6	-0.28
Portuguese	0	0.0	0	0.0	0.00	0	0.0	0	0.0	0.00
German	4	1.2	0	0.0	0.22	3	1.9	0	0.0	0.25
Spanish	2	0.6	0	0.0	0.16	2	1.2	0	0.0	0.20
Dutch	5	1.5	2	4.9	-0.20	2	1.2	2	4.9	-0.25
Belgian	4	1.2	0	0.0	0.22	3	1.9	0	0.0	0.25
French	3	0.9	1	2.4	-0.12	1	0.6	1	2.4	-0.18
none / other	4	1.2	0	0.0	0.22	3	1.9	0	0.0	0.25
Month participant was born										
January	22	6.7	0	0.0	0.53	7	4.3	0	0.0	0.38
February	12	3.7	4	9.8	-0.26	6	3.7	4	9.8	-0.30
March	33	10.1	3	7.3	0.09	17	10.5	3	7.3	0.04
April	19	5.8	3	7.3	-0.07	7	4.3	3	7.3	-0.18
May	24	7.3	3	7.3	0.00	11	6.8	3	7.3	-0.08
June	28	8.6	1	2.4	0.28	11	6.8	1	2.4	0.16
July	27	8.3	1	2.4	0.27	12	7.4	1	2.4	0.18
August	22	6.7	5	12.2	-0.20	12	7.4	5	12.2	-0.23
September	24	7.3	4	9.8	-0.09	10	6.2	4	9.8	-0.19
October	23	7.0	8	19.5	-0.39	12	7.4	8	19.5	-0.44
November	48	14.7	2	4.9	0.34	29	17.9	2	4.9	0.34
December	45	13.8	7	17.1	-0.10	28	17.3	7	17.1	-0.10
Family situation										
parents married	310	94.8	34	82.9	0.38	150	92.6	34	82.9	-0.26
parents divorced	2	0.6	1	2.4	-0.16	2	1.2	1	2.4	-0.12
parents seperated	2	0.6	4	9.8	-0.49	1	0.6	4	9.8	-0.51
parents remarried	3	0.9	1	2.4	-0.12	2	1.2	1	2.4	-0.12
foster family	1	0.3	0	0.0	0.11	1	0.6	0	0.0	0.14
widowed	9	2.8	1	2.4	0.02	6	3.7	1	2.4	0.03

Annex D: Figures

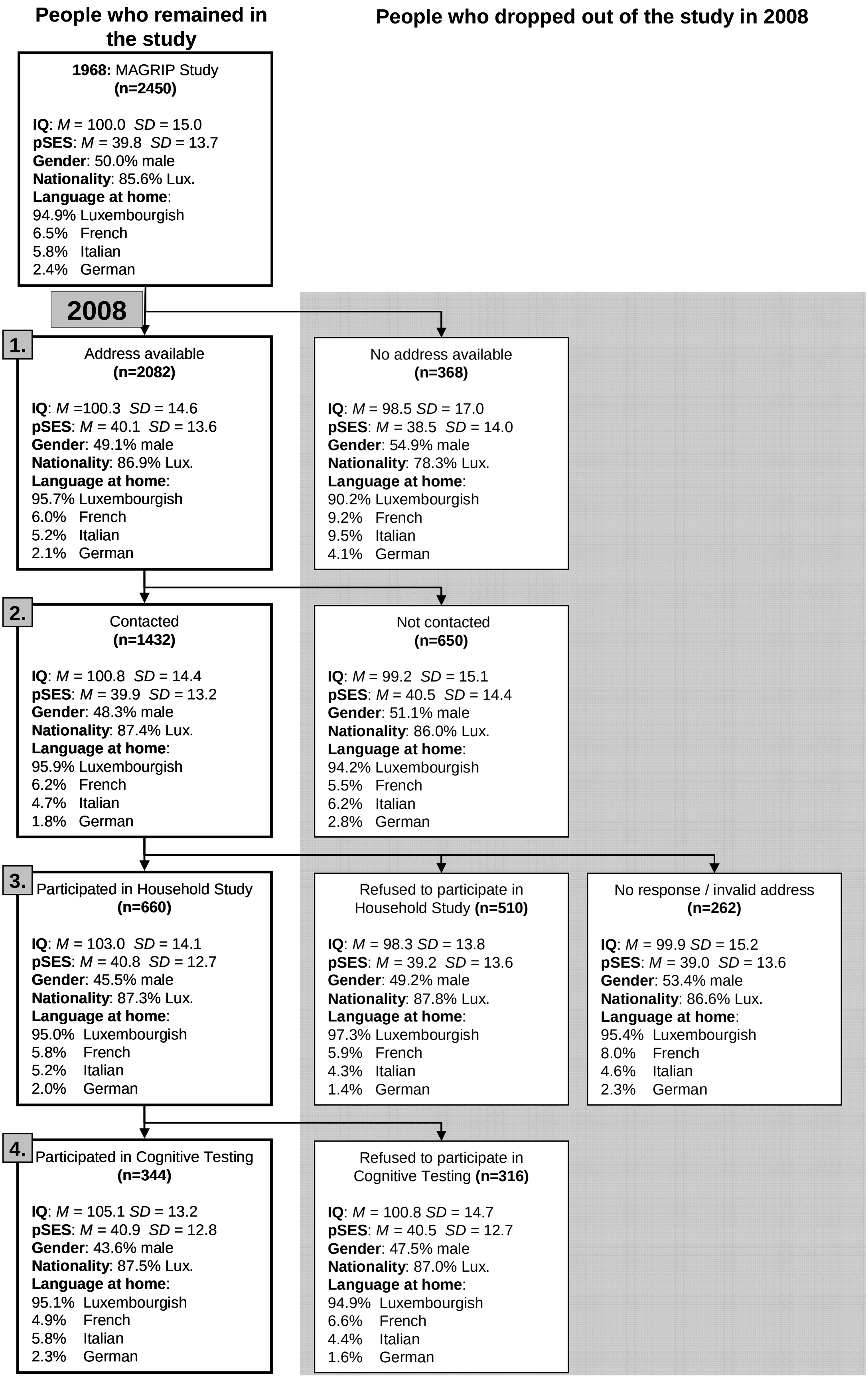

Figure 9

Flowchart of the multistage sampling procedure of the current study including information on childhood characteristics for each subsample. IQ = general cognitive ability at age 12, pSES = parental socioeconomic status at age 12 measured on the ISEI scale (Ganzeboom et al., 1992; Ganzeboom & Treiman, 1996). Description of the multistage sampling procedure: Stage 1: In 2008, all available addresses of the former participants were identified via their social security number. For this procedure, official permission was obtained from the national commission of data protection in Luxembourg (CNPD, Commission nationale pour la protection des données). The most frequent reasons for no available address were that these participants had either died or moved out of the country. Stage 2: Due to budgetary reasons of the current research project, it was not possible to contact all of the former participants for whom addresses were available. Thus, a random sample was drawn from the available addresses. This sample was stratified for gender and region of residence in 1968. Stage 3: In a household study, participants were visited at home by trained interviewers, and data was collected on health, subjective wellbeing, educational and occupational paths. Stage 4: A subsample of participants who took part in the household study also volunteered to complete the intelligence test. Specifically, data were first collected in a group setting from 227 participants. To increase the sample size, 117 further participants who were not able to attend the group testing were visited at home by trained assessors who administered the intelligence test individually. Note that the test administration procedure of the group setting and the individual assessment strictly followed the standardization requirements that were given in the test manual to ensure comparability of results.

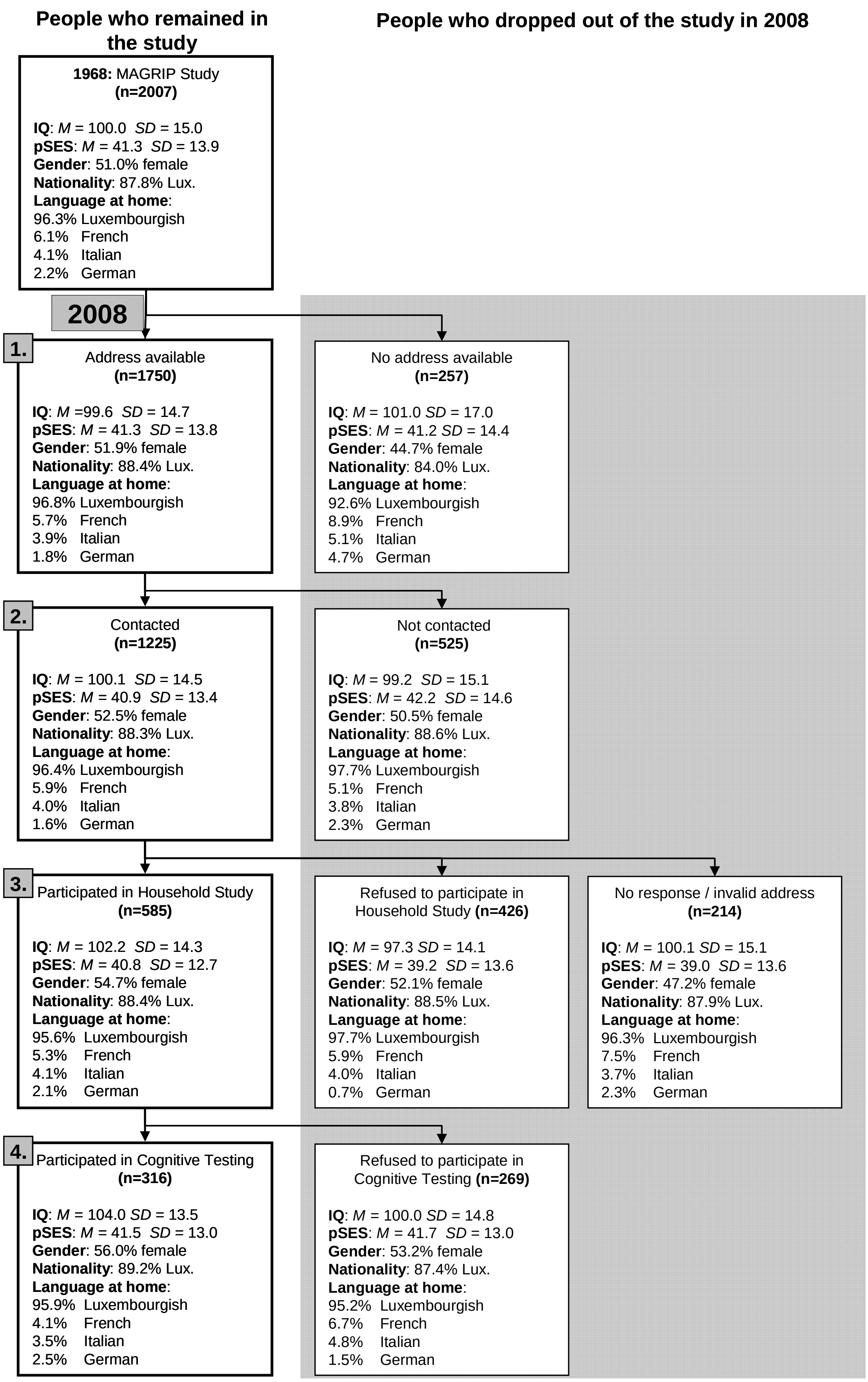
People who remained in the study
People who dropped out of the study in 2008
1968: MAGRIP Study
(n=2007)
IQ: M = 100.0 SD = 15.0
pSES: M = 41.3 SD = 13.9
Gender: 51.0% female
Nationality: 87.8% Lux.
Language at home:
96.3% Luxembourgish
6.1% French
4.1% Italian
2.2% German
2008
1.
Address available
(n=1750)
IQ: M =99.6 SD = 14.7
pSES: M = 41.3 SD = 13.8
Gender: 51.9% female
Nationality: 88.4% Lux.
Language at home:
96.8% Luxembourgish
5.7% French
3.9% Italian
1.8% German
No address available
(n=257)
IQ: M = 101.0 SD = 17.0
pSES: M = 41.2 SD = 14.4
Gender: 44.7% female
Nationality: 84.0% Lux.
Language at home:
92.6% Luxembourgish
8.9% French
5.1% Italian
4.7% German
2.
Contacted
(n=1225)
IQ: M = 100.1 SD = 14.5
pSES: M = 40.9 SD = 13.4
Gender: 52.5% female
Nationality: 88.3% Lux.
Language at home:
96.4% Luxembourgish
5.9% French
4.0% Italian
1.6% German
Not contacted
(n=525)
IQ: M = 99.2 SD = 15.1
pSES: M = 42.2 SD = 14.6
Gender: 50.5% female
Nationality: 88.6% Lux.
Language at home:
97.7% Luxembourgish
5.1% French
3.8% Italian
2.3% German
3.
Participated in Household Study
(n=585)
IQ: M = 102.2 SD = 14.3
pSES: M = 40.8 SD = 12.7
Gender: 54.7% female
Nationality: 88.4% Lux.
Language at home:
95.6% Luxembourgish
5.3% French
4.1% Italian
2.1% German
Refused to participate in Household Study (n=426)
IQ: M = 97.3 SD = 14.1
pSES: M = 39.2 SD = 13.6
Gender: 52.1% female
Nationality: 88.5% Lux.
Language at home:
97.7% Luxembourgish
5.9% French
4.0% Italian
0.7% German
No response / invalid address
(n=214)
IQ: M = 100.1 SD = 15.1
pSES: M = 39.0 SD = 13.6
Gender: 47.2% female
Nationality: 87.9% Lux.
Language at home:
96.3% Luxembourgish
7.5% French
3.7% Italian
2.3% German
4.
Participated in Cognitive Testing
(n=316)
IQ: M = 104.0 SD = 13.5
pSES: M = 41.5 SD = 13.0
Gender: 56.0% female
Nationality: 89.2% Lux.
Language at home:
95.9% Luxembourgish
4.1% French
3.5% Italian
2.5% German
Refused to participate in Cognitive Testing (n=269)
IQ: M = 100.0 SD = 14.8
pSES: M = 41.7 SD = 13.0
Gender: 53.2% female
Nationality: 87.4% Lux.
Language at home:
95.2% Luxembourgish
6.7% French
4.8% Italian
1.5% German

Figure 10

Flowchart of the multistage sampling procedure of the current study in 2008 including information on childhood characteristics for each subsample. IQ = general cognitive ability at age 12, pSES = parental socioeconomic status at age 12 measured on the ISEI scale (Ganzeboom et al., 1992; Ganzeboom & Treiman, 1996). Description of the multistage sampling procedure: Stage 1: In 2008 first, all available addresses of the former participants were identified via their social security ID. For this procedure, official permission was obtained from the national commission of data protection in Luxembourg (CNPD, Commission nationale pour la protection des données). Stage 2: A random sample was drawn from the available addresses this sample was stratified for gender and region of residence in 1968. Due to budgetary reasons of the current research project, it was not possible to contact all of the former participants for whom the addresses were available. Stage 3: In a household study, participants were visited at home by trained interviewers, and data was collected on health, subjective wellbeing, educational and occupational paths. Stage 4: Data on cognitive tests was collected from participants who took part in the household study[8]. Specifically, first data was collected from 212 participants in a group setting. To enhance the sample size, 103 further participants, who were not available at the dates of the group testing, were visited at home. All cognitive tests were assessed by trained assessors and the test taking procedure strictly followed the standardization of the test manual.

[8] Of those 316 participants at this stage, data was missing for one person on the school type variable. This person had to be excluded from the structural models, thus leaving a longitudinal sample of 315 participants for the current study.

Annex E: Curriculum Vitae

CURRICULUM VITAE
DANIELA SANGME SCHALKE-MANDOUX

„Der Lebenslauf ist in der Online-Version aus Gründen des Datenschutzes nicht enthalten“

„Der Lebenslauf ist in der Online-Version aus Gründen des Datenschutzes nicht enthalten“

References

Ackerman, P. L., & Lohman, D. F. (2003). Education and g. In H. Nyborg (Ed.), *The scientific study of general intelligence: Tribute to Arthur R. Jensen* (pp. 275–292). Oxford, UK: Pergamon/Elsevier Science.

Adey, P., Csapo, B., Demetriou, A., Hautamäki, J., & Shayer, M. (2007). Can we be intelligent about intelligence? Why education needs the concept of plastic general ability. *Educational Research Review*, *2*(2), 75–97. http://doi.org/10.1016/j.edurev.2007.05.001

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *19*(6), 716–723. http://doi.org/10.1109/TAC.1974.1100705

Alexander, K. L., Entwisle, D. R., & Dauber, S. L. (2003). *On the Success of Failure: A Reassessment of the Effects of Retention in the Primary School Grades* (2 edition). Cambridge, UK: Cambridge University Press.

Alexander, K. L., Entwisle, D. R., & Horsey, C. S. (1997). From First Grade Forward: Early Foundations of High School Dropout. *Sociology of Education*, *70*(2), 87–107. http://doi.org/10.2307/2673158

Anastasi, A., & Urbina, S. (1997). *Psychological Testing* (Vol. 7). Upper Saddle River, NJ: Prentice Hall.

Arbuckle, J. L. (1996). Full information estimation in the presence of incomplete data. In G. A. Marcoulides & R. E. Schumacker (Eds.), *Advanced Structural Equation Modeling* (pp. 243–277). Mahwah, NJ: Lawrence Erlbaum Associates.

Austin, P. C. (2011). Optimal caliper widths for propensity-score matching when estimating differences in means and differences in proportions in observational studies. *Pharmaceutical Statistics*, *10*(2), 150–161. http://doi.org/10.1002/pst.433

Bacher, J. (2002). Statistisches Matching: Anwendungsmöglichkeiten, Verfahren und ihre praktische Umsetzung in SPSS [Statistical matching: potential applications, methods and their practical implementation in SPSS]. *ZA-Information / Zentralarchiv für Empirische Sozialforschung*, (51), 38–66.

Balinsky, B. (1941). An analysis of the mental factors of various age groups from nine to sixty. *Genetic Psychology Monographs*, *23*, 191–234.

Ball, K., Berch, D. B., Helmers, K. F., Jobe, J. B., Leveck, M. D., Marsiske, M., … Willis, S. L. (2002). Effects of Cognitive Training Interventions With Older Adults. *JAMA : The*

Journal of the American Medical Association, *288*(18), 2271–2281. http://doi.org/10.1001/jama.288.18.2271

Baltes, P. B. (1987). Theoretical propositions of life-span developmental psychology: On the dynamics between growth and decline. *Developmental Psychology*, *23*(5), 611–626. http://doi.org/10.1037/0012-1649.23.5.611

Baltes, P. B., Cornelius, S. W., Spiro, A., Nesselroade, J. R., & Willis, S. L. (1980). Integration versus differentiation of fluid/crystallized intelligence in old age. *Developmental Psychology*, *16*(6), 625–635. http://doi.org/10.1037/0012-1649.16.6.625

Baltes, P. B., & Lindenberger, U. (1997). Emergence of a powerful connection between sensory and cognitive functions across the adult life span: A new window to the study of cognitive aging? *Psychology and Aging*, *12*(1), 12–21. http://doi.org/10.1037/0882-7974.12.1.12

Baltes, P. B., Staudinger, U., & Lindenberger, U. (1999). Lifespan psychology: Theory and application to intellectual functioning. *Annual Review of Psychology*, *50*, 471–507. http://doi.org/10.1146/annurev.psych.50.1.471

Bamberg, M., Dickes, P., & Schaber, G. (1977). *Etude MAGRIP: Premier rapport de synthèse [The MAGRIP Study: First report and summary]*.

Beauducel, A., Liepmann, D., Felfe, J., & Nettelnstroth, W. (2007). The impact of different measurement models for fluid and crystallized intelligence on the correlation with personality traits. *European Journal of Psychological Assessment*, *23*(2), 71–78. http://doi.org/10.1027/1015-5759.23.2.71

Becker, M., Lüdtke, O., Trautwein, U., Köller, O., & Baumert, J. (2012). The differential effects of school tracking on psychometric intelligence: Do academic-track schools make students smarter? *Journal of Educational Psychology*, *104*(3), 682–699. http://doi.org/10.1037/a0027608

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, *107*(2), 238–246. http://doi.org/10.1037/0033-2909.107.2.238

Bentler, P. M. (1995). *EQS structural equations program manual*. Encino, CA: Multivariate Software.

Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin*, *88*(3), 588–606. http://doi.org/10.1037/0033-2909.88.3.588

Bollen, K. A. (1989). *Structural equations with latent variables*. New York, NY: Wiley.

Borkenau, P., & Liebler, A. (1993). Convergence of stranger ratings of personality and intelligence with self-ratings, partner ratings, and measured intelligence. *Journal of Personality and Social Psychology*, *65*(3), 546–553. http://doi.org/10.1037/0022-3514.65.3.546

Bradley, R. H., & Corwyn, R. F. (2002). Socioeconomic status and child development. *Annual Review of Psychology*, *53*(1), 371–399. http://doi.org/10.1146/annurev.psych.53.100901.135233

Brody, N. (1997). Intelligence, schooling, and society. *American Psychologist*, *52*(10), 1046–1050. http://doi.org/10.1037/0003-066X.52.10.1046

Brody, N. (2000). History of theories and measurement of intelligence. In R. J. Sternberg (Ed.), *Handbook of Intelligence* (1st ed., pp. 3–15). Cambridge, UK: Cambridge University Press.

Brunner, M., Nagy, G., & Wilhelm, O. (2012). A tutorial on hierarchically structured constructs. *Journal of Personality*. http://doi.org/10.1111/j.1467-6494.2011.00749.x

Byrd, R. S., & Weitzman, M. L. (1994). Predictors of early grade retention among children in the United States. *Pediatrics*, *93*(3), 481–487.

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*(3), 456–466. http://doi.org/10.1037//0033-2909.105.3.456

Byrne, B. M., & Stewart, S. M. (2006). Teacher’s corner: The MACS approach to testing for multigroup invariance of a second-order structure: A walk through the process. *Structural Equation Modeling*, *13*(2), 287–321. http://doi.org/10.1207/s15328007sem1302_7

Cahan, S., & Cohen, N. (1989). Age versus schooling effects on intelligence development. *Child Development*, *60*(5), 1239–49. http://doi.org/10.2307/1130797

Carroll, J. B. (1989). The Carroll model a 25-year retrospective and prospective view. *Educational Researcher*, *18*(1), 26–31. http://doi.org/10.3102/0013189X018001026

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

Cattell, R. B. (1967). The theory of fluid and crystallized general intelligence checked at the 5-6 year-old level. *The British Journal of Educational Psychology*, *37*(2), 209–224. http://doi.org/10.1111/j.2044-8279.1967.tb01930.x

Cattell, R. B. (1971). *Abilities: Their structure, growth, and action*. Oxford, England: Houghton Mifflin.

Cattell, R. B. (1987). *Intelligence: Its structure, growth and action*. Amsterdam: Elsvier.

Ceci, S. J. (1991). How much does schooling influence general intelligence and its cognitive components? A reassessment of the evidence. *Developmental Psychology*, *27*(5), 703–722. http://doi.org/10.1037/0012-1649.27.5.703

Ceci, S. J. (1999). Schooling and intelligence. In S. J. Ceci & W. M. Williams (Eds.), *The nature-nurture debate: The essential readings. Essential Readings in Developmental Psychology* (pp. 167–175). Oxford: Blackwell.

Ceci, S. J., & Papierno, P. B. (2005). The rhetoric and reality of gap closing: When the "have-nots" gain but the "haves" gain even more. *American Psychologist*, *60*(2), 149–160. http://doi.org/10.1037/0003-066X.60.2.149

Ceci, S. J., & Williams, W. M. (1997). Schooling, intelligence, and income. *American Psychologist*, *52*(10), 1051–1058. http://doi.org/10.1037/0003-066X.52.10.1051

Chen, X., Liu, C., Zhang, L., Shi, Y., & Rozelle, S. (2010). Does taking one step back get you two steps forward? Grade retention and school performance in poor areas in rural China. *International Journal of Educational Development*, *30*(6), 544–559. http://doi.org/10.1016/j.ijedudev.2009.12.002

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*, *9*(2), 233–255. http://doi.org/10.1207/S15328007SEM0902_5

Cliffordson, C., & Gustafsson, J. E. (2008). Effects of age and schooling on intellectual performance: Estimates obtained from analysis of continuous variation in age and length of schooling. *Intelligence*, *36*(2), 143–152. http://doi.org/10.1016/j.intell.2007.03.006

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*(1), 155–159. http://doi.org/10.1037/0033-2909.112.1.155

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/ correlation analysis for the behavioral sciences* (3rd ed.). Lawrence Erlbaum Associates.

Cole, D. A., & Maxwell, S. E. (2003). Testing mediational models with longitudinal data: Questions and tips in the use of structural equation modeling. *Journal of Abnormal Psychology*, *112*(4), 558–577. http://doi.org/10.1037/0021-843X.112.4.558

Conley, J. J. (1984). The hierarchy of consistency: A review and model of longitudinal findings on adult individual differences in intelligence, personality and self-opinion.

Personality and Individual Differences, *5*(1), 11–25. http://doi.org/10.1016/0191-8869(84)90133-8

Deary, I. J. (2010). Cognitive epidemiology: Its rise, its current issues, and its challenges. *Personality and Individual Differences*, *49*(4), 337–343. http://doi.org/10.1016/j.paid.2009.11.012

Deary, I. J., Egan, V., Gibson, G. J., Austin, E. J., Brand, C. R., & Kellaghan, T. (1996). Intelligence and the differentiation hypothesis. *Intelligence*, *23*(2), 105–132. http://doi.org/10.1016/S0160-2896(96)90008-2

Deary, I. J., Strand, S., Smith, P., & Fernandes, C. (2007). Intelligence and educational achievement. *Intelligence*, *35*(1), 13–21. http://doi.org/10.1016/j.intell.2006.02.001

Deary, I. J., Whalley, L. J., Lemmon, H., Crawford, J. R., & Starr, J. M. (2000). The stability of individual differences in mental ability from childhood to old age: follow-up of the 1932 Scottish Mental Survey. *Intelligence*, *28*(1), 49–55. http://doi.org/10.1016/S0160-2896(99)00031-8

Deary, I. J., Whiteman, M. C., Starr, J. M., Whalley, L. J., & Fox, H. C. (2004). The impact of childhood intelligence on later life: following up the Scottish Mental Surveys of 1932 and 1947. *Journal of Personality and Social Psychology*, *86*(1), 130–147. http://doi.org/10.1037/0022-3514.86.1.130

DeShon, R. P. (2004). Measures are not invariant across groups without error variance homogeneity. *Psychology Science*, *46*(1), 137–149.

Dickens, W. T., & Flynn, J. R. (2001). Heritability estimates versus large environmental effects: The IQ paradox resolved. *Psychological Review*, *108*(2), 346. http://doi.org/10.1037/0033-295X.108.2.346

Dickes, P. (2011). Vorwort/préface [Foreword]. In M. Brunner & R. Martin (Eds.), *Die MAGRIP-Studie (1968-2009). Wie beeinflussen sozio-kognitive Merkmale von Kindern im Grundschulalter und ihre Bildungswege ihr späteres Leben als Erwachsene in Luxemburg? [The MAGRIP study (1968-2009). How socio-cognitive childhood characteristics in primary school influence educational pathways in later adulthood in Luxembourg?]* (pp. 6–9). Luxemburg: Universität Luxemburg, Forschungseinheit EMACS.

DiPrete, T. A., & Eirich, G. M. (2006). Cumulative Advantage as a Mechanism for Inequality: A Review of Theoretical and Empirical Developments - Annual Review of Sociology, 32(1):271. *Annual Review of Sociology*, *32*, 271–297. http://doi.org/10.1146/annurev.soc.32.061604.123127

Dunlap, W. P., Cortina, J. M., Vaslow, J. B., & Burke, M. J. (1996). Meta-analysis of experiments with matched groups or repeated measures designs. *Psychological Methods*, *1*(2), 170–177. http://doi.org/10.1037/1082-989X.1.2.170

Edwards, J. R., & Bagozzi, R. P. (2000). On the nature and direction of relationships between constructs and measures. *Psychological Methods*, *5*(2), 155–174. http://doi.org/10.1037//1082-989X.5.2,155

Eichorn, D. H., Hunt, J. V., & Honzik, M. P. (1981). Experience, personality, and IQ: Adolescence to middle age. In D. H. Eichorn, J. A. Clausen, N. Haan, M. P. Honzik, & P. H. Mussen (Eds.), *Present and past in middle life* (pp. 89–116). New York, NY: Academic Press.

Escorial, S., Juan-Espinosa, M., García, L. F., Rebollo, I., & Colom, R. (2003). Does g variance change in adulthood? Testing the age de-differentiation hypothesis across sex. *Personality and Individual Differences*, *34*(8), 1525–1532. http://doi.org/10.1016/S0191-8869(02)00133-2

European Commission. (2008). Improving competences for the 21st century: An agenda for European cooperation on schools. Retrieved from http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2008:0425:FIN:EN:PDF

European Commission. (2011). *Grade retention during compulsory education in Europe: Regulations and statistics* (1st ed.). Brussels: EACEA — Education, Audiovisual and Culture Executive Agency. Retrieved from http://bookshop.europa.eu/en/grade-retention-during-compulsory-education-in-europe-pbEC3110651/

European Commission. (2014). *Population ageing in Europe*. Luxembourg: Publications Office of the European Union.

Ferrer, E., & McArdle, J. J. (2004). An experimental analysis of dynamic hypotheses about cognitive abilities and achievement from childhood to early adulthood. *Developmental Psychology*, *40*(6), 935–952. http://doi.org/10.1037/0012-1649.40.6.935

Flynn, J. R. (1987). Massive IQ gains in 14 nations: What IQ tests really measure. *Psychological Bulletin*, *101*(2), 171–191. http://doi.org/10.1037/0033-2909.101.2.171

Ganzeboom, H. B. G., De Graaf, P. M., Treiman, D. J., & De Leeuw, J. (1992). A standard international socio-economic index of occupational status. *Social Science Research*, *21*(1), 1–56. http://doi.org/10.1016/0049-089X(92)90017-B

Ganzeboom, H. B. G., & Treiman, D. J. (1996). Internationally comparable measures of occupational status for the 1988 International Standard Classification of Occupations. *Social Science Research*, *25*(3), 201–239. http://doi.org/10.1006/ssre.1996.0010

Ganzeboom, H. B. G., & Treiman, D. J. (2003). Three Internationally Standardised Measures for Comparative Research on Occupational Status. In J. H. P. Hoffmeyer-Zlotnik & C. Wolf, *Advances in Cross-National Comparison: A European Working Book for Demographic and Socio-economic Variables* (pp. 159–193). New York: Springer Science & Business Media.

Garrett, H. E. (1946). A developmental theory of intelligence. *American Psychologist*, *1*(9), 372–378. http://doi.org/10.1037/h0056380

Ghisletta, P., & De Ribaupierre, A. (2005). A dynamic investigation of cognitive dedifferentiation with control for retest: Evidence from the Swiss Interdisciplinary Longitudinal Study on the Oldest Old. *Psychology and Aging*, *20*(4), 671–682. http://doi.org/10.1037/0882-7974.20.4.671

Ghisletta, P., & Lindenberger, U. (2003). Age-based structural dynamics between perceptual speed and knowledge in the berlin aging study: direct evidence for ability dedifferentiation in old age. *Psychology and Aging*, *18*(4), 696–713. http://doi.org/10.1037/0882-7974.18.4.696

Glaser, R. (1976). Components of a psychology of instruction: Toward a science of design. *Review of Educational Research*, 1–24. http://doi.org/10.3102/00346543046001001

Glaser, R. (1982). Instructional psychology: Past, present, and future. *American Psychologist; American Psychologist*, *37*(3), 292. http://doi.org/10.1037/0003-066X.37.3.292

Goos, M., Van Damme, J., Onghena, P., Petry, K., & de Bilde, J. (2013). First-grade retention in the Flemish educational context: Effects on children's academic growth, psychosocial growth, and school career throughout primary education. *Journal of School Psychology*, *51*(3), 323–347. http://doi.org/10.1016/j.jsp.2013.03.002

Gottfredson, D. C., Fink, C. M., & Graham, N. (1994). Grade retention and problem behavior. *American Educational Research Journal*, *31*(4), 761–784. http://doi.org/10.3102/00028312031004761

Gottfredson, L. S. (1997a). Mainstream science on intelligence: An editorial with 52 signatories, history and bibliography. *Intelligence*, *24*(1), 13–23. http://doi.org/10.1016/S0160-2896(97)90011-8

Gottfredson, L. S. (1997b). Why g matters: The complexity of everyday life. *Intelligence*, *24*, 79–132. http://doi.org/10.1016/S0160-2896(97)90014-3

Gottfredson, L. S. (1998). The general intelligence factor. *Scientific American Presents*, 24–29.

Gottfredson, L. S. (2002). Where and why g matters: Not a mystery. *Human Performance*, *15*(1), 25–46. http://doi.org/10.1207/S15327043HUP1501&02_03

Gottfredson, L. S. (2003). The Challenge and Promise of Cognitive Career Assessment. *Journal of Career Assessment*, *11*(2), 115 –135. http://doi.org/10.1177/1069072703011002001

Gottfredson, L. S., & Deary, I. J. (2004). Intelligence predicts health and longevity, but why? *Current Directions in Psychological Science*, *13*(1), 1–4. http://doi.org/10.1111/j.0963-7214.2004.01301001.x

Gottfredson, L. S., & Saklofske, D. H. (2009). Intelligence: Foundations and issues in assessment. *Canadian Psychology*, *50*(3), 183–195. http://doi.org/10.1037/a0016641

Gustafsson, J.-E., & Undheim, J. O. (1992). Stability and change in broad and narrow factors of intelligence from ages 12 to 15 years. *Journal of Educational Psychology*, *84*(2), 141–149. http://doi.org/10.1037/0022-0663.84.2.141

Haertel, G. D., Walberg, H. J., & Weinstein, T. (1983). Psychological models of educational performance: A theoretical synthesis of constructs. *Review of Educational Research*, *53*(1), 75–91. http://doi.org/10.3102/00346543053001075

Hansen, B. B. (2004). Full matching in an observational study of coaching for the SAT. *Journal of the American Statistical Association*, *99*(467), 609–618. http://doi.org/10.1198/016214504000000647

Hansen, B. B., & Bowers, J. (2008). Covariate balance in simple, stratified and clustered comparative studies. *Statistical Science*, 219–236. http://doi.org/10.1214/08-STS254

Härnqvist, K. (1968). Relative changes in intelligence from 13 to 18: I. Background and methodology. *Scandinavian Journal of Psychology*, *9*(1), 50–64. http://doi.org/10.1111/j.1467-9450.1968.tb00519.x

Hattie, J. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement* (1st ed.). New York, USA: Routledge.

Hayslip Jr, B., & Sterns, H. L. (1979). Age differences in relationships between crystallized and fluid intelligences and problem solving. *Journal of Gerontology*, *34*(3), 404–414. http://doi.org/10.1093/geronj/34.3.404

Heckman, J. J. (2000). Policies to foster human capital. *Research in Economics*, *54*(1), 3–56. http://doi.org/10.1006/reec.1999.0225

Heckman, J. J. (2006). Skill formation and the economics of investing in disadvantaged children. *Science*, *312*(5782), 1900–1902. http://doi.org/10.1126/science.1128898

Heckman, J. J. (2008). Schools, skills, and synapses. *Economic Inquiry*, *46*(3), 289–324. http://doi.org/10.1111/j.1465-7295.2008.00163.x

Hertzog, C., & Schaie, K. W. (1986). Stability and change in adult intelligence: I. Analysis of longitudinal covariance structures. *Psychology and Aging*, *1*(2), 159–171. http://doi.org/10.1037/0882-7974.1.2.159

Ho, D., Imai, K., King, G., & Stuart, E. (2007a). Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. *Political Analysis*, *15*(3), 199–236. http://doi.org/10.1093/pan/mpl013

Ho, D., Imai, K., King, G., & Stuart, E. (2007b). MatchIt: nonparametric preprocessing for parametric causal inference. *Journal of Statistical Software*. http://doi.org/10.18637/jss.v042.i08

Holmes, T. H., & Matthews, K. M. (1984). The effects of nonpromotion on elementary and junior high school pupils: A meta-analysis. *Review of Educational Research*, *54*(2), 225–236. http://doi.org/10.3102/00346543054002225

Hommel, B., Li, K. Z. H., & Li, S.-C. (2004). Visual search across the life span. *Developmental Psychology*, *40*(4), 545–558. http://doi.org/10.1037/0012-1649.40.4.545

Horn, J. L., & McArdle, J. J. (2007). Understanding human intelligence since Spearman. In R. Cudeck & R. C. MacCallum (Eds.), *Factor analysis at 100* (pp. 205–247). Mahwah, NJ: Lawrence Erlbaum Associates.

Horn, J. L., McArdle, J. J., & Mason, R. (1983). When is invariance not invariant: A practical scientist’s look at the ethereal concept of factor invariance. *The Southern Psychologist*, *1*(4), 179–188.

Horn, J. L., & Noll, J. (1997). Human cognitive capabilities: Gf-Gc theory. In D. P. Flanagan, J. L. Genshaft, & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues.* (pp. 53–91). New York, NY: Guilford Press.

Horn, W. (1962). *Leistungsprüfsystem (L-P-S). Handanweisung [Achievement Test Battery (L-P-S). Test manual (1st edition)]* (1. ed.). Göttingen: Hogrefe.

Horn, W. (1983). *Leistungsprüfsystem (L-P-S). Handanweisung [Achievement Test Battery (L-P-S). Test manual (2nd edition)]* (2. ed.). Göttingen: Hogrefe.

Hughes, J. N., Chen, Q., Thoemmes, F. J., & Kwok, O. (2010). An investigation of the relationship between retention in first grade and performance on high stakes tests in third grade. *Educational Evaluation and Policy Analysis*, *32*(2), 166–182. http://doi.org/10.3102/0162373710367682

Hughes, J. N., Kwok, O., & Im, M. H. (2013). Effect of Retention in First Grade on Parents' Educational Expectations and Children's Academic Outcomes. *American Educational Research Journal*, *50*(6), 1336–1359. http://doi.org/10.3102/0002831213490784

Hu, L., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods*, *3*(4), 424–453. http://doi.org/10.1037//1082-989X.3.4.424

Hülsheger, U. R., Maier, G. W., Stumpp, T., & Muck, P. M. (2006). Vergleich kriteriumsbezogener Validitäten verschiedener Intelligenztests zur Vorhersage von Ausbildungserfolg in Deutschland [Comparison of criterion-related validities of different intelligence tests for the prediction of training success in Germany]. *Zeitschrift Für Personalpsychologie*, *5*(4), 145–162. http://doi.org/10.1026/1617-6391.5.4.145

Hunt, E. (2011). *Human Intelligence.* New York, NY, US: Cambridge University Press.

Hunter, J. E., & Hunter, R. F. (1984). Validity and utility of alternative predictors of job performance. *Psychological Bulletin*, *96*(1), 72–98. http://doi.org/10.1037/0033-2909.96.1.72

Husén, T., & Tuijnman, A. (1991). The contribution of formal schooling to the increase in intellectual capital. *Educational Researcher*, *20*(7), 17–25. http://doi.org/10.3102/0013189X020007017

Im, M. H., Hughes, J. N., Kwok, O., Puckett, S., & Cerda, C. A. (2013). Effect of retention in elementary grades on transition to middle school. *Journal of School Psychology*, *51*(3), 349–365. http://doi.org/10.1016/j.jsp.2013.01.004

International Labour Organization (ILO). (1968). Bureau of Statistics, work unit of the Policy Integration Department. Retrieved September 13, 2015, from http://www.ilo.org/public/english/bureau/stat/isco/isco68/major.htm

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

Jimerson, S. R. (1999). On the failure of failure: Examining the association between early grade retention and education and employment outcomes during late adolescence. *Journal of School Psychology*, *37*(3), 243–272. http://doi.org/10.1016/S0022-4405(99)00005-9

Jimerson, S. R. (2001). Meta-analysis of grade retention research: Implications for practice in the 21st century. *School Psychology Review*, *30*(3), 420–437.

Jimerson, S. R., Anderson, G. E., & Whipple, A. (2002). Winning the battle and losing the war: Examining the relation between grade retention and dropping out of high school. *Psychology in the Schools*, *39*(4), 441–457. http://doi.org/10.1002/pits.10046

Jimerson, S. R., Carlson, E., Rotert, M., Egeland, B., & Sroufe, L. A. (1997). A prospective, longitudinal study of the correlates and consequences of early grade retention. *Journal of School Psychology*, *35*(1), 3– 25. http://doi.org/10.1016/S0022-4405(96)00033-7

Jimerson, S. R., Ferguson, P., Whipple, A. D., Anderson, G. E., & Dalton, M. J. (2002). Exploring the association between grade retention and dropout: A longitudinal study examining socio-emotional, behavioral, and achievement characteristics of retained students. *California School Psychologist*, *7*, 51–62. http://doi.org/10.1007/BF03340889

Kangas, J., & Bradway, K. (1971). Intelligence at middle age: A thirty-eight year follow-up. *Developmental Psychology*, *5*(2), 333–337. http://doi.org/10.1037/h0031471

Kan, K.-J., Kievit, R. A., Dolan, C., & van der Maas, H. L. J. (2011). On the interpretation of the CHC factor Gc. *Intelligence*, *39*(5), 292–302. http://doi.org/10.1016/j.intell.2011.05.003

Kolan, M. J., & Brennan, R. L. (1995). *Test equating: Methods and practices*. New York, NY: Springer-Verlag.

Kratzmeier, H. (1979). *RAVEN-Matrizen-Test – Standard Progressive Matrices (Manual)[RAVEN Matrices Test – Standard Progressive Matrices (Manual)]*. Weinheim, Germany: Beltz.

Kuelz, A. K., Hohagen, F., & Voderholzer, U. (2004). Neuropsychological performance in obsessive-compulsive disorder: a critical review. *Biological Psychology*, *65*(3), 185–236. http://doi.org/10.1016/j.biopsycho.2003.07.007

Kultusministerkonferenz. (2005a). *Bildungsstandards der Kultusministerkonferenz - Erläuterungen zur Konzeption und Entwicklung* (Vol. 1). Munich: Luchterhand.

Kultusministerkonferenz. (2005b). *Bildungsstandards im Fach Deutsch für den Primarbereich (Jahrgangsstufe 4) Beschluss vom 15.10.2004* (Vol. 1). Munich: Luchterhand.

Kultusministerkonferenz. (2005c). *Bildungsstandards im Fach Mathematik für den Primarbereich (Jahrgangsstufe 4) Beschluss vom 15.10.2004* (Vol. 1). Munich: Luchterhand.

Kvist, A. V., & Gustafsson, J. E. (2008). The relation between fluid intelligence and the general factor as a function of cultural background: A test of Cattell's Investment theory. *Intelligence*, *36*(5), 422–436. http://doi.org/10.1016/j.intell.2007.08.004

Larsen, L., Hartmann, P., & Nyborg, H. (2008). The stability of general intelligence from early childhood to middle-age. *Intelligence*, *36*(1), 29–34. http://doi.org/10.1016/j.intell.2007.01.001

Liepmann, D., Beauducel, A., Brocke, B., & Amthauer, R. (2001). *Intelligenz-Struktur-Test 2000 R. Manual*. Göttingen: Hogrefe.

Lindenberger, U. (2001). Lifespan theories of cognitive development. In N. J. Smelser & P. B. Baltes (Eds.), *International encyclopedia of the social and behavioral sciences* (pp. 8848–8854). Oxford, UK: Elsevier.

Lindenberger, U., & von Oertzen, T. (2006). Variability in cognitive aging: From taxonomy to theory. In E. Bialystok & F. I. M. Craik (Eds.), *Lifespan cognition: Mechanisms of change* (pp. 297–314). New York, NY: Oxford University Press.

Li, S.-C., & Baltes, P. B. (2006). Cognitive developmental research from lifespan perspectives: The challenge of integration. In E. Bialystok & F. I. M. Craik (Eds.), *Lifespan cognition: Mechanisms of change* (pp. 344–363). New York, NY: Oxford University Press.

Li, S.-C., Lindenberger, U., Hommel, B., Aschersleben, G., Prinz, W., & Baltes, P. B. (2004). Transformations in the couplings among intellectual abilities and constituent cognitive processes across the life span. *Psychological Science*, *15*(3), 155 –163. http://doi.org/10.1111/j.0956-7976.2004.01503003.x

Little, R. J. A., & Rubin, D. B. (2002). *Statistical analysis with missing data* (2nd ed.). New York, NY: John Wiley.

Little, T. D. (1997). Mean and covariance structures (MACS) analyses of cross-cultural data: Practical and theoretical issues. *Multivariate Behavioral Research*, *32*(1), 53–76. http://doi.org/10.1207/s15327906mbr3201_3

Lubinski, D. (2004). Introduction to the special section on cognitive abilities: 100 years after Spearman's (1904) "'General intelligence,' objectively determined and measured". *Journal of Personality and Social Psychology*, *86*(1), 96–111. http://doi.org/10.1037/0022-3514.86.1.96

Lubke, G. H., Dolan, C. V., Kelderman, H., & Mellenbergh, G. J. (2003). On the relationship between sources of within-and between-group differences and measurement invariance in the common factor model. *Intelligence*, *31*(6), 543–566. http://doi.org/10.1016/S0160-2896(03)00051-5

Mazzonna, F., & Peracchi, F. (2012). *Aging, cognitive abilities and retirement* (CEIS Research Paper No. 152). Tor Vergata University, CEIS. Retrieved from http://econpapers.repec.org/paper/rtvceisrp/152.htm

McArdle, J. J., Ferrer-Caja, E., Hamagami, F., & Woodcock, R. W. (2002). Comparative longitudinal structural analyses of the growth and decline of multiple intellectual abilities over the life span. *Developmental Psychology*, *38*(1), 115–142. http://doi.org/10.1037/0012-1649.38.1.115

McCoy, A. R., & Reynolds, A. J. (1999). Grade retention and school performance: An extended investigation. *Journal of School Psychology*, *37*(3), 273–298. http://doi.org/10.1016/S0022-4405(99)00012-6

McDonald, L. R., & Bean, L. C. (1992). Thinking of retaining a student? Try one or more of the twenty-five alternatives to retention. *Education*, *112*(4), 567–570.

McGrew, K. (2009). CHC theory and the human cognitive abilities project: Standing on the shoulders of the giants of psychometric intelligence research. *Intelligence*, *37*(1), 1–10. http://doi.org/10.1016/j.intell.2008.08.004

Meisels, S. J., & Liaw, F. R. (1993). Failure in grade: Do retained students catch up? *Journal of Educational Research*, *87*(2), 69–77. http://doi.org/10.1080/00220671.1993.9941169

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, *58*(4), 525–543. http://doi.org/10.1007/BF02294825

Meredith, W., & Horn, J. L. (2001). The role of factorial invariance in modeling growth and change. In L. M. Collins & A. G. Sayer (Eds.), *New methods for the analysis of change* (pp. 203–240). Washington, DC: American Psychological Association.

Millsap, R. E., & Olivera-Aguilar, M. (2012). Investigating measurement invariance using confirmatory factor analysis. In R. Hoyle (Ed.), *Handbook of structural equation modeling* (pp. 380–392). New York, NY: Guilford Press.

Ming, K., & Rosenbaum, P. R. (2000). Substantial gains in bias reduction from matching with a variable number of controls. *Biometrics*, *56*(1), 118–124. http://doi.org/10.1111/j.0006-341X.2000.00118.x

Molenaar, D., Dolan, C. V., Wicherts, J. M., & van der Maas, H. L. J. (2010). Modeling differentiation of cognitive abilities within the higher-order factor model using moderated factor analysis. *Intelligence*, *38*(6), 611–624. http://doi.org/10.1016/j.intell.2010.09.002

Morgan, S. L., & Winship, C. (2007). *Counterfactuals and causal inference: Methods and principles for social research* (1st ed.). Cambridge, United Kingdom: Cambridge University Press.

Muthén, L. K., & Muthén, B. O. (1998). *Mplus user's guide. Fifth edition*. Los Angeles, CA: Muthén & Muthén.

Neisser, U., Boodoo, G., Bouchard Jr, T. J., Boykin, A. W., Brody, N., Ceci, S. J., … Urbina, S. (1996). Intelligence: Knowns and unknowns. *American Psychologist*, *51*(2), 77–101. http://doi.org/10.1037/0003-066X.51.2.77

Neubauer, A. C., Fink, A., & Schrausser, D. G. (2002). Intelligence and neural efficiency: The influence of task content and sex on the brain–IQ relationship. *Intelligence*, *30*(6), 515–536. http://doi.org/10.1016/S0160-2896(02)00091-0

Niklason, L. B. (1984). Nonpromotion: A pseudoscientific solution. *Psychology in the Schools*, *21*(4), 485–499. http://doi.org/10.1002/1520-6807(198410)21:4<485::AID-PITS2310210415>3.0.CO;2-Z

Nisbet, J. D. (1957). Intelligence and age: Retesting with twenty-four years' interval. *British Journal of Educational Psychology*, *27*, 190–198.

Nisbett, R. E., Aronson, J., Blair, C., Dickens, W., Flynn, J., Halpern, D. F., & Turkheimer, E. (2012). Intelligence: New findings and theoretical developments. *American Psychologist*, *67*(2), 130–159. http://doi.org/10.1037/a0026699

OECD (Ed.). (1999). *Classifying educational programmes. Manual for ISCED-97 implementation on OECD countries*. Paris, France: OECD Publishing. Retrieved from http://www.oecd.org/dataoecd/41/42/1841854.pdf

Owens, W. (1966). Age and mental abilities: A second adult follow-up. *Journal of Educational Psychology, 57*(6), 311–325. http://doi.org/10.1037/h0023962

Owings, W. A., & Kaplan, L. S. (2001). *Alternatives to retention and social promotion.* Bloomington, Indiana: Phi Delta Kappa International.

Pagani, L., Tremblay, R. E., Vitaro, F., Boulerice, B., & McDuff, P. (2001). Effects of grade retention on academic performance and behavioral development. *Development and Psychopathology*, *13*(2), 297–315. http://doi.org/10.1017/S0954579401002061

Phelps, L., Dowdell, N., Rizzo, F. G., Ehrlich, P., & Wilczenski, F. (1992). Five to ten years after placement: The long-term efficacy of retention and pre-grade transition. *Journal* *of Psychoeducational Assessment*, *10*(2), 116–123. http://doi.org/10.1177/073428299201000203

Pierson, L. H., & Connell, J. P. (1992). Effect of grade retention on self-system processes, school engagement, and academic performance. *Journal of Educational Psychology*, *84*(3), 300–307. http://doi.org/10.1037/0022-0663.84.3.300

Pushkar Gold, D., Andres, D., Etezadi, J., Arbuckle, T., Schwartzman, A., & Chaikelson, J. (1995). Structural equation model of intellectual change and continuity and predictors of intelligence in older men. *Psychology and Aging*, *10*(2), 294–303. http://doi.org/10.1037/0882-7974.10.2.294

Reinert, G. (1970). Comparative factor analytic studies of intelligence through the human life-span. In L. R. Goulet & P. B. Baltes (Eds.), *Life-span developmental psychology: Research and theory* (pp. 468– 485). New York, NY: Academic Press.

Renkl, A. (1996). Vorwissen und Schulleistung. In J. Möller & O. Köller (Eds.), *Emotionen, Kognitionen und Schulleistung* (pp. 175–190). Weinheim: Beltz.

Reynolds, A. J. (1992). Grade retention and school adjustment: An explanatory analysis. *Educational Evaluation and Policy Analysis*, *14*(2), 101–121. http://doi.org/10.2307/1164496

Rosenbaum, P. R., & Rubin, D. B. (1985). Constructing a Control Group Using Multivariate Matched Sampling Methods That Incorporate the Propensity Score. *The American Statistician*, *39*(1), 33–38. http://doi.org/10.2307/2683903

Rubin, D. B., & Thomas, N. (2000). Combining Propensity Score Matching with Additional Adjustments for Prognostic Covariates. *Journal of the American Statistical Association*, *95*(450), 573–585. http://doi.org/10.1080/01621459.2000.10474233

Rumberger, R. (1995). Dropping Out of Middle School: A Multilevel Analysis of Students and Schools. *American Educational Research Journal*, *32*(3), 583–625. http://doi.org/10.3102/00028312032003583

Rust, J. O., & Wallace, K. A. (1993). Effects of grade level retention for four years. *Journal of Instructional Psychology*, *20*(2), 162–166.

Salthouse, T. A. (1996). The processing-speed theory of adult age differences in cognition. *Psychological Review*, *103*(3), 403–428. http://doi.org/10.1037/0033-295X.103.3.403

Sandoval, J. (1984). First-Grade Promotional Practices: Implications for the School Psychologist's Role: A Symposium. *Psychology in the Schools*, *21*(4), 457–76.

Satorra, A., & Bentler, P. M. (1999). A scaled difference chi-square test statistic for moment structure analysis. Retrieved June 8, 2011, from http://preprints.stat.ucla.edu/260/chisquare.pdf

Scarr, S., & McCartney, K. (1983). How people make their own environments: A theory of genotype→ environment effects. *Child Development*, *54*(2), 424–435. http://doi.org/10.2307/1129703

Schaie, K. W. (2005). *Developmental influences on adult intelligence: The Seattle longitudinal study*. New York: Oxford University Press.

Schaie, K. W., & Hofer, S. M. (2001). Longitudinal studies in research on aging. In J. E. Birren & K. W. Schaie (Eds.), *Handbook of the psychology of aging* (5th ed., pp. 55–77). San Diego, CA: Academia.

Schaie, K. W., Maitland, S. B., Willis, S. L., & Intrieri, R. C. (1998). Longitudinal invariance of adult psychometric ability factor structures across 7 years. *Psychology and Aging*, *13*(1), 8–20. http://doi.org/10.1037/0882-7974.13.1.8

Schaie, K. W., & Strother, C. R. (1968). A cross-sequential study of age changes in cognitive behavior. *Psychological Bulletin*, *70*(6), 671–680. http://doi.org/10.1037/h0026811

Schalke, D., Brunner, M., Geiser, C., Preckel, F., Keller, U., Spengler, M., & Martin, R. (2012). Stability and change in intelligence from age 12 to age 52: Results from the Luxembourg MAGRIP study. *Developmental Psychology*. http://doi.org/10.1037/a0030623

Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research Online*, *8*(2), 23–74.

Schipolowski, S., Wilhelm, O., & Schroeders, U. (2014). On the nature of crystallized intelligence: The relationship between verbal ability and factual knowledge. *Intelligence*, *46*, 156–168. http://doi.org/10.1016/j.intell.2014.05.014

Schmidt, F. L., & Hunter, J. (2004). General Mental Ability in the World of Work: Occupational Attainment and Job Performance. *Journal of Personality and Social Psychology*, *86*(1), 162–173. http://doi.org/10.1037/0022-3514.86.1.162

Schooler, C., & Mulatu, M. S. (2001). The reciprocal effects of leisure time activities and intellectual functioning in older people: A longitudinal analysis. *Psychology and Aging*, *16*(3), 466–482. http://doi.org/10.1037/0882-7974.16.3.466

Schooler, C., Mulatu, M. S., & Oates, G. (1999). The continuing effects of substantively complex work on the intellectual functioning of older workers. *Psychology and Aging, 14*(3), 483–506. http://doi.org/10.1037/0882-7974.14.3.483

Schwartzman, A. E., Gold, D., Andres, D., Arbuckle, T. Y., & Chaikelson, J. (1987). Stability of intelligence: A 40-year follow-up. *Canadian Journal of Psychology*, *41*(2), 244–256. http://doi.org/10.1037/h0084155

Shavit, Y., & Featherman, D. L. (1988). Schooling, tracking, and teenage intelligence. *Sociology of Education*, 42–51. http://doi.org/10.2307/2112308

Shea, D. L., Lubinski, D., & Benbow, C. P. (2001). Importance of assessing spatial ability in intellectually talented young adolescents: A 20-year longitudinal study. *Journal of Educational Psychology*, *93*(3), 604–614. http://doi.org/10.1037/0022-0663.93.3.604

Simon, T., & Binet, A. (1905). Méthodes nouvelles pour le diagnostic du niveau intellectuel des anormaux. *L'année Psychologique*, *11*(1), 191–244. http://doi.org/10.3406/psy.1904.3675

Simon, T., & Binet, A. (1916). New methods for the diagnosis of the intellectual level of subnormals. In *Development of intelligence in children (the Binet-Simon Scale) (E.S. Kite, Trans.* (pp. 37–90). Baltimore: Williams & Wilkins.

Snow, R. E. (1996). Aptitude development and education. *Psychology, Public Policy, and Law*, *2*(2/4), 536–560. http://doi.org/10.1037/1076-8971.2.3-4.536

Steiger, J. H. (1990). Structural model evaluation and modification: An interval estimation approach. *Multivariate Behavioral Research*, *25*(2), 173–180. http://doi.org/10.1207/s15327906mbr2502_4

Stelzl, I., Merz, F., Ehlers, T., & Remer, H. (1995). The effect of schooling on the development of fluid and cristallized intelligence: A quasi-experimental study. *Intelligence*, *21*(3), 279–296. http://doi.org/10.1016/0160-2896(95)90018-7

Strenze, T. (2007). Intelligence and socioeconomic success: A meta-analytic review of longitudinal research. *Intelligence*, *35*(5), 401–426. http://doi.org/10.1016/j.intell.2006.09.004

Sturm, W., & Büssing, A. (1982). Ein Vergleich von HAWIE und LPS bei der psychometrischen Einzelfalldiagnostik neurologischer Patienten [A comparison of HAWIE and LPS for psychometric individual case diagnosis of neurological patients]. *Diagnostica*, (28), 348–359.

Sturm, W., Willmes, K., & Horn, W. (1993). *Leistungsprüfsystem für 50-bis 90-Jährige (LPS 50+). Handanweisung [Achievement Test Battery for 50 to 90 year olds (LPS 50+). Test manual (1st edition)]* (1. ed.). Göttingen, Germany: Hogrefe.

Tent, L. (1965). Das Leistungsprüfsystem von Horn bei Schülern des 4. Volksschuljahres [The Performance Test System in grade 4 students]. *Psychologische Beiträge, 8*, 564–595.

Tent, L. (1969). *Die Auslese von Schülern für weiterführende Schulen [The selection of students to secondary schools]*. Göttingen, Germany: Hogrefe.

Tewes, U. (1991). *Hamburg-Wechsler-Intelligenztest für Erwachsene–Revision (HAWIE-R) [Hamburg Wechsler Intelligence Test for Adults–Revision (HAWIE-R)]*. Göttingen, Germany: Hogrefe.

Thoemmes, F. J. (2011). An SPSS R menu for propensity score matching. *Available at Arxiv. Org*.

Thoemmes, F. J. (2012). Propensity score matching in SPSS. *arXiv Preprint arXiv:1201.6385*. Retrieved from http://arxiv.org/abs/1201.6385

Thoemmes, F. J., & Kim, E. S. (2011). A systematic review of propensity score methods in the social sciences. *Multivariate Behavioral Research*, *46*(1), 90–118. http://doi.org/10.1080/00273171.2011.540475

Thurstone, L. L. (1938). *Primary mental abilities*. Chicago, IL: University Press.

Tucker-Drob, E. M. (2009). Differentiation of cognitive abilities across the life span. *Developmental Psychology*, *45*(4), 1097–1118. http://doi.org/10.1037/a0015864

Tucker-Drob, E. M. (2011). Neurocognitive functions and everyday functions change together in old age. *Neuropsychology*, *25*(3), 368–377. http://doi.org/10.1037/a0022348

Tucker-Drob, E. M., & Salthouse, T. A. (2008). Adult age trends in the relations among cognitive abilities. *Psychology and Aging*, *23*(2), 453–460. http://doi.org/10.1037/0882-7974.23.2.453

Tucker-Drob, E. M., & Salthouse, T. A. (2011). Individual differences in cognitive aging. In T. Chamorro-Premuzic, S. von Stumm, & A. Furnham (Eds.), *The Wiley-Blackwell handbook of individual differences* (pp. 242–267). Oxford, UK: Wiley-Blackwell.

Tuddenham, R. D., Blumenkrantz, J., & Wilkin, W. R. (1968). Age changes on AGCT: A longitudinal study of average adults. *Journal of Consulting and Clinical Psychology*, *32*(6), 659–663. http://doi.org/10.1037/h0026583

UNESCO (Ed.). (2006). *International Standard Classification of Education: ISCED 1997* (2nd ed.). UNESCO. Retrieved from http://www.uis.unesco.org/Library/Documents/isced97-en.pdf

van der Maas, H. L. J., Dolan, C. V., Grasman, R. P. P. P., Wicherts, J. M., Huizenga, H. M., & Raijmakers, M. E. J. (2006). A dynamical model of general intelligence: The positive manifold of intelligence by mutualism. *Psychological Review*, *113*(4), 842–861. http://doi.org/10.1037/0033-295X.113.4.842

Weiss, E. M., Kemmler, G., Deisenhammer, E. A., Fleischhacker, W. W., & Delazer, M. (2003). Sex differences in cognitive functions. *Personality and Individual Differences*, *35*(4), 863–875. http://doi.org/10.1016/S0191-8869(02)00288-X

Whalley, L. J., & Deary, I. J. (2001). Longitudinal cohort study of childhood IQ and survival up to age 76. *British Medical Journal*, *322*, 1–5. http://doi.org/10.1136/bmj.322.7290.819

Widaman, K. F., & Reise, S. P. (1997). Exploring the measurement invariance of psychological instruments: Applications in the substance use domain. In M. Windle & S. G. West (Eds.), *The science of prevention: Methodological advances from alcohol and substance abuse research* (pp. 281–324). Washington, DC: American Psychological Association.

Widaman, K. F., & Thompson, J. S. (2003). On specifying the null model for incremental fit indices in structural equation modeling. *Psychological Methods*, *8*(1), 16–37. http://doi.org/10.1037/1082-989X.8.1.16

Wothke, W. (2000). Longitudinal and multi-group modelling with missing data. In T. D. Little, K. U. Schnabel, & J. Baumert (Eds.), *Modeling longitudinal and multilevel data: Practical issues, applied approaches, and specific examples* (pp. 219–240). Hillsdale, NJ: Erlbaum.

Wu, W., West, S. G., & Hughes, J. N. (2008). Effect of retention in first grade on children`s achievement trajectories over 4 years: A piecewise growth analysis using propensity score matching. *Journal of Educational Psychology*, *100*(4), 727– 740. http://doi.org/10.1037/a0013098

Wu, W., West, S. G., & Hughes, J. N. (2010). Effect of Grade Retention in First Grade on Psychosocial Outcomes. *Journal of Educational Psychology*, *102*(1), 135–152. http://doi.org/10.1037/a0016664

Zelinski, E. M., & Lewis, K. L. (2003). Adult age differences in multiple cognitive functions: Differentiation, dedifferentiation, or process-specific change?. *Psychology and Aging*, *18*(4), 727–745. http://doi.org/10.1037/0882-7974.18.4.727