

MASTERARBEIT | MASTER'S THESIS

Titel | Title

Cross-temporale Stabilität von psychometrischem g: Eine
longitudinale Untersuchung von kognitiver Testleistung

verfasst von | submitted by
Jana Mirella Wurzer BSc

angestrebter akademischer Grad | in partial fulfilment of the requirements for the degree of
Master of Science (MSc)

Wien | Vienna, 2024

Studienkennzahl lt. Studienblatt | Degree
programme code as it appears on the
student record sheet:

UA 066 840

Studienrichtung lt. Studienblatt | Degree
programme as it appears on the student
record sheet:

Masterstudium Psychologie

Betreut von | Supervisor:

Ass.-Prof. Mag. Dr. Jakob Pietschnig Privatdoz.

Danksagung

Vorab möchte ich Univ.-Prof. Mag. Dr. Jakob Pietschnig für die Betreuung dieser Arbeit und die Vermittlung an die fliegerpsychologische Ambulanz des Heeresspitals in Stammersdorf danken. Ebenfalls Mag.a. Dr.in Elisabeth Zeilinger, durch die diese Zusammenarbeit der Masterarbeit entstanden ist, danke Lisi.

Ein weiterer Dank gilt der fliegerpsychologischen Ambulanz und Hofrat Mag. Michael Mikas und Mag.a. Bettina Wieland für die Zusammenarbeit und Bereitstellung der Daten.

Der wohl größte Dank geht an meine Familie. Danke Mama für jedes Korrektur lesen im Studium, deine endlose Geduld bei meiner Kommasetzung und jeder Präsentation, die du sicherlich hunderte Mal vor allen anderen Leuten gehört hast, ohne dich jemals zu beschweren. Danke für deine Empathie bei kleinen und größeren Krisen und deine Erfahrung.

Danke Papa, dass du mich damals 8h lang nach Wien, und in väterlicher Fürsorge viel zu früh zum Aufnahmetest gefahren hast. Für jeden Umzug, bei dem du geholfen hast. Danke, für deinen rationalen Rat und dass du jeden noch so kleinen Erfolg als mein größter Fan mit mir und Schaumwein gefeiert hast.

Danke Nicolas für deine Gedanken und Anmerkungen zu all meinen Fragen, deinen Mut und dein Vertrauen. Für jede Diskussion und für deine Illustrationen, die meine Präsentationen gerettet haben. Ich hab dich lieb! Danke Thea, für alles! Ohne deine Unterstützung hätte ich viel früher aufgegeben. Schauen wir mal, was wird. Danke Zoe und Lotta für all euren aufbauenden Worte und eure Liebe. Danke Tim.

Danke für euren Mut und eure Unterstützung!

Einleitung

Der Flynn Effekt

Intelligenz hat in unserer Gesellschaft einen hohen Stellenwert. Als allgemeine geistige Fähigkeit umfasst sie neben abstraktem und schlussfolgerndem Denken, Planen, Problemlösen und dem Verstehen komplexer Ideen, das Lernen aus Erfahrungen (Gottfredson, 1997, S. 13), um Herausforderungen in einer sich ständig verändernden Welt zu bewältigen. Die Bedeutung der Intelligenz zeigt sich besonders in einem positiven Zusammenhang mit einer Vielzahl von wünschenswerten individuellen Lebensergebnissen, wie beispielsweise dem beruflichen und akademischen Erfolg (Kuncel & Hezlett, 2010; Lang et al., 2020), dem Einkommen (Zagorsky, 2007), oder der Gesundheit und Lebenserwartung (Deary, 2009) von Personen. Auch auf gesellschaftlicher Ebene zeigt der nationale Intelligenzquotient (IQ) positive Korrelationen mit dem Wirtschaftswachstum und Wohlstand, sowie dem Grad der Demokratie eines Landes (Lynn & Vanhanen, 2012). Es ist daher nicht verwunderlich, dass Ergebnisse (inter-)nationaler Vergleiche kognitiver Leistungen, wie beispielsweise in Form der *PISA*-Studie, einen Einfluss auf politische Entscheidungsprozesse haben (Pietschnig et al., 2018).

Vor dem Hintergrund dieser Bedeutung, wird die beobachtete Steigerung durchschnittlicher Leistungen in Intelligenztests im Verlauf des 20. Jahrhunderts, wie sie weltweit in zahlreichen Studien beobachtet wurde, als eines der „bemerkenswertesten Phänomene“ der Intelligenzforschung beschrieben (Neisser et al., 1996, S. 96).

James R. Flynn bemerkte anhand der Stanford-Binet und Wechsler-Intelligenztests, dass die durchschnittlichen IQ-Werte der Amerikaner:innen zwischen 1932 und 1978 um 0.33 Punkte pro Jahr, über den gesamten Zeitraum also insgesamt 13.8 IQ-Punkte, gestiegen waren (Flynn, 1984). Herrnstein und Murray benennen das Phänomen der IQ-Veränderungen über Generationen hinweg in ihrem umstrittenen Buch *The Bell-Curve* (Herrnstein & Murray, 1994, S. 307) erstmals *Flynn Effekt*.

Der Flynn Effekt beschreibt, unter der Annahme eines gleichbleibend starken Anstiegs, dass eine Person in einer Kohorte einen höheren Testwert erzielt, wenn ihr Rohwert mit der Norm einer früheren Kohorte, als mit der eigenen verglichen wird. Mit anderen Worten, überschätzt ein Test die Leistung einer Person umso mehr, desto mehr Zeit zwischen der Normierung und der Anwendung des Testinstruments liegt. Nimmt man an, dass die IQ-Zuwächse mit 0.3 IQ-Punkten pro Jahr zunehmen, dann würde das über 50 Jahre gesehen, 15 IQ-Punkte ausmachen, was einer Standardabweichung des Intelligenzquotienten entspricht.

Eine Stichprobe, welche 1950 vollkommen den Leistungen der Amerikaner:innen entsprach, wäre im Jahr 2000 dementsprechend nicht mehr repräsentativ (Flynn, 2006).

Die Bedeutung einer korrekten Normierung eines Tests zeigt sich besonders deutlich anhand der Todesstrafe in den USA (Flynn, 2006). In einigen US-Bundesstaaten wird die Todesstrafe nicht vollstreckt, wenn bei der angeklagten Person eine intellektuelle Beeinträchtigung vorliegt, die durch einen Intelligenzquotienten (IQ) von zumindest zwei Standardabweichungen unterhalb des Bevölkerungsdurchschnitts, also einem IQ-Wert von 70 oder weniger, definiert wird. Aufgrund des Flynn Effekts, schneidet eine Person je nach dem Jahr, in welchem der Test normiert wurde, unterschiedlich gut ab. Eine veraltete Norm maskiert dann die „wahre“ Leistung der Person und überschätzt diese unter Umständen, was zu einem besseren Abschneiden führt und gegebenenfalls das Todesurteil für die Person bedeuten kann (Flynn, 2006).

Aufgrund des breiten Einsatzes von psychometrischen Tests in psychologischen Feldern, wie der Pädagogik, Entwicklungspsychologie, Diagnostik und klinischen Anwendungen, aber auch darüber hinaus im akademischen Kontext, bei Hochschulzulassungen und Aufnahmetests, der Musterung im militärischen Kontext, beruflicher Eignungsdiagnostik und Personalauswahl, bis hin zur Fahrtauglichkeit einer Person, ist eine genaue Normierung des Testinstruments unabdingbar. Die Entscheidung bezüglich der Unterstützung von Lernschwierigkeiten oder sonderpädagogischer Betreuung wird oftmals ebenfalls auf der Basis des Ergebnisses eines psychometrischen Tests getroffen. So kann eine veraltete Norm auch hier aufgrund des Flynn Effekts die Leistung der Person überschätzen und einen möglichen Förderbedarf durch ein besseres Abschneiden der Person maskieren, was dazu führen kann, dass die Person schlimmstenfalls trotz einer vorliegenden Schwäche oder eines Bedarfs keine Unterstützung bekommt.

Obwohl erst bei Schaie und Strother (1968) die Veränderungen nicht mehr auf eine präzisere Standardisierung der neuen Skalen, Stichprobenfehler oder statistische Artefakte (Merrill, 1938; Tuddenham, 1948) zurückgeführt wurde, sondern tatsächlich als einen systematischen Unterschied zwischen den Kohorten und somit als Kohorteneffekt interpretiert wurde, erschien bereits 1936 in den USA ein erster Bericht, über den Anstieg des Mittelwerts im Minnesota College Aptitude Test von 33.7 im Jahr 1929 auf 42.3 im Jahr 1932 (Runquist, 1936). Im darauffolgenden Jahr wurden bereits zwei weitere Studien veröffentlicht, die über eine IQ-Steigerung berichteten, gefolgt von weiteren acht Studien in den folgenden 35 Jahren zwischen 1938 und 1973 (Lynn, 2013). So bestätigte zum Beispiel Tuddenham (1948) die

Ergebnisse und zeichnete in den Jahren 1917 bis 1943 einen durchschnittlichen Anstieg von 4.4 IQ-Punkten pro Dekade bei amerikanischen Wehrpflichtigen auf. Der durchschnittliche Anstieg der zehn Studien von 3.65 IQ-Punkten pro Dekade, stimmt in etwa mit der später von Flynn berichteten Zunahme von 3.0 IQ-Punkten pro Dekade überein (Flynn, 1984).

Dabei handelt es sich bei den Intelligenz-Anstiegen um ein globales Phänomen, das sich nicht nur in Anstiegen in westlichen Industrienationen, sondern auch außerhalb der USA und Europa in Ländern des globalen Südens, wie Kenia (Daley et al., 2003), Domenica (Meisenberg et al., 2005), Brasilien (Colom et al., 2007), Sudan (Khaleefa, 2009), Saudi-Arabien (Batterjee et al., 2013), Süd-Korea (te Nijenhuis et al., 2012), aber auch in China (Liu & Lynn, 2013) und der Türkei (Kagıtcıbası & Biricik, 2011; Rindermann et al., 2013) zeigt.

Die Ergebnisse einer ersten formalen Meta Analyse mit fast vier Millionen Teilnehmenden aus 31 Ländern stützen die von Flynn (1984) berichteten Zunahmen von 3.0 IQ-Punkten pro Dekade ebenfalls. Die Meta Analyse zeigt über den Zeitraum von über einem Jahrhundert von 1909 bis 2013 einen durchschnittlichen Anstieg von 0.28 IQ-Punkten pro Jahr, also 2.8 IQ-Punkten pro Dekade (Pietschnig & Voracek, 2015). Dabei handelt es sich bei den Zuwächsen nicht um einen linearen Anstieg, sondern kommt im Verlauf des 20. Jahrhunderts immer wieder zu einem Wechsel von stärkeren und schwächeren Zunahmen (Pietschnig & Voracek, 2015).

Obwohl der Flynn Effekt einen globalen Anstieg zeigt, unterscheidet sich die Stärke der Anstiege in verschiedenen Ländern. Bereits bei Flynn's (1987) systematischer Arbeit zeigt sich ein erster Hinweis auf unterschiedlich starke Anstiege zwischen 5 und 25 IQ-Punkten innerhalb einer Generation über 14 Ländern, wie Österreich, Deutschland, Niederlande, Frankreich, Japan, Kanada und dem Vereinigten Königreich (Flynn, 1987).

Aus einer Meta Analyse geht hervor, dass vor allem Österreich, Frankreich, Deutschland, Israel, Japan, Kenia, die Niederlande und Spanien starke Zuwächse über den Zeitraum des letzten Jahrhunderts verzeichnen (Pietschnig & Voracek, 2015). Während es in Australien, Brasilien, Irland, Neuseeland, dem Vereinigten Königreich und Amerika zu schwächeren IQ-Anstiegen über diesen Zeitraum kommt (Pietschnig & Voracek, 2015). In den letzten Jahren zeigen sich zudem zunehmend differenziertere Muster, die neben weiteren Anstiegen insbesondere in skandinavischen Ländern wie Dänemark (Teasdale & Owen, 2008), Finnland (Dutton & Lynn, 2013) und Estland (Korgesaar, 2013) auf eine Stagnation der IQ-Anstiege bis zu einer Umkehr des Flynn Effekts hinweisen (Pietschnig & Voracek, 2015).

Für einen geschlechterspezifischen Unterschied in den IQ-Anstiegen über die Zeit gibt es keine Evidenz (Pietschnig et al., 2011).

Der g-Faktor

Mittlerweile besteht eine Einigkeit darüber, dass es sich bei den beobachteten Intelligenz-Anstiegen um einen Kohorteneffekt handelt und nicht um statistische Fehler oder Zufälligkeiten (Schaie & Strother, 1968). Die Frage nach der Intelligenz-Ebene, auf der es zu den Anstiegen kommt, geht aus dem Vorliegen eines negativen Zusammenhangs zwischen den beobachteten Anstiegen der IQ-Werte (Flynn Effekt) und dem allgemeinen Faktor der Intelligenz (auch psychometrisches g oder g-Faktor genannt) hervor. Dieser Zusammenhang zeigt sich in zahlreichen Studien im Verlauf des letzten Jahrhunderts (Pietschnig & Voracek, 2015; te Nijenhuis & Van der Flier, 2013; Must et al., 2003, Rushton, 1999; nicht bei Colom et al., 2001) und wird als „one of the most striking and consistent results“ (Pietschnig et al., 2023, S. 1) beschrieben. Er deutet darauf hin, dass die Anstiege nicht auf eine Verbesserung im allgemeinen Faktor der Intelligenz zurückzuführen sind, sondern auf eine Verbesserung in den spezifischen kognitiven Fähigkeiten hinweisen (Pietschnig & Voracek, 2015).

Die Idee eines allgemeinen Faktors der Intelligenz wurde noch vor der Veröffentlichung des ersten Intelligenztests durch den Psychologen Alfred Binet und dem Arzt Théodore Simon im Jahr 1905 (*Binet-Simon-Test*, Binet & Simon, 1904) von Charles E. Spearman formuliert. Bereits 1904 postulierte Spearman mit dem *Zwei-Faktoren-Modell* ein Intelligenzmodell, das Intelligenz in einen Faktor der allgemeinen Intelligenz und spezifische kognitive Fähigkeiten unterteilt. Das Modell beruht auf der Grundlage seiner Methode der Faktorenanalyse und nimmt an, dass ein kognitiver Test immer von zwei Faktoren beeinflusst wird (Jensen, 1998).

Der erste Faktor, der Generalfaktor der Intelligenz (*general intelligence*), den Spearman als g-Faktor bezeichnet, beschreibt die latente Fähigkeit, welche allen kognitiven Tests gemeinsam ist. Die Annahme eines solchen Faktors geht dabei auf die Beobachtung zurück, dass die Ergebnisse verschiedener Intelligenztests eine positive Korrelation miteinander aufweisen, also Personen, die in einem Leistungstest eine gute Leistung zeigen, tendenziell auch in weiteren Tests gute Leistungen erzielen. Dieses Phänomen einer positiven Korrelation zwischen mentalen Tests wird *positiver Manifold* genannt (Jensen, 1992). Die mathematische Methode der Faktorenanalyse ermöglicht es, das Ausmaß, in welchem jede Variable mit dem zugrundeliegenden Faktor aller Variablen korreliert ist und damit die g-Ladung der Variablen zu bestimmen (Jensen, 1998).

Der zweite Faktor ist ein testspezifischer Faktor (*s*), der für jeden Test individuell ist (Jensen, 1998). In späteren Analysen der Theorie wird jedoch deutlich, dass anders als von Spearman vorhergesagt, durch *g* allein nicht alle Korrelationen zwischen den verschiedenen Tests erklärt werden können, was schließen lässt, dass es noch weitere Faktoren der Intelligenz außer *g* und *s* geben muss. Die Idee des gemeinsamen *g*-Faktors wird deshalb weiterentwickelt und durch Modelle mit mehreren Faktoren ersetzt (Jensen, 1998). Eines dieser Modelle ist das derzeit führende und weitgehend akzeptierte *Cattell-Horn-Carroll-Intelligenzmodell (CHC)* (Schneider & McGrew, 2018). Es verbindet Spearmans *g*-Faktor mit dem Ansatz einer hierarchischen Anordnung von Faktoren von John B. Carroll (Carroll, 1993), sowie der Unterteilung der Fähigkeiten in *fluide* und *kristalline* Intelligenz von Raymond B. Cattell und John L. Horn (Horn & Cattell, 1966).

Die *kristalline Intelligenz (Gc)* ist dabei die Fähigkeit, bereits aus Erfahrungen erworbenes (Fakten-)Wissen, anzuwenden und zeigt sich beispielsweise im Leseverständnis oder Vokabeltests (Horn & Cattell, 1966). Als Intelligenzdomäne hängt sie stark mit dem *g*-Faktor zusammen (Johnson et al., 2004). Die *fluide Intelligenz (Gf)* wiederum umfasst, als weitgehend angeborene Fähigkeit, Denkprozesse unabhängig von Vorwissen, wie abstraktes Denken, logisches Schlussfolgern oder Problemlösen (Horn & Cattell, 1966). Sie zeigt sich besonders in Matrizentests oder Aufgaben, die räumliches Denken erfordern.

Auf der höchsten Ebene (*Stratum III*) des CHC-Modells befindet sich der allgemeine Faktor der Intelligenz. Der *g*-Faktor erklärt oft einen erheblichen Anteil der Varianz kognitiver Fähigkeiten und macht in umfassenden kognitiven Testbatterien bei repräsentativen Stichproben Erwachsener, oft fast die Hälfte der Varianz aus (Deary, 2012; Schneider & McGrew, 2018). Trotz einer hohen Übereinstimmung zwischen Intelligenztestergebnissen und dem allgemeinen Faktor der Intelligenz, unterscheiden sich der Intelligenzquotient (IQ) und der *g*-Faktor grundlegend voneinander. Während der *g*-Faktor auf der Korrelation zwischen einzelnen kognitiven Aufgaben oder Subtests basiert, ergibt sich der Gesamt-IQ, als Metrik der Intelligenz, durch die Summation der Leistungen verschiedener Intelligenztestaufgaben (Colom et al., 2002). Der Gesamt-IQ kann jedoch als Träger des Konstrukts der allgemeinen Intelligenz (general intelligence, *g*-Faktor) gesehen werden, der neben einem Anteil des *g*-Faktors jedoch auch spezifische Fähigkeiten beinhaltet (Colom et al., 2002).

Die darunterliegende Ebene des Modells (*Stratum II*) besteht aus verschiedenen Gruppen, die in unterschiedlichen Intelligenztests und je nach Autor:in unterschiedlich benannt werden, jedoch in der Regel die Domänen *fluide Intelligenz (Gf)* und *kristalline Intelligenz*

(Gc), sowie weitere Faktoren wie *Wahrnehmung, Gedächtnis, verbale Fähigkeiten* oder *schlussfolgerndes Denken* enthalten. In der letzten Ebene (*Stratum I*) werden für den jeweiligen Test direkt erfassbare und messbare spezifische Fähigkeiten aufgeführt (Deary, 2012; Schneider & McGrew, 2018).

Domänenspezifischer Effekt

Im Verlauf des 20. Jahrhunderts zeigt sich, dass die Zuwächse der IQ-Werte nicht über alle Intelligenzdomänen gleichstark ausfallen (Flynn, 1987). Bereits 1942 wurde ein Unterschied in den Zuwächsen zwischen *verbalen* und *non-verbalen* Fähigkeiten bei Kindern im Alter von 10 bis 14 Jahren auf Hawaii festgestellt (Smith, 1942). Während bei den *verbalen* Fähigkeiten ein altersspezifischer Effekt gezeigt wurde und der Anstieg der *nonverbalen* Fähigkeiten über die Altersgruppen hinweg etwa gleichstark ausfiel, unterschieden sich beide Domänen auch in der Stärke der Zuwächse. Die *nonverbalen* Fähigkeiten wiesen einen deutlich stärkeren Zuwachs von 6.0 IQ-Punkten pro Dekade, im Gegensatz zu dem Anstieg von durchschnittlich 2.6 IQ-Punkten der *verbalen* Fähigkeiten auf (Smith, 1942). Dieser domänenspezifische Unterschied zwischen fluider und kristalliner Intelligenz wird in zahlreichen folgenden Studien bestätigt (Lynn, 2013).

Unterschiedlich starke Zuwächse in einzelnen Intelligenzdomänen gehen auch aus einer Meta Analyse zum Flynn Effekt hervor (Pietschnig & Voracek, 2015). Über verschiedene Studien zeigt sich ein wesentlich stärkerer Anstieg der fluiden Intelligenz, die weniger g-Faktor geladen ist, im Vergleich zur kristallinen Intelligenz, die eine hohe Ladung mit dem g-Faktor zeigt (Vergleich 4.1 vs. 2.1 IQ-Punkte pro Jahrzehnt). Die Ergebnisse lassen einen domänenspezifischen Flynn Effekt vermuten (Pietschnig & Voracek, 2015). Als Ursache für diese Unterschiede in der Stärke der Zuwächse zwischen den Intelligenzdomänen wurde unter anderem die Verwendung obsoleter Items in Intelligenztests untersucht. Da obsolete Items, welche veraltetes Wissen abfragen, vor allem die tatsächliche Leistung der kristallinen Intelligenz maskieren, besteht die Möglichkeit, dass die tatsächlichen domänenspezifischen Unterschiede im Flynn Effekt zwischen kristalliner und fluider Domäne möglicherweise viel kleiner sind, als bisher angenommen wurde (Hörtl, 2019; Pietschnig et al., 2019).

Ein weiterer, später ausführlicher erwähnte Ansatz, ist die Theorie der sozialen Multiplikatoren nach Dickens & Flynn (2010), die die Unterschiede in den Domänen unter anderem damit begründet, dass verschiedene kognitive Fähigkeiten unterschiedlich stark von

sozialen und kulturellen Einflüssen geprägt sind (Dickens & Flynn, 2010). Diese könnten somit maßgeblich die domänenspezifischen Unterschiede beeinflussen.

Eine umfassende Analyse domänenspezifischer Veränderungen gemäß dem *Cattell-Horn-Carroll-Modell (CHC)* wurde in einer Studie an 36 deutschsprachigen weitgehend bevölkerungsrepräsentativen Stichproben durchgeführt (Lazaridis et al., 2022). In Bezug auf die Stratum-II-Domäne zeigten sich unterschiedliche Effektmuster des Flynn Effekts. Positive Veränderungen waren in den Bereichen bereichsspezifisches Wissen (domain-specific knowledge), des Verständniswissens (comprehension-knowledge) und der Lerneffizienz (learning-efficiency) zu verzeichnen, während bei der Verarbeitungsgeschwindigkeit (processing speed), sowie dem Lesen und Schreiben (reading and writing) keine bedeutsamen Veränderungen auftraten. Die Arbeitsgedächtniskapazität (working memory capacity) zeigte einen negativen Effekt. Hinsichtlich fluiden Denkens (fluid reasoning), Reaktions- und Entscheidungsgeschwindigkeit (reaction and decision speed), sowie quantitativen Wissen (quantitative knowledge) und visueller Verarbeitung (visual processing) zeigten sich mehrdeutige Ergebnisse (Lazaridis et al., 2022). Diese Befunde unterschiedlicher Effekte über die verschiedenen Intelligenzdomänen hinweg könnten laut Autor:innen einerseits auf eine bisher unentdeckte Domänenspezifität des Flynn Effekts hindeuten, die über die Unterschiede in den Zuwächsen kristalliner und fluiden Intelligenz hinausgehen, andererseits auch auf eine mögliche Stagnation des Flynn Effekts in einigen Domänen hinweisen, wie sie bereits im skandinavischen Raum berichtet wird (z.B. Dutton & Lynn, 2013; Teasdale & Owen, 2008; Sundet et al., 2004) hinweisen (Lazaridis et al., 2022).

Altersspezifischer Effekt

Unterschiede in den Anstiegen zeigen sich ebenfalls über das Lebensalter hinweg. Dabei fungiert das Alter als eine Moderatorvariable, die sich auf die unterschiedliche Stärke der Anstiege auswirkt (Flynn, 2010). Dabei gehören Wheeler (1942) und Smith (1942) zu den Ersten, die über einen Altersunterschied im Flynn Effekt berichteten (Lynn, 2013). Wheeler (1942) berichtete unterschiedlich starke Zuwächse bei Kindern aus Tennessee im Alter von sechs bis 16 Jahren von 1930 bis 1940. In der Altersgruppe der sechs- bis Neunjährigen stieg der IQ im Durchschnitt um bis zu 10.3 Punkten, während bei den Zehnjährigen ein durchschnittlicher Anstieg von 7.2 IQ-Punkten zu verzeichnen war. Die größte Steigerung wurde bei den 11-jährigen Kindern beobachtet, mit einem durchschnittlichen Anstieg von 13.9 IQ-Punkten, gefolgt von einer geringeren Zunahme von 8.7 IQ-Punkten bei den Zwölfjährigen. Bei den 13- bis 14-jährigen Jugendlichen wurde im Durchschnitt eine Zunahme von 10.2 IQ-

Punkten festgestellt, während 16-jährige Jugendliche im Durchschnitt einen Anstieg von 6.5 IQ-Punkten verzeichneten.

Smith (1942) zeigte eine ähnliche Altersdifferenz zwischen 10- bis 14-jährigen Kindern auf Hawaii, bei denen der Anstieg der *verbalen* Fähigkeiten in der Altersgruppe der Zehnjährigen am größten war (2.4 IQ-Punkte pro Dekade) und über die Altersgruppen bis zu den 14-jährigen (1.6. IQ-Punkte pro Dekade) immer weiter abnahm. Während sich im Anstieg der *non-verbalen* Fähigkeiten über die Altersgruppen von 10 bis 14 Jahren nahezu kein Unterschied zeigte.

Neuere Studien zeigen jedoch insbesondere im Bereich der räumlichen und fluiden Intelligenz höhere Anstiege bei Erwachsenen im Vergleich zu Jugendlichen und Kindern, während sich für die kristalline Intelligenz kein Alterseffekt zeigte (Pietschnig & Voracek, 2015).

Ursachen für die IQ-Anstiege (Flynn Effekt)

In der Forschung zum Flynn Effekt haben zahlreiche Wissenschaftler:innen Theorien zu potenziellen Ursachen für die bemerkenswerten Anstiege der IQ-Werte aufgestellt. Die Erklärungen sind vielfältig und reichen dabei von umweltbezogenen, über biologischen Faktoren bis hin zu hybriden, integrierenden Modellen. Aufgrund der engen Verbindung des Effekts mit historischen Ereignissen (Pietschnig & Voracek, 2015) und der besonderen Stärke der Zuwächse, werden vor allem umweltbedingte Einflüsse als Erklärung in Betracht gezogen (te Nijenhuis & van der Flier, 2013).

Einige Theorien, wie beispielsweise die einer abnehmenden Familiengröße (Zajonc & Mullally, 1997), des technologischen Fortschritts (Neisser, 1997), Hybrid vigor (Mingorini, 2007) oder Veränderungen im Testverhalten aufgrund vermehrter Testerfahrung (Brand, 1987) und die genomischen Prägung (Storfer, 1999), weisen dabei eine geringe empirische Fundierung oder geringe Effekte auf. Im Folgenden werden die derzeit führenden Erklärungsansätze des Flynn Effekts und ihre empirische Evidenz vorgestellt.

Bildung

Ein auf den ersten Blick naheliegender Umweltfaktor ist die Bildung. Ein erster Zusammenhang zeigt sich bereits in den ersten Studien zum Flynn Effekt, in denen über einen Einfluss der schulischen Bildung auf die Entwicklung der Intelligenz berichtet wird (Tuddenham, 1948).

Auch Flynn erwähnt die Bedeutung der Bildung (1984, 2007) als maßgeblichen Faktor für die IQ-Anstiege, was von vielen Wissenschaftspersonen befürwortet wird (Teasdale & Owen, 1994, 2005; Jensen, 1998; Meisenberg et al., 2006).

Der Bildungseinfluss (Ceci & Williams, 1997) zeigt sich beispielsweise durch die negative Korrelation zwischen IQ und Alter der Kinder, welche gar nicht oder nur unregelmäßig die Schule besuchten. Die fehlende Bildung führte zu einer Abnahme der kognitiven Fähigkeiten und des IQs der Kinder mit steigendem Alter. So zeigt sich bei dem Vergleich der Leistungen mit Kindern, die einen regelmäßigen Schulbesuch hatten, eine Abnahme von etwa 5 IQ-Punkten pro Jahr ohne Schulbildung (Ceci & Williams, 1997).

Die Erklärung der Anstiege durch eine Zunahme der Bildung stimmt jedoch nicht mit der Beobachtung eines domänenspezifischen Unterschieds in den Zunahmen zwischen fluider und kristalliner Intelligenz (Pietschnig & Voracek, 2015) überein. Demnach müssten in der kristallinen Intelligenz, die sich auf wissensbasierte Fragen bezieht, stärkere Zuwächse verzeichnet werden. In der kristallinen Intelligenz zeigen sich zwar erhebliche Zuwächse, im Vergleich zu den Zuwächsen der fluiden Intelligenz, die kein Vorwissen voraussetzt, fallen diese jedoch geringer aus (Pietschnig & Voracek, 2015).

Eine Untersuchung im deutschsprachigen Raum zeigt in einer zeitübergreifenden Analyse keine überzeugenden Hinweise für einen Flynn Effekt in der kognitiven Entwicklung der drei bis sechsjährigen, was die Theorie der Bildung als wichtiger Faktor für die IQ-Anstiege unterstützt (Pietschnig et al., 2021). Wenn Bildung jedoch der einzige Faktor wäre, müsste die statistische Kontrolle der Bildung demnach einen Einfluss auf die beobachteten Zuwächse haben. Eine Studie zeigt jedoch, dass die Zuwächse auch nach der Kontrolle für den höchsten Bildungsabschluss weiterhin erhalten bleiben (Pietschnig et al., 2013). Demnach scheint Bildung nicht der einzige erklärende Faktor zu sein.

Ernährung

Ein hybrider Faktor, der ebenfalls als Erklärung in Betracht gezogen wird, ist die Ernährung. Die Ernährungssituation hat sich über das 20. Jahrhundert, dem Zeitraum der beobachteten IQ-Anstiege, in entwickelten Ländern erheblich verbessert, was unter anderem in einer Zunahme durchschnittlicher Körpergröße und des Kopfumfangs gut ernährter Personen resultierte (Lynn, 2009). Der gefundene Zusammenhang zwischen schlechter Ernährung und niedrigeren IQ-Werten führt zu der Annahme, dass die höheren Testleistungen mit einer verbesserten

Ernährung erklärt werden könnten (Hunt, 2011). Eine verbesserte prä- und postnatale Ernährung könnte demnach durch die daraus resultierende bessere Entwicklung des Zentralnervensystems die kognitive Entwicklung fördern und damit zu einem Anstieg der Intelligenz über das letzte Jahrhundert beigetragen haben (Lynn, 2009; Pietschnig & Voracek, 2015).

In den Intelligenzanstiegen zeigt sich jedoch ein Alterseffekt, der darauf hindeutet, dass es zu stärkeren Anstiegen durchschnittlicher Leistung mit steigendem Alter kommt (Pietschnig & Voracek, 2015), demzufolge Ernährung ebenfalls nicht als alleinige Ursache für den Flynn Effekt verantwortlich sein kann (Pietschnig & Voracek, 2015; Lynn, 2009).

Pathogener Stress

Ein weiterer hybrider Faktor stellt die zunehmende medizinische Versorgung und geringere Belastung durch Umweltgifte dar, welche einen positiven Einfluss auf die Gehirnentwicklung und dadurch auch auf die Intelligenz zu haben scheinen. Infektionskrankheiten gelten als signifikanter negativer Prädiktor für den durchschnittlichen IQ (Eppig et al., 2010; Lynn & Vanhanen, 2012). Eine Abnahme der Infektionskrankheiten durch eine verbesserte medizinische Versorgung, die ebenso wie die verbesserte Ernährung mit der Entwicklung vieler Länder über das 20. Jahrhundert einher ging, kann somit zu den Anstiegen der IQ-Werte beitragen haben. Auch das Wegfallen weiterer pathogener Stressoren, wie der erhebliche Rückgang der nachgewiesenen Bleikonzentration im Blut der Amerikaner:innen, durch das Verbot von Bleifarben und bleihaltigem Benzin seit den 1970er Jahren, gilt als bedeutender Prädiktor für den Anstieg von 4 bis 5 IQ-Punkten (Kaufmann et al., 2014; Nevin, 2000).

Soziale Multiplikatoren

Die hybride Theorie der sozialen Multiplikatoren (Dickens & Flynn, 2001) kombiniert die Ansätze umweltbedingter und genetischer Faktoren. Hierbei wird von einem, sich gegenseitig beeinflussenden Zusammenhang der Faktoren ausgegangen. Bereits etwas über dem Durchschnitt liegende intellektuelle Fähigkeiten führen demnach zu einer verbesserten Umweltbedingung der Person, indem beispielsweise Begabungen der Person durch Unterricht oder eine verbesserte (Schul-)Bildung gefördert werden. Dies wiederum führt zu einer Verbesserung der individuellen Leistung, was wiederum zu einer (noch) vorteilhafteren Umwelt beitragen kann (Dickens & Flynn, 2001).

Obwohl diese Theorie gut zu den Beobachtungen der IQ-Zuwächse passt, die unabhängig vom psychometrischen g sind, sowie den starken Zuwächsen in der fluiden Intelligenz (Pietschnig & Voracek, 2015), legt die empirische Untersuchung des Modells, sowie das Fehlen ausreichender empirischer Daten nahe, dass die Theorie zwar einen gewissen Teil der beobachteten Anstiege erklären kann, aber nicht alles. Auch die Stärke der Gen-Umwelt-Interaktion für derartige Anstiege in einem so kurzen Zeitraum wird hinterfragt (Loehlin, 2002).

Life history Model

Die Theorie der Geschwindigkeit des Lebens (Woodley, 2012) versucht dabei viele der oben genannten Intelligenz-fördernden Faktoren, wie geringerer pathogener Stress (Eppig et al., 2010), oder einer Verbesserung von Bildung (Ceci & Williams, 1997) und Ernährung (Lynn, 2009), welche Teile der Intelligenz-Anstiege erklären können, in einem gemeinsamen Modell zu verbinden.

Der Theorie nach unterscheiden sich Individuen in ihrer Geschwindigkeit des Lebenslaufs. Personen mit einer langsameren Lebensgeschwindigkeit (Slow Life History Speed) wiederum zeichnen sich durch weniger Sexualpartner:innen, eine geringe Anzahl an Nachkommen und eine spätere Elternschaft aus. Sie befinden sich meistens in einer sichereren, vorhersagbareren Umwelt, in der geringer pathogener Stress und weniger Risikofaktoren wie mangelnde Ernährung als Stressoren wirken, wodurch ihnen mehr Ressourcen für die Differenzierung kognitiver Fähigkeiten zu Verfügung stehen (Woodley, 2012). Während Individuen mit einer schnellen Lebensgeschwindigkeit (Life History Speed) in einer Umwelt mit vielen Risikofaktoren für eine hohe Sterblichkeit gut angepasst sind und dadurch einen Vorteil im Überleben haben, stehen ihnen gleichzeitig weniger Ressourcen für die Entwicklung und Differenzierung kognitiver Fähigkeiten zur Verfügung (Woodley, 2012). Somit trägt jeder Faktor, der das Risiko einer hohen Sterblichkeit verringert (wie eine verbesserte Ernährung oder weniger pathogener Stress) und damit eine langsamere Lebensgeschwindigkeit fördert, zu einer höheren Wahrscheinlichkeit der Differenzierung und Entwicklung der kognitiven Fähigkeiten bei (Woodley, 2012).

Insgesamt kann gesagt werden, dass keine der Theorien der Ursachen des Flynn Effekts allein genug Erklärungskraft hat. Die momentan wahrscheinlichste Erklärung besteht aus einem Zusammenwirken der verschiedenen Ursachen.

Aktuelle Entwicklung des Flynn Effekts

Die Zunahme der IQ-Werte folgt wie bereits erwähnt, keinem durchgehend linearen Trend, sondern zeigt über das letzte Jahrhundert hinweg einen nicht-linearen Verlauf, in dem sich die Stärke der Zuwächse über die Zeit verändert (Pietschnig & Voracek, 2015).

Ein linearer Anstieg der durchschnittlichen Testwerteergebnisse würde bedeuten, dass jede nachfolgende Generation die vorherige in Bezug auf ihre Leistungsfähigkeit übertreffen würde. Damit einhergehend würden Kohorten-Unterschiede deutlicher werden, je größer der zeitliche Abstand zwischen ihnen ist (Lazaridis et al., 2022). Nach einem anfänglichen geringen Anstieg von 0.8 IQ-Punkten pro Dekade im frühen 20. Jahrhundert, kommt es in den 1920er Jahren jedoch zu einem bemerkenswert starken Anstieg von 7.2 IQ-Punkten pro Jahrzehnt und einem anschließenden Rückgang auf 2.1 IQ-Punkten pro Jahrzehnt von 1935 bis 1947. Anschließend steigt der Anstieg wieder auf 3 IQ-Punkte pro Dekade bis zum Jahr 1974 an (Pietschnig & Voracek, 2015).

Über weite Teile des 20. Jahrhunderts schien der Flynn Effekt größtenteils positiv zu verlaufen. Neuere Studien deuten jedoch auf eine zunehmend vielschichtige Entwicklung und differenziertere Muster des Flynn Effekts hin, die sich sowohl über die verschiedenen Länder als auch Intelligenzbereiche zeigen (Pietschnig & Voracek, 2015, Pietschnig & Gittler, 2015; Lazaridis et al., 2022).

Länderspezifischer Effekt

Während einige Länder weiterhin anhaltende Zuwächse verzeichnen (Flynn, 2012; Daley et al., 2003; Meisenberg et al., 2005), zeigen Daten aus anderen Ländern bereits robuste Anzeichen für schwächer werdende Anstiege (Pietschnig & Voracek, 2015). Insbesondere in skandinavischen Ländern wie Norwegen (Sundet et al., 2004), Dänemark (Teasdale & Owen, 2008), Finnland (Dutton & Lynn, 2013), Estland (Korgesaar, 2013), sowie den Niederlanden (Woodley & Meisenberg, 2013) und dem Vereinigten Königreich (Shayer & Ginsburg, 2007; 2009) deuten Berichte auf einen negativen Trend und eine Umkehr des Flynn Effekts hin (vgl. Dutton et al., 2016, für einen umfassenden Überblick). Ähnliche Ergebnisse wurden auch in Frankreich berichtet. Eine neuere Studie zeigt jedoch, dass die berichtete Abnahme von 3.8 IQ-Punkten in Frankreich (Dutton & Lynn, 2015) durch die unterschiedliche Funktionsweise der Items im Laufe der Zeit verzerrt wird und der suggerierte negative Flynn Effekt auf statistische Artefakte zurückzuführen ist (Gonthier et al., 2021).

Es ist anzumerken, dass die Evidenz für eine Umkehr des Flynn Effekts nicht eindeutig ist, da es in einigen Ländern und teilweise auch in den gleichen Ländern, aber unterschiedlichen Domänen, durchaus zu weiteren Zuwächse kommt.

In Norwegen und Schweden wurde bereits früh über eine Stagnation von Zuwächsen berichtet (Emanuelsson et al, 1993; Sundet et al., 2004), die auf einen Höhepunkt der Anstiege in den 1990er Jahren hindeuten. Bei männlichen norwegischen Wehrpflichtigen kommt es zwischen Mitte der 1990er Jahren und dem Jahr 2000 zu einem Ausbleiben der IQ-Anstiege (Sundet et al., 2004). Ähnliches zeigt sich auch in dänischen Wehrpflichtigen zwischen 1959 und 2004, bei denen sich ebenfalls ein Höhepunkt der Anstiege in den 1990er Jahren verzeichnet und anschließend zu einem geringen Rückgang auf das Level vor dem Jahr 1991 kommt (Teasdale & Owen, 2005).

Domänenspezifischer Effekt

Daran angeschlossen zeigt auch eine Meta Analyse im deutschsprachigen Raum für räumliche Wahrnehmung zwischen 1977 und 2014 einen umgekehrt u-förmigen Verlauf und nach einem anfänglichen Anstieg der Werte einen Höhepunkt in den 1990er Jahren und einen anschließenden Rückgang (Pietschnig & Gittler, 2015).

Wie bereits im domänenspezifischen Effekt erwähnt, zeigt eine weitere Studie aus dem deutschsprachigen Raum unterschiedliche Effektmuster über die verschiedenen Intelligenz-Domänen, die je nach Domäne sowohl positive (bereichsspezifisches Wissen, Verständniswissen und Lerneffizienz), stagnierende (Verarbeitungsgeschwindigkeit, Lesen und Schreiben) als auch negativen (Arbeitsgedächtniskapazität) und mehrdeutigen Ergebnissen (fluidem Denken, Reaktions- und Entscheidungsgeschwindigkeit sowie quantitatives Wissen und visuelle Verarbeitung) Flynn Effekte zeigen (Lazaridis et al., 2022). Neben einem Hinweis auf unentdeckte Domänenspezifität, die über die Unterschiede in kristalliner und fluider Intelligenz hinausgehen, können die Ergebnisse auch auf eine mögliche Stagnation des Flynn Effekts hinweisen (Lazaridis et al., 2022).

Erklärungsansätze gegenwärtig negativer Entwicklungen

Vor dem Hintergrund der zunehmenden Berichte über stagnierende und rückläufige IQ-Testergebnisse aus verschiedenen Ländern (z.B. Dutton & Lynn, 2013, 2015; Sundet et al., 2004; Teasdale & Owen, 2005; 2008; Shayer & Ginsburg, 2007; 2009), gewinnt das Phänomen eines gegenwärtig negativen Flynn Effekts, insbesondere aufgrund der weitreichenden

Auswirkungen, zunehmend an Interesse. Dieser Entwicklung widmen sich bereits erste Erklärungstheorien, die versuchen den Rückgang im Kontext der früheren Zuwächse und Unterschieden zwischen den Ländern zu erklären.

Geschlechterverhältnis

Aufgrund der Tatsache, dass viele Studien zur Untersuchung des Flynn Effekts auf Daten von Wehrpflichtigen basieren (z. B. Tuddenham, 1948, Teasdale & Owen, 2005), und auch die ersten Gruppenintelligenztests, wie der *Army*-Alpha und *Army*-Beta Test der ökonomischen Gruppentestung US-amerikanischer Rekruten dienten, beruht ein Großteil der bisherigen Daten zum Flynn Effekt auf vorwiegend männlichen Stichproben.

Demnach wird vermutet, dass Leistungsunterschiede zwischen den Geschlechtern durch Veränderungen im Geschlechterverhältnis möglicherweise Veränderungen in den Intelligenz-Anstiegen zur Folge haben könnten (Dutton et al., 2016). Doch trotz geschlechterspezifischer Unterschiede, wie einem Vorteil von Männern in der allgemeinen visuellen Verarbeitung oder weiblichen Vorteilen bei der allgemeinen Verarbeitungsgeschwindigkeit, zeigen sich keine Unterschiede in der allgemeine Intelligenz zwischen den Geschlechtern (Reynolds et al., 2022).

Darüber hinaus besteht sowohl die dänische (Teasdale & Owen, 2008), als auch die finnische Stichprobe (Dutton & Lynn, 2013), an welchen negative Flynn Effekte gezeigt wurden, nur aus Männern (Dutton et al., 2016).

Auch bei deutschsprachigen Testpersonen zeigte sich kein Einfluss des Geschlechts auf die generationsübergreifenden Veränderungen im Flynn Effekt (Pietschnig et al., 2011). Das Geschlechterverhältnis der Stichproben ist somit wahrscheinlich kein entscheidender Faktor für die Veränderungen in den IQ-Werten über die Zeit (Dutton et al., 2016).

Migration

Angeichts der bisherigen Anstiege der IQ-Werte und der verstärkten Migrationsbewegung der letzten Jahrzehnte besonders im europäischen Raum, wurde für den beobachteten Rückgang der Intelligenzanstiege Migrationseffekte in Betracht gezogen. Diese These basiert auf der Beobachtung einiger Studien, dass der durchschnittliche IQ nicht-westlicher ethnischer Minderheiten deutlich niedriger als der europäische Durchschnitt ist (Lynn, 2006). Eine Studie an der dänischen Bevölkerung zeigt außerdem, dass der Unterschied zwischen Zugewanderten und der dänischen Bevölkerung fast 14 IQ-Punkte beträgt (Kirkegaard, 2013). Auf dieser Basis

wird die Theorie aufgestellt, dass es durch die Immigration von Menschen aus Ländern mit niedrigerem durchschnittlichen IQ in Länder, deren Durchschnitts-IQ höher ist, zu einem Rückgang des durchschnittlichen IQs der Einreiseländer kommt und die negativen Veränderungen im Flynn Effekt auf Migrationseffekte zurückzuführen sind.

Aus wissenschaftlicher Sicht erweist sich diese Begründung für eine Abnahme des Flynn Effekts jedoch aus mehreren Gründen als nicht schlüssig. Zum einen sind die Auswirkungen der Migration auf den nationalen IQ des Einreiselandes nicht von Dauer, da Leistungsunterschiede zwischen den Bevölkerungsgruppen über die Zeit abnehmen (te Nijenhuis et al., 2004; Pietschnig et al., 2018). Zum anderen ist der Einfluss der Migration in einigen Stichproben zu gering, um einen maßgeblichen Beitrag für die Abnahme des Durchschnitt-IQs der Einreiseländer zu liefern (Dutton et al., 2016; Dutton & Lynn, 2013). So können in einigen Stichprobe in Finnland (Dutton & Lynn, 2013) die Rückgänge der IQ-Werte beispielsweise nicht vollständig durch die Zuwanderung erklärt werden, da sich für einen signifikanten Einfluss auf den Flynn Effekt die Anzahl der eingewanderten Personen, als zu gering erweist (Dutton et al., 2016). In neueren Studien zeigen sich für den Einfluss von Migration auf den Flynn Effekt ebenfalls keine Evidenz (Pietschnig et al., 2018).

Dysgenische Prozesse

Um die Abnahme der durchschnittlichen IQ-Werte zu erklären, werden auch genetische Faktoren als mögliche Ursache in Betracht gezogen. Die Theorie der dysgenischen Prozesse beruht auf der Beobachtung eines negativen Zusammenhangs zwischen Intelligenz und Fruchtbarkeit in westlichen Ländern (z.B. Lynn, 2011), der bereits bei Sir Francis Galton (1869) postuliert wurde. Die Annahme ist dabei, dass die Vererbung ungünstiger oder nachteiliger Eigenschaften oder Merkmalen zu einer Verringerung der kognitiven Fähigkeiten führt, die gepaart mit der höheren Fruchtbarkeit weniger intelligenter Eltern zur Abnahme der Intelligenz der Population beiträgt. Demnach wäre eigentlich ein deutlicher Rückgang der Intelligenzwerte seit Beginn des 20. Jahrhunderts zu erwarten gewesen. Doch entgegen der Annahme zeigen sich deutliche Anstiege der IQ-Werte über das letzte Jahrhundert (z.B. Pietschnig & Voracek, 2015).

Eine mögliche Erklärung dafür könnte die Maskierung der dysgenischen Prozesse der Intelligenz durch den starken Anstieg der Umweltqualität und damit einhergehende Erhöhung des phänotypischen IQs darstellen. Erst mit Erreichen einer gewissen Grenze in der

Verbesserung der Umweltqualität, durch beispielsweise eine bessere Bildung und Ernährung, könnte der Rückgang der genetischen Intelligenz sichtbar werden (Dutton & Lynn, 2015).

Eine empirische Untersuchung der dysgenischen Prozesse zeigt uneindeutige Ergebnisse mit kleinen eher unbedeutenden Effekten (Pietschnig et al., 2018). Lediglich bei der kristallinen Intelligenz zeigt sich ein trivialer Effekt eines negativen Zusammenhangs der Fertilität mit den durchschnittlichen Intelligenzveränderungen, während sich bei der fluiden Intelligenz ein positiver Zusammenhang der Anzahl dokumentierter Geburten und der Stärke der Zuwächse zeigt. Schlussfolgernd kann auf Grundlage der Ergebnisse nicht von einem bedeutenden Effekt der Fruchtbarkeit als Erklärung für die Veränderungen des durchschnittlichen IQ ausgegangen werden (Pietschnig et al., 2018).

Mortalität

In Anlehnung an dysgenische Prozesse wird die verringerte Sterblichkeit als weitere Theorie für die Umkehr des Flynn Effekts postuliert. Demzufolge ist die Abnahme der durchschnittlichen IQ-Werte unter anderem auf eine verringerte Kindersterblichkeit zurückzuführen, die durch eine verbesserte medizinische Versorgung der modernen Medizin bedingt ist (Nyborg, 2012). Unter dem vorliegenden Zusammenhang von physischer und psychischer Fitness, führt die verringerte Sterblichkeit folglich dazu, dass mehr Menschen mit geringerer physischer und damit auch psychischer Fitness das fortpflanzungsfähig Alter erreichen und ihre Gene an die nächste Generation weitergeben können. Langfristig könnte dies zu einer Abnahme der durchschnittlichen IQ-Werte in der Bevölkerung führen.

Gemeinsam mit einer hohen Fruchtbarkeit nicht-westlicher Immigrant:innen wird angenommen, dass es durch diese beiden Faktoren sukzessiv zu einer Abnahme der durchschnittlichen Intelligenz kommt (Nyborg, 2012). Eine systematische Übersichtsarbeit zeigt für Einflüsse von Sterblichkeit auf die abnehmenden durchschnittlichen IQ-Werte ebenfalls nur wenig Evidenz. Eine höhere nationale Mortalitätsrate bei Kleinkindern, die der Theorie zufolge mit höheren IQ-Zuwächsen einhergehen sollte, zeigt sich nicht (Pietschnig et al., 2018).

Die Verbreitung von Theorien, wie die der Migration, Fertilität und Mortalität als mögliche Erklärungsansätze für abnehmende IQ-Werte der Gesellschaft, erfordert insbesondere aufgrund ihrer gesellschaftspolitischen Auswirkungen und ihres Potentials, Vorurteile und

Rassismus gegenüber Zuwander:innen zu bestärken, solide empirische Evidenz, welche gegenwärtig für dieses Phänomen nicht existieren.

Weitere Überlegungen, wie ein erhöhtes Alter der Mutter, wurden ebenfalls in Betracht gezogen. Dabei zeigt sich jedoch ein negativer Zusammenhang zwischen der Abnahme der IQ-Anstiege und dem Alter der Mutter, wodurch diese folglich als Ursache unplausibel erscheint (Dutton et al., 2016).

Alternative Erklärungen

Es ist festzuhalten, dass der Rückgang der IQ-Anstiege verschiedene Ursachen haben kann. Eine Annahme ist, dass die Verbesserungen in spezifischen kognitiven Fähigkeiten bereits eine Grenze erreicht haben und sich Deckeneffekte zeigen, da weitere Differenzierungen und Zuwächse nicht mehr möglich sind (Pietschnig & Voracek, 2015; Pietschnig et al., 2018).

Aufgrund des großen Einflusses von umweltbedingten Faktoren auf den Flynn Effekt, ist es ebenso wahrscheinlich, dass es mit der Zeit zu abnehmenden Erträgen von IQ-steigernden Faktoren, wie beispielsweise der Bildung (Ceci & Williams, 1997), der Ernährung (Lynn, 2009), oder geringerem pathogenen Stress (Eppig et al., 2010) kommt. Weitere Verbesserungen haben dann nicht mehr den gleichen Effekt auf die Steigerung, wie anfänglich, was eine Stagnation des Effekts erklären würde.

Die Annahme geht mit der Beobachtung eines systematischen Literatur Reviews einher, in welchem die Länder, in denen sich bereits ein negativer Flynn Effekt zeigt, wie Finnland (Dutton & Lynn, 2013), Niederlande (Woodley & Meisenberg, 2013), Dänemark (Teasdale & Owen, 2008), Großbritannien (Shayer & Ginsburg, 2007; 2009) oder Norwegen (Sundet et al., 2004) zu den wohlhabendsten Nationen gehören in denen Faktoren wie Bildung (Ceci & Williams, 1997), Ernährung (Lynn, 2009), oder pathogenen Stress (Eppig et al., 2010) bereits eine Obergrenze erreicht haben könnten und sich in einer Stagnation der Zuwächse bemerkbar machen könnte (Dutton et al., 2016).

Die negative Verbindung von g-Faktor und IQ-Zuwächsen

Eine weitere Hypothese zur Erklärung der aktuellen Veränderungen in den durchschnittlichen Intelligenztestwerten basiert auf dem bereits erwähnten konsistenten Befunden eines negativen Zusammenhangs der beobachteten Anstiege durchschnittlicher IQ-Werte mit g (z.B. Pietschnig & Voracek, 2015, te Nijenhuis & Van der Flier, 2013; Rushton, 1999). Diese negative

Korrelation bedeutet, dass stärkere Anstiege der IQ-Werte mit einer Abnahme des psychometrischen g einhergehen. Demnach ist die Zunahme der durchschnittlichen IQ-Werte nicht auf die Zunahme der allgemeinen Intelligenz (g) zurückzuführen, sondern weist vielmehr auf eine Verbesserung der spezifischen kognitiven Fähigkeiten hin (Pietschnig & Voracek, 2015). Dieser Zusammenhang klingt auf den ersten Blick widersprüchlich, kann aber gut anhand des Sportwettbewerbs eines Zehn-Kampfs veranschaulicht werden.

Allen zehn Leistungen des Wettbewerbs liegt faktorenanalytisch ein gemeinsamer Faktor zugrunde, den man, als die allgemeine oder Leichtathletik-Fitness bezeichnen könnte. Die einzelnen Leistungen in den verschiedenen Disziplinen (wie 100-Meter-Lauf, Weitsprung, Kugelstoßen oder Hürdenlauf) korrelieren dabei miteinander. Wenn eine Zehnkämpferin beispielsweise im Hochsprung herausragende Leistungen erzielt, wird sie wahrscheinlich auch in den übrigen Disziplinen tendenziell gut abschneiden. Fokussiert sich ihr Training jedoch auf die Disziplin des Kugelstoßens, wird sich ihre Leistung in dieser Disziplin im Laufe der Zeit verbessern, was zu einem besseren Gesamtergebnis führt. Die Ergebnisse der einzelnen Disziplinen korrelieren dann jedoch nicht mehr so stark miteinander und ein hohes Ergebnis im Kugelstoßen wird dann nicht mehr zwangsläufig auf hohe Ergebnisse in den anderen Disziplinen hinweisen. Folglich nimmt der zugrundeliegende gemeinsame Faktor, der auf der Korrelation beruht, ab. Gleichzeitig kommt es nur so lange zu einem Anstieg des Gesamtergebnisses, solange durch das Training des Kugelstoßens, die anderen Disziplinen nicht vernachlässigt werden. Eine Differenzierung der Fähigkeiten führt also nur so lange zu einem besseren Ergebnis, solange das gute Abschneiden in einzelnen Disziplinen, das schlechtere Abschneiden in anderen Disziplinen ausgleicht. Eine erhöhte Fähigkeitsdifferenzierung könnte demnach dazu beitragen, dass einerseits der gemeinsame Faktor schwächer wird, als auch dass über die Zeit durch eine zunehmende Differenzierung der Fähigkeiten die ursprüngliche Verbesserung des Gesamtergebnisses in eine Abnahme dieses umschlägt.

Übertragen auf die Intelligenzebene, geht es dabei nicht um das Training einer einzelnen Person, sondern über die Generationen hinweg (Pietschnig, 2021; Pietschnig et al., 2023). Eine Differenzierung kognitiver Fähigkeiten lässt sich plausibel durch die immer spezifischeren Anforderungen der Gesellschaft erklären. Dies könnte als Reaktion auf die zunehmenden Forderungen nach Spezialisierung im Alltag und Beruf erfolgen (Pietschnig, 2021).

Die Annahme eines Rückgangs im g-Faktor über das letzte Jahrhundert wird darüber hinaus durch Ergebnisse aus Reaktionszeitstudien gestützt. Studien zufolge steht die

Reaktionszeit in Zusammenhang mit dem allgemeinen Faktor der Intelligenz (Johnson & Deary, 2011). Eine psychometrischen Meta Analyse zeigt über den Zeitraum von 1889 bis 2004 eine Korrelation zwischen einfacher Reaktionszeit und dem g-Faktor, die auf einen Rückgang von -1.16 IQ-Punkten pro Jahrzehnt oder -13.35 IQ-Punkten seit dem viktorianischen Zeitalter hinweist, wodurch von den Autoren geschlossen wurde, dass Viktorianer:innen in Bezug auf den allgemeinen Faktor der Intelligenz wesentlich klüger waren als die moderne westliche Bevölkerungen (Woodley et al., 2013) und eine Abnahme des g-Faktors in den Industrieländern im letzten Jahrhundert beobachtet werden (Woodley et al., 2013; Woodley et al., 2014; Woodley of Menie et al., 2015).

Dazu muss jedoch angemerkt werden, dass die Befunde der Studie umstritten sind und die Zunahme in der Reaktionszeit auch auf alternative Erklärungen zurückgeführt werden kann, wie beispielsweise Unterschiede in der Reaktionszeitmessung im Verlauf der Zeit, wodurch ein studienübergreifender Vergleich der Reaktionszeit nicht aussagekräftig ist (z.B. Dodonova & Dodonov, 2013; Nettelbeck, 2014).

Aufgrund eines nachgewiesenen starken Zusammenhangs von *räumlicher Orientierung* mit dem psychometrischen g (Johnson & Bouchard, 2005), legt eine Meta Analyse aus dem deutschsprachigen Raum nahe, dass die Ergebnisse bezüglich räumlicher Wahrnehmung und Orientierung, welche einen umgekehrt u-förmigen Verlauf der IQ-Werte von 1977 bis 2014 mit einem Höhepunkt in den 1990er Jahren zeigen, einen Hinweis für eine Abnahme des psychometrischen g seit den 1990er Jahren sein könnten (Pietschnig & Gittler, 2015). Demnach könnte der Hochpunkt der Kurve als Zeitpunkt interpretiert werden, an welchem die positiven Faktoren, die zu einer IQ-Steigerung beigetragen haben, von einem negativen Trend in g übertroffen werden (Pietschnig & Gittler, 2015).

Relevanz der Aktuelle Studie

Die vorliegende Studie ist ein direkter Test einer cross-temporal schwächer werdenden positiven Kupplung verschiedener Subtestergebnisse als Konsequenz höherer Fähigkeitsdifferenzierung.

Aktuelle Studie

Während generationsübergreifende Leistungsveränderungen auch durch querschnittliche Testung anhand originaler und restandardisierter Tests an derselben Stichprobe untersucht werden können, wird in der folgenden Studie eine longitudinale Untersuchung des

psychometrischen g im Zusammenhang mit den generationsübergreifenden IQ-Veränderungen des Flynn Effekts anhand einer Stichprobe österreichischer Militärpilot:innen und Fluglots:innen über den Zeitraum von 16, 21 und 25 Jahren zwischen 1992 bis 2016 durchgeführt. Die Stichprobe ist aufgrund der strengen Selektierung eine Stichprobe besonders leistungstarker Personen (High Achiever).

Forschungsfragen

Die ersten Forschungsfrage zielt darauf ab, herauszufinden, ob in der Stichprobe österreichischer Militärpilot:innen und Fluglots:innen über den Zeitraum 1992 bis 2016 ein Anstieg der durchschnittlichen IQ-Werte (Flynn Effekt) in den einzelnen Subtests beobachtet werden kann. Vor dem Hintergrund globaler Anstiege und starken Zuwächse in Österreich (vgl. Pietschnig & Voracek, 2015), sowie neueren Studien zu differenzierteren Effektmustern, die Flynn Effekte in verschiedene Richtungen zeigen (Pietschnig & Gittler, 2015; Lazaridis et al, 2022), wird angenommen, dass sich auch in dieser Stichprobe signifikante Leistungsunterschiede über die Zeit hinweg zeigen.

Hypothese 1: H₁: In der Stichprobe kommt es zu signifikanten Anstiegen der durchschnittlichen IQ-Werte in einzelne Subtests über den Zeitraum (1992-2016).

Für die Beantwortung werden Testergebnisse des *Wilde*-Intelligenztests (Jäger & Althoff, 1983) einer Stichprobe im Zeitraum von den 25 Jahren zwischen 1992 bis 2016 mittels linearer Regressionsanalysen in zeitliche Beziehung gesetzt. Ein positiver Flynn Effekt ergibt sich aus einem signifikant positiven Anstieg von IQ-Punkten pro Jahr. Ein negativer Effekt wiederum durch eine signifikante Abnahme der IQ-Punkte pro Jahr.

Die zweite Hauptfragestellung bezieht sich auf die Stabilität des psychometrischen g im Zusammenhang der negativen Korrelation von psychometrischem g und dem Flynn Effekt (z.B. Pietschnig & Voracek, 2015). Auf der Grundlage verschiedener Studien zur Abnahme der Fähigkeiten, die im Zusammenhang mit dem psychometrischen g stehen, wie der Reaktionszeit (Woodley et al., 2013; Woodley et al., 2014) oder der räumlichen Wahrnehmung (Pietschnig & Gittler, 2015), wird angenommen, dass sich in der Stichprobe eine Abnahme des positiven Manifolds zeigt, die eine Erklärung des negativen Zusammenhangs von Flynn Effekt und psychometrischen g liefern kann.

Hypothese 2: H₁: In der Stichprobe nimmt die Stabilität des psychometrischen g über den Zeitraum (1991-2016) ab.

Für die Beantwortung der zweiten Frage wird der Zeitraum von 25 Jahren zwischen 1992 bis 2016 in verschiedene Zeiträume (Zeitcluster) eingeteilt und die mittels einer Hauptkomponentenanalyse (PCA) die gemeinsame latente Struktur der Subtests, R^2 (gemeinsam erklärte Varianz) ermittelt und die Veränderung über die Zeit untersucht. Dabei ist die Annahme, dass die positive Kupplung (gemeinsam erklärte Varianz) über die Zeit hinweg abnimmt. Dies könnte durch die Zunahme der Fähigkeitsdifferenzierung erklärt werden. Die Umkehr des Flynn Effekts könnte durch die gestiegene Fähigkeitsdifferenzierung erklärt werden, die sich in einer Abnahme des durchschnittlichen IQs widerspiegelt.

Aus den beiden ersten Fragestellungen ergibt sich eine dritte Fragestellung, die überprüft, ob die angenommenen IQ-Anstiege (*Hypothese 1*) wie in der Literatur bereits gezeigt (z.B. Pietschnig & Voracek, 2015), negativ mit dem g-Faktor korrelieren und sich für die High-Ability-Stichprobe ebenfalls ein negativer Zusammenhang von Flynn Effekt und g-Faktor zeigt.

Hypothese 3: H₁: In der Stichprobe zeigt sich ein negativer Zusammenhang zwischen den Anstieg der durchschnittlichen IQ-Werte in einzelne Subtests über den Zeitraum (1992-2016) und entsprechenden g-Ladungen der Subtests.

Für die Beantwortung der dritten Frage werden die Veränderungen der IQ-Werte der Subtests (*Hypothese 1*) mit den entsprechenden g-Ladungen der Subtests der Faktorenanalyse (*Hypothese 2*) verglichen. Dabei wird zwischen den Variablen eine *Pearson*-Korrelation gerechnet. Die Subtests, in welchen sich die stärksten Flynn Effekte zeigt, müssten der Theorie nach die geringsten Ladungen mit dem g-Faktor zeigen. Demnach würde eine negative Korrelation zwischen g-Ladungen und IQ-Anstiegen für die *Hypothese 3* eines negativen Zusammenhangs von Flynn Effekt und psychometrischem g, wie er sich in der Literatur zeigt (z.B. Pietschnig & Voracek, 2015), sprechen. Ein positiver Zusammenhang würde der bisherigen Literatur diesbezüglich widersprechen.

Methoden

Stichprobe

Die vorliegende Studie, ist eine sekundäre Analyse archivierter Intelligenztestdaten, die im Rahmen des Auswahlverfahrens von Militärpilot- und Fluglots:innen für das österreichischen Bundesheers gewonnen wurden. Die Stichprobe umfasst dabei die Daten der Personen, die das Auswahlverfahren für angehende Militärpilot:innen und Fluglots:innen im Heeresspital in der Van-Swieten-Kaserne in Stammersdorf (Wien) erfolgreich absolviert haben und in dieser

Funktion beim österreichischen Bundesheer zum Zeitpunkt der Datengewinnung im April 2023 im Dienst waren.

Insgesamt waren im Archiv des österreichischen Bundesheer Datenmappen von 276 Personen verfügbar, die im Zeitraum von 1978 bis 2021 erhoben wurden. Die Hauptanalyse dieser Studie stützt sich aus Gründen der Vergleichbarkeit ausschließlich auf Daten desselben Intelligenztests. In diesem Fall handelt es sich um die Daten des *Wilde*-Intelligenztests (Jäger & Althoff, 1983), welche den Zeitraum von 1992 bis 2016 abdecken, da dieser Test die größte Stichprobe und den längsten Zeitraum umfasst. Weitere Daten wurden nicht berücksichtigt, da die Datenmengen der jeweiligen Intelligenztests begrenzt waren. Für die Analyse wurde sich beim Vorliegen mehrere Testzeitpunkte (bis auf einige Ausnahmen) auf den ersten Zeitpunkt der Datengewinnung bezogen. Die Daten wurden vom Heeresspital zur Verfügung gestellt und anonymisiert damit die Vertraulichkeit der Daten gewährleistet wird.

Die Gesamtstichprobe umfasst $N = 204$ deutschsprachige, gesunde Teilnehmer:innen der Militärpilot:innen und Fluglots:innen des österreichischen Bundesheers. Durch die strengen Auswahlkriterien des Bundesheers, die neben psychologische Voruntersuchung, auch körperliche und geistige Fitness sowie praktische fliegerische Eignung und medizinische Gesundheit voraussetzen, ist die Zusammensetzung der Stichprobe sehr homogen und setzt sich aus besonders leistungsstarken Personen (High Achiever) zusammen. Etwa ein zwei Drittel (133 Personen, 65.2%) der Stichprobe sind dabei Militärpilot:innen und ein Drittel (69 Personen (33.8 %) Fluglots:innen. Das Alter der Testpersonen lag zwischen 18 und 37 Jahren ($M = 21.32$, $SD = 3.79$). Das Geschlecht wurde dichotom erfasst und die Verteilung des Geschlechts sehr einseitig mit 198 (97.1 %) Männern und 6 Frauen (2.9 %). Das Bildungsniveau war folgendermaßen verteilt: Zwei Personen (1.0%) gaben Bildungsgrad zwei an, was einem Pflichtschulabschluss oder abgeschlossenen Realschule entspricht (9-10 Schuljahre). 29 Personen (14.2%) gaben Bildungsgrad drei an, welcher einer abgeschlossenen Fachschule oder Berufsausbildung entspricht (10-12 Jahre Ausbildung). 171 Personen (83.8%) gaben Bildungsgrad vier an, was einer höheren Schule mit Abitur/ Matura entspricht (12-13 Jahre Ausbildung) und eine Person (0.5%) gab Bildungsgrad fünf an, was einem Universitäts- bzw. Hochschulabschluss entspricht. Eine weitere Person (0.5%) hat keine Angaben zum Bildungsgrad gemacht.

Ausschluss und Besonderheiten bei Daten

Eine Person wurde ausgeschlossen, bei der weder ein Geburtsdatum noch ein Testdatum erkennbar waren. Eine weitere Person, bei der das Testdatum nicht eindeutig war. Weitere fünf Personen wurden ausgeschlossen, da in den Unterlagen keine Rohwerte angegeben waren. Drei Personen wurden ausgeschlossen, bei denen ein Subtest einen Rohwert hatte, der höher als die Anzahl der Items des Subtests war. In vier Fällen wurde bei Personen die Daten der zweiten Testung herangezogen, da diese entweder die einzigen verfügbaren Daten waren oder mehr Daten als die erste Voruntersuchung enthielten.

Den meisten Personen wurde die *Langform* des Tests vorgelegt. Einigen Teilnehmenden bearbeiteten einige Subtests des Intelligenztests jedoch als *Kurzform*. Bei zwei Personen wurde sich in dem Fall für die Daten einer späteren Testung mit der *Langform* entschieden, bei weiteren zwei Personen wurden die entsprechenden Werte der *Kurzform* verwendet und in einer entsprechenden Tabelle kodiert. Bei einer Person wurden die Daten der ersten Voruntersuchung herangezogen, obwohl diese in einigen Subtests die *Kurzform*-Version beinhaltete, da diese deutlich mehr Daten enthielt als die spätere Testung, die die *Langform*-Version genutzt hatte.

Bei neun Personen wurde in zumindest einem Subtest höhere Rohwerte erzielt als der Höchstwert der Normwerttabelle. In diesem Fall wurde der höchste Wert genommen und dies mitprotokolliert. Bei einer Person waren sogar zwei Rohwerte höher als der Maximum-Standardwert. Da es sich bei diesen Personen um Deckeneffekte des Tests oder Fehler bei der Auswertung handeln kann.

Messinstrumente

Die verwendeten Subskalen sind aus dem deutschen *Wilde*-Intelligenztest (WIT) (Jäger & Althoff, 1983), welcher sich vom Aufbau und Inhalt an den *Primary Mental Abilities* (PMA) von Louis L. Thurstone orientiert (Thurstone, 1938) und nicht auf der Annahme eines übergeordneten allgemeinen Faktors (g-Faktor) basiert. Demzufolge lassen sich die kognitiven Fähigkeiten grundsätzlich auf sieben voneinander unabhängigen Primärfaktoren zurückführen: Schlussfolgerndes Denken (*Reasoning*), Räumliches Denken (*Space*), Rechnerisches Denken (*Number*), sprachliches Denken (*Verbal Comprehension*), Merkfähigkeit (*Memory*), Flüssigkeit sprachlicher Einfälle (*Word Fluency*) und raschem Erkennen von Details (*Perceptual Speed*) (Thurstone, 1938).

Der Gesamttest ist für Jugendliche und Erwachsene von 14 bis 38 Jahren konzipiert mit einer Testdauer von ca. 4 Stunden. Die Split-half-Reliabilität des Gesamttests beträgt laut dem Manual (nach Rulon & Spearman-Brown) .98. Die Retest-Reliabilität für den Gesamttest liegt bei .93 für die Langform und .92 für die Kurzform. Strukturvergleiche legen nahe, dass eine hohe Konstruktvalidität gegeben ist. Ebenso zeigen Untersuchungen zur Kriteriumsvalidität deutliche Zusammenhänge zwischen den Testergebnissen und den entsprechenden Kriterien auf (Jäger & Althoff, 1983).

Der *Wilde-Intelligenztest* (Jäger & Althoff, 1983) besteht aus 15 verschiedenen Subtests. In Abhängigkeit von den Proband:innen wurde bei der vorliegenden Stichprobe mindestens drei Subtests (*Analogien*, *Zahlenmerken* und *Wortgewandtheit*) und maximal zwölf (ohne *Satzergänzung*, *Grundrechnen* und *Spiegelbilder*) der 15 Subtests durchgeführt. Der Subtest *Satzergänzungen* dient lediglich der Eingewöhnung und wurde den Teilnehmenden in der Voruntersuchung nicht präsentiert. Auf den Subtest *Grundrechnen* wurde im Rahmen der Voruntersuchung ebenfalls verzichtet. Der Subtest *Spiegelbilder* wurde durch einen *dreidimensionalen Würfeltest (3dW)* ersetzt, welcher kein Subtest des *Wilde-Intelligenztest* (WIT) ist, und somit ebenfalls nicht in die Analyse miteinfließt. Folgende Subtests des *Wilde-Intelligenztests* (Jäger & Althoff, 1983, S. 6) wurden für die Analyse herangezogen:

Abwicklungen. Der Subtest Abwicklungen erfasst das räumliche Vorstellen und beinhaltet 20 Items, die jeweils eine Faltvorlage zeigen, die aus mehreren Flächen mit verschiedenen Zeichen besteht. Aus 5 vorgegebenen Körpern soll für jedes Item derjenige herausgefunden werden, der sich aus der vorgegeben Faltvorlage herstellen lässt. Die Split-Half-Reliabilität (nach Rulon) beträgt .75 und die Retest-Reliabilität .74. Die innere Konsistenz für die Langform (nach Kuder-Richardson) .82 (Jäger & Althoff, 1983).

Analogien. Der Subtest Analogien prüft das schlussfolgende Denken von sprachlichem Material und umfasst 25 Items, bei welchen jeweils auf der linken Seite eines Gleichheitszeichens zwei Wörter mit einer bestimmten Beziehung zueinander vorgegeben werden. Auf der rechten Seite befindet sich ein weiteres Wort, zu dem aus einer Auswahl von fünf Wörtern das Wort gewählt werden soll, welches zum rechten Wort eine analoge Beziehung hat (z.B. Schaf: Wolle = Vogel:? a) Flügel b) Nest c) Adler d) Federn e) Tier. Die richtige Antwort ist d) Federn). Die Split-Half-Reliabilität (nach Rulon) beträgt .87 und die Retest-Reliabilität .86. Die innere Konsistenz für die Langform (nach Kuder-Richardson) .91 (Jäger & Althoff, 1983).

Beobachtung. Der Subtest Beobachtungen erfasst die Geschwindigkeit und Genauigkeit der Wahrnehmung und umfasst 42 Items, in welchen jeweils drei sehr ähnliche Zeichnungen von Schemagesichter gezeigt werden, von denen zwei identisch sind. Die Aufgabe besteht darin, das dritte Gesicht zu identifizieren, das sich durch eine klar erkennbare Einzelheit von den anderen beiden unterscheidet. Die Split-Half-Reliabilität (nach Rulon) beträgt .94 und die Retest-Reliabilität .73. Die innere Konsistenz für die Langform (nach Kuder-Richardson) .94 (Jäger & Althoff, 1983).

Buchstabenreihen. Der Subtest Buchstabenreihen umfasst 15 Items, die das schlussfolgernde Denken in Bezug auf „sinnfreies“ Material erfassen. Der Testperson wird dabei eine Folge von Buchstaben präsentiert, die nach einer bestimmten Regel angeordnet sind. Aus fünf Buchstabenpaaren ist dasjenige zu identifizieren, dass die Reihe auf logische Weise fortführt (z.B. die Buchstabenreihe a h b h c h d h e h mit folgenden Lösungsmöglichkeiten 1) a h; 2) e h; 3) b h; 4) f h; 5) g h. Die richtige Antwort ist Lösung 4) f h, da die Buchstabenreihe dem Alphabet folgt und immer zwischen zwei Buchstaben ein h eingeschoben ist. Die Split-Half-Reliabilität (nach Rulon) beträgt .85 und die Retest-Reliabilität .72. Die innere Konsistenz für die Langform (nach Kuder-Richardson) .89 (Jäger & Althoff, 1983).

Eingekleidete Rechenaufgabe. Der Subtest Eingekleidete Rechenaufgaben umfasst 18 Items zur Überprüfung des rechnerischen Denkens. Der Testperson werden verbal eingekleidete Rechenaufgaben vorgegeben, deren Lösungen in einem Antwortbogen in Ziffern eingetragen werden soll (z.B. Ein Schüler erhält täglich 30 Pfennig Taschengeld. Wieviel Pfennig bekommt er dann in 7 Tagen? Die richtige Antwort sind 210 Pfennig). Die Split-Half-Reliabilität (nach Rulon) beträgt .86 und die Retest-Reliabilität .84. Die innere Konsistenz für die Langform (nach Kuder-Richardson) .86 (Jäger & Althoff, 1983).

Gedächtnis. Der Subtest Gedächtnis erfasst anhand 40 Items das mittelfristige Gedächtnis. Dabei werden Lebensläufen von zwei fiktiven Personen vorgegeben, einschließlich ihrer Geburts- und Wohnorte, Geburtsdaten, Telefonnummern sowie Fotos von ihnen, ihren Eltern und Ehepartner:innen. Nach einer Einprägungszeit von sechs Minuten folgen weitere Subtests, die etwa eine Stunde dauern. Im darauffolgenden dreiteiligen Reproduktionstest (W= 13 Items, 5 Antwortmöglichkeiten vorgegeben, 1 richtig; K=11 Items, Sätze mit Details zu den Geschichten, wobei 5 Satzteile unterstrichen sind und einer davon falsch ist und als solcher erkannt werden soll; R=16 Items, teilweise Auswahlmöglichkeiten teilweise Antworten auf Fragen im freien Format) sollen Teile der Lebensläufe und Fotos wiedererkannt, unterschieden

und reproduziert werden. Die Split-Half-Reliabilität (nach Rulon) beträgt .85 und die Retest-Reliabilität .66 (Jäger & Althoff, 1983).

Gleiche Wortbedeutung. Der Subtest Gleiche Wortbedeutungen erfasst in 20 Items das Wortverständnis im engeren Sinn. Aus fünf vorgegeben Wörtern das herausgesucht werden, das dem Sinn eines vorgegeben Wort am ähnlichsten ist (z.B. für das Wort „Kopf“: a) Haar b) Haupt c) Körper d) Mensch e) Hut). Die Split-Half-Reliabilität (nach Rulon) beträgt .85 und die Retest-Reliabilität .86. Die innere Konsistenz für die Langform (nach Kuder-Richardson) .85 (Jäger & Althoff, 1983).

Schätzen. Der Subtest Schätzen erfasst in 21 Items das rechnerische Denken und insbesondere die Vertrautheit mit dem System Zahlen und seine Beziehungen, sowie das Abschätzen von Größenordnungen. In diesem Subtest soll aus fünf Lösungsoptionen aufgrund rechnerischer Überlegungen oder Schätzungen die Lösung einer Schätz- oder Rechenaufgabe gewählt werden (z.B. die Lösung der Rechenaufgabe $118492 - 3684 - 2106 - 4768 = ?$ mit den Antwortalternativen a) 107621 b) 104527 c) 107934 d) 98743 e) 99825. Die richtige Antwort ist c) 107934). Die Split-Half-Reliabilität (nach Rulon) beträgt .81 und die Retest-Reliabilität .76. Die innere Konsistenz für die Langform (nach Kuder-Richardson) .83 (Jäger & Althoff, 1983).

Sprichwörter. Der Subtest Sprichwörter erfasst in 20 Items das Sinnesverständnis für komplexere sprachliche Inhalte. Dabei sollen jeweils fünf Sprichwörter betrachtet werden, aus denen die beiden auszuwählen sind, die dem Sinn eines vorgegeben Sprichwortes am besten entsprechen (z.B. das Sprichwort „Gesundheit ist der größte Reichtum“. Zwei der nun folgenden fünf Sprichwörter haben ungefähr den gleichen Sinn. Welche sind es? a) Jeder ist seines Glückes Schmied b) Nichts ist so wertvoll wie die Gesundheit c) Der eine stirbt, der andere lebt d) Schmiede das Eisen, so lange es heiß ist e) Gesundheit ist wie Gold. Die richtigen Antworten sind b) und e). Die Split-Half-Reliabilität (nach Rulon) beträgt .84 und die Retest-Reliabilität .88. Die innere Konsistenz für die Langform (nach Kuder-Richardson) .87 (Jäger & Althoff, 1983).

Wortgewandtheit. Der Subtest Wortgewandtheit erfasst die sprachliche Flüssigkeit anhand assoziativer Worteinfälle. Dabei ist keine maximale Itemanzahl vorgegeben, der Rohwert der Anzahl an genannten Worten der Versuchsperson Rohwert entspricht. Dafür werden Anfangs- und Endbuchstaben vorgegeben und innerhalb einer Minute möglichst viele Wörter niedergeschrieben, die mit denselben Anfangsbuchstaben beginnen und Endbuchstaben

enden (z.B. Wörter, die mit dem Buchstaben B anfangen und N enden, wie die Wörter Bahn, Bohnen, Bühnen, Brücken). Die Split-Half-Reliabilität (nach Spearman-Brown) beträgt .93 und die Retest-Reliabilität .85 (Jäger & Althoff, 1983).

Zahlenmerken. Der Subtest Zahlenmerken erfasst die Merkfähigkeit und das unmittelbare Behalten. Bei diesem Subtest ergibt sich der Rohwert aus der Anzahl richtig gemerkter Zahlen. Dabei wird eine verbale Folge von fünf bis elf einstellige Zahlen in gleichen Zeitabständen vorgegeben, die unmittelbar nach der mündlichen Darbietung schriftlich festgehalten werden sollen. Die Split-Half-Reliabilität (nach Rulon) beträgt .95 und die Retest-Reliabilität .17 (Jäger & Althoff, 1983).

Zahlenreihen. Der Subtest Zahlenreihen erfasst in 15 Items das schlussfolgernde Denken bei Zahlenmaterial und gibt eine Folge von Zahlen vor, die nach einer Regel aufgebaut ist. Auf dem Antwortbogen sollen die Ziffern der Zahl durchgestrichen werden, welche als nächstes in der Reihe folgen müsste (z.B. 2 5 8 11 14 17? Jede Zahl ist um 3 größer als die vorhergehende, dementsprechend ist die richtige Antwort 20). Die Split-Half-Reliabilität (nach Rulon) beträgt .89 und die Retest-Reliabilität .82. Die innere Konsistenz für die Langform (nach Kuder-Richardson) .90 (Jäger & Althoff, 1983).

Durchführung

Es handelt sich bei der gegenwärtigen Analyse um eine sekundäre Untersuchung bereits vorhandener Intelligenztestdaten des *Wilde-Intelligenztests* (WIT) von Jäger & Althoff (1983). Dafür wurde für die Fragestellungen auf Daten einzelner Subtests, die während des Auswahlverfahrens für Militärpilot:innen (und Fluglots:innen) des Bundesheers durchgeführt wurden, zurückgegriffen. Soziodemografische Daten, wie *Bildungsniveau*, *Alter* und *Geschlecht* der Person wurden ebenfalls aus den analogen Testaufzeichnungen entnommen. Es waren nicht für alle Subtests und Personen Itemlevel-Daten verfügbar, die das Verhalten der Personen in den einzelnen Testaufgaben zeigen, sondern lediglich Subtestscores, also Rohwerte für die einzelnen Subtests, die eine zusammengefasste Messung der Leistung der Person darstellen.

Statistische Analyse

Vor der statistischen Analyse wurden verschiedene Datenverarbeitungsschritte durchgeführt. Die analogen Rohwerte jedes Subtests, die die Anzahl an gelösten Aufgaben widerspiegeln, wurden nach der Digitalisierung in einer test- und altersspezifischen Normwerttabelle in

Standardwerte kodiert und anschließend in IQ-Werte umgewandelt (Normwertrechner: Lenhard & Lenhard, 2015). Für die Analyse der drei Hypothesen, ob in der Stichprobe ein Flynn Effekt vorliegt (*Hypothese 1*), inwiefern sich die Stabilität des psychometrischen g über die Zeit hinweg verändert (*Hypothese 2*) und dem negativen Zusammenhang zwischen IQ-Anstiegen und psychometrischem g (*Hypothese 3*), wurden die IQ-Werte verwendet.

Subtests

In der Analyse sind nur die Subtests enthalten, die in der Voruntersuchung des Aufnahmeverfahrens zur Anwendung kamen. Die Personen unterscheiden sich in der Anzahl absolvierter Subtests und haben zwischen drei und zwölf Subtests absolviert. Die Analyse der Subtests in der vorliegenden Stichprobe zeigt, dass von den insgesamt 204 Personen 168 Personen (82.4 %) sämtliche zwölf Subtests bearbeiteten. Zwölf weitere Personen (5.9 %) führten zehn der zwölf Subtests durch (Kumulativ n = 180 Personen, die zehn Subtests durchgeführt haben). Drei weitere (1.5 %) Personen bearbeiteten sieben Subtests (insgesamt n = 183). Eine Person (0.5 %) absolvierten sechs der Subtests (insgesamt n = 184). 20 Personen (9.8 %) absolvierten drei Subtests (insgesamt n = 204). Für eine Übersicht sind in *Tabelle 1* die Anzahl der absolvierten Subtests und ihre Verteilung aufgeführt.

Tabelle 1

Anzahl und Verteilung der absolvierten Subtests

Anzahl Subtests	Subtests												nSub	nKum
	Abwicklungen	Analogien	Beobachtungen	Buchstabenreihen	Eingekleidete Rechenaufgaben	Gedächtnis	Gleiche Wortgewandtheit	Schätzen	Sprichwörter	Wortgewandtheit	Zahlenmerken	Zahlenreihen		
12	x	x	x	x	x	x	x	x	x	x	x	x	168	168
10		x	x	x		x	x	x	x	x	x	x	12	180
7		x	x	x	x	x				x	x		3	183
6		x		x	x	x				x	x		1	184
3		x								x	x		20	204

In *Tabelle 2* zeigt sich, dass von 1992 bis 2007 Daten für alle zwölf durchgeführten Subtests vorliegen. Ab dem Jahr 2008 wurden die Subtests *Eingekleidete Rechenaufgaben* und *Abwicklungen* nicht mehr durchgeführt. Ab 2013 wurden nur noch die drei Subtests *Analogien*, *Zahlenmerken* und *Wortgewandtheit* durchgeführt.

Tabelle 2

Anzahl an Daten pro Subtest und Jahr

N in Subtests pro Jahr												
Jahr	Abwicklungen	Analogien	Beobachtungen	Buchstabenreihen	Eingekleidete Rechenaufgaben	Gedächtnis	Gleiche Wortbedeutung	Schätzen	Sprichwörter	Wortgewandtheit	Zahlenmerken	Zahlenreihen
1992	5	8	7	8	8	8	5	5	5	8	8	5
1993	15	16	16	16	16	16	15	15	15	16	16	15
1994	13	13	13	13	13	13	13	13	13	13	13	13
1995	2	2	2	2	2	2	2	2	2	2	2	2
1996	6	6	6	6	6	6	6	6	6	6	6	6
1997	8	8	8	8	8	8	8	8	8	8	8	8
1998	16	16	16	16	16	16	16	16	16	16	16	16
1999	13	14	13	13	13	13	13	13	13	14	14	13
2000	12	12	12	12	12	12	12	12	12	12	12	12
2001	15	15	15	15	15	15	15	15	15	15	15	15
2002	11	11	11	11	11	11	11	11	11	11	11	11
2003	12	12	12	12	12	12	12	12	12	12	12	12
2004	12	12	12	12	12	12	12	12	12	12	12	12
2005	8	8	8	8	8	8	8	8	8	8	8	8
2006	8	8	8	8	8	8	8	8	8	8	8	8
2007	11	11	11	11	11	11	11	11	11	11	11	11
N	167	172	170	171	171	171	167	167	167	172	172	167
2008	-	9	7	7	-	7	7	7	7	9	9	7
2009	-	4	1	1	-	1	1	1	1	4	4	1

2010	-	2	2	2	-	2	2	2	2	2	2	2	
2011	-	7	2	2	-	2	2	2	2	7	7	2	
2012	-	1	1	1	-	1	1	1	1	1	1	1	
N		195	183	184		184	180	180	180	195	195	180	
2013	-	2	-	-	-	-	-	-	-	2	2	-	
2014	-	1	-	-	-	-	-	-	-	1	1	-	
2015	-	4	-	-	-	-	-	-	-	4	4	-	
2016	-	2	-	-	-	-	-	-	-	2	2	-	
N		167	204	183	184	171	184	180	180	180	204	204	180

Um festzustellen, ob ein Flynn Effekt vorliegt (*Hypothese 1*) wurde eine lineare Regressionsanalyse, mit den IQ-Werten des ersten Termins der Voruntersuchung als abhängige Variable (*AV*) und das Testjahr als unabhängige Variable (*UV*), für jeden Subtests durchgeführt. Dabei wurden drei Analysen gerechnet: Eine Analyse, die alle 12 Subtests beinhaltet (1992-2007). Eine Analyse die 10 Subtest beinhaltet (1992-2012). Und zuletzt eine Analyse mit drei Subtests (gesamte Stichprobe, 1992-2012). Für jede Analyse ergibt sich für jeden Subtest ein Regressionskoeffizient, der die Veränderung von IQ-Punkten pro Jahr angibt.

Für die zweite Forschungsfrage (*Hypothese 2*) zur Stabilität des psychometrischen g wurde sich die Veränderung der gemeinsam erklärten Varianz des ersten Faktors (g-Faktor) einer Faktorenanalyse angeschaut. Zunächst die Stichprobe für jede der drei Analysen (12 Subtests, 10 Subtests, 3 Subtests) in zwei annähernd gleichgroße Zeitspannen (Zeitcluster) aufgeteilt (*Analyse 12 Subtests*: 1992-1999 vs. 2000-2007; *Analyse 10 Subtests*: 1992-2001 vs. 2002-2012; *Analyse 3 Subtests*: 1992-2001 vs. 2002-2016) und für beide Spannen eine Hauptkomponentenanalyse (PCA) gerechnet. Anschließend wurde in allen drei Analysen die gemeinsam erklärte Varianz, in Form des Bestimmtheitsmaß (R^2) der beiden Hauptkomponentenanalysen, deskriptiv miteinander verglichen.

Für eine genauere Untersuchung der Veränderung von R^2 wurde, um Jahre mit begrenzten Daten (s. *Tabelle 2* ab 2008) zu berücksichtigen, nicht für jedes Jahr eine gemeinsame Varianz des ersten Faktors berechnet, sondern die Stichprobe in jeder der *drei* Analysen anhand der Jahre des Tests in Zeiträume (Zeitcluster) unterteilt. Für jeden Zeitcluster wurde in der Analyse eine eigene Hauptkomponentenanalyse gerechnet und die gemeinsam erklärte Varianz des ersten Faktors (R^2) für diese Zeitspanne extrahiert (s. *Tabellen 3, 4 und 5*).

Um die Frage nach dem Zusammenhang zwischen der gemeinsam erklärten Varianz (R^2) der Zeitabschnitte (psychometrisches g) und dem Erhebungszeitraum zu beantworten, wurden *Spearman*-Rangkorrelationen gerechnet, um eine mögliche Veränderung des latenten Konstrukts g, dem gemeinsamen Faktor, zu ermitteln. Weiters wurden für jede Analyse die innere Konsistenz der Items (Reliabilität) mittels McDonald's Omega (ω) berechnet.

Für die dritte Forschungsfrage (*Hypothese 3*) des negativen Zusammenhangs der IQ-Anstiege und dem psychometrischem g, wurden eine *Pearson*-Korrelation der Regressionsanstiege in den einzelnen Subtests der ersten Hypothese mit den g-Ladungen der Subtests in der Hauptkomponentenanalyse der zweiten Hypothese gerechnet. Eine negative Korrelation spricht für die *Hypothese 3* eines negativen Zusammenhangs von Flynn Effekt und psychometrischem g, wie er sich in der Literatur zeigt. Ein positiver Zusammenhang würde der bisherigen Literatur diesbezüglich widersprechen.

Tabelle 3

Faktorenanalysen (1992-2007), 12 Subtests, 16 Jahre, 4 Cluster

	Cluster	Zeitraum	Jahre	N
Überblick- Analyse	1	1992-1999	8	78-83
	2	2000-2007	8	89
Analyse	1	1992-1995	4	35-39
	2	1996-1999	4	43-44
	3	2000-2003	4	50
	4	2004-2007	4	39

Tabelle 4

Faktorenanalysen (1992-2012) 10 Subtests, 21 Jahre, 5 Cluster

	Cluster	Zeitraum	Jahre	N
Überblick-Analyse	1	1992-2001	10	105-110

	2	2002-2012	11	75-85
Analyse	1	1992-1995	4	35-39
	2	1996-1999	4	43-44
	3	2000-2003	4	50
	4	2004-2007	4	39
	5	2008-2012	5	13-23

Tabelle 5

Faktorenanalysen (1992-2016), 3 Subtests, 25 Jahre, 5 Cluster

	Cluster	Zeitraum	Jahre	N
Überblick-Analyse	1	1992-2001	9	110
	2	2002-2016	15	94
Analyse	1	1992-1995	4	39
	2	1996-1999	4	44
	3	2000-2003	4	50
	4	2004-2007	4	39
	5	2008-2016	9	32

Ergebnisse

Für die Durchführung der statistischen Analyse wurde die Software *SPSS Statistics* und die grafischen Darstellungen der linearen Regressionen das Statistikprogramm *RStudio* verwendet. Der Code für die in *RStudio* erstellten Grafiken wurde in OSF (Open Science: https://osf.io/9h548/?view_only=83dd59b4e4444aa9826a9aff2d384d2f) gestellt. In den

folgenden *Tabellen 6-8* werden die Ergebnisse der linearen Regressionsanalysen mit dem Jahr des Tests (UV) und den IQ-Werten der Subtests (AV) dargestellt.

Analyse 1: Zwölf Subtests (1992-2007)

Die Analyse der linearen Regressionen der zwölf Subtests ergab in Bezug auf das *Jahr des Tests* (UV) und den *IQ-Werten der Subtests* (AV) im Zeitraum 1992 bis 2007 sieben signifikante Ergebnisse. Liegt eine (signifikant) positive lineare Regression vor, entspricht das einem linearen Anstieg der IQ-Punkte des Subtests über die Zeit. Eine signifikant positive Regressionsanalyse lag bei den folgenden Subtests vor (*Tabelle 6*).

Tabelle 6

Analyse 12 Subtests: Ergebnisse Lineare Regression 12 Subtests (1992-2007)

Subtest	N	Regressions- koeffizient	Standard -fehler	Beta	T	Signifikanz
Abwicklungen	167	1.396	0.193	0.490	7.216	<0.001**
Analogien	172	0.094	0.173	0.042	0.545	0.587
Beobachtungen	170	0.427	0.193	0.169	2.218	0.028*
Buchstabenreihen	171	0.391	0.172	0.173	2.279	0.024*
Eingekleidete Rechenaufgaben	171	2.371	0.260	0.575	9.129	<0.001**
Gedächtnis	171	0.842	0.198	0.311	4.253	<0.001**
Gleiche Wortbedeutung	167	0.089	0.162	0.043	0.549	0.584
Schätzen	167	0.482	0.212	0.174	2.274	0.024*
Sprichwörter	167	0.256	0.143	0.138	1.791	0.075
Wortgewandtheit	172	0.262	0.188	0.107	1.398	0.164
Zahlenmerken	172	0.305	0.231	0.101	1.323	0.187

Zahlenreihen	167	0.690	0.230	0.227	2.994	0.003**
--------------	-----	-------	-------	-------	-------	---------

* $p < 0.05$; ** $p < 0.01$

Der Subtest *Eingekleidete Rechenaufgaben*, der das rechnerische Denken prüft, zeigt über den Zeitraum von 16 Jahren einen starken signifikanten Anstieg von 2.371 IQ-Punkte pro Jahr.

Auch der IQ-Wert des Subtests *Abwicklungen*, welcher das räumliche Vorstellungsvermögen erfasst, steigt in der Zeitspanne von 1992 bis 2007 mit jedem Jahr signifikant um 1.396 IQ-Punkte.

Der IQ-Wert im Subtest *Gedächtnis*, für die Erfassung des mittelfristige Gedächtnisses, steigt signifikant um 0.842 IQ-Punkte pro Jahr.

Der Subtest *Zahlenreihen*, welcher schlussfolgerndes Denken bei Zahlenmaterial erfasst, zeigt im selben Zeitraum einen signifikanten Anstieg von 0.690 IQ-Punkte pro Jahr.

Im Subtest *Schätzen*, welcher das rechnerische Denken und den Umgang mit dem Zahlensystem testet, zeigt sich ebenfalls ein signifikanter Anstieg von 0.482 IQ-Punkten pro Jahr.

Im Subtest *Beobachtungen*, welcher die Wahrnehmungsgeschwindigkeit und Wahrnehmungsgenauigkeit testet, steigt der IQ-Wert des Subtests signifikant um 0.427 IQ-Punkte pro Jahr an.

Der Subtest *Buchstabenreihen*, der das schlussfolgernde Denken in Bezug auf „sinnfreiem“ Material testet zeigt sich ein signifikanter Anstieg um 0.391 IQ-Punkte.

Die Subtests *Zahlenmerken* (0.305 IQ-Punkte pro Jahr), *Wortgewandtheit* (0.262 IQ-Punkte), *Sprichwörter* (0.256 IQ-Punkte), *Analogien* (0.094 IQ-Punkte) und *Gleiche Wortbedeutung* (0.089 IQ-Punkte) zeigen ebenfalls Anstiege, diese sind jedoch nicht signifikant.

Analyse 2: Zehn Subtests (1992-2012)

In der Analyse der zehn Subtests kam es in den linearen Regressionen in Bezug auf das *Jahr des Tests* und den IQ-Werten der Subtests von 1992 bis 2012 zu zwei signifikanten Subtests: *Buchstabenreihen* und *Gedächtnis* (Tabelle 7).

Tabelle 7*Analyse 10 Subtests: Ergebnisse Lineare Regression 10 Subtests (1992-2012)*

Subtest	N	Regressions- koeffizient	Standard -fehler	Beta	T	Signifikanz
Analogien	195	-0.060	0.135	-0.032	-0.441	0.660
Beobachtungen	183	0.154	0.168	0.068	0.919	0.360
Buchstabenreihen	184	0.298	0.147	0.149	2.030	0.044*
Gedächtnis	184	0.862	0.171	0.351	5.054	<0.001**
Gleiche Wortbedeutung	180	0.089	0.138	0.048	0.645	0.520
Schätzen	180	0.174	0.184	0.070	0.942	0.347
Sprichwörter	180	0.154	0.124	0.093	1.241	0.216
Wortgewandtheit	195	0.181	0.144	0.090	1.254	0.211
Zahlenmerken	195	0.180	0.184	0.070	0.978	0.329
Zahlenreihen	180	0.369	0.201	0.136	1.831	0.069

*p< 0.05; ** p< 0.01

Auch in der Analyse des Subtests bezüglich des mittelfristigen *Gedächtnisses* ist das *Jahr der Testung* signifikanter Prädiktor für den durchschnittlichen IQ-Wert. Der durchschnittliche IQ-Wert des Subtests erhöht sich über den Zeitraum von 21 Jahren um 0.862 IQ-Punkte pro Jahr.

Der Subtest *Buchstabenreihen* wird ebenfalls in der Analyse signifikant von dem *Jahr der Testung* beeinflusst. Mit jedem Jahr, das vergeht, steigt der durchschnittliche IQ-Wert des Subtests signifikant um 0.298 IQ-Punkte an.

Alle Regressionskoeffizienten der Subtests sind unabhängig der Signifikanz dabei positiv, außer der Regressionskoeffizient des Subtest *Analogien*, der ein negatives Vorzeichen hat. Die Subtests *Zahlenreihen* (0.369 IQ-Punkte pro Jahr), *Wortgewandtheit* (0.181 IQ-Punkte), *Zahlenmerken* (0.180 IQ-Punkte), *Schätzen* (0.174 IQ-Punkte), *Beobachtungen* (0.154

IQ-Punkte) und *Sprichwörter* (0.154 IQ-Punkte) zeigen ebenfalls Anstiege, sind jedoch nicht signifikant. Der Subtest *Analogien* (-0.060 IQ-Punkte) hat als einziger Subtest einen negativen Regressionskoeffizient.

Analyse 3: Drei Subtests (1992-2016)

Die Analyse der linearen Regressionen der drei Subtests ergab in Bezug auf das *Jahr des Tests* und den IQ-Werten der Subtests im Zeitraum 1992 bis 2016 zwei signifikanten Ergebnisse (Tabelle 8).

Tabelle 8

Analyse 3 Subtests: Ergebnisse Lineare Regression drei Subtests (1992-2016)

Subtest	N	Regressions- koeffizient	Standard -fehler	Beta	T	Signifikanz
Analogien	204	-0.297	0.122	-0.169	-2.440	0.016*
Wortgewandtheit	204	0.319	0.126	0.175	2.528	0.012*
Zahlenmerken	204	0.165	0.162	0.071	1.016	0.311

* $p < 0.05$; ** $p < 0.01$

Bei dem Subtest *Analogien*, der das schlussfolgernde Denken bei sprachlichem Material überprüft, zeigt sich ein signifikant negativer Einfluss des *Jahrs der Testung* auf den IQ-Wert. Über den Zeitraum von 25 Jahren sinkt mit jedem Jahr, das vergeht der durchschnittliche IQ-Wert des Subtests um -0.297 IQ-Punkte, was einer Abnahme von -2.97 IQ-Punkte pro Jahrzehnt entspricht.

Bei dem Subtest *Wortgewandtheit*, der die quantitative Seite der Flüssigkeit assoziierter Worteinfälle überprüft, zeigt sich einen signifikanten positiver Einfluss des *Jahrs der Testung* auf den IQ-Wert. Mit jedem Jahr, das vergeht steigt der durchschnittliche IQ-Wert des Subtests um 0.319 IQ-Punkte.

Der Subtest *Zahlenmerken* zeigt in der Analyse über 25 Jahre einen nicht signifikanten Anstieg von 0.165 IQ-Punkten pro Jahr.

Die Subtests *Gleiche Wortbedeutung*, *Sprichwörter* und *Zahlenmerken* zeigen in keiner der drei Analysen einen signifikanten Zusammenhang zwischen IQ-Werten und dem Jahr des Tests, zeigen aber dennoch Anstiege der IQ-Werte über den Zeitraum (1992-2016). Bei den anderen Subtests kommt es zumindest in einer Analyse zu einem signifikanten Anstieg oder einer Abnahme über die Zeit (*Tabelle 9*).

Tabelle 9

Anstiege IQ-Punkte in den Subtests pro Jahr (Dekade)

IQ-Punkte Anstiege der Subtests pro Jahr (pro Dekade)			
Subtests	Analyse: 12 Subtests (1992-2007) N = 167-172	Analyse: 10 Subtests (1992-2012) N = 180-195	Analyse: 3 Subtests (1992-2016) N = 204
Abwicklungen	1.396 (13.96)**	-	-
Analogien	0.094 (0.94)	-0.060 (-0.60)	-0.297 (2.97)*
Beobachtungen	0.427 (4.27)*	0.154 (1.54)	-
Buchstabenreihen	0.391 (3.91)*	0.298 (2.98)*	-
Eingekleidete Rechenaufgaben	2.371 (23.71)**	-	-
Gedächtnis	0.842 (0.842)**	0.862 (8.62)**	-
Gleiche Wortbedeutung	0.089 (0.89)	0.089 (0.89)	-
Schätzen	0.482 (4.82) *	0.174 (1.74)	-
Sprichwörter	0.256 (2.56)	0.154 (1.54)	-
Wortgewandtheit	0.262 (2.62)	0.181 (1.81)	0.319 (3.19)*

CROSS-TEMPORALE STABILITÄT DES PSYCHOMETRISCHEN G

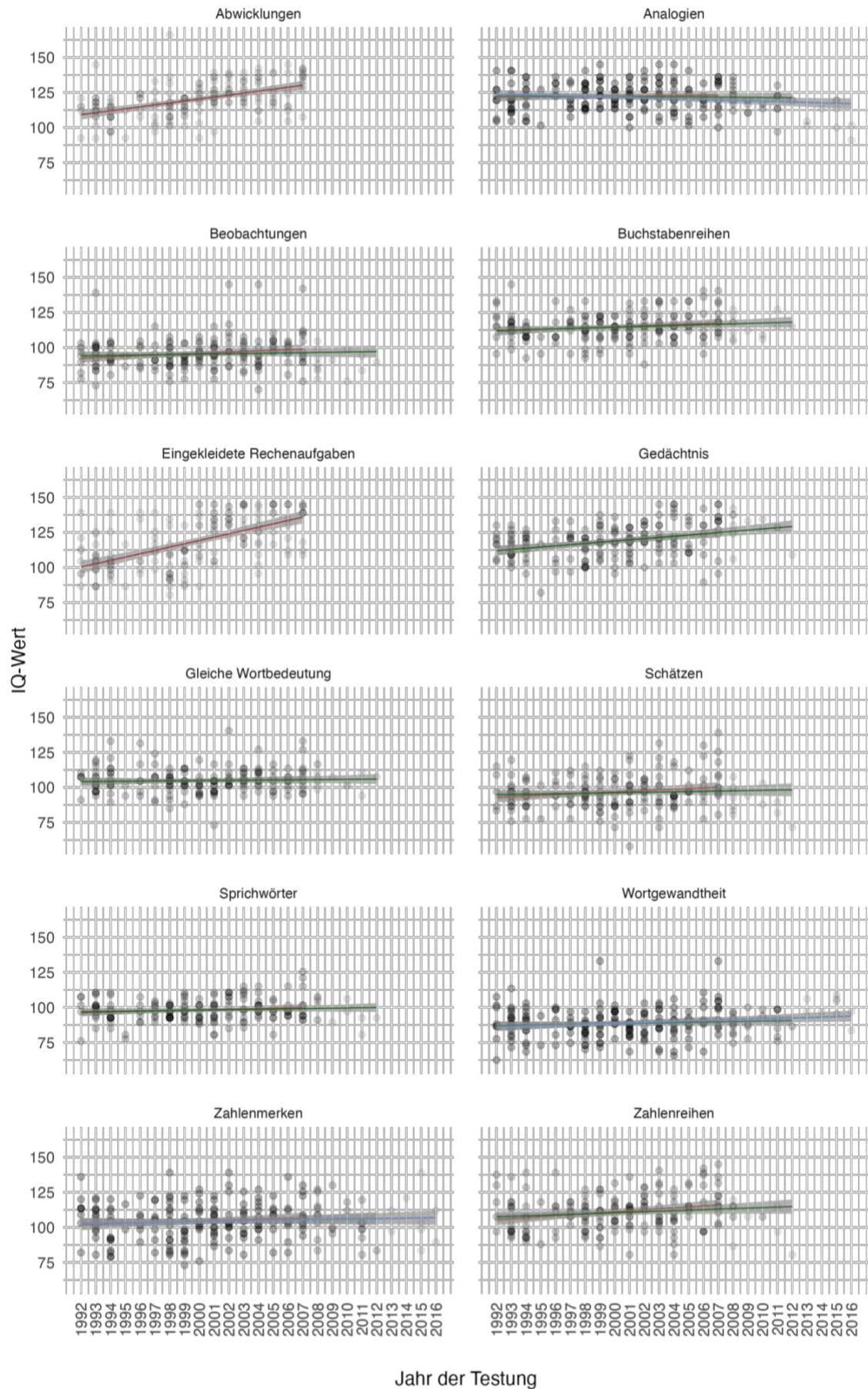
Zahlenmerken	0.305 (3.05)	0.180 (1.80)	0.165 (1.65)
Zahlenreihen	0.690 (6.90)**	0.369 (3.69)	-

* $p < 0.05$; ** $p < 0.01$

Grafik 1

Abbildung. Streudiagramm der IQ-Werte der Subtests über die drei Analysen (1992-2016)

CROSS-TEMPORALE STABILITÄT DES PSYCHOMETRISCHEN G



Analysen

— Analyse 12 Subtests — Analyse 10 Subtests — Analyse 3 Subtests

Stabilität des psychometrischen g (Hauptkomponentenanalyse, PCA)

Für jeden Cluster jeder Analyse, wie sie in der statistischen Analyse festgelegt wurden, wurde eine Hauptkomponentenanalyse (PCA) gerechnet. In den *Tabellen 10-12* ist die gemeinsam erklärte Varianz (R^2) des ersten gemeinsamen Faktors jeder Hauptkomponentenanalyse (PCA) und die Reliabilität anhand der McDonald's Omega-Werte (ω) für jeden der drei Analysen angegeben. Die Richtung, Stärke und Signifikanz der Veränderung der erklärten Varianz (R^2) über die Zeitcluster hinweg wurde anschließend in allen drei Analysen mithilfe einer *Spearman-Rho* Rangkorrelation festgestellt.

Tabelle 10

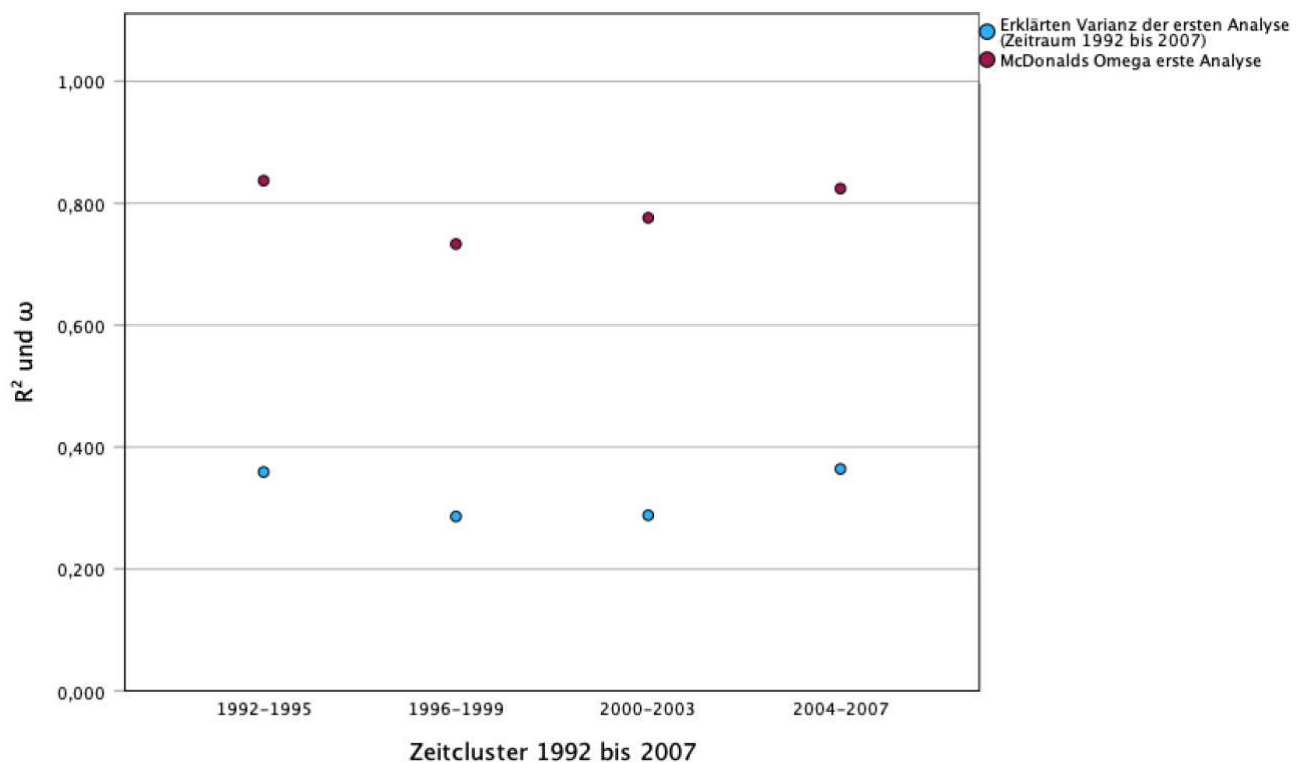
Analyse 12 Subtests: Faktorenanalysen Zeitraum 1992-2007, 16 Jahre, 4 Cluster

	Cluster	R^2	McDonald's Omega (ω)	Zeitraum	Jahre
Überblick-Analyse	1	0.307	0.788	1992-1999	8
	2	0.324	0.804	2000-2007	8
Analyse	1	0.359	0.837	1992-1995	4
	2	0.286	0.733	1996-1999	4
	3	0.288	0.776	2000-2003	4
	4	0.364	0.824	2004-2007	4

Dabei sind die R^2 -Werte der *Analyse der 12 Subtests* weitgehend konstant und schwanken zwischen 0.286 und 0.364. Die *Spearman-Rho* Rangkorrelation für den Zeitraum 1992 bis 2007 (*12 Subtests*) zwischen R^2 der 12 Subtests und den vier Zeitclustern zeigt eine nicht signifikante positive Korrelation von $r = 0.400$ ($p = 0.600$). Die McDonald's Omega-Werte (ω) schwanken zwischen 0.733 und 0.837. Für McDonald's Omega zeigt sich in einer *Spearman-Rho* Rangkorrelation eine nicht signifikante negative Korrelation $r = -0.200$ ($p = 0.800$). Die R^2 - und McDonald's-Omega-Werte werden in *Grafik 2* dargestellt.

Grafik 2

Analyse 12 Subtests: Veränderung von g und McDonald's Omega (1992-2007)

**Tabelle 11**

Analyse 10 Subtests: Faktorenanalysen Zeitraum 1992-2012, 21 Jahre, 5 Cluster

	Cluster	R ²	McDonald's Omega (ω)	Zeitraum	Jahre
Überblick-Analyse	1	0.290	0.714	1992-2001	10
	2	0.353	0.791	2002-2012	11
Analyse	1	0.356	0.792	1992-1995	4
	2	0.298	0.665	1996-1999	4
	3	0.290	0.707	2000-2003	4
	4	0.424	0.839	2004-2007	4

5 0.301 0.654 2008-2012 5

Die R^2 -Werte der *Analyse der 10 Subtests* sind ebenfalls weitgehend konstant und schwanken über die Zeit zwischen 0.290 und 0.424. Die *Spearman*-Rho Rangkorrelation für den Zeitraum 1992 bis 2012 (*10 Subtests*) zwischen den R^2 über die 10 Subtests und den fünf Zeitclustern zeigt eine nicht signifikante positive Korrelation von $r = 0.100$ ($p = 0.873$). Für McDonald's Omega zeigt sich in einer *Spearman*-Rho Rangkorrelation eine nicht signifikante negative Korrelation $r = -0.300$ ($p = 0.624$). Die R^2 - und McDonald's-Omega-Werte werden in *Grafik 3* dargestellt.

Grafik 3

Analyse 10 Subtests: Diagramm zur Veränderung von psychometrischem g (1992-2012)

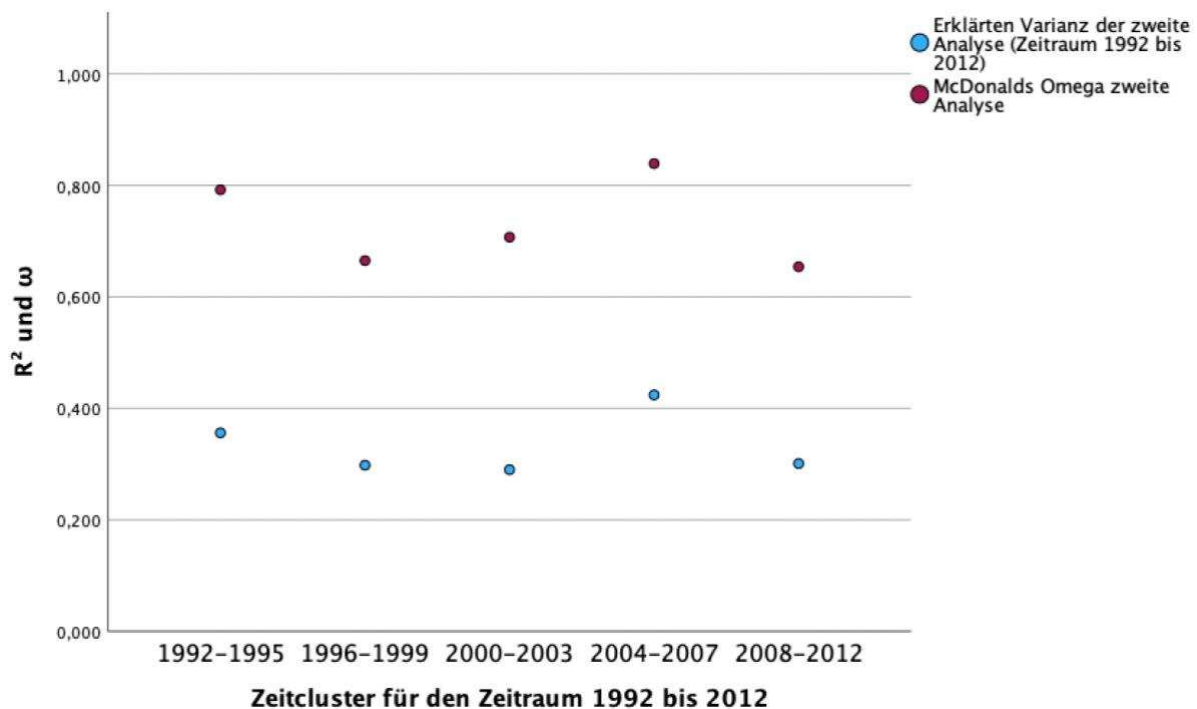


Tabelle 12

Analyse 3 Subtests: Faktorenanalysen für den Zeitraum 1992-2016, 25 Jahre, 5 Cluster

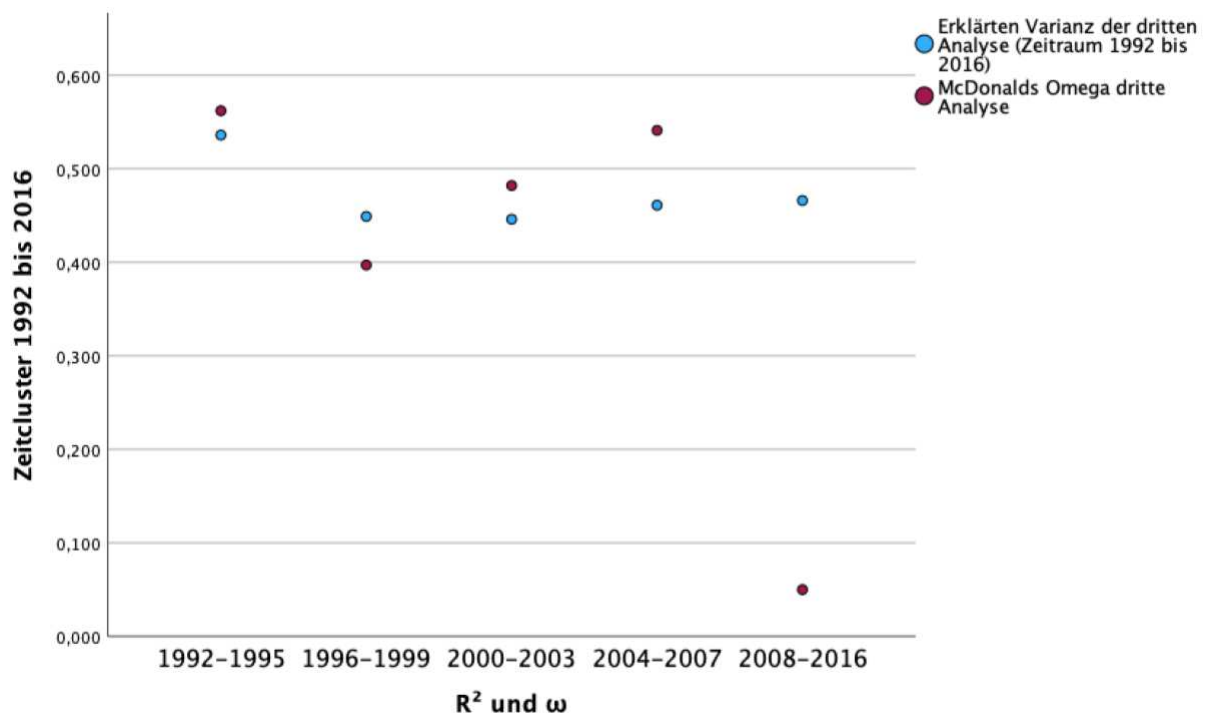
	Cluster	R^2	McDonald's Omega (ω)	Zeitraum	Jahre
Überblick-Analyse	1	0.458	0.391	1992-2001	9

	2	0.394	0.416	2002-2016	15
Analyse	1	0.536	0.562	1992-1995	4
	2	0.449	0.397	1996-1999	4
	3	0.446	0.482	2000-2003	4
	4	0.461	0.541	2004-2007	4
	5	0.466	-	2008-2016	9

Die R^2 -Werte der *Analyse der 3 Subtests* sind weitgehend konstant und schwanken zwischen 0.446 und 0.536. Die *Spearman-Rho* Rangkorrelation für den Zeitraum 1992 bis 2016 (3 *Subtests*) zwischen den R^2 über die 3 Subtests und den fünf Zeitclustern zeigt eine nicht signifikante negative Korrelation $r = -0.100$ ($p = 0.873$). Für McDonald's Omega zeigt sich in einer *Spearman-Rho* Rangkorrelation eine nicht signifikante negative Korrelation $r = -0.600$ ($p = 0.285$). Die R^2 - und McDonald's-Omega-Werte werden in *Grafik 4* dargestellt.

Grafik 4

Analyse 3 Subtests: Diagramm zur Veränderung von psychometrischen g (1992-2016),



Für die Analyse über die 3 Subtests (1992-2016) wird für den letzten Zeitcluster keine Reliabilität anhand eines McDonald's Omega berichtet, da der letzten Datenpunkt ein Ausreißer (McDonald's Omega (ω) = 0.050) darstellt.

In *Tabelle 13* werden die IQ-Punkt Veränderungen und g-Ladungen der einzelnen Subtests aufgeführt, aus denen die Korrelationen für den Zusammenhang zwischen Flynn Effekt und psychometrischem g gerechnet werden.

Tabelle 13

Abbildung 12. Korrelationen der IQ-Punkt-Veränderungen mit der g-Ladung der Subtests

	Analyse: 12 Subtests (1992-2007) N = 167-172		Analyse: 10 Subtests (1992-2012) N = 180-195		Analyse: 3 Subtests (1992-2016) N = 204	
	<i>R² des ersten Faktors: 0.329</i>		<i>R² des ersten Faktors: 0.333</i>		<i>R² des ersten Faktors: 0.423</i>	
Subtests	IQ-Punkt Anstiege	Ladungen auf ersten Faktor	IQ-Punkt Anstiege	Ladungen	IQ-Punkt Anstiege	Ladungen
Abwicklungen	1.396**	0.523	-	-	-	-
Analogien	0.094	0.548	-0.060	0.618	-0.297*	0.700
Beobachtungen	0.427*	0.530	0.154	0.547	-	-
Buchstabenreihen	0.391*	0.638	0.298*	0.601	-	-
Eingekleidete Rechenaufgaben	2.371**	0.674	-	-	-	-
Gedächtnis	0.842 **	0.525	0.862**	0.475	-	-

Gleiche Wortbedeutung	0.089	0.467	0.089	0.532	-	-
Schätzen	0.482*	0.661	0.174	0.653	-	-
Sprichwörter	0.256	0.568	0.154	0.628	-	-
Wortgewandtheit	0.262	0.574	0.181	0.593	0.319*	0.653
Zahlenmerken	0.305	0.382	0.180	0.354	0.165	0.593
Zahlenreihen	0.690**	0.706	0.369	0.695	-	-

* $p < 0.05$; ** $p < 0.01$

Aus einer Faktorenanalyse der einzelnen Analysen (12, 10 und 3 Subtests) über die Zeiträume (1992-2007, 1992-2012, 1992-2016) gehen Ladungen des ersten Faktors auf die einzelnen Subtests hervor (*Tabelle 13*). Diese g-Ladungen wurden mit den Anstiegen der linearen Regression aus *Hypothese 1* korreliert. Daraus ergeben sich für die Analyse über die 12 Subtests (1992-2007) eine nicht signifikante positive Korrelation zwischen g-Ladungen und IQ-Anstiegen ($r = 0.365$; $p = 0.243$). Für die beiden anderen Analysen ergeben sich nicht signifikante negative Korrelationen. Für die Analyse über die 10 Subtests (1992-2012) $r = -0.214$ ($p = 0.553$) und für die Analyse über die 3 Subtests (1992-2016) ein $r = -0.670$ ($p = 0.532$).

Diskussion

Das Ziel der Studie war es, über einen festgelegten Zeitraum Veränderungen des allgemeinen Intelligenzfaktors, dem psychometrischen g, in Zusammenhang mit den Veränderungen der IQ-Testergebnisse des Flynn Effekts anhand kognitiver Testleistungen an einer österreichischen Stichprobe besonders leistungsstarken Personen (High-Ability-Stichprobe) zu untersuchen.

Obwohl zahlreiche Studien über den Verlauf des 20. Jahrhunderts einen negativen Zusammenhang zwischen Flynn Effekt und psychometrischem g zeigten (z.B. Pietschnig & Voracek, 2015; te Nijenhuis & Van der Flier, 2013), wurde diesem im Kontext der Veränderungen der Effektmuster des Flynn Effekts bisher wenig Beachtung geschenkt. Dabei wurde die Hypothese aufgestellt, dass den Veränderungen des Flynn Effekts eine Verringerung

der positiven Kupplung zwischen den einzelnen Subtests zugrunde liegen könnte, was wiederum zu einem niedrigeren psychometrischem g über die Zeit führen würde.

Basierend auf den Daten zeigt sich ein überwiegend positiver Flynn Effekt über die verschiedenen Subtests und untersuchten Zeiträume hinweg. In der Stichprobe kommt es zu unterschiedlich starken signifikanten IQ-Anstiegen und in einem Subtest zu einer signifikanten Abnahme der IQ-Werte über die Zeit (Pietschnig & Voracek, 2015; Lazaridis et al., 2022). Allerdings zeigt sich in keiner Analyse der Stichprobe besonders leistungstarker Personen Hinweise für eine Abnahme von g. Sowohl die erklärten Varianzen (R^2), als auch die Reliabilitäten (McDonald's Omega, ω) zeigen keine signifikanten Veränderungen über den untersuchten Zeitraum von 1992 bis 2016.

Ebenfalls zeigen die *Pearson*-Korrelationen zwischen IQ-Anstiegen und der g-Ladungen der einzelnen Subtests für keine Analyse der besonders leistungstarken Stichprobe signifikante Ergebnisse.

Flynn Effekt

Es zeigt sich ein bemerkenswert hoher Anstieg im Subtest *Eingekleidete Rechenaufgaben*, welcher das rechnerische Denken testet, von 2.371 IQ-Punkten pro Jahr. Die insgesamt besonders hohen Werte und Deckeneffekte können durch die besondere Leistungsstärke der Stichprobe erklärt werden. Die ungewöhnlich starke Zunahme über den untersuchten Zeitraum, von 2.371 IQ-Punkten pro Jahr kann jedoch nicht allein durch die besonderen Eigenschaften der Stichprobe erklärt werden.

Die signifikante Zunahme von 1.396 IQ-Punkten pro Jahr im Subtest *Abwicklungen*, der räumliches Vorstellen umfasst, in der Analyse über 12 Subtests (1992-2007) steht im Kontrast zu den Ergebnissen einer Meta Analyse, die von 1977 bis 2014 eine Abnahme der räumlichen Wahrnehmung ab den 1990er Jahren im deutschsprachigen Raum zeigt (Pietschnig & Gittler, 2015). Jedoch ist anzumerken, dass ab dem Jahr 2008 der Subtest *Abwicklungen* nicht mehr durchgeführt wurde und die Daten nur einen kleinen Ausschnitt der Entwicklung des räumlichen Vorstellungsvermögens in einer leistungstarken Stichprobe zeigen. Möglicherweise können die Ergebnisse dieser Untersuchung auch ein Hinweis auf Unterschiede in den Entwicklungsmustern des Flynn Effekts zwischen besonders begabten Personen und der durchschnittlichen Bevölkerung sein.

Die festgestellte Abnahme der IQ-Punkte in der Analyse über 3 Subtests (1992-2016) im Subtest *Analogien*, der das schlussfolgernde Denken von sprachlichem Material prüft, von -0.297 IQ-Punkten könnte auf verschiedene Faktoren zurückzuführen sein. Diese könnte einerseits Anzeichen für eine Stagnation oder sogar Umkehr des Flynn Effekts in der High-Ability-Stichprobe sein, die bereits in einigen Ländern berichtet wird (z.B. Norwegen, Sundet et al., 2004; Dänemark, Teasdale & Owen, 2008). Andererseits könnte es für einen Vorliegen eines domänenspezifischen Flynn Effekts in der High-Ability-Stichprobe sprechen.

Eine weitere mögliche Erklärung für den Rückgang der IQ-Testergebnisse in *Analogien* könnte mit veralteten Items in Verbindung stehen. Wie bereits erwähnt, können obsoleete Items, die veraltetes Wissen abfragen, besonders die Leistung der kristallinen Intelligenz maskieren (Hörtl, 2019; Pietschnig et al., 2019). Es besteht die Möglichkeit, dass Wörter wie „*Siechtum*“, „*Tuschkasten*“, oder „*Bücherbrett*“ des Subtests *Analogien* im *Wilde-Intelligenztest* (Jäger & Althoff, 1983) über den Verlauf der Zeit (1992-2016) nicht mehr verstanden werden. Da das Verstehen der Worte Voraussetzung ist, um eine Analogie dazu bilden zu können beeinflusst ein veränderter Sprachgebrauch das Abschneiden im Subtest maßgeblich. Wenn die Worte von der Stichprobe nicht mehr verstanden werden, kann eine beobachtete Abnahme der IQ-Werte im Subtest unter anderem auf eine vorliegende Itemobsoleszenz, dem Vorliegen veralteter Testaufgaben, zurückgeführt werden (vgl. Hörtl, 2019; Pietschnig et al., 2019).

Weiters könnten verschiedene Intelligenz-Domänen unterschiedlich stark von sozialen und kulturellen Einflüssen geprägt sein (siehe Soziale Multiplikatoren, Dickens & Flynn, 2010). Ein negativer Flynn Effekt könnte darauf hindeuten, dass fördernde Umweltfaktoren für das schlussfolgernde Denken von sprachlichem Material in der Stichprobe zurückgegangen sein könnten, oder Faktoren, die einen Rückgang begünstigen einen größeren Einfluss auf die kognitiven Fähigkeiten haben (Dickens & Flynn, 2010).

Psychometrisches g und Stabilität

Aufgrund der Hypothese einer zunehmenden Fähigkeitsdifferenzierung sind wir davon ausgegangen, dass es im Zeitraum von 1992 bis 2016 in der vorliegenden Stichprobe besonders leistungsstarker Personen (High-Ability-Stichprobe) zu einer Abnahme des g-Faktors in diesem Zeitraum kommt. Diese Veränderung könnte plausibel durch eine gestiegene Fähigkeitsdifferenzierung in der Bevölkerung, erklärt werden. In Kombination mit abnehmenden Erträgen und Deckeneffekten IQ-steigernder Umweltfaktoren wie Bildung (Lynn, 2009) und Ernährung (Ceci & Williams, 1997) könnte dies als Erklärungsansatz für

einen möglichen Rückgang der IQ-Anstiege dienen. Allerdings zeigen sich basierend auf den Daten österreichischer Militärpilot:innen und Fluglots:innen keine Hinweise für eine Abnahme von g.

Für die Analyse über 12 Subtests (1992-2007), als auch für die Analyse über 10 Subtests (1992-2012) zeigen sich nicht signifikante positive Veränderungen, während sich in der Analyse über 3 Subtests eine nicht signifikante negative Veränderung der erklärten Varianz (R^2) zeigt. Die McDonald's Omegas (ω) zeigen über alle drei Analysen hinweg nicht signifikante negative Veränderungen.

Die inkonsistenten Befunde deuten darauf hin, dass es im Wesentlichen in der Stichprobe besonders leistungsstarker Personen, keine Veränderungen in g zu geben scheint. Durch die Besonderheit der Stichprobe, die auf die hohen Anforderungen und speziellen Auswahlkriterien des österreichischen Bundesheers sehr leistungsstark und in ihrer Varianz sehr homogen ist (z.B. *Geschlecht*: männlich = 198 (97.1%) Personen; $M = 21.32$ Jahre $SD = 3.79$ Jahren), könnten die Ergebnisse darauf hindeuten, dass im Kontext der vorgeschlagenen Fähigkeitsdifferenzierung (Pietschnig, 2021; Pietschnig & Voracek, 2015; Pietschnig, 2018; Pietschnig & Gittler, 2015), die Theorie eines abnehmenden g in Bezug auf die High-Ability-Stichprobe möglicherweise nicht zutrifft.

Negativer Zusammenhang von Flynn Effekt und psychometrischem g

Als dritter Punkt galt es in der Stichprobe den negativen Zusammenhang der IQ-Wert Anstiege mit dem psychometrischen g zu untersuchen, den zahlreichen Studien gezeigt haben (z.B. Pietschnig & Voracek, 2015; te Nijenhuis & Van der Flier, 2013; Rushton, 1999). Die *Pearson*-Korrelation zwischen IQ-Anstiegen und der g-Ladungen der einzelnen Subtests zeigt für keine Analysen der besonders leistungsstarken Stichprobe signifikante Ergebnisse. Da die IQ-Werte über die Zeit ansteigen und sich ein überwiegend positiver Flynn Effekt zeigt, während das psychometrische g über den Zeitraum 1992 bis 2016 keine signifikanten Änderungen zeigt, deuten die Ergebnisse nicht darauf hin, dass in der vorliegenden Stichprobe der Flynn Effekt nicht negativ mit g assoziiert ist.

Limitationen

Die Interpretation der Ergebnisse ist durch nachfolgende Limitationen eingeschränkt. Die Studie basiert auf dem *Wilde*-Intelligenztest (Jäger & Althoff, 1983), der als deutschsprachiger Test auf den Primärfaktoren von Thurstone beruht, die mittlerweile durch hierarchische

Modelle eines übergeordneten g-Faktors (z.B. im *CHC*-Modell von Schneider & McGrew, 2018) ersetzt wurden.

Die Datenanalyse beinhaltet nur Daten, die zum Zeitpunkt der Digitalisierung (April, 2023) im Daten-Archiv des österreichischen Bundesheers, im Heeresspital Stammersdorf in Wien, zur Verfügung standen. Dies schließt alle Personen ein, die den Aufnahmetest zur Militärpilot:in oder Fluglots:in erfolgreich bestanden haben, die Ausbildung abgeschlossen und sich zur gegenwärtigen Datenerhebung noch im Dienst befunden haben. Durch die Abhängigkeit der Daten von diesen Faktoren, lagen insbesondere in den letzten Jahren der Analyse (s. *Tabelle 2*) weniger Daten vor. Allerdings ergibt sich durch die spezielle Stichprobe auch eine Stärke, da sie ermöglicht Veränderungen im oberen Bereich des Intelligenzspektrums genauer zu untersuchen und spezifische Trend oder Muster bei besonders begabten Personen aufzudecken, die in einer bevölkerungsrepräsentativen Stichprobe möglicherweise weniger deutlich wären.

Eine weitere Limitation ist, dass bei neun Personen mindestens ein erzielter Rohwert höher war als der Höchstwert der Normwerttabelle (*Eingekleidete Rechenaufgaben* $n = 6$; *Zahlenreihen* $n = 1$; *Beobachtungen* $n = 1$; *Gedächtnis* $n = 1$). Dies kann daran liegen, dass die High-Ability-Stichprobe besser abschneidet, als die Stichprobe anhand derer genormt wurde. Da besonders die Personen über 30 Jahren deutlich besser abschneiden als die Werte der Normwerttabelle, zeigt die besondere Leistungsfähigkeit der Stichprobe des Bundesheers insbesondere in höheren Altersklassen.

Während die besonders hohen Werte im Subtest *Eingekleidete Rechenaufgaben* und entsprechende Deckeneffekte auf die besondere Leistungsstärke der Stichprobe zurückgeführt werden können, kann die starke Zunahme von 2.371 IQ-Punkten pro Jahr jedoch nicht allein durch die besonderen Eigenschaften der Stichprobe erklärt werden. Insbesondere vor dem Hintergrund, dass vor der Analyse drei Personen ausgeschlossen wurden, bei welchen im Subtest *Eingekleidete Rechenaufgaben* höhere Rohwerte als die Anzahl vorgegebener Items im *Wilde-Intelligenztests* (Jäger & Althoff, 1983), erzielt wurden (der Rohwert ergibt sich aus der Anzahl der gelösten Items). In Rücksprache mit dem Bundesheer besteht die Möglichkeit, dass es bei der Datenerhebung über den Zeitraum zu Änderungen kam. Eine genaue Nachvollziehbarkeit der Testdurchführung und möglichen Veränderungen ist jedoch nicht mehr möglich, wodurch die bemerkenswerten Anstiege des Subtests mit Vorsicht zu interpretieren sind.

Ebenfalls wurde drei Personen eine leicht abweichende Testversion (*Kurzform*) vorgelegt. Bei Ausschluss der Personen ergeben sich jedoch nahezu numerisch identische Ergebnisse, wodurch diese Limitation die Ergebnisse nicht gefährden sollten.

Die Geschlechterverteilung der Stichprobe sollte für die Interpretation der Flynn Effekte kein Problem darstellen, da sich im deutschsprachigen Raum kaum Hinweise für geschlechtsspezifische Unterschiede in den IQ-Anstiegen gezeigt haben (Pietschnig et al., 2011).

Conclusio

Es zeigt sich ein starker positiver Flynn Effekt in einer Stichprobe besonders leistungstarker Personen (High-Ability Individuals) in den beobachteten Zeiträumen von 1992 bis 2016. Allerdings zeigt sich keine konklusive Evidenz für eine Änderung der Stärke der positiven Kupplung über die Zeit. Die Ergebnisse stützen die aufgestellte These einer zunehmende Fähigkeitsdifferenzierung für die Gruppe besonders leistungstarker Personen (High Achiever) nicht eindeutig und können ein erster Hinweis gegen die Theorie einer zunehmenden Fähigkeitsdifferenzierung und Abnahme des psychometrischen g bei besonders leistungstarken Personen (High-Ability-Sample) sein. Die Analyse des negativen Zusammenhangs von IQ-Anstiegen und psychometrischen g zeigen darüber hinaus ebenfalls keine signifikanten Ergebnisse. Da der IQ ansteigt, während g gleichbleibt, kann dies jedoch nicht als Hinweis darauf gedeutet werden, dass in der Stichprobe der Flynn Effekt nicht negativ mit g assoziiert ist.

Die vorliegende Studie trägt zur bestehenden Literatur bei, indem sie eine direkte Untersuchung der Veränderung des psychometrischen g-Faktors im Zusammenhang mit den Anstiegen des durchschnittlichen IQs an einer besonders leistungstarken Stichprobe (High-Ability-Sample) durchführt.

Weitere Untersuchungen an kognitiv besonders leistungstarken Personen (High-Ability-Sample) könnten dazu beitragen, die gefundenen Ergebnisse zur Stabilität des g-Faktors kritisch zu prüfen und wertvolle Einblicke in diese spezielle Personengruppe zu gewinnen. Zukünftige Studien könnten zeigen, ob die beobachteten Veränderungen des g-Faktors in dieser herausragenden Stichprobe im Vergleich zur Allgemeinbevölkerung signifikante Unterschiede aufweisen und ob die gefundenen Ergebnisse tatsächlich replizierbar sind.

Literatur

- Batterjee, A. A., Khaleefa, O., Ali, K. & Lynn, R. (2013). An Increase of Intelligence in Saudi Arabia, 1977–2010. *Intelligence*, 41(2), 91–93. <https://doi.org/10.1016/j.intell.2012.10.011>
- Binet, A. & Simon, T. (1904). Méthodes nouvelles pour le diagnostic du niveau intellectuel des anormaux. *Annee Psychologique*, 11(1), 191–244. <https://doi.org/10.3406/psy.1904.3675>
- Brand, C. (1987). Keeping up with the times. *Nature*, 328(6133), 761. <https://doi.org/10.1038/328761a0>
- Ceci, S. J. & Williams, W. M. (1997). Schooling, intelligence, and income. *American Psychologist*, 52(10), 1051–1058. <https://doi.org/10.1037/0003-066x.52.10.1051>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences (2nd ed.)*. Hillsdale, N.J.: L. Erlbaum Associates.
- Colom, R., Abad, F., García, L. F. & Juan-Espinosa, M. (2002). Education, Wechsler's full scale IQ, and G. *Intelligence*, 30(5), 449–462. [https://doi.org/10.1016/s0160-2896\(02\)00122-8](https://doi.org/10.1016/s0160-2896(02)00122-8)
- Daley, T. C., Whaley, S. E., Sigman, M., Espinosa, M. & Neumann, C. G. (2003). IQ on the rise. *Psychological Science*, 14(3), 215–219. <https://doi.org/10.1111/1467-9280.02434>
- Deary, I. J. (2012). Intelligence. *Annual Review of Psychology*, 63(1), 453–482. <https://doi.org/10.1146/annurev-psych-120710-100353>
- Deary, I. J. (Ed.). (2009). Intelligence, health, and death: The emerging field of cognitive epidemiology [Special issue]. *Intelligence*, 37.

- Dickens, W. T., & Flynn, J. R. (2001). Herability estimates versus large environmental effects: The IQ paradox resolved. *Psychological Review*, 108, 346-369. Doing:10.1037//0033-295X.108.2.346
- Dickinson, M. & Hiscock, M. (2010). Age-related IQ decline is reduced markedly after adjustment for the Flynn effect. *Journal of Clinical and Experimental Neuropsychology*, 32(8), 865–870. <https://doi.org/10.1080/13803391003596413>
- Dodonova, Y. A. & Dodonov, Y. S. (2013). Is there any evidence of historical slowing of reaction time? No, unless we compare apples and oranges. *Intelligence*, 41(5), 674–687. <https://doi.org/10.1016/j.intell.2013.09.001>
- Dutton, E. & Lynn, R. (2013). A negative Flynn effect in Finland, 1997–2009. *Intelligence*, 41(6), 817–820. <https://doi.org/10.1016/j.intell.2013.05.008>
- Dutton, E., & Lynn, R. (2015). A negative Flynn Effect in France, 1999 to 2008-9. *Intelligence*, 51, 67-70. <https://doi.org/10.1016/j.intell.2015.05.005>
- Dutton, E., van der Linden, D., & Lynn, R. (2016). The negative Flynn effect: A systematic literature review. *Intelligence*, 59, 1163-169 <https://doi.org/10.1016/j.intell.2016.10.002>
- Eppig, C., Fincher, C. L. & Thornhill, R. (2010). Parasite prevalence and the worldwide distribution of cognitive ability. *Proceedings of the Royal Society B: Biological Sciences*, 277(1701), 3801–3808. <https://doi.org/10.1098/rspb.2010.0973>
- Flynn J.R. (1987). Massive IQ gains in 14 nations: What IQ tests really measure. *Psychological Bulletin*, 101, 171-191.

- Flynn, J. R. (2006). Tethering the Elephant: capital cases, IQ, and the Flynn effect. *Psychology, Public Policy and Law*, 12(2), 170–189. <https://doi.org/10.1037/1076-8971.12.2.170>
- Flynn, J. R. (2007). *What is intelligence? Beyond the Flynn effect*. Cambridge: Cambridge University Press.
- Flynn, J. R. (2010). Problems with IQ gains: the huge vocabulary gap. *Journal of Psychoeducational Assessment*, 28(5), 412–433. <https://doi.org/10.1177/0734282910373342>
- Flynn, J.R. (1984). The mean IQ of Americans: Massive gains 1932 to 1978. *Psychological Bulletin*, 95, 29-51.
- Gonthier, C., Grégoire, J. & Besançon, M. (2021). No negative Flynn effect in France: Why variations of intelligence should not be assessed using tests based on cultural knowledge. *Intelligence*, 84, 101512. <https://doi.org/10.1016/j.intell.2020.101512>
- Gottfredson, L. S. (1997). Mainstream Science on Intelligence: an editorial with 52 signatories, history, and bibliography. *Intelligence*, 24(1), 13–23. [https://doi.org/10.1016/s0160-2896\(97\)90011-8](https://doi.org/10.1016/s0160-2896(97)90011-8)
- Herrnstein, R. J. & Murray, C. A. (1994). *The Bell Curve: Intelligence and Class Structure in American Life*.
- Hötl, Franziska (2019). Einflüsse von Itemobsoleszenz auf generationsspezifische Testnormverschiebungen. Masterarbeit, Universität Wien. Fakultät für Psychologie.
BetreuerIn: Pietschnig, Jakob

- Horn, J. L. & Cattell, R. B. (1966). Refinement and test of the theory of fluid and crystallized general intelligences. *Journal of Educational Psychology*, 57(5), 253–270. <https://doi.org/10.1037/h0023816>
- Hunt, E. (2011). *Human intelligence*. Cambridge University Press.
- Jäger, A. O. & Althoff, K. (1983). *Der Wilde-Intelligenz-Test (WIT). Ein Strukturdiagnostikum*. Verlag für Psychologie.
- Jensen, A. R. (1992). Understanding g in terms of information processing. *Educational Psychology Review*, 4(3), 271–308. <https://doi.org/10.1007/bf01417874>
- Jensen, A. R. (1998). *The G factor: The Science of Mental Ability*. Praeger.
- Kaufman, A. S., Zhou, X., Reynolds, M. R., Kaufman, N. L., Green, G. P. & Weiss, L. G. (2014). The possible societal impact of the decrease in U.S. blood lead levels on adult IQ. *Environmental Research*, 132, 413–420. <https://doi.org/10.1016/j.envres.2014.04.015>
- Johnson, W. & Bouchard, T. J. (2005). The structure of human intelligence: it is verbal, perceptual, and image rotation (VPR), not fluid and crystallized. *Intelligence*, 33(4), 393–416. <https://doi.org/10.1016/j.intell.2004.12.002>
- Johnson, W. & Deary, I. J. (2011). Placing inspection time, reaction time, and perceptual speed in the broader context of cognitive ability: the VPR Model in the Lothian Birth Cohort 1936. *Intelligence*, 39(5), 405–417. <https://doi.org/10.1016/j.intell.2011.07.003>
- Kâğıtçıbaşı, Ç. & Biricik, D. (2011). Generational Gains on the Draw-a-Person IQ Scores: A three-decade comparison from Turkey. *Intelligence*, 39(5), 351–356. <https://doi.org/10.1016/j.intell.2011.06.001>

Kaufman, A. S., Zhou, X., Reynolds, M. R., Kaufman, N. L., Green, G. P. & Weiss, L. G.

(2014). The possible societal impact of the decrease in U.S. blood lead levels on adult

IQ. *Environmental Research*, 132, 413–

420. <https://doi.org/10.1016/j.envres.2014.04.015>

Kersting, M. (2007). Wenn Tests in die Jahre kommen. Probleme des Einsatzes überalterter

Testverfahren. In C. Lorei (Hrsg.), *Polizei und Psychologie* (S. 565-577). Frankfurt:

Verlag für Polizeiwissenschaft.

Kersting, M., Althoff, K. & Jäger, A. O. (2008). *WIT-2: Wilde-Intelligenz-Test 2* (1. Auflage

2008). Hogrefe.

Khaleefa, O., Sulman, A. & Lynn, R. (2009). An increase of intelligence in Sudan, 1987–

2007. *Journal of Biosocial Science*, 41(2), 279–

283. <https://doi.org/10.1017/s0021932008003180>

Kirkegaard, E. O. W. (2013). Predicting immigrant IQ from their countries of origin, and

Lynn's national IQs: a case study from Denmark. *Mankind Quarterly*, 54(2), 151–

167. <https://doi.org/10.46469/mq.2013.54.2.2>

Kuncel, N. R., & Hezlett, S. A. (2010). Fact and fiction in cognitive ability testing for

admissions and hiring decisions. *Current Directions in Psychological Science*, 19(6),

339–345. <https://doi.org/10.1177/0963721410389459>

Lang, J. W. B. & Kell, H. J. (2020). General mental ability and specific abilities: Their

relative importance for extrinsic career success. *Journal of Applied*

Psychology, 105(9), 1047–1061. <https://doi.org/10.1037/apl0000472>

- Lazaridis, A., Vetter, M. & Pietschnig, J. (2022). Domain-specificity of Flynn effects in the CHC-model: Stratum II test score changes in Germanophone samples (1996–2018). *Intelligence*, 95, 101707. <https://doi.org/10.1016/j.intell.2022.101707>
- Lenhard, W. & Lenhard, A. (2015). *Normwertrechner*. verfügbar unter: <https://www.psychometrica.de/normwertrechner.html>. Psychometrica. DOI: 10.13140/RG.2.1.4592.5363
- Liu, J. & Lynn, R. (2013). An increase of intelligence in China 1986–2012. *Intelligence*, 41(5), 479–481. <https://doi.org/10.1016/j.intell.2013.06.017>
- Lynn, R. (1982). IQ in Japan and the United States shows a growing disparity. *Nature*, 297(5863), 222–223. <https://doi.org/10.1038/297222a0>
- Loehlin, J. C. (2002). The IQ paradox: Resolved? Still an open question. *Psychological Review*, 109, 754-758; discussion 764-771.
- Lynn, R. (2009). What has caused the Flynn effect? Secular increases in the development quotients of infants. *Intelligence*, 37, 16-24. <https://doi.org/10.1016/j.intell.2008.07.008>
- Lynn, R. (2011). *Dysgenics: Genetic deterioration in modern populations*. London: Ulster Institute for Social Research.
- Lynn, R. (2013). Who discovered the Flynn Effect? A review of early studies of the secular increase of intelligence. *Intelligence*, 41(6), 765–769. <https://doi.org/10.1016/j.intell.2013.03.008>
- Lynn, R. & Irwing, P. (2004). Sex Differences on the Progressive Matrices: A Meta-analysis. *Intelligence*, 32(5), 481–498. <https://doi.org/10.1016/j.intell.2004.06.008>

- Lynn, R. & Vanhanen, T. (2012). National IQs: A review of their educational, cognitive, economic, political, demographic, sociological, epidemiological, geographic and climatic correlates. *Intelligence*, 40(2), 226–234. <https://doi.org/10.1016/j.intell.2011.11.004>
- Meisenberg, G., Lawless, E., Lambert, E. & Newton, A. V. (2005). The Flynn Effect in the Caribbean: Generational change of cognitive test performance in Dominica. *Mankind Quarterly*, 46(1), 29–69. <https://doi.org/10.46469/mq.2005.46.1.2>
- Meisenberg, G., Lawless, E., Lambert, E. & Newton, A. V. (2006). The social ecology of intelligence on a Caribbean Island. *Mankind Quarterly*, 46(4), 395–433. <https://doi.org/10.46469/mq.2006.46.4.1>
- Merrill, M. A. (1938). The significance of IQ's on the revised Stanford-Binet scales. *Journal of Educational Psychology*, 29(9), 641–651. <https://doi.org/10.1037/h0057523>
- Mingroni, M. A. (2007). Resolving the IQ paradox: heterosis as a cause of the Flynn effect and other trends. *Psychological Review*, 114(3), 806–829. <https://doi.org/10.1037/0033-295x.114.3.806>
- Must, O., Must, A. & Raudik, V. (2003). The secular rise in IQs: In Estonia, the Flynn effect is not a Jensen effect. *Intelligence*, 31(5), 461–471. [https://doi.org/10.1016/s0160-2896\(03\)00013-8](https://doi.org/10.1016/s0160-2896(03)00013-8)
- Neisser, U. (1997). Rising scores on intelligence test. *American scientists*, 85, 440-447.
- Neisser, U., Boykin, A. W., Brody, N., Ceci, S. J., Loehlin, J. C., Perloff, R., Sternberg, R. J. & Urbina, S. (1996). Intelligence: Knowns and unknowns. *American Psychologist*, 51(2), 77–101. <https://doi.org/10.1037/0003-066x.51.2.77>

- Nettelbeck, T. (2014). Smarter but slower? A comment on Woodley, Te Nijenhuis & Murphy (2013). *Intelligence*, 42, 1–4. <https://doi.org/10.1016/j.intell.2013.09.006>
- Nevin, R. (2000). How lead exposure relates to temporal changes in IQ, violent crime, and unwed pregnancy. *Environmental Research*, 83(1), 1–22. <https://doi.org/10.1006/enrs.1999.4045>
- Nyborg, H. (2012). The Decay of Western Civilization: double relaxed Darwinian selection. *Personality and Individual Differences*, 53(2), 118–125. <https://doi.org/10.1016/j.paid.2011.02.031>
- Pietschnig, J. & Voracek, M. (2015). One Century of Global IQ Gains. *Perspectives on Psychological Science*, 10(3), 282–306. <https://doi.org/10.1177/1745691615577701>
- Pietschnig, J., Deimann, P., Hirschmann, N., & Kastner-Koller, U. (2021). The Flynn effect in Germanophone preschoolers (1996-2018): Small effects, erratic directions, and questionable interpretations. *Intelligence*, 86, Article 101544.
- Pietschnig, J., Gittler, G., Höttl, F., Tran, U. S., & Voracek, M. (2019). Smaller Flynn effects for crystallized intelligence may be rooted in item obsolescence: Results from archival data and a direct test of generational IQ gains. Twentieth Annual Conference of the International Society for Intelligence Research (ISIR), 13. -15.07.2018, Minneapolis, MN.
- Pietschnig, J., Oberleiter, S., Toffalini, E. & Giofrè, D. (2023). Reliability of the G factor over time in Italian INVALSI data (2010-2022): What can Achievement-G tell us about the Flynn effect? *Personality and Individual Differences*, 214, 112345. <https://doi.org/10.1016/j.paid.2023.112345>

- Pietschnig, J., Tran, U. S. & Voracek, M. (2013). Item-response Theory Modeling of IQ gains (the Flynn Effect) on crystallized intelligence: Rodgers' hypothesis Yes, Brand's hypothesis perhaps. *Intelligence*, 41(6), 791–801. <https://doi.org/10.1016/j.intell.2013.06.005>
- Pietschnig, J., Voracek, M. & Formann, A. K. (2010). Pervasiveness of the IQ Rise: A Cross-Temporal Meta-Analysis. *PLOS ONE*, 5(12), e14406. <https://doi.org/10.1371/journal.pone.0014406>
- Pietschnig, J., Voracek, M. & Formann, A. K. (2011). Female Flynn effects: No sex differences in generational IQ gains. *Personality and Individual Differences*, 50(5), 759–762. <https://doi.org/10.1016/j.paid.2010.12.019>
- Pietschnig, J., Voracek, M., & Gittler, G. (2018). Is the Flynn effect related to migration? Meta-analytic evidence for correlates of stagnation and reversal of generational IQ test score changes. *Politische Psychologie*, 6, 276-283
- Rindermann, H., Schott, T. & Baumeister, A. (2013). FLynn Effect in Turkey: A comment on Kagitcibasi and Biricik (2011). *Intelligence*, 41(3), 178–180. <https://doi.org/10.1016/j.intell.2013.02.003>
- Rushton, J. P. (1997). Racial research and final solutions. *Society*, 34(3), 78–82. <https://doi.org/10.1007/s12115-997-1011-8>
- Rushton, J. P. (1999). Secular gains in IQ are not related to the g factor and inbreeding depression—Unlike BlackWhite differences: A reply to Flynn. *Personality and Individual Differences*, 26, 381–389. doi:10.1016/S01918869(98)00148-2
- Schaie, K. W. (1994). The course of adult Intellectual Development. *American Psychologist*, 49(4), 304–313. <https://doi.org/10.1037/0003-066x.49.4.304>

Schaie, K. W., & Strother, C. R. (1968). A cross-sequential study of age changes in cognitive behaviour. *Psychological Bulletin*, 70, 651-680. Doing:10.1037/h0026811

Schneider, W. J., & McGrew, K. S. (2018). The Cattell–Horn–Carroll theory of cognitive abilities. In D. P. Flanagan & E. M. McDonough (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 73–163). The Guilford Press.

Shayer, M., & Ginsburg, D. (2009). Thirty years on—A large anti-Flynn Effect? 11: 13-year olds. Piagetian tests of formal operations norms 1976–2006/7. *British Journal of Educational Psychology*, 79, 409–418. <https://doi.org/10.1348/978185408x383123>

Shayer, M., Ginsburg, D., & Coe, R. (2007). Thirty years on - a large anti-Flynn effect? The Piagetian test *Vol. & Heaviness* norms 1975-2003. *British Journal of Educational Psychology*, 77(1), 25–41. <https://doi.org/10.1348/000709906X96987>

Sundet, J. M., Barlaug, D. G. & Torjussen, T. M. (2004). The end of the Flynn effect? A study of secular trends in mean intelligence test scores of Norwegian conscripts during half a century. *Intelligence*, 32(4), 349–362. <https://doi.org/10.1016/j.intell.2004.06.004>

te Nijenhuis, de Jong, M.-J., Evers, A., & van der Flier, H. (2004). Are cognitive differences between immigrant and majority groups diminishing? *European Journal of Personality*, 18(5), 405–434. <https://doi.org/10.1002/per.511>

te Nijenhuis, J. T. & Van Der Flier, H. (2013). Is the Flynn effect on G?: a meta-analysis. *Intelligence*, 41(6), 802–807. <https://doi.org/10.1016/j.intell.2013.03.001>

te Nijenhuis, J. T., Cho, S. H., Murphy, R. & Lee, K. H. (2012). The Flynn effect in Korea: large gains. *Personality and Individual Differences*, 53(2), 147–151. <https://doi.org/10.1016/j.paid.2011.03.022>

Teasdale, T. W. & Owen, D. R. (2008). Secular declines in cognitive test scores: a reversal of the Flynn effect. *Intelligence*, 36(2), 121–126. <https://doi.org/10.1016/j.intell.2007.01.007>

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

Tuddenham, R. D. (1948). Soldier intelligence in World Wars I and II. *American Psychologist*, 3(2), 54–56. <https://doi.org/10.1037/h0054962>

Wheeler, L. R. (1942). A comparative study of the intelligence of East Tennessee mountain children. *Journal of Educational Psychology*, 33, 321–334.

Woodley of Menie, M. A., te Nijenhuis, J., & Murphy, R. (2015). Do variable signal luminances and confounded stimuli contribute to slowing simple RT and cross study heterogeneity? A response to Parker (2014). *Intelligence*, 49, 23–24. <https://doi.org/10.1016/j.intell.2014.12.003>

Woodley, M. A. (2012). A life history model of the Lynn–Flynn effect. *Personality and Individual Differences*, 53(2), 152–156. <https://doi.org/10.1016/j.paid.2011.03.028>

Woodley, M. A., Nijenhuis, J. T. & Murphy, R. (2013). Were the Victorians cleverer than us? The decline in general intelligence estimated from a meta-analysis of the slowing of simple reaction time. *Intelligence*, 41(6), 843–850. <https://doi.org/10.1016/j.intell.2013.04.006>

Woodley, M. A., Nijenhuis, J. T. & Murphy, R. (2014). Is there a dysgenic secular trend towards slowing simple reaction time? Responding to a quartet of critical commentaries. *Intelligence*, 46, 131–147. <https://doi.org/10.1016/j.intell.2014.05.012>

Zagorsky, J. L. (2007). Do you have to be smart to be rich? The impact of IQ on wealth, income and financial distress. *Intelligence*, 35(5), 489–501.

<https://doi.org/10.1016/j.intell.2007.02.003>

Zajonc, R. B. & Mullally, P. R. (1997). Birth order: Reconciling conflicting effects. *American Psychologist*, 52(7), 685–699. <https://doi.org/10.1037/0003-066x.52.7.685>

Anhang

Zusammenfassung

Der bemerkenswerte globale Anstieg durchschnittlicher IQ-Werte über Generationen hinweg, heute als Flynn Effekt bekannt, zeigt sowohl länder- als auch domänenspezifische Muster, die sich in jüngster Zeit zunehmend differenzieren. Neben Berichten über weitere Zunahmen in einigen Ländern, zeigen sich insbesondere in skandinavischen Ländern Hinweise auf eine Stagnation bis Umkehr des Effekts. Eine Hypothese zur Erklärung dieser Veränderungen ist, dass eine zunehmende Differenzierung der kognitiven Fähigkeiten zu einer Abnahme des psychometrischen g führen könnte, was sich in einer Stagnation oder sogar einem Rückgang der IQ-Werte bemerkbar machen könnte. Um die Stabilität des psychometrischen g zu testen, wurden Intelligenztestdaten des Wilde-Intelligenztests von 204 deutschsprachigen, kognitiv besonders leistungsstarken Militärpilot:innen und Fluglots:innen des österreichischen Bundesheers ($M = 21.32$ Jahre, $SD = 3.79$ Jahre; 6 weiblich, 198 männlich) untersucht.

Die Ergebnisse von drei Analysen von 1992 bis 2016 zeigen signifikant positive Flynn Effekte zwischen 0.298 IQ-Punkten im Subtest *Buchstabenreihen* bis 2.372 IQ-Punkte im Subtest *Eingekleidete Rechenaufgaben* pro Jahr, sowie einen signifikant negativen Flynn Effekt von -0.297 IQ-Punkten im Subtest *Analogien*. Eine Hauptkomponentenanalyse zeigt keine Hinweise auf eine signifikante Abnahme von g. Sowohl die erklärten Varianzen (R^2) als auch die Reliabilitäten (McDonald's Omega ω) zeigen keine signifikante Veränderung. Die Analyse des negativen Zusammenhangs von IQ-Anstiegen und psychometrischen g zeigen darüber hinaus ebenfalls keine signifikanten Ergebnisse.

Abstract

Englisch

The remarkable global rise in average IQ scores across generations, now known as the Flynn effect, shows both country- and domain-specific patterns that have recently become increasingly differentiated. In addition to reports of further increases in some countries, reports from Scandinavian countries in particular show evidence of a stagnation to reversal of the effect. One hypothesis to explain these changes is that an increasing differentiation of cognitive abilities could lead to a decrease in psychometric g, which could be reflected in a stagnation or even a decline in IQ scores. In order to test the stability of psychometric g, intelligence test data from the Wilde Intelligence Test of 204 German-speaking, cognitively high-performing military pilots and air traffic controllers of the Austrian Armed Forces (M = 21.32 years, SD = 3.79 years; 6 female, 198 male) were examined.

The results of three analyses from 1992 to 2016 show significant positive Flynn effects ranging from 0.298 IQ points in the subtest *Buchstabenreihen* to 2.372 IQ points in the subtest *Eingekleidete Rechenaufgaben* per year, and a significant negative Flynn effect of -0.297 IQ points in the subtest *Analogien*. A principal component analysis shows no evidence of a significant decrease in g. Both the explained variances (R²) and the reliabilities (McDonald's omega ω) show no significant change in g. Furthermore, the analysis of the negative correlation between IQ increases and psychometric g also shows no significant results.

R-Code

```
library(tidyverse)
```

```
library(haven)
```

```
library(here)
```

Masterarbeit Daten

```
path <- here("data/Masterarbeit_Daten_final_ausgeschlosseneDaten.sav")
```

```
data <- read_sav(path)
```

Alle Jahre von 1992 bis 2016 erstellen

```
all_years <- seq(1992, 2016)
```

Funktion zum Füllen von fehlenden Jahren mit NA-Werten

```
fill_missing_years <- function(data_subset) {
```

```
  data_subset %>%
```

```
    complete(jahr = all_years) %>%
```

```
    replace_na(list(iq = NA))
```

```
}
```

reshape data -> make it long data (Alle Spalten mit)

```
data_filtered <- data %>%
```

```
  pivot_longer(
```

```
    cols = matches("_IQ$"),
```

```
    names_to = "test_name",
```

```
    values_to = "iq"
```

```
) %>%
```

```
filter(!str_detect(test_name, "2007|2012|MW|SO")) %>%
```

```
select(
```

```
  Jahr,
```

```
  test_name,
```

```
  iq
```

```
) %>%
```

```
rename(
```

```
  jahr = Jahr
```

```
) %>%
```

```
filter(!is.na(iq))
```

```
test_periods <- c(
```

```
  "max_year" = 2007,
```

```
  "max_year" = 2012,
```

```
  "max_year" = 2016
```

```
)
```

```
data_per_period <- map(seq_along(test_periods), function(i){
```

```
  max_year <- test_periods[[i]]
```

```
  data_max_year <- data_filtered %>%
```

```

filter(jahr <= max_year) %>%

group_by(test_name) %>%

filter(max(jahr) == max_year) %>%

fill_missing_years()

data_max_year[["max_year"]] <- paste0("max_", max_year)

return(data_max_year)

}) %>% setNames(test_periods) %>% bind_rows()

```

compute one lm for each test and time period

```

d <- data_per_period %>%

group_by(test_name, max_year) %>%

nest() %>%

mutate(

lm = map(data, ~lm(iq~jahr, data = .x)),

p_value = map_dbl(lm, ~summary(.x)$coefficients[2,4]),

slope = map_dbl(lm, ~.x$coefficients[[2]])

)

d %>%

select(-lm) -> d_cleaned

```

```
##### plot
```

```
order_ <- c("AW_IQ", "AL_IQ", "BO_IQ","BR_IQ","ER_IQ", "GDT_IQ", "GW_IQ",
"SCH_IQ", "SW_IQ","WG_IQ", "ZM_IQ","ZN_IQ")
```

```
data_ordered <- data_per_period %>%
```

```
  mutate(
```

```
    test_name = as.factor(test_name),
```

```
    test_name = fct_relevel(test_name, order_)
```

```
  ) %>%
```

```
  mutate(test_name = recode(test_name,
```

```
    "AW_IQ" = "Abwicklungen",
```

```
    "AL_IQ" = "Analogien",
```

```
    "BO_IQ" = "Beobachtungen",
```

```
    "BR_IQ" = "Buchstabenreihen",
```

```
    "ER_IQ" = "Eingekleidete Rechenaufgaben",
```

```
    "GDT_IQ" = "Gedächtnis",
```

```
    "GW_IQ" = "Gleiche Wortbedeutung",
```

```
    "SCH_IQ" = "Schätzen",
```

```
    "SW_IQ" = "Sprichwörter",
```

```
    "WG_IQ" = "Wortgewandtheit",
```

```
    "ZM_IQ" = "Zahlenmerken",
```

"ZN_IQ" = "Zahlenreihen"

))

data_ordered %>%

ggplot(aes(jahr, iq, color = max_year)) +

geom_point(color = "#000000", alpha = .1) +

geom_smooth(method = "lm", se = TRUE, size = 0.3) + #Standard Errors

scale_color_manual(

values = c(max_2007 = "firebrick", max_2012 = "darkgreen", max_2016 =
"cornflowerblue"),

labels = c(max_2007 = "Analyse 12 Subtests", max_2012 = "Analyse 10 Subtests",
max_2016 = "Analyse 3 Subtests")

) +

guides(color = guide_legend(title = "Analysen", title.position = "top", hjust = .5)) +

labs(

x = "Jahr der Testung",

y = "IQ-Wert",

title = "IQ-Wert",

subtitle = "nach Subtest und Testperiode"

) +

scale_x_continuous(breaks = seq(1992, 2016, by = 1), labels = seq(1992, 2016, by = 1)) +

scale_y_continuous(breaks = seq(0, 150, by = 25), name = "IQ-Wert") +

theme_minimal() +

```
theme(  
  legend.position = "bottom",  
  axis.text.x = element_text(angle = 90, vjust = 0.5, hjust = 1),  
  axis.title.y = element_text(margin = margin(t = 20)), # Änderung hier  
  panel.spacing = unit(1, "lines"),  
  axis.title.x = element_text(margin = margin(t = 20)) # Änderung hier  
) +  
  
facet_wrap(~ test_name, ncol = 2, scales = "fixed") -> p  
  
ggsave("~/Desktop/plot.png", p, height = 18)
```