



MASTERARBEIT | MASTER'S THESIS

Titel | Title

Der Flynn Effekt und das psychometrische g: Eine
Querschnittsstudie des Intelligenz-Struktur-Tests (IST)

verfasst von | submitted by
Melanie Hummer 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:

Assoz. Prof. Mag. Dr. Jakob Pietschnig Privatdoz.

Danksagung

An dieser Stelle möchte ich allen herzlich danken, die mich während der Entstehung dieser Arbeit und meines gesamten Studiums unterstützt haben.

Mein besonderer Dank gilt Ass.-Prof. Mag. Dr. Jakob Pietschnig für die Betreuung meiner Masterarbeit. Er stand mir stets mit Rat und Tat zur Seite, bot wertvolle Unterstützung bei allen Fragen und Unsicherheiten und gab mir konstruktives Feedback, das zur Verbesserung meiner Arbeit beigetragen hat.

Ein ebenso großer Dank geht an meine Familie, die mich während meines Studiums immer gestärkt und begleitet hat. Ohne ihre Unterstützung wäre dieser Weg nicht möglich gewesen.

Mein besonderer Dank gilt auch meinem Mann Thomas, der meine Arbeit sorgfältig Korrektur gelesen und mich in vielerlei Hinsicht begleitet hat. Sein emotionaler und motivierender Beistand war für mich während des gesamten Studiums von unschätzbarem Wert.

Nicht zuletzt danke ich allen Teilnehmer*innen meiner Studie bedanken, die sich die Zeit genommen haben, an meiner Untersuchung mitzuwirken. Ihr Beitrag war für den Erfolg dieser Arbeit unerlässlich.

Einleitung

Die Intelligenz steht in enger Verbindung mit positiven Auswirkungen in verschiedenen Lebensbereichen. Ein höherer Intelligenzquotient (IQ) wird häufig mit beruflichem und akademischem Erfolg, besserer Bildung (Lang & Kell, 2020), höherem Einkommen und höheren sozioökonomischen Status assoziiert (Strenze, 2007). Neben diesen beruflichen und bildungsbezogenen Vorteilen bestehen auch Zusammenhänge zwischen dem IQ und verschiedenen Gesundheitsaspekten. Dazu zählen unter anderem eine bessere allgemeine Gesundheit und eine höhere Lebenserwartung (Deary et al., 2021). Welche Auswirkungen Intelligenz auf gesellschaftlicher Ebene hat, lässt sich indirekt anhand der PISA-Studie beobachten. Obwohl die PISA-Studie schulische Leistung (school achievement) misst und nicht direkt Intelligenz erfasst, gilt die schulische Leistung als guter Indikator für Intelligenz (z. B. Kriegbaum et al., 2018; Pokropek et al., 2022). Die Ergebnisse solcher Vergleichsstudien von Schüler*innen beeinflussen maßgeblich bildungspolitische Entscheidungen (Pietschnig et al., 2018). Darüber hinaus werden kognitive Tests unter anderem bei der Personalauswahl, bei Universitätszulassungen, in der klinischen Diagnostik sowie zur Feststellung von Lernschwierigkeiten oder Förderbedarf eingesetzt.

Dies verdeutlicht das breite Spektrum an Anwendungsmöglichkeiten solcher Tests und unterstreicht die Notwendigkeit, dass die Normierungen kognitiver Tests stets aktuell gehalten werden sollten. In diesem Zusammenhang spielt der Flynn Effekt eine zentrale Rolle, ein Phänomen, das den Anstieg der durchschnittlichen IQ-Werte in der Allgemeinbevölkerung beschreibt (Flynn, 1984, 2007). Weltweit stiegen die IQ-Werte zwischen 1909 und 2013 durchschnittlich um etwa drei IQ-Punkte pro Dekade (Pietschnig & Voracek, 2015). Diese Zuwächse verlaufen jedoch nicht linear, es gab Phasen mit stärkeren Zuwächsen und solche mit geringeren. Darüber hinaus variiert der Flynn Effekt je nach Land und Intelligenzbereich (Pietschnig & Voracek, 2015).

Veraltete Normen können dazu führen, dass die kognitiven Fähigkeiten von Personen überschätzt werden oder potenzielle Defizite unentdeckt bleiben. Zum Beispiel führt ein Test, der 1960 normiert wurde, im Vergleich zu einem Test auf Basis einer Stichprobe aus dem Jahr 2000 zu deutlich besseren Ergebnissen. Eine Gruppe von Proband*innen, die in einem aktuellen Test einen IQ von 100 erzielte, hätte in einem älteren Test einen IQ von 112 erreicht (Flynn, 2019). Die Verwendung veralteter Normen kann in der Diagnostik weitreichende Folgen haben. Eine Intelligenzminderung könnte unentdeckt bleiben und damit eine notwendige Förderung verwehrt werden. Häufig wird ein Grenzwert von 70 IQ-Punkten als Indikator für eine Einschränkung der kognitiven Fähigkeiten herangezogen. Wenn eine Person einen Test

durchführt, dessen Normierung 20 Jahre zurückliegt und der Flynn Effekt den IQ um 0.3 Punkte pro Jahr erhöht, würde diese Person einen IQ von 73 erzielen. Bei aktuellen Normen würde sich dieser Wert auf 67 reduzieren, was auf eine Einschränkung der kognitiven Fähigkeiten hinweist (Trahan et al., 2014). In seltenen Fällen könnten veraltete Normen jedoch auch dazu führen, dass Personen fälschlicherweise Zugang zu Förderprogrammen für Hochbegabte erhalten (Flynn, 2019). Dies hat weitreichende Implikationen für die Normierung und Anwendung kognitiver Tests in verschiedenen Bereichen und unterstreicht die Notwendigkeit, stets aktuelle Normen zu verwenden.

Der Flynn Effekt

Bereits im Jahr 1936 gab es erste Anzeichen für den Flynn Effekt in Amerika. In einer Studie von Runquist (1936) wurden Steigerungen in einem Leistungstest von 1929 auf 1932 beobachtet. Diese Beobachtung legte nahe, dass die durchschnittliche Leistungsfähigkeit gestiegen war. Später analysierte Tuddenham (1948) US-Militärdaten aus dem Ersten und Zweiten Weltkrieg und stellte fest, dass die Rohwerte der Testergebnisse der Soldaten des Zweiten Weltkriegs deutlich höher waren als die der Soldaten des Ersten Weltkriegs. Trotz dieser frühen Erkenntnisse wurden die steigenden Werte in den Testergebnissen in der Mitte des 20. Jahrhunderts als Stichprobenfehler oder statistische Artefakte interpretiert (z. B. Merrill, 1938; Tuddenham, 1948).

Schaie und Strother (1968) interpretierten erstmals die beobachteten IQ-Zuwächse als Kohorteneffekte. Diese Erkenntnis deutete darauf hin, dass die Veränderungen in den Ergebnissen möglicherweise systematischen Einflüssen unterliegen. In den 1980er Jahren begann James R. Flynn systematisch IQ-Veränderungsmuster zu untersuchen und leistete damit einen wesentlichen Beitrag zur Erforschung dieses Phänomens. Seine Arbeit dokumentierte Zuwächse der IQ-Werte in den USA zwischen 1932 und 1978, basierend auf Daten verschiedener bekannter Intelligenztests. Flynn stellte fest, dass die durchschnittlichen IQ-Werte jährlich um etwa 0.3 Punkte anstiegen, was zu einem Gesamtzuwachs von 13.8 Punkten führte (Flynn, 1984).

Diese Ergebnisse regten die wissenschaftliche Gemeinschaft dazu an, den Flynn Effekt intensiver zu erforschen und führten zu weitreichenden Diskussionen über die Bedeutung, Ursachen und Konsequenzen dieser IQ-Steigerungen. Den Begriff Flynn Effekt prägten Herrnstein und Murray (1994) in ihrem kontroversen Buch „The Bell Curve“. Seitdem hat sich der Flynn Effekt als bedeutendes Phänomen etabliert, das die Veränderungen des IQ über Generationen hinweg beschreibt (Flynn, 1984, 2007).

Länderspezifischer Flynn Effekt

Der Flynn Effekt wurde nicht nur in den Vereinigten Staaten beobachtet, sondern auch in vielen anderen Ländern weltweit (z. B. Kenia: Daley et al., 2003; China: Liu & Lynn, 2013). Diese globale Beobachtung verdeutlicht, dass der Anstieg der durchschnittlichen IQ-Werte eine weit verbreitete Entwicklung darstellt. Allerdings zeigt der Flynn Effekt unterschiedliche Ausprägungen und Muster in verschiedenen Ländern, Intelligenzbereichen und Zeiträumen (Pietschnig & Voracek, 2015), was darauf hindeutet, dass dieser Effekt nicht einheitlich verläuft.

In einer umfassenden Metaanalyse von Pietschnig und Voracek (2015), die den Flynn Effekt in verschiedenen Ländern im Zeitraum von 1909 bis 2013 untersuchten, wurde ein durchschnittlicher Anstieg des IQs um 2.8 Punkte pro Dekade festgestellt. Diese Zunahmen waren weitgehend konsistent, obwohl die Stärke der Zuwächse je nach Land unterschiedlich war. Einige Länder wiesen stärkere Anstiege auf, darunter beispielsweise Japan, Kenia, Deutschland, Österreich, Spanien und die Niederlande (Pietschnig & Voracek, 2015). Andere Länder wie Australien, Brasilien, das Vereinigte Königreich, die USA und Irland verzeichneten hingegen geringere Zuwächse (Pietschnig & Voracek, 2015).

Seit 1995 wurde in skandinavischen Ländern eine Stagnation oder sogar ein Rückgang der durchschnittlichen IQ-Werte beobachtet, ein Phänomen, das als negativer Flynn Effekt bezeichnet wird (Flynn & Shayer, 2018). Ähnliche Entwicklungen wurden auch in anderen europäischen Ländern, wie etwa in Frankreich festgestellt (Dutton & Lynn, 2015). Die Ergebnisse von Dutton und Lynn (2015) sollten jedoch mit Vorsicht interpretiert werden. Gonthier et al. (2021) fanden heraus, dass der beobachtete Rückgang möglicherweise auf ein Differential Item Functioning (DIF) zurückzuführen ist. DIF bedeutet, dass eine Gruppe von Teilnehmer*innen schlechter abschneidet, nicht weil ihre Fähigkeiten geringer sind, sondern weil die Fragen (Items) für diese Gruppe schwieriger geworden sind (z. B. Zumbo, 2007). Etwa ein Viertel der Items war laut Gonthier et al. (2021) für die aktuellere Stichprobe schwieriger als in der ursprüngliche Normgruppe, und zwar bei gleichem Fähigkeitsniveau. Daher gibt es laut den Autor*innen keinen Hinweis auf einen tatsächlichen Rückgang der kognitiven Fähigkeiten in Frankreich (Gonthier et al., 2021).

Aktuelle Hinweise aus deutschsprachigen Ländern deuten darauf hin, dass die visuell-räumlichen Fähigkeiten in bestimmten Regionen Europas abnehmen (Pietschnig & Gittler, 2015). Zudem gibt es erste Anzeichen für eine Stagnation oder Umkehrung des Flynn Effekts in Nordamerika (Dworak et al., 2023). Obwohl die Evidenz für eine Umkehrung des Flynn Effekts bisher nicht eindeutig ist, deuten erste Anzeichen auf eine solche Entwicklung hin.

Domänenspezifischer Flynn Effekt

Der domänenspezifische Flynn Effekt beschreibt die Beobachtung, dass sich Veränderungen der IQ-Werte im Laufe der Zeit nicht gleichermaßen auf alle kognitiven Bereiche verteilen. Bestimmte kognitive Domänen sind stärker oder schwächer vom Flynn Effekt betroffen. Diese Bereiche werden oft in fluide und kristalline Intelligenz unterteilt (Cattell, 1963). Die fluide Intelligenz bezieht sich auf logisches Denken oder Problemlösen, ohne dass spezifisches Vorwissen benötigt wird. Oft wird diese mit Matrizen tests oder dem räumlichen Denken in Verbindung gebracht. Im Gegensatz dazu umfasst die kristalline Intelligenz bereits erlerntes Wissen, das durch Erfahrung und Bildung erworben wird (Cattell, 1963). Diese wird oft in Vokabeltests oder wissensbasierten Fragen gemessen. Über mehrere Studien hinweg wurde beobachtet, dass die fluide Intelligenz im Vergleich zur kristallinen Intelligenz stärkere Zuwächse aufweist (z. B. Nisbett et al., 2012; Pietschnig & Voracek, 2015). Eine Metaanalyse von Pietschnig und Voracek (2015) ergab, dass der Flynn Effekt in der fluiden Intelligenz um 4.1 IQ-Punkte pro Dekade zunahm, während der Anstieg in der kristallinen Intelligenz bei 2.1 Punkten pro Dekade lag. Diese Unterschiede könnten möglicherweise auf einen domänenspezifischen Flynn Effekt hinweisen.

Es ist auch denkbar, dass die Aufgaben, also die Items, heute schwieriger zu beantworten sind als früher, da sie möglicherweise veraltetes Wissen abfragen (Gonthier & Grégoire, 2022). Ein Beispiel hierfür ist die ältere Version des Intelligenz-Struktur-Tests, in der in einem Subtest abgefragt wird, welches der wichtigste Bestandteil eines Fernsehers ist, damals war dies die Bildröhre (Amthauer, 1970). Heute ist dieses Wissen nicht mehr relevant, was dazu führen kann, dass die Items für heutige Testteilnehmer*innen aufgrund von Differential Item Functioning (DIF) schwieriger zu beantworten sind (Gonthier & Grégoire, 2022). Dies könnte die Ergebnisse des Flynn Effekts in Bezug auf die kristalline Intelligenz beeinflussen.

Darüber hinaus haben neuere Studien auch einen Rückgang der fluiden Intelligenz festgestellt, was auf Veränderungen in den broad abilities (Stratum II; Schneider & McGrew, 2018) des CHC-Modells hinweist. Oberleiter, Patzl, et al. (2024) fanden im Zeitraum von 2012 bis 2022 bei deutschen Schüler*innen der Sekundarstufe einen Rückgang im figuralen Denken. Dabei zeigte sich, dass diese Abnahmen nicht linear verliefen (Oberleiter, Patzl, et al., 2024), ähnlich wie die Zunahmen im 20. Jahrhundert (Pietschnig & Voracek, 2015). Die Ergebnisse weiterer Studien deuten ebenfalls auf Rückgänge in diesem Bereich hin (Bratsberg & Rogeberg, 2018; Dworak et al., 2023). Diese Ergebnisse sind besonders interessant, da das figurale Denken einen Bereich der fluiden Intelligenz darstellt, in dem typischerweise über die Zeit ein starker positiver Flynn Effekt beobachtet wurde (Oberleiter, Patzl, et al., 2024).

Der domänenspezifische Flynn Effekt thematisiert häufig die Differenzierung zwischen kristalliner und fluider Intelligenz, Lazaridis et al. (2022) gehen jedoch über diese Unterscheidung hinaus. Die Autor*innen untersuchten Veränderungen in den Testergebnissen im Kontext des Cattell-Horn-Carroll Modells (CHC) (Schneider & McGrew, 2018) anhand einer deutschsprachigen Stichprobe im Zeitraum von 1996 bis 2018. Lazaridis et al. (2022) fokussierten sich auf die Stratum II Domänen und wie sich der Flynn Effekt in den verschiedenen Bereichen manifestiert. Es zeigte sich ein positiver Flynn Effekt im fachspezifischen Wissen (Gkn), in der Effizienz des Lernens (Gl) und im Verständniswissen (Gc). Hinsichtlich der Verarbeitungsgeschwindigkeit (Gs) und dem Lesen und Schreiben (Grw) traten keine Veränderungen auf. Ein negativer Flynn Effekt wurde bei der Arbeitsgedächtniskapazität (Gwm) verzeichnet. Mehrdeutige Veränderungen ergaben sich beim schlussfolgernden Denken (Gf), bei der Reaktions- und Entscheidungsgeschwindigkeit (Gt), beim quantitativen Denken (Gq) und bei der visuellen Verarbeitung (Gv) (Lazaridis et al., 2022). Vor allem beim schlussfolgernden Denken zeigt sich typischerweise ein starker Flynn Effekt, was in dieser Studie nicht der Fall war. Nach Lazaridis et al. (2022) deuten diese Ergebnisse auf bisher unbekannte Eigenschaften des Flynn Effekts hin, da häufig nur zwischen kristalliner und fluider Intelligenz unterschieden wird. Dies legt nahe, dass über diese beiden Intelligenzdomänen hinaus wichtige Details möglicherweise übersehen wurden, da die domänenspezifische IQ-Entwicklung komplexer sein könnte als bisher angenommen (Oberleiter, Patzl, et al., 2024). Die Befunde dieser Studie zeigten zudem, dass der Flynn Effekt nicht einheitlich über die verschiedenen Intelligenzbereiche hinweg verläuft (Lazaridis et al., 2022). Während im 20. Jahrhundert vorwiegend Zuwächse in der fluiden Intelligenz festgestellt wurden (z. B. Pietschnig & Voracek, 2015), zeigten die Ergebnisse von Lazaridis et al. (2022) ambivalente Ergebnisse.

Alters- und geschlechtsspezifischer Flynn Effekt

Die Veränderungen der IQ-Werte im Laufe der Zeit sind nicht nur domänenspezifisch, sondern können auch je nach Alter variieren. In der Metaanalyse von Pietschnig und Voracek (2015) wurden stärkere Zuwächse für Erwachsene im Vergleich zu Kindern gefunden. Die höheren IQ-Zuwächse im Alter werden häufig einer besseren Bildung zugeschrieben. Unter dieser Annahme sollten sich diese Effekte besonders in der kristallinen Intelligenz zeigen (Flynn, 2010). Allerdings wurden deutliche und starke Effekte für die fluide und räumliche Intelligenz beobachtet, während der Einfluss des Alters auf die Veränderungen in der kristallinen Intelligenz als vernachlässigbar eingestuft wurde (Pietschnig & Voracek, 2015).

Zudem zeigten Personen mit geringeren Fähigkeiten, insbesondere in der unteren Hälfte der Verteilung, größere IQ-Zuwächse als diejenigen im oberen Bereich (Oberleiter, Fries, et al., 2024; Pietschnig et al., 2013). Dies unterstützt die Idee von Rodgers (1998), dass sich die Fähigkeitsverteilungen verengen.

Eine weitere Frage bei der Untersuchung des Flynn Effekts ist, ob das Geschlecht einen signifikanten Einfluss auf den Flynn Effekt hat. Studien haben gezeigt, dass es zwar Unterschiede zwischen den Geschlechtern in den IQ-Werten der einzelnen Subtests gibt. Jedoch scheint das Geschlecht keinen Einfluss auf den Flynn Effekt zu haben (Dutton et al., 2016; Pietschnig et al., 2011).

Das psychometrische g

Um die vielseitigen Aspekte des Flynn Effekts besser zu verstehen, ist es auch wichtig, das Konzept des psychometrischen g zu berücksichtigen. Das psychometrische g (auch g-Faktor genannt) bezieht sich auf die allgemeinen kognitiven Fähigkeiten einer Person. Es basiert auf der Beobachtung, dass die Ergebnisse in den verschiedenen kognitiven Tests oder Subtests in der Regel positiv miteinander korrelieren. Das bedeutet, dass eine Person, die in einem Subtest gut abschneidet, tendenziell auch in anderen Subtests gute Leistungen erbringt. Diese positive Kupplung wird als psychometrisches g angesehen (Pietschnig & Voracek, 2015).

Als Erstes unterschied Charles Spearman in seinem *Zwei-Faktoren-Modell* zwischen der allgemeinen Intelligenz und spezifischen Fähigkeiten (Spearman, 2010). Obwohl dieses Modell einen wichtigen Grundstein in der Intelligenzforschung legte, gilt es mittlerweile als veraltet. Heute findet das umfassende und integrative *Cattell-Horn-Carroll-Modell* breite Zustimmung (Schneider & McGrew, 2018). Dieses Modell integriert die Drei-Stratum-Theorie der kognitiven Fähigkeiten (Carroll, 1993), sowie die Unterscheidung der fluiden und kristallinen Intelligenz von Horn und Cattell (1966). Das CHC-Modell besteht aus drei Dimensionen, die eine hierarchische Struktur aufweisen. Stratum III bildet den allgemeinen Faktor der Intelligenz ab (psychometrisches g), der oft mehr als die Hälfte der Gesamtvarianz ausmacht (Deary, 2012; Schneider & McGrew, 2018). Stratum II umfasst die „broad abilities“, wie beispielsweise fluide und kristalline Intelligenz, visuelle Verarbeitung und das Kurzzeitgedächtnis. Auf der untersten Ebene, Stratum I, befinden sich die „narrow abilities“. Dies sind spezifische Fähigkeiten, die durch unterschiedliche Tests und Aufgaben gemessen werden können (Schneider & McGrew, 2018). Diese Struktur, insbesondere bezogen auf das Stratum III, schafft einen Rahmen, um die Beziehung zwischen den allgemeinen kognitiven Fähigkeiten (psychometrisches g) und dem Flynn Effekt genauer zu untersuchen.

Zusammenhang psychometrisches g und Flynn Effekt

Die Zuwächse im IQ im Laufe der Zeit werden in den meisten Studien negativ mit dem psychometrischem g in Verbindung gebracht (Must et al., 2003; Pietschnig & Voracek, 2015; Te Nijenhuis, 2013; Woodley et al., 2013; nicht bei Colom et al., 2001). Dies deutet darauf hin, dass Tests mit geringer g-Ladung tendenziell höhere Flynn Effekte zeigen als Tests mit hoher g-Ladung. Eine hohe g-Ladung eines Tests oder Subtests bedeutet, dass dieser eine gute Vorhersagekraft für die allgemeinen kognitiven Fähigkeiten hat (Flynn, 2018).

Verschiedene IQ-Bereiche weisen unterschiedliche Beziehungen zum psychometrischen g auf, dies bedeutet, dass Subtests unterschiedlich stark auf g laden. Subtests mit höheren g-Ladungen zeigen tendenziell einen geringeren Flynn Effekt und umgekehrt (Pietschnig & Voracek, 2015; Te Nijenhuis, 2013; Woodley et al., 2014). Eine Untersuchung von Te Nijenhuis & Van Der Flier (2013) ergab, dass die Korrelation zwischen dem Flynn Effekt und den g-Ladungen der Subtests bei $-.38$ liegt. Es wurde ein stärkerer Anstieg der fluiden Intelligenz, die weniger hoch auf dem psychometrischem g lädt, im Vergleich zur kristallinen Intelligenz, die stärker auf g lädt, beobachtet (Pietschnig & Voracek, 2015). Das bedeutet, dass fluide Aufgabengruppen tendenziell einen höheren Flynn Effekt zeigen, aber eine geringere Vorhersagekraft für allgemeine kognitive Fähigkeiten besitzen. Es ist daher möglich, dass Intelligenzbereiche mit niedrigen g-Ladungen zunehmen, während das psychometrische g stabil bleibt oder sogar abnimmt (Pietschnig, 2016).

Mögliche Ursachen für den Flynn Effekt

In der Forschung besteht noch Uneinigkeit darüber, welche Theorie den Flynn Effekt am besten erklärt. Verschiedene Erklärungsansätze werden in Betracht bezogen, die den Flynn Effekt aus unterschiedlichen Perspektiven beleuchten. Im Folgenden werden mögliche Erklärungen für den Flynn Effekt in umweltbedingte, biologische und hybride Faktoren für IQ-Zuwächse unterteilt (Pietschnig & Voracek, 2015).

Umweltfaktoren

Bildung. Der Flynn Effekt wird häufig durch die Zunahme der Bildungsjahre und verbesserte Bildungssysteme, insbesondere in Industrieländern, erklärt (z. B. Teasdale & Owen, 2005). Ein wesentlicher Faktor hierbei ist die kristalline Intelligenz, da diese eng mit dem Bildungsniveau verknüpft ist. Höhere Bildungsabschlüsse werden oft mit einer besseren kristallinen Intelligenz in Verbindung gebracht (z. B. Kaufman et al., 2009; Rehnberg et al., 2024). Personen mit höheren Bildungsabschlüssen zeigen in der Regel bessere Leistungen bei

Aufgaben, die kristalline Komponenten wie Wortschatz- oder Allgemeinwissensfragen beinhalten. Allerdings erklärt Bildung allein nicht die teilweise sehr großen Zuwächse in der fluiden Intelligenz, die nicht auf erworbenes Wissen zurückzuführen sind. Die Ergebnisse einer Studie zeigten, dass die IQ-Zuwächse auch nach Kontrolle der Variable Bildung erheblich blieben (Pietschnig et al., 2013). Dies deutet darauf hin, dass zusätzliche Faktoren eine Rolle spielen müssen.

Interessanterweise wurden größere IQ-Zuwächse bei Erwachsenen beobachtet, was die Bedeutung der Bildung als Erklärungsansatz unterstützt. Die steigende Anzahl an Bildungsjahren könnte die höheren IQ-Werte in dieser Gruppe erklären. Da jedoch auch bei jüngeren Personen IQ-Zuwächse festgestellt wurden, kann Bildung den Flynn Effekt nicht vollständig erklären (Pietschnig & Voracek, 2015). Diese Ergebnisse legen nahe, dass Bildung zwar einen Einfluss auf den Flynn Effekt hat, aber nicht die alleinige Ursache ist. Besonders bei der fluiden Intelligenz müssen andere Faktoren eine entscheidende Rolle spielen, da deren Zuwächse nicht durch eine bessere Bildung erklärt werden können (Pietschnig & Voracek, 2015).

Weitere Umweltfaktoren. Eine weitere mögliche Erklärung für den Flynn Effekt könnte die zunehmende Exposition gegenüber moderner Technologie in den letzten Jahrzehnten sein. Es wird vermutet, dass der verstärkte Umgang mit technologischen Geräten und Anwendungen zu Trainingseffekten führen könnte, die die fluide Intelligenz fördern (Clark et al., 2016). Die regelmäßige Nutzung von Computern, Smartphones und anderen Geräten könnte Fähigkeiten wie Problemlösen und Informationsverarbeitung verbessern. Jedoch stehen die beobachteten IQ-Zuwächse in nicht industrialisierten Ländern, die weniger vom technologischen Fortschritt betroffen sind (Kenia: Daley et al., 2003), im Widerspruch zu dieser Erklärung. Zudem können die Rückgänge im IQ nicht durch technologische Entwicklungen erklärt werden, da diese Entwicklungen im Widerspruch zum allgemeinen technologischen Fortschritt stehen. Ebenso lassen sich die frühen Anstiege im IQ nicht durch den Einfluss der Technologie erklären, da zu dieser Zeit die Nutzung technologischer Geräte noch begrenzt war. Insgesamt gibt es nur wenig Evidenz für den Einfluss von Technologie auf den Flynn Effekt (Pietschnig & Voracek, 2015).

Das Testverhalten, vor allem das Raten, wurde als potenzieller Umweltfaktor für die Erklärung des Flynn Effekts diskutiert (z. B. Brand et al., 1989). Die zugrundeliegende Argumentation besagt, dass der vermehrte Einsatz von Multiple-Choice-Antwortformaten in den psychometrischen Tests das Rateverhalten begünstigt. Da Multiple-Choice-Formate besonders in Tests zur fluiden Intelligenz verwendet werden, wird angenommen, dass ein

erhöhtes Raten zu IQ-Zuwächsen beitragen könnte (Pietschnig & Voracek, 2015). Jedoch gibt es Einschränkungen dieser Erklärung. Erstens sind die Zuwächse in der fluiden Intelligenz nicht vollständig durch Raten erklärbar. Zweitens zeigt sich der Flynn Effekt auch in anderen Testformaten, einschließlich solcher, die die kristalline Intelligenz messen, bei denen Wissen eine zentrale Rolle spielt (Pietschnig & Voracek, 2015).

Weiters wurde in westlichen Ländern eine abnehmende Familiengröße mit einer Steigerung der kognitiven Fähigkeiten in Verbindung gebracht (Zajonc & Mullally, 1997). Allerdings ist diese Erklärung eher unwahrscheinlich, da der Zusammenhang nicht eindeutig nachgewiesen werden konnte und die Evidenz insgesamt begrenzt ist (Pietschnig & Voracek, 2015; Wichman et al., 2006).

Biologische Erklärungen

Als möglicher biologischer Erklärungsansatz wird das Konzept des Hybrid Vigor (Heterosis) betrachtet. Dieser Ansatz beschreibt die Fortpflanzung von genetisch unähnlichen Individuen, was zu einer erhöhten Allel-Heterozygotie und einer verringerten Homozygotie führt (Pietschnig & Voracek, 2015). Das bedeutet, dass die IQ-Zuwächse durch die zunehmende Mobilität von Individuen erklärt werden. Diese Mobilität führte dazu, dass Ehen unter Verwandten sowie Ehen innerhalb derselben Gruppe seltener wurden (Mingroni, 2007). Jedoch scheint die Relevanz dieser biologischen Erklärung begrenzt, da sie das hohe Ausmaß und die hohe Geschwindigkeit der beobachteten IQ-Zuwächse nicht vollständig berücksichtigt. Hybrid Vigor allein kann die Steigerung der kognitiven Fähigkeiten nicht ausreichend erklären (Pietschnig & Voracek, 2015).

Hybride Faktoren

Ernährung. Ein weiterer möglicher Erklärungsansatz für den Flynn Effekt ist die Rolle der Ernährung, insbesondere der perinatalen Ernährung. Eine verbesserte Ernährung kann zu zahlreichen positiven Effekten führen, einschließlich einer möglichen Steigerung der kognitiven Fähigkeiten. Schlechte Ernährung wird mit niedrigen Leistungen bei IQ-Tests in Verbindung gebracht (z. B. Hunt, 2011). Eine Studie von Focosi Mazzei et al. (2019) aus Brasilien fand eine signifikante negative Korrelation zwischen Mangelernährung und den kognitiven Fähigkeiten von Schüler*innen. In Regionen mit hoher Prävalenz von Mangelernährung entwickelten sich die kognitiven Fähigkeiten von Kindern oft nicht vollständig. Die Forscher*innen stellten fest, dass Personen in solchen Regionen mit niedriger Intelligenz tendenziell weniger Zugang zu Nahrungsmitteln haben, die für eine gesunde

Ernährung empfohlen werden (Focosi Mazzei et al., 2019). Vor allem die Leistungen im Bereich der fluiden Intelligenz sind von unzureichender Ernährung betroffen (Lynn, 2009).

Dennoch erscheint es unplausibel, Ernährung als die einzige Ursache für den Flynn Effekt anzusehen. Während die IQ-Zuwächse bei Kindern diese Erklärung unterstützen, sprechen die stärkeren Zuwächse bei Erwachsenen für zusätzliche Faktoren, die erst später im Leben wirksam werden (Pietschnig & Voracek, 2015). Dies legt nahe, dass neben der Ernährung noch weitere Ursachen in Betracht gezogen werden müssen, um den Flynn Effekt umfassend zu verstehen.

Soziale Multiplikatoren. Die Theorie der sozialen Multiplikatoren von Dickens und Flynn aus dem Jahre 2001, besagt, dass IQ-Zuwächse durch eine Wechselwirkung zwischen genetischen und Umweltfaktoren erklärt werden können. Bereits geringfügige Umweltvorteile können die individuelle Leistung verbessern. Diese gesteigerte Leistung trägt zu noch vorteilhafteren Umwelten bei, die wiederum die Leistung weiter erhöhen (Dickens & Flynn, 2001). Das heißt, wenn Menschen in einer Umgebung leben, die kleine Vorteile bietet (z. B. besserer Zugang zu Bildung) kann dies ihre Leistung verbessern. Diese Verbesserung führt zu einem Rückkopplungseffekt, die erhöhte Leistung schafft noch günstigere Umweltbedingungen, was die Entwicklung weiter fördert. Dies führt zu einem Multiplikatoreffekt, da diese Steigerungen die Umwelt positiv beeinflussen und sich die Vorteile weiterverbreiten (Dickens & Flynn, 2001). Auf diese Weise können kleine anfängliche Umweltvorteile durch einen Multiplikatoreffekt große, weitreichende Auswirkungen auf die allgemeine Leistungsfähigkeit und Intelligenz in einer Gesellschaft haben (Dickens & Flynn, 2001; Pietschnig & Voracek, 2015). Diese Theorie erklärt, wie die Steigerungen der IQ-Werte zustande kommen könnten. Insbesondere die Zuwächse bei der fluiden Intelligenz, die weniger stark vom psychometrischen g beeinflusst werden, lassen sich durch die Theorie gut nachvollziehen (Pietschnig & Voracek, 2015).

Allerdings ist die Theorie schwer empirisch zu belegen und es gibt daher nur wenig unterstützende Evidenz für ihre Gültigkeit. Darüber hinaus erklärt dieses Modell vermutlich nur einen Teil der beobachteten IQ-Zuwächse, da es andere Einflussfaktoren unberücksichtigt lässt (Pietschnig & Voracek, 2015).

Pathogener Stress. Durch verbesserte medizinische Versorgung, bessere hygienische Bedingungen und eine geringere Belastung durch Krankheitserreger könnten insbesondere die Zuwächse in der fluiden Intelligenz erklärt werden. Diese verbesserten Rahmenbedingungen

ermöglichen es Kindern, ihre kognitiven Ressourcen stärker für die geistige Entwicklung zu nutzen, anstatt sie für die Bekämpfung von Krankheitserregern aufzuwenden (Pietschnig & Voracek, 2015).

Allerdings sollten die IQ-Zuwächse bereits in der frühen Kindheit sichtbar sein. Diese Theorie allein kann die stärkeren Zuwächse im Erwachsenenalter nur schwer erklären, daher sollten weitere Faktoren als Ursachen für den Flynn Effekt in Betracht gezogen werden (Pietschnig & Voracek, 2015).

Life history speed. Das Life history speed Modell (Woodley, 2012) integriert mehrere Faktoren, die als Ursachen für den Flynn Effekt in Frage kommen (z. B. Bildung, Ernährung, pathogener Stress). Personen mit einer langsamen Lebensgeschwindigkeit haben in der Regel weniger sexuelle Partner*innen im Laufe ihres Lebens, werden später Eltern und haben weniger Nachkommen im Vergleich zu Menschen mit einer schnellen Lebensgeschwindigkeit. Menschen mit einer langsamen Lebensgeschwindigkeit leben meist in stabileren und sichereren Umfeldern und investieren mehr Zeit in die persönliche Entwicklung, in die Erziehung der Kinder und in die Entfaltung kognitiver Fähigkeiten (Woodley, 2012). Umgebungen mit weniger pathogenem Stress, besserer Ernährung und besseren Bildungschancen fördern eine langsamere Lebensgeschwindigkeit. Das Modell legt somit nahe, dass eine langsamere Lebensgeschwindigkeit die kognitive Entwicklung positiv beeinflusst (Woodley, 2012), dies könnte insbesondere die Zuwächse in der fluiden Intelligenz erklären. Darüber hinaus berücksichtigt dieses Modell die unregelmäßigen Muster des Flynn Effekts, da die Stärke der IQ-Veränderungen von Land zu Land und über die Zeit hinweg variiert (Pietschnig & Voracek, 2015).

Insgesamt bietet das Life history speed Modell eine plausible Erklärung für den Flynn Effekt. Es bedarf jedoch weiterer Forschung, um die Erklärungskraft dieses Modells für den Flynn Effekt umfassend zu evaluieren (Pietschnig & Voracek, 2015).

Weitere hybride Faktoren. Weitere hybride Erklärungsansätze umfassen die Zunahme der Variabilität des IQ (z. B. Colom et al., 2005; Teasdale & Owen, 2005), die Verringerung der Bleiexposition (Kaufman et al., 2014) sowie die genomische Prägung (Storfer, 1999).

Keiner der bisher diskutierten Theorien kann den Flynn Effekt allein vollständig erklären. Vielmehr erscheint es plausibler, einen multikausalen Ansatz zu verfolgen, der verschiedene Einflussfaktoren kombiniert (Pietschnig, 2016).

Mögliche Ursachen für den negativen Flynn Effekt

In einigen Studien gibt es erste Hinweise darauf, dass der Flynn Effekt schwächer wird, stagniert oder sogar umkehrt (z. B. Bratsberg & Rogeberg, 2018; Dworak et al., 2023; Lazaridis et al., 2022; Oberleiter, Patzl, et al., 2024). Für diese Trends gibt es verschiedene Theorien, die als mögliche Erklärungen in Betracht gezogen werden.

Migration. Ein Ansatz zur Erklärung dieser Trends ist der Migrationseffekt. Dieser basiert auf der Annahme, dass Menschen aus nicht-westlichen Ländern, in denen die durchschnittlichen kognitiven Fähigkeiten niedriger sind, in westliche Länder migrieren, in denen die durchschnittlichen kognitiven Fähigkeiten höher sind. Diese Migration könnte zu einem Rückgang der IQ-Werte in den aufnehmenden Ländern führen (Dutton & Lynn, 2013). Zudem deuteten die Ergebnisse von Woodley Of Menie et al. (2018) darauf hin, dass Migration den IQ-Rückgang besonders stark beeinflusst, wenn die Messung eine höhere g-Ladung aufweist. Jedoch fanden Pietschnig et al. (2018) heraus, dass Migration keine bedeutsamen Einflüsse auf den Flynn Effekt hat, was der Migrationshypothese als wesentliche Ursache widerspricht. Möglicherweise sind migrationsbedingte Veränderungen eher vorübergehend und wirken sich nicht langfristig auf die kognitiven Veränderungen der Bevölkerung aus (Te Nijenhuis et al., 2004). Außerdem könnte die Anzahl der zugewanderten Personen zu gering sein, um den negativen Flynn Effekt vollständig durch Migration zu erklären (Dutton et al., 2016).

Dysgenische Prozesse. Genetische Faktoren werden ebenfalls zur Erklärung des negativen Flynn Effekts herangezogen. Eine Annahme in diesem Kontext ist, dass dysgenische Prozesse zu einem Rückgang der kognitiven Fähigkeiten führen könnten. Hierbei wird ein negativer Zusammenhang zwischen Intelligenz und Fertilität postuliert (Lynn, 2011). Diese Theorie besagt, dass in Bevölkerungsgruppen mit niedrigen durchschnittlichen kognitiven Fähigkeiten tendenziell höhere Fertilitätsraten beobachtet werden. Wenn weniger intelligente Eltern mehr Nachkommen bekommen, könnte dies langfristig zu einem Rückgang des durchschnittlichen IQ führen (Lynn, 2011). Eine Metaanalyse ergab jedoch ambivalente Ergebnisse: Während eine triviale negative Assoziation von Fertilität und der kristallinen Intelligenz festgestellt wurde, zeigte sich bei der Gesamtskala und der fluiden Intelligenz ein nicht-trivialer positiver Effekt (Pietschnig et al., 2018).

Mortalität. Ein weiterer Erklärungsansatz für den Rückgang des Flynn Effekts ist die Reduzierung der Mortalität, insbesondere der Kindersterblichkeit, in westlichen Ländern (Nyborg, 2012). Eine verbesserte medizinische Versorgung, die Reduzierung von Krankheiten und bessere hygienische Bedingungen haben die Sterblichkeit, besonders in ärmeren Bevölkerungsgruppen, gesenkt (Lynn, 2011). Dadurch erreichen mehr Menschen das fortpflanzungsfähige Alter. Diese Entwicklungen könnten die natürliche Selektion verändert haben, indem sie die Überlebenschancen in diesen Gruppen erhöht und somit möglicherweise eine Weitergabe geringerer kognitiver Fähigkeiten begünstigt haben. Infolgedessen könnte dies zu einem Rückgang des durchschnittlichen IQs in der Bevölkerung führen (Nyborg, 2012). Auch hier zeigten die Ergebnisse eine gewisse Uneindeutigkeit. Es wurde festgestellt, dass eine niedrigere Mortalität negativ mit der räumlichen Wahrnehmung assoziiert war, während bei der Gesamtskala und der fluiden Intelligenz positive Effekte beobachtet wurden (Pietschnig et al., 2018).

Insgesamt gibt es also wenig klare Evidenz dafür, dass Migration, Fertilität und Mortalität signifikante Auswirkungen auf den Flynn Effekt haben.

Weitere Erklärungsansätze. Eine weitere Erklärung für den Rückgang der IQ-Zuwächse ist das Auftreten von Deckeneffekten bei den Faktoren, die ursprünglich zum Anstieg des IQ beigetragen haben. Das bedeutet, dass die Wirksamkeit von Faktoren wie einer verbesserten Ernährung nach einer gewissen Zeit nachgelassen hat, was möglicherweise die neuen Trends des Flynn Effekts erklärt (Pietschnig et al., 2018). Zudem könnten die abnehmenden IQ-Zuwächse auf das Nachlassen positiver Effekte zurückzuführen sein. So könnten Verbesserungen in der Bildung oder die Verringerung des pathogenen Stresses zunehmend weniger Erträge bringen. Dadurch werden die Grenzen weiterer Steigerungen sichtbar, da die bisherigen Verbesserungen in den Umweltfaktoren nur noch geringe Auswirkungen auf eine positive Intelligenzentwicklung haben (Pietschnig & Voracek, 2015).

Zudem wurde das mütterliche Alter als mögliche Erklärung für den Rückgang der Intelligenz in Betracht gezogen. Allerdings zeigte sich eine negative Korrelation zwischen dem durchschnittlichen Alter der Mutter in den jeweiligen Ländern und dem IQ-Rückgang. Daher erscheint diese Erklärung nicht überzeugend zu sein (Dutton et al., 2016).

Theorie der Fähigkeitsdifferenzierung

Der negative Zusammenhang zwischen dem psychometrischem g und dem Flynn Effekt (z. B. Pietschnig & Voracek, 2015; Te Nijenhuis, 2013; Woodley & Madison, 2013) könnte

ebenfalls eine Erklärung für diese Entwicklungen bieten. IQ-Zuwächse in spezifischen Fähigkeiten können nicht unbegrenzt ansteigen und haben möglicherweise eine Grenze erreicht. Wenn die Faktoren, die den IQ steigern, nicht mehr ausreichen, um den negativen Zusammenhang zwischen dem psychometrischem g und dem Flynn Effekt zu überdecken, könnte es zu einer Abschwächung, Stagnation oder sogar Umkehrung des Flynn Effekts kommen (Pietschnig, 2016). Dies bedeutet, dass Abnahmen im psychometrischen g sichtbar werden, sobald die Zuwächse in spezifischen Fähigkeiten zum Stillstand kommen, etwa durch Deckeneffekte der IQ-fördernden Umweltfaktoren (Pietschnig et al., 2021). Bisherige Steigerungen in den spezifischen Fähigkeiten könnten somit Rückgänge im psychometrischen g verdeckt haben (Pietschnig et al., 2018).

Diese Idee stimmt mit den beobachteten Rückgängen der Reaktionszeiten im letzten Jahrhundert überein, da die Reaktionszeit eng mit dem psychometrischem g korreliert ist (z. B. Woodley et al., 2013). Eine Verlangsamung der Reaktionszeiten könnte demnach auf eine Verringerung der allgemeinen kognitiven Fähigkeiten hindeuten. Allerdings ist es wichtig, diese Ergebnisse kritisch zu hinterfragen. Dodonova und Dodonov (2013) sowie Nettelbeck (2014) betonen, dass die unterschiedlichen Messmethoden und -instrumente, die in den verschiedenen Untersuchungen verwendet wurden, die Vergleichbarkeit und Zuverlässigkeit der Ergebnisse beeinträchtigen können. Zudem müsste die Verlangsamung der Reaktionszeit auf genetisch bedingte Rückgänge der allgemeinen kognitiven Fähigkeiten zurückzuführen sein, was jedoch nicht nachgewiesen wurde (Nettelbeck, 2014). Aufgrund der methodischen Ungenauigkeiten ist eine Vergleichbarkeit zwischen der in Woodley et al. (2013) herangezogenen Studien nicht gegeben und die Beurteilung der Reaktionszeiten erweist sich als wenig zuverlässig (Dodonova & Dodonov, 2013; Nettelbeck, 2014).

Pietschnig und Gittler (2015) fanden signifikante Abnahmen der Leistungen in der räumlichen Wahrnehmung über einen Zeitraum von 38 Jahren. Da die räumliche Wahrnehmung stark mit dem psychometrischen g assoziiert ist, deutet dieser Rückgang auf eine Abnahme der allgemeinen kognitiven Fähigkeiten hin. Dies wirft die Frage auf, warum es zu Beginn einen Anstieg der Leistungen gab. Eine plausible Erklärung könnte sein, dass IQ-steigernde Faktoren wie beispielsweise verbesserte Bildung und Ernährung zunächst den Rückgang im psychometrischem g maskierten, bis ein Sättigungspunkt erreicht wurde (Pietschnig & Gittler, 2015). Weitere Studien haben ebenfalls Rückgänge in der räumlichen Wahrnehmung festgestellt (Lazaridis et al., 2022; Oberleiter, Patzl, et al., 2024).

In Anbetracht dieser Entwicklungen zeigten die meisten Untersuchungen, dass der Flynn Effekt in Tests und Subtests, die stark mit dem psychometrischem g (allgemeine kognitiven

Fähigkeiten) korrelieren, weniger ausgeprägt war (z. B. Te Nijenhuis & Van Der Flier, 2013; Woodley et al., 2014). Diese Erkenntnis sowie der beobachtete negative Zusammenhang zwischen dem psychometrischen g und dem Flynn Effekt (z. B. Pietschnig & Voracek, 2015; Te Nijenhuis, 2013; Woodley & Madison, 2013) könnten auf eine Abnahme der positiven Kupplung der Intelligenz hindeuten.

Unsere Fähigkeiten, unser Wissen und unsere Anforderungen werden immer spezifischer, je nachdem, in welchem Beruf wir arbeiten oder worin die Interessen einer Person liegen. Die Wahl einer bestimmten Schule kann bereits als eine Form der Spezialisierung betrachtet werden. Verschiedene Bildungseinrichtungen bieten spezialisierte Ausbildungsprogramme, die zu einer Differenzierung der Fähigkeiten beitragen (Oberleiter, Fries, et al., 2024). Daher kann angenommen werden, dass sich spezifische Fähigkeiten aufgrund einer verstärkten Spezialisierung der Individuen stärker entwickeln. Diese Spezialisierung schwächt die Interkorrelation zwischen den Subtests, was g-basierte Rückgänge erklären kann (Pietschnig et al., 2023; Pietschnig & Voracek, 2015). Das bedeutet, dass der negative Flynn Effekt eher auf eine zunehmende Spezialisierung aufgrund veränderter Umweltanforderungen an kognitive Fähigkeiten zurückzuführen ist, als auf einen Rückgang der allgemeinen kognitiven Fähigkeiten (Oberleiter, Fries, et al., 2024).

Die zunehmende Spezialisierung könnte zunächst einen positiven Flynn Effekt hervorgebracht haben, gefolgt von einer Phase der Stagnation und schließlich einem Rückgang. Erste Hinweise auf eine Abschwächung der positiven Kupplung wurden von Oberleiter, Fries, et al. (2024) im Zeitraum von 2005 bis 2024 festgestellt. Die Ergebnisse dieser Studie zeigten, dass die Stärke der positiven Kupplung abgenommen hat, während es gleichzeitig Zuwächse in einzelnen Bereichen gab (Oberleiter, Fries, et al., 2024). Dies legt nahe, dass moderne Umgebungen zwar die Entwicklung spezifischer Fähigkeiten begünstigen, jedoch weniger die allgemeinen kognitiven Fähigkeiten fördern. Infolgedessen könnte dies die positive Kupplung zwischen verschiedenen Intelligenzbereichen verringern.

Es ist jedoch auch möglich, dass das psychometrische g stabil bleibt, während sich spezifische kognitive Fähigkeiten verändern. Es gab Hinweise darauf, dass sich das psychometrische g über die Zeit hinweg nicht verändert hat und somit zeitliche Stabilität aufweisen könnte (z. B. Armstrong & Woodley, 2014).

Zusammenfassend lässt sich sagen, dass die zunehmende Spezialisierung der Bevölkerung Auswirkungen auf deren Fähigkeiten haben könnte. Die beobachteten Veränderungen der positiven Kupplung der Intelligenz sind besonders aufschlussreich, wenn es darum geht, die Ursachen des Flynn Effekts zu verstehen (Pietschnig et al., 2023). Das Verständnis der g-

basierten Veränderungen ist entscheidend, da sie Hinweise auf Rückgänge oder die Stabilität des psychometrischen g geben. Dies impliziert, dass ein Teil der Veränderungen im IQ durch eine zunehmende Fähigkeitsdifferenzierung erklärt werden könnte. Während sich die Menschen zunehmend spezialisieren, nehmen die allgemeinen kognitiven Fähigkeiten weniger stark zu, stagnieren oder gehen sogar zurück.

Aktuelle Studie

Die aktuelle Studie zielt darauf ab, den Flynn Effekt in Bezug auf das psychometrische g zu untersuchen. Dabei stellt sich die Frage, ob die beobachteten IQ-Zuwächse tatsächlich auf eine Steigerung der allgemeinen kognitiven Fähigkeiten (g) zurückzuführen sind oder ob sie eher Verbesserungen in spezifischen Fähigkeiten widerspiegeln. In der folgenden Querschnittsstudie werden den in Österreich lebenden Proband*innen sechs Subtests des Intelligenz-Struktur-Test 70 (IST-70; Amthauer, 1970) und der revidierten Version (IST-2000-R; Amthauer et al., 2007) vorgelegt. Dies ermöglicht es, die kognitiven Leistungen der Teilnehmer*innen mit den Normen aus den Jahren 1970 und 2000 direkt miteinander zu vergleichen. Diese Studie dient dazu, das Verständnis des Flynn Effekts und seiner Ursachen zu vertiefen, indem die Veränderungen in der Stärke der positiven Kupplung der Intelligenz im Kontext einer zunehmenden Spezialisierung untersucht werden.

Fragestellungen/ Hypothesen

Die erste Fragestellung dieser Studie bezieht sich darauf, ob in der Stichprobe ein Flynn Effekt auftritt. Basierend auf den Ergebnissen vorheriger Studien wird angenommen, dass es signifikante Leistungsunterschiede in den verschiedenen Subtests gibt (z. B. Lazaridis et al., 2022; Oberleiter, Patzl, et al., 2024; Pietschnig & Voracek, 2015). Ein positiver Flynn Effekt tritt auf, wenn die IQ-Werte im älteren Test (IST-70) höher sind als in der revidierten Version (IST-2000-R). Umgekehrt liegt ein negativer Flynn Effekt vor, wenn die IQ-Werte in der revidierten Version höher sind als im originalen Test. Die Veränderungen der Normtabellen im Laufe der Zeit können dazu führen, dass neuere Tests im Vergleich zu älteren Tests entweder niedrigere oder höhere Durchschnittswerte aufweisen, was auf die unterschiedlichen Normierungsstichproben zurückzuführen ist. Ein Beispiel hierfür ist die Aufgabengruppe „Würfelaufgaben“. Im IST-70 (Amthauer, 1970) erzielt eine Person, die zehn Aufgaben korrekt löst, einen IQ von 100. Im IST-2000-R (Amthauer et al., 2007) erzielt die gleiche Person für dieselbe Leistung einen IQ von 97. Das bedeutet, dass es in diesem Fall einfacher ist, im älteren

Test höhere Ergebnisse zu erzielen, was auf einen positiven Flynn Effekt hindeutet (Flynn, 1984).

H1: Es gibt signifikante Unterschiede zwischen den Werten des IST-70 und des IST-2000-R.

Die zweite Fragestellung überprüft, ob es signifikante Unterschiede im psychometrischem g gibt. Basierend auf bisherigen Erkenntnissen wird angenommen, dass das psychometrische g im Laufe der Zeit entweder zurückgegangen ist (Oberleiter, Fries, et al., 2024) oder sich nicht verändert hat (Armstrong & Woodley, 2014). Diese Annahme würde die Theorie der Fähigkeitsdifferenzierung unterstützen, nach der spezifische Fähigkeiten zugenommen haben, während die allgemeinen kognitiven Fähigkeiten entweder stabil geblieben oder zurückgegangen sind (z. B. Woodley & Madison, 2013).

H2: Es gibt signifikante Unterschiede im psychometrischem g zwischen dem IST-70 und IST-2000-R.

Die dritte Fragestellung befasst sich mit den g -Ladungen der einzelnen Subtests und erforscht, ob diese unterschiedlich stark mit dem allgemeinen Faktor (g) assoziiert sind. Es wird erwartet, dass Subtests mit hohen g -Ladungen einen geringeren Flynn Effekt aufweisen als Subtests mit niedrigeren g -Ladungen (z. B. Pietschnig & Voracek, 2015; Te Nijenhuis & Van Der Flier, 2013). Insbesondere sollten Subtests, die der kristallinen Intelligenz zugeordnet werden und eine hohe g -Ladung aufweisen, einen niedrigeren Flynn Effekt haben als Subtests, die der fluiden Intelligenz zugeordnet werden und eine niedrigere g -Ladung aufweisen (Pietschnig & Voracek, 2015).

H3: Subtests, die höhere g -Ladungen aufweisen, zeigen einen signifikant geringeren Flynn Effekt als Subtests, die niedrigere g -Ladungen aufweisen.

Methoden

Die Hypothesen und Methoden wurden vor der Datenerhebung im „Open Science Framework“ (OSF; <https://osf.io/y29ar>) präregistriert.

Stichprobe

Die Proband*innen wurden über Convenience Sampling rekrutiert, wobei persönliche Kontakte und soziale Medien genutzt wurden. Zusätzlich kam das Schneeballsystem zum Einsatz, bei dem die Teilnehmenden gebeten wurden, die Studie weiterzuempfehlen, um so

weitere Teilnehmer*innen zu gewinnen. Teilnahmeberechtigt waren alle Personen im Alter von 18 bis 60 Jahren, entsprechend den Vorgaben des IST-70 (Amthauer, 1970). Ein weiteres Kriterium für die Teilnahme waren sehr gute Deutschkenntnisse, da viele Aufgaben stark sprachbasiert sind. Eine unzureichende Sprachkompetenz könnte die kognitive Leistung einer Person verzerren, indem der Test nicht die tatsächlichen Fähigkeiten, sondern lediglich Sprachbarrieren misst. Vor der Anmeldung zur Studie wurden die Teilnehmer*innen über diese Anforderung informiert und konnten nach eigenem Ermessen entscheiden, ob sie teilnehmen möchten.

Vor Beginn der Teilnahme wurden die Testpersonen mittels eines Informed Consent über die Freiwilligkeit der Teilnahme, die Vertraulichkeit ihrer Daten und über die Möglichkeit, die Studie jederzeit ohne Konsequenzen abubrechen, informiert. Die Teilnahme wurde nicht vergütet. Auf Wunsch konnten die Proband*innen eine Rückmeldung ihrer Testergebnisse erhalten.

Insgesamt konnten 154 Personen rekrutiert werden, darunter 92 (59.7%) Frauen, 60 (39%) Männer und 2 (1.3%) diverse Personen. Das Alter der Teilnehmenden lag zwischen 18 und 60 Jahren ($M = 25.91$; $SD = 9.53$). Die Mehrheit der Personen hatte Matura oder Abitur (109; 70.8%), gefolgt von Personen mit einem Bachelorabschluss (25; 16.2%), einem Master oder vergleichbarem Abschluss (11; 7.1%), einem Lehrabschluss (6; 3.9%), einem Pflichtschulabschluss (2; 1.3%) und einem PhD oder Doktorat (1; 0.6%).

Messinstrumente

Der Intelligenz-Struktur-Test (Amthauer et al., 2007) ist ein psychometrisches Testverfahren, welches im deutschsprachigen Raum weit verbreitet ist und zur Erfassung verschiedener kognitiver Bereiche eingesetzt wird. Für die Studie wurden sechs Subtests verwendet, die sowohl im IST-2000-R (Amthauer et al., 2007) als auch im IST-70 (Amthauer, 1970) in gleicher Form vorliegen. Das bedeutet, dass die Gestaltung dieser Subtests in beiden Testversionen gleich ist, insbesondere hinsichtlich der Art der Fragestellungen und der Antwortmöglichkeiten. Diese Übereinstimmung ist wichtig, um eine direkte Vergleichbarkeit der Testversionen zu gewährleisten. Die verwendeten Subtests umfassen Satzergänzungen, Analogien, Gemeinsamkeiten, Zahlenreihen, Figurenauswahl und Würfelaufgaben. Laut Manual werden Satzergänzungen, Analogien und Gemeinsamkeiten der verbalen Intelligenz zugeordnet, Zahlenreihen der numerischen Intelligenz und Figurenauswahl sowie Würfelaufgaben der figuralen Intelligenz (Amthauer et al., 2007).

Jede dieser Aufgabengruppen (Subtests) umfasst 20 Items. Aufgaben, die in der ursprünglichen und der revidierten Version unverändert geblieben sind, werden als identische Items bezeichnet. Items, die in der revidierten Version vollständig durch neue Aufgaben ersetzt wurden, werden als komplett ersetzte Items klassifiziert. Bei modifizierten Items wurden lediglich einzelne Distraktoren und/oder die Lösung geändert, während die Fragestellung selbst unverändert blieb. Um die Testzeit effizient zu gestalten, wurden identische Items nur einmal präsentiert, während komplett ersetzte und modifizierte Items der beiden Testversionen vorgegeben wurden. Dadurch variiert die Anzahl der Items zwischen den einzelnen Subtests.

Satzergänzungen (SE). Der Subtest Satzergänzungen wird der kristallinen Intelligenz zugeordnet, da für das Lösen Vorwissen erforderlich ist. Jede Aufgabe besteht aus einem Satz, der vervollständigt werden muss. Es gibt insgesamt fünf Antwortmöglichkeiten, wobei diejenige ausgewählt werden soll, die den Satz richtig ergänzt. Ein Beispiel wäre: „Das Gegenteil von Hoffnung ist ...? a) Trauer b) Verzweiflung c) Elend d) Liebe e) Hass“ (Amthauer et al., 2007, S.3, Grundmodul B). In diesem Fall wäre die richtige Antwort „Verzweiflung“. In diesem Subtest wurden im IST-2000-R zwei Aufgaben modifiziert, indem entweder die Distraktoren und/oder die Lösungen angepasst wurden. Zehn Aufgaben wurden komplett ersetzt, um die Anzahl der wissensbezogenen Fragen in der älteren Version zu verringern. Darüber hinaus wurde die Reihenfolge der Aufgaben erheblich geändert. All diese Veränderungen führten zu einer überwiegend neuen Aufgabengruppe im IST-2000-R. Aufgrund dieser Anpassungen umfasst die Aufgabengruppe in dieser Studie insgesamt 32 Items, wobei die interne Konsistenz (Cronbachs Alpha) für diese Aufgabengruppe relativ gering war ($\alpha = .54$). Im IST-2000-R lag die Split-Half-Reliabilität bei .73 (Amthauer et al., 2007) und im IST-70 bei .91 (Amthauer, 1970).

Analogien (AN). Der Subtest Analogien wird der fluiden Intelligenz zugeordnet, da er darauf abzielt, logische Verbindungen zwischen Wörtern herzustellen. Er misst schlussfolgerndes Denken anhand sprachlichen Materials. Dabei werden drei Begriffe vorgegeben, zwischen dem ersten und zweiten Begriff besteht eine Beziehung. Die Aufgabe besteht darin, zwischen dem dritten Begriff und einer der fünf Antwortmöglichkeiten eine ähnliche Beziehung zu identifizieren. Ein Beispiel für diesen Subtest wäre: „dunkel : hell = nass : ? a) Regen b) Tag c) feucht d) Wind e) trocken“ (Amthauer et al., 2007, S.5, Grundmodul B). In diesem Fall wäre die richtige Antwort „trocken“. Bei zwei Items wurde eine Modifikation im IST-2000-R vorgenommen, indem Distraktoren verändert wurden. Daher umfasst diese Aufgabengruppe 22 Items und weist eine geringe interne Konsistenz auf ($\alpha = .58$). Laut dem

Manuel des IST-2000-R beträgt die Split-Half-Reliabilität .74 (Amthauer et al., 2007), während sie im IST-70 auf .86 geschätzt wurde (Amthauer, 1970).

Gemeinsamkeiten (GE). Dieser Subtest umfasst sowohl fluide als kristalline Komponenten. Er misst die Fähigkeit, Beziehungen zwischen den Begriffen herzustellen. Hierbei werden sechs Wörter vorgegeben und es müssen jene zwei Wörter gefunden werden, die einen gemeinsamen Oberbegriff haben. Eine Beispielaufgabe wäre: „a) Messer b) Butter c) Zeitung d) Brot e) Zigarre f) Armband“ (Amthauer et al., 2007, S.7, Grundmodul B). Die richtige Lösung für dieses Beispiel wäre „Butter und Brot“, da beides Nahrungsmittel sind. In der revidierten Version des Subtests wurden drei Items angepasst, dies betrifft sowohl die richtige Lösung als auch die Distraktoren. Dies führte zu insgesamt 23 Aufgaben in diesem Subtest mit einer moderaten internen Konsistenz ($\alpha = .63$). Im IST-2000-R wurde die Split-Half-Reliabilität auf .80 geschätzt (Amthauer et al., 2007), während sie in der älteren Version bei .96 lag (Amthauer, 1970).

Zahlenreihen (ZR). Der Subtest Zahlenreihen zielt darauf ab, logische Beziehungen zwischen Zahlen herzustellen. Dieser Subtest wird der fluiden Intelligenz zugeordnet, da kein Vorwissen benötigt wird. Es werden Zahlenreihen vorgegeben, die nach einer bestimmten Regel aufgebaut sind und die nächste Zahl in dieser Folge soll gefunden werden (z. B.: „2, 4, 6, 8, 10, 12, 14, ?“; Lösung: 16; Amthauer et al., 2007, S.7, Grundmodul B). Abgesehen von einer geänderten Aufgabenreihenfolge wurden bei den Zahlenreihen keine weiteren Veränderungen im IST-2000-R vorgenommen. Dieser Subtest umfasst somit 20 Items und weist eine hohe interne Konsistenz auf ($\alpha = .89$). Laut Manual des IST-2000-R wurde die Split-Half-Reliabilität auf .94 geschätzt (Amthauer et al., 2007), ähnlich wie beim IST-70, hier lag sie bei .96 (Amthauer, 1970).

Figurenauswahl (FA). Dieser Subtest misst die Fähigkeit zur räumlichen Wahrnehmung anhand zweidimensionaler Figuren. Die Aufgabe besteht darin, die Proportionen von Flächen zu erfassen. In der obersten Zeile werden fünf Figuren gezeigt (z. B. Kreis, Dreieck, etc.). Die Aufgaben zeigen diese Figuren in mehrere Stücke zerschnitten. Durch das Zusammensetzen der einzelnen Stücke soll herausgefunden werden, welche Figur aus der oberen Zeile dabei entsteht. Im IST-2000-R wurde die Reihenfolge der Figuren und der Antwortmöglichkeiten verändert, während die Items selbst unverändert blieben. Der Subtest besteht aus insgesamt 20 Items und weist eine gute Reliabilität auf ($\alpha = .73$). Die Split-Half-Reliabilität beim IST-2000-R lag bei .80 (Amthauer et al., 2007) und beim IST-70 bei .84 (Amthauer, 1970).

Würfelaufgaben (WÜ). Der Subtest Würfelaufgaben misst die Fähigkeit der räumlichen Wahrnehmung und der mentalen Rotation von Würfeln. In der obersten Reihe werden fünf

Würfel präsentiert, die auf jeder der sechs Seiten unterschiedliche Muster aufweisen. Drei dieser Muster sind sichtbar. Die Aufgaben zeigen einen Würfel in einer veränderten Lage und es gilt zu bestimmen, welcher Würfel aus der obersten Reihe dargestellt wird. Während sich die Reihenfolge der Items und der vorgegebenen Würfel geändert hat, bleiben die Aufgaben selbst im IST-70 und IST-2000-R unverändert. Diese Aufgabengruppe beinhaltet daher 20 Items und zeigte eine gute interne Konsistenz ($\alpha = .79$). Beim IST-2000-R betrug die Split-Half-Reliabilität .85 (Amthauer et al., 2007), ähnlich wie in der originalen Version, bei der sie .89 betrug (Amthauer, 1970).

Design/ Durchführung

Wie bereits erwähnt, handelt es sich bei dieser Studie um eine Querschnittsstudie, bei der die Versuchspersonen zu einem einzigen Zeitpunkt getestet wurden. Um eine doppelte Bearbeitung identischer Aufgaben zu vermeiden, wurden ausschließlich non-redundante Items verwendet, also solche, die sich zwischen den Testversionen unterscheiden. Auf dieser Grundlage wurde ein neues Testheft erstellt, das die Aufgaben aus dem IST-70 (Amthauer, 1970) und dem IST-2000-R (Amthauer et al., 2007) kombiniert. Um mögliche Reihenfolgeeffekte zu minimieren und die Verfälschung der Ergebnisse durch solche Effekte zu verhindern, wurden zwei Testhefte (A und B) erstellt. In Testheft A wurden zuerst alle Aufgaben des IST-2000-R präsentiert, gefolgt von den Aufgaben, die nur im IST-70 enthalten sind, wobei die Reihenfolge der Aufgaben des IST-2000-R beibehalten wurde. In Testheft B wurden dagegen zuerst die Aufgaben des IST-70 präsentiert, gefolgt von denen des IST-2000-R, wobei die Reihenfolge der Items des IST-70 berücksichtigt wurde. Die Zuteilung der Testhefte an die Teilnehmer*innen erfolgte zufällig.

Die Durchführung der Studie erstreckte sich von Mitte Dezember 2023 bis Mitte April 2024. Da es sich beim IST-70 und dem IST-2000-R um ein Papier-Bleistift-Verfahren handelt, wurden die Testungen in privaten Räumlichkeiten als auch in Räumlichkeiten der Universität Wien durchgeführt.

Für jede Aufgabengruppe wurden gemäß den Vorgaben im Manual von Amthauer et al. (2007) spezifische Bearbeitungszeiten festgelegt. Diese Zeitvorgaben wurden verwendet und entsprechend angepasst, wenn zusätzliche Items aus dem IST-70 integriert wurden. Die Bearbeitungszeiten für die einzelnen Subtests ergeben sich wie folgt: Für die Aufgabengruppe Satzergänzungen war ursprünglich eine Bearbeitungszeit von sechs Minuten für 20 Items vorgesehen. Durch das Hinzufügen von zwölf weiteren Items erhöhte sich die Zeit auf neun Minuten und 36 Sekunden. Bei den Analogien betrug die vorgesehene Bearbeitungszeit für 20

Items sieben Minuten. Mit der Hinzugabe von zwei zusätzlichen Items aus dem IST-70 verlängerte sich die Zeit auf sieben Minuten und 42 Sekunden. Für den Subtest Gemeinsamkeiten war eine Bearbeitungszeit von acht Minuten festgelegt. Durch das Hinzufügen von drei weiteren Aufgaben stieg die Zeit auf neun Minuten und zwölf Sekunden. Die Subtests Zahlenreihen, Figurenauswahl und Würfelaufgaben wiesen in der originalen sowie in der revidierten Version identische Aufgaben auf. Daher wurden keine zusätzlichen Items integriert und infolgedessen blieben die Bearbeitungszeiten unverändert, zehn Minuten für die Zahlenreihen, sieben Minuten für die Figurenauswahl und neun Minuten für die Würfelaufgaben. Insgesamt dauerte die Testung mit den dazugehörigen Instruktionen in etwa eine Stunde.

Die Tests wurden in Form von Gruppentestungen durchgeführt, wobei die Anzahl der Teilnehmer*innen zwischen zwei und sechs Personen variierte. Vor jeder Testdurchführung wurden die Testhefte auf eventuelle Markierungen oder Verunreinigungen überprüft, um sicherzustellen, dass jede Person dieselben Bedingungen vorfand. Vor Beginn der Testung wurden alle Teilnehmenden gebeten, eine Einverständniserklärung zu unterschreiben. Diese Erklärung informierte sie umfassend über Datenschutz, die Anonymität der Teilnahme, die Freiwilligkeit der Teilnahme, die Möglichkeit, die Studie jederzeit abubrechen sowie über den allgemeinen Ablauf der Studie. Im Anschluss daran erläuterte der*die Testleiter*in den Teilnehmenden die allgemeinen Informationen über den Testablauf.

Vor Beginn eines jeden Subtests wurde dieser gemäß den Anweisungen von Amthauer et al. (2007) erklärt und durch Beispiele veranschaulicht. Die Teilnehmenden konnten bei Unklarheiten vor Beginn der Testzeit Fragen stellen. Alle Proband*innen begannen gleichzeitig mit dem jeweiligen Subtest. Erst nachdem alle Teilnehmenden die Aufgabengruppe beendet hatten oder die Zeit abgelaufen war, wurde mit der Erklärung der nächsten Aufgabengruppe fortgefahren. Sämtliche Instruktionen und Erklärungen wurden standardisiert, um die Vergleichbarkeit der Testergebnisse sicherzustellen. Diese Standardisierung gewährleistet, dass jede Erklärung und jede Instruktion bei allen Testdurchführungen konstant blieb und somit die Ergebnisse nicht durch unterschiedliche Anweisungen beeinflusst werden.

Nach Abschluss des Tests wurden soziodemografische Daten wie Geschlecht, Alter und höchster Bildungsabschluss erfasst. Darüber hinaus hatten die Teilnehmer*innen die Möglichkeit, Fragen zu stellen.

Statistische Analyse

Vor der Durchführung der statistischen Analysen mussten die Daten zunächst transformiert werden. Die Rohwerte der einzelnen Subtests wurden berechnet, wobei jede korrekte Antwort mit einem Punkt bewertet wurde, während falsche oder nicht eindeutige Antworten keinen Punkt erhielten. Ausgelassene Items wurden ebenfalls als inkorrekt gewertet. Anschließend wurden die Rohwerte mithilfe der Normwerttabelle des IST-70 (Amthauer, 1970) und des IST-2000-R (Amthauer et al., 2007) in Standardwerte umgewandelt. Auf diese Weise wurde für jede*n Proband*in und jeden Subtest sowohl der Original- als auch der revidierten Version ein Standardwert ermittelt. Um die Werte vergleichbar zu machen, wurden für alle Proband*innen die Normen der 21- bis 25-Jährigen herangezogen. Auf Basis dieser Standardwerte wurden anschließend die IQ-Werte berechnet. Darüber hinaus wurde ein Gesamt-IQ-Wert ermittelt, der den Durchschnitt der sechs Aufgabengruppen darstellt.

Für alle statistischen Analysen wurden die IQ-Werte herangezogen. Da die ursprünglichen und revidierten Testnormen 30 Jahre auseinanderliegen, deuten höhere IQ-Werte im ursprünglichen Test auf einen Zuwachs hin, während niedrigere Werte auf einen Rückgang im jeweiligen Subtest hindeuten. Ein Zuwachs impliziert, dass die Normierungsstichprobe des älteren Tests bessere Werte erzielt hat, während ein Rückgang darauf hinweist, dass die Normierungsstichprobe im neueren Test höhere Werte aufwies.

Zur Überprüfung der ersten Hypothese, ob in der Stichprobe ein Flynn Effekt vorliegt, wurde eine 2 x 6 ANOVA mit Messwiederholung durchgeführt. Die Faktoren dieser ANOVA waren das Standardisierungsjahr (1970 vs. 2000) und die Subtests (Satzergänzungen vs. Analogien vs. Gemeinsamkeiten vs. Zahlenreihen vs. Figurenauswahl vs. Würfelaufgaben). Da die ANOVA keine Auskunft über die Richtung eines Effekts liefert, wurden anschließend paarweise *t*-Tests für jeden Subtest durchgeführt. Die Richtung der Unterschiede wurde mithilfe von Cohen's *d* interpretiert. Ein positiver Wert von Cohen's *d* deutet darauf hin, dass die Leistung im älteren Test besser war, was auf einen positiven Flynn Effekt hinweist. Ein negatives Cohen's *d* zeigt an, dass die Leistung im neueren Test besser war, was auf einen negativen Flynn Effekt hinweist. Die Cohen's *d*-Werte wurden nach folgender Formel berechnet: $d = (M1 - M2) / [(SD1 + SD2) / 2]$. Anschließend wurden die Cohen's *d*-Werte in IQ-Veränderungen pro Dekade (DIQ) umgerechnet, wobei folgende Formel verwendet wurde: $DIQ (\text{Intervall}) = [(d \times 15) / \text{Intervall}] \times 10$ (siehe Lazaridis et al., 2022).

Um die zweite Hypothese zu prüfen, ob Unterschiede im psychometrischen *g* bestehen, wurde für beide Testversionen eine Hauptkomponentenanalyse durchgeführt. Anschließend

wurden Konfidenzintervalle für das R^2 der beiden Tests berechnet sowie ein p -Wert ermittelt, um festzustellen, ob die Unterschiede zwischen den beiden Testversionen signifikant sind.

Zur Überprüfung der dritten Hypothese, ob ein negativer Zusammenhang zwischen den g -Ladungen der einzelnen Subtests und dem Flynn Effekt besteht, wurde ein neuer Datensatz erstellt. Dieser enthielt die g -Ladungen der Subtests aus dem IST-2000-R und dem IST-70 sowie die Cohen's d -Werte der Subtests. Anschließend wurde eine Pearson-Korrelation berechnet, um den Zusammenhang zu analysieren.

Zusätzlich wurde explorativ untersucht, ob das Geschlecht einen Einfluss auf den Flynn Effekt hat. Diese Analyse war nicht vorab geplant und daher nicht präregistriert. Da die Anzahl der Personen mit der Angabe „divers“ zu gering war, wurden in dieser Analyse nur Männer und Frauen berücksichtigt. Hierzu wurde eine $2 \times 6 \times 2$ ANOVA mit Messwiederholung durchgeführt, bei der das Standardisierungsjahr (1970 vs. 2000) und die Subtests (Satzergänzungen vs. Analogien vs. Gemeinsamkeiten vs. Zahlenreihen vs. Figurenauswahl vs. Würfelaufgaben) als Innersubjektfaktoren und das Geschlecht (männlich vs. weiblich) als Zwischensubjektfaktor definiert wurden.

Für alle Analysen wurde ein Signifikanzniveau von .05 festgelegt. Die Effektgrößen wurden gemäß Cohen (1988) eingestuft: Für das partielle Eta-Quadrat (η_p^2) gilt als untere Grenze für kleine Effekte ein Wert von .01, für mittlere Effekte .06 und für große Effekte ein Wert von .14. Für Cohen's d gelten folgende untere Grenzwerte: 0.2 für kleine Effekte, 0.5 für mittlere Effekte und 0.8 für große Effekte. Für den Pearson-Korrelationskoeffizient (r) gilt als untere Grenze für kleine Effekte .1, für mittlere Effekte .3 und für große Effekte .5.

Zur Analyse der Daten wurde primär SPSS 29 (IBM SPSS Statistics, 2020) genutzt. Ergänzend kam die statistische Software R 4.4.0 (R Core Team, 2024) für spezifische Analysen zum Einsatz. Für die Berechnung der Konfidenzintervalle des R^2 wurde das R-Paket „psychometric“ (Fletcher, 2023) verwendet.

Ergebnisse

Unterschiede in den IQ-Werten: IST-70 vs. IST-2000-R

Die durchgeführte 2×6 ANOVA mit Messwiederholung, bei der die Faktoren Standardisierungsjahr und Subtests berücksichtigt wurden, zeigte einen signifikanten Haupteffekt des Standardisierungsjahres ($F(1, 153) = 183.23, p < .001, \eta_p^2 = .55$). Für den Haupteffekt der Subtests sowie die Interaktion zwischen Standardisierungsjahr und Subtests wurde die Annahme der Sphärizität verletzt, daher wurde die Greenhouse-Geisser-Korrektur

angewandt. Der Haupteffekt der Subtests war signifikant ($F(4.24, 648.43) = 12.10, p < .001, \eta^2_p = .07$). Auch die Interaktion zwischen Standardisierungsjahr und Subtests zeigte einen signifikanten Effekt ($F(1.74, 266.01) = 266.32, p < .001, \eta^2_p = .64$).

Zusätzlich wurden gepaarte t -Tests durchgeführt, um signifikante Unterschiede zwischen den einzelnen Subtests zu überprüfen und um die Richtung des Effekts zu bestimmen. Im Subtest Satzergänzungen wurde eine signifikante Differenz im Mittelwert von -9.45 IQ-Punkten ($t(153) = -11.26, p < .001, d = -0.834$) zwischen der originalen und der revidierten Version festgestellt. Für den Subtest Analogien ergab der t -Test einen signifikanten Unterschied von 3.17 IQ-Punkten ($t(153) = 12.36, p < .001, d = 0.291$). Im Subtest Zahlenreihen zeigte sich eine signifikante Differenz von 10.29 IQ-Punkten ($t(153) = 53.94, p < .001, d = 0.649$). Bei der Figurenauswahl betrug die durchschnittliche Differenz 4.19 IQ-Punkte ($t(153) = 20.26, p < .001, d = 0.284$) und bei den Würfelaufgaben lag der Unterschied bei 4.91 IQ-Punkten ($t(153) = 25.82, p < .001, d = 0.310$). Auch die Gesamtskala zeigte einen signifikanten Effekt mit einem Unterschied im Mittelwert von 2.13 IQ-Punkten ($t(153) = 13.54, p < .001, d = 0.237$). Der Subtest Gemeinsamkeiten zeigte keinen signifikanten Unterschied, mit einer Durchschnittsdifferenz von -0.29 IQ-Punkten ($t(153) = -0.98, p = .327, d = -0.024$). Die deskriptiven Statistiken sind in Tabelle 1 aufgeführt, während die Cohen's d -Werte der einzelnen Subtests sowie die IQ-Veränderungen pro Dekade in Tabelle 2 dargestellt werden.

Tabelle 1*Deskriptive Statistiken der IQ-Werte für die Subtests und die Gesamtskala*

	Min	Max	<i>M</i>	<i>SD</i>
<i>Satzergänzungen</i>				
IST-70	77.5	127	101.29	10.80
IST-2000-R	74.5	140.5	110.73	11.85
<i>Analogien</i>				
IST-70	74.5	131.5	110.23	10.12
IST-2000-R	71.5	134.5	107.06	11.67
<i>Gemeinsamkeiten</i>				
IST-70	62.5	125.5	104.79	11.51
IST-2000-R	67	134.5	105.08	12.52
<i>Zahlenreihen</i>				
IST-70	70	137.5	118.21	16.69
IST-2000-R	65.5	127	107.92	15.03
<i>Figurenauswahl</i>				
IST-70	59.5	140.5	108.13	15.86
IST-2000-R	64	136	103.95	13.60
<i>Würfelaufgaben</i>				
IST-70	58	143.5	112.40	16.82
IST-2000-R	59.5	137.5	107.49	14.85
<i>Gesamt</i>				
IST-70	81.5	128.5	109.17	9.07
IST-2000-R	80.25	125.25	107.04	8.92

Anmerkung. N = 154

Tabelle 2*Übersicht der Cohen 's d-Werte und IQ-Veränderungen pro Dekade (DIQ)*

	<i>d</i>	DIQ
Satzergänzungen	-0.834*	-4.17
Analogien	0.291*	1.46
Gemeinsamkeiten	-0.024	-0.12
Zahlenreihen	0.649*	3.25
Figurenauswahl	0.284*	1.42
Würfelaufgaben	0.310*	1.55
Gesamt	0.237*	1.19

Anmerkung. $N = 154$. DIQ (Intervall) = $[(d \times 15) / 30] \times 10$. * $p < .01$ (zweiseitig).

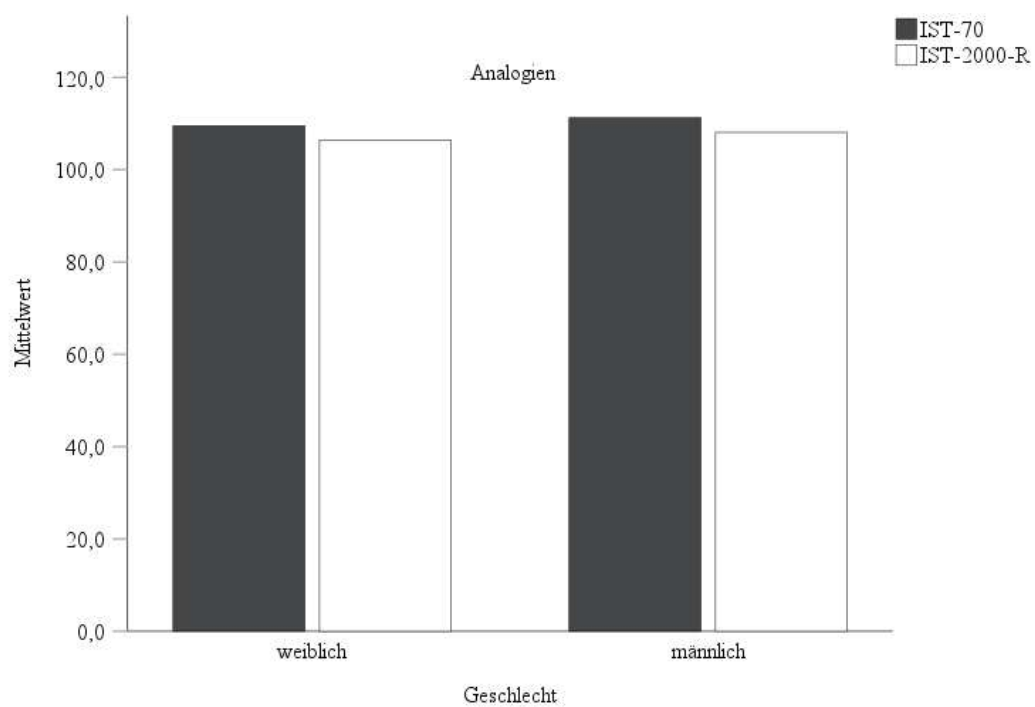
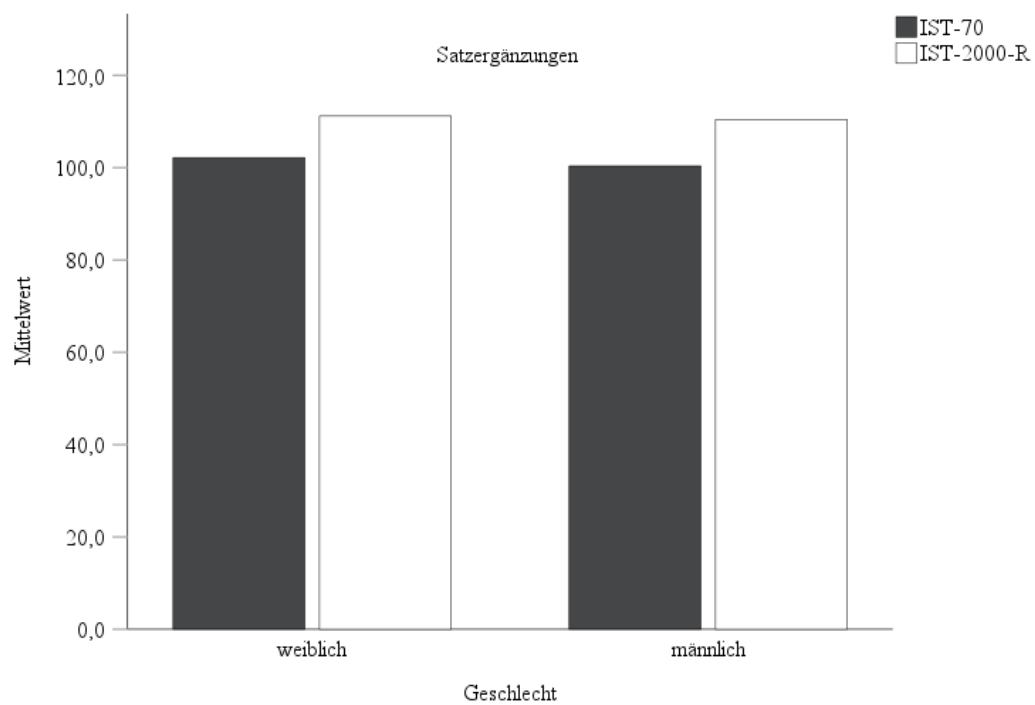
Einfluss des Geschlechts auf den Flynn Effekt

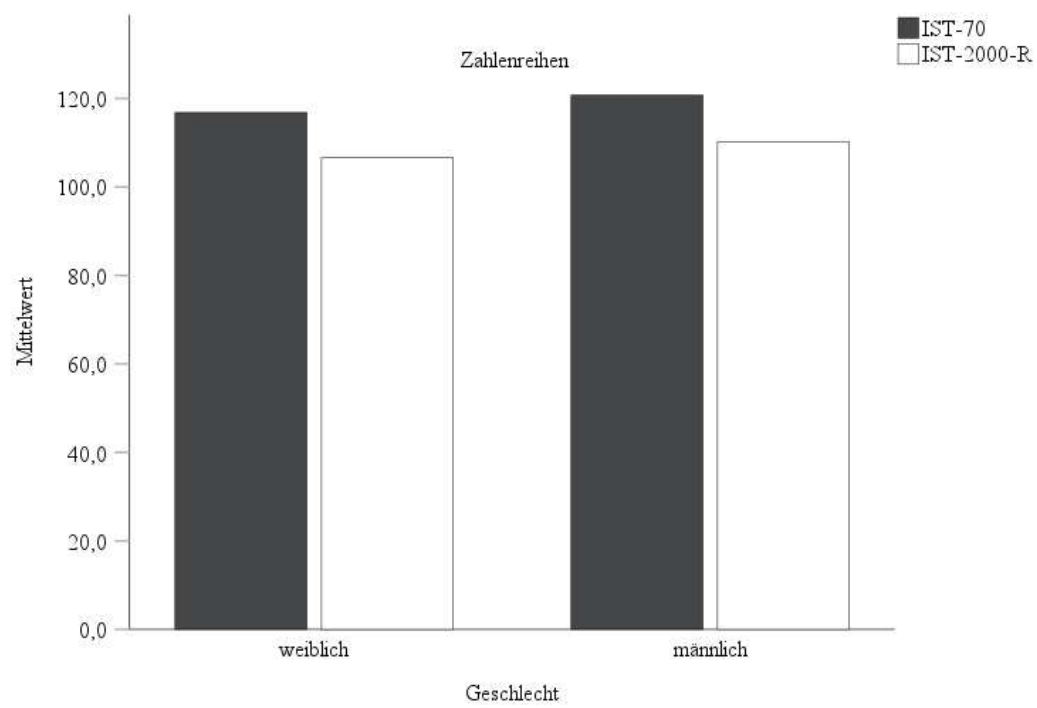
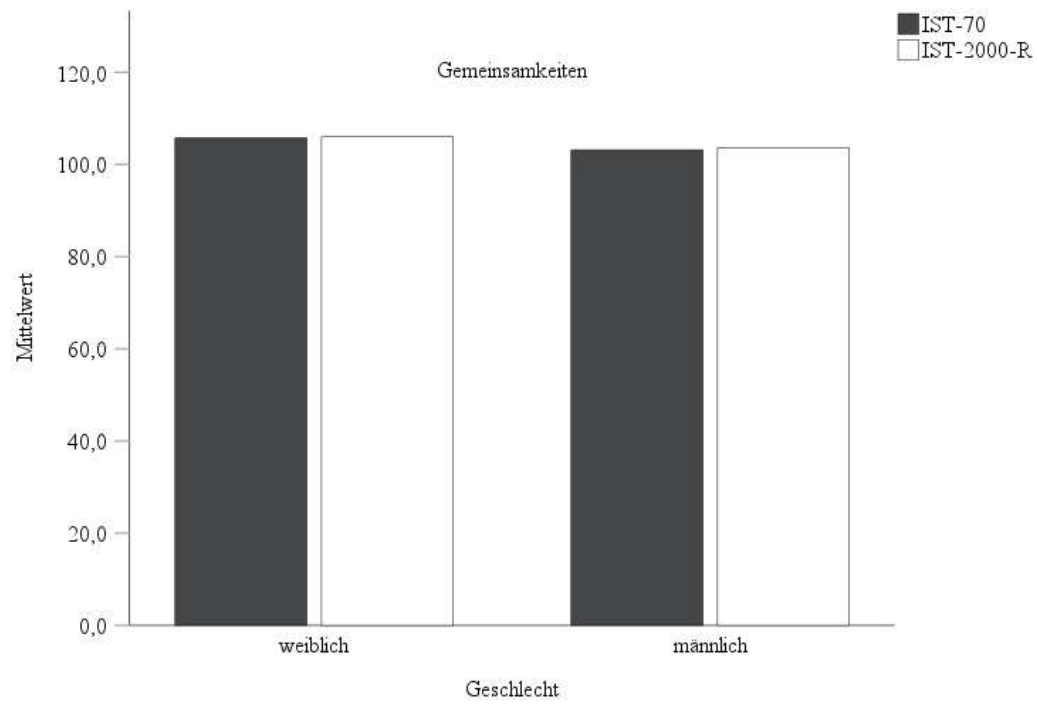
Diese Analyse war explorativ und es wurde eine 2 x 6 x 2 ANOVA mit Messwiederholung durchgeführt. Dabei wurde das Standardisierungsjahr und die Subtests als Innersubjektfaktoren und das Geschlecht als Zwischensubjektfaktor definiert. Der Box-Test zeigte, dass die Homogenität der Kovarianzmatrizen erfüllt war ($p = .374$). Gemäß dem Levene-Test war die Homogenität der Fehlervarianzen zwischen den Gruppen für alle Variablen gegeben ($p > .05$). Allerdings zeigte der Mauchly-Test für den Haupteffekt Standardisierungsjahr und die Interaktion zwischen Standardisierungsjahr und Subtests eine Verletzung der Sphärizität. Zur Korrektur dieser Verletzung wurde die Greenhouse-Geisser-Korrektur angewandt.

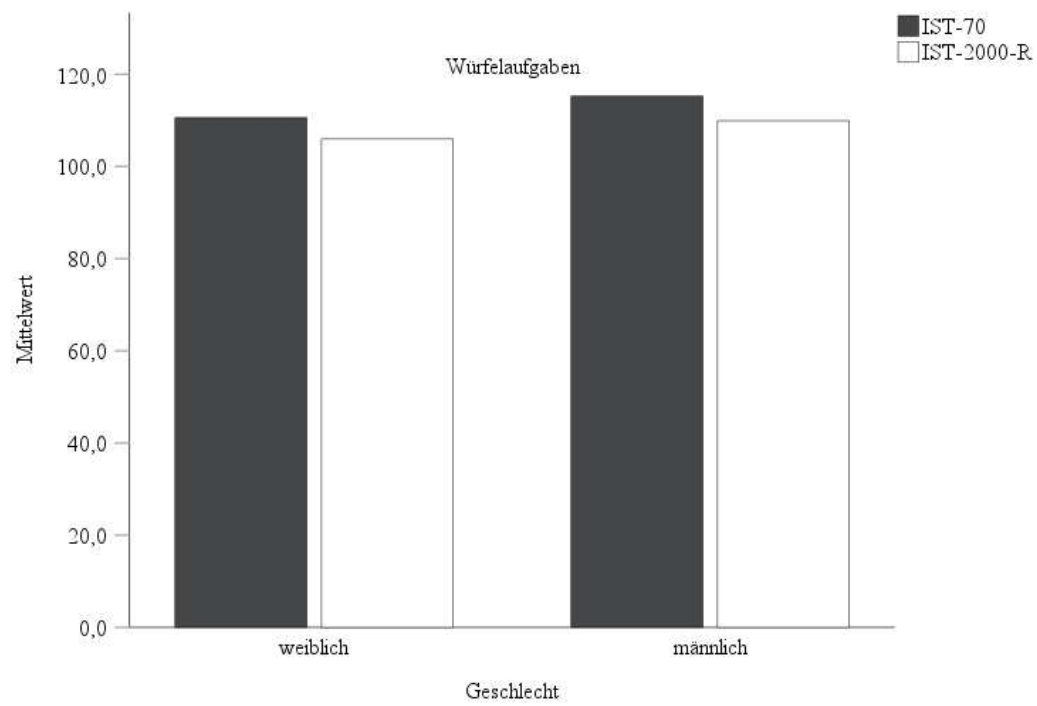
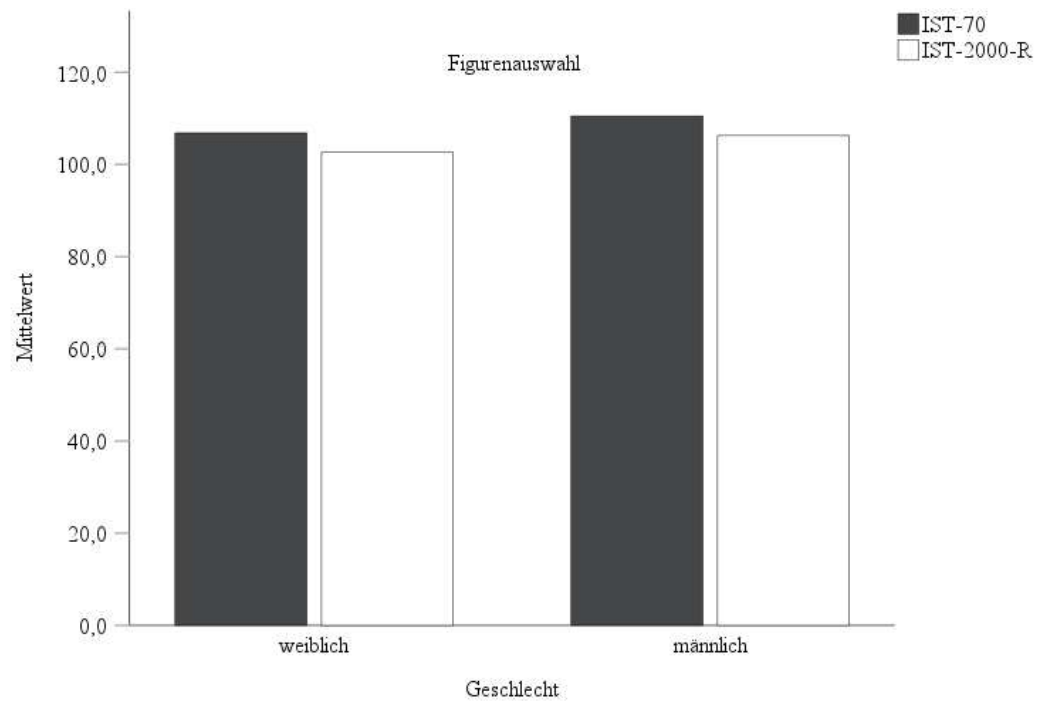
Die Analyse ergab signifikante Haupteffekte sowohl für das Standardisierungsjahr ($F(1, 150) = 169.71, p < .001, \eta^2_p = .53$) als auch für die Subtests ($F(5, 750) = 13.38, p < .001, \eta^2_p = .08$). Die Interaktion zwischen Standardisierungsjahr und Geschlecht erwies sich als nicht signifikant ($F(1, 150) = 0.01, p = .907, \eta^2_p = .001$). Eine signifikante Interaktion wurde jedoch für Subtests und Geschlecht festgestellt ($F(4.25, 637.27) = 2.55, p = .035, \eta^2_p = .02$). Der Haupteffekt des Geschlechts als Zwischensubjektfaktor war nicht signifikant ($F(1, 150) = 1.15, p = .285, \eta^2_p = .008$). Abbildung 1 illustriert die Mittelwertsunterschiede zwischen den Geschlechtern für die verschiedenen Subtests und der Gesamtskala.

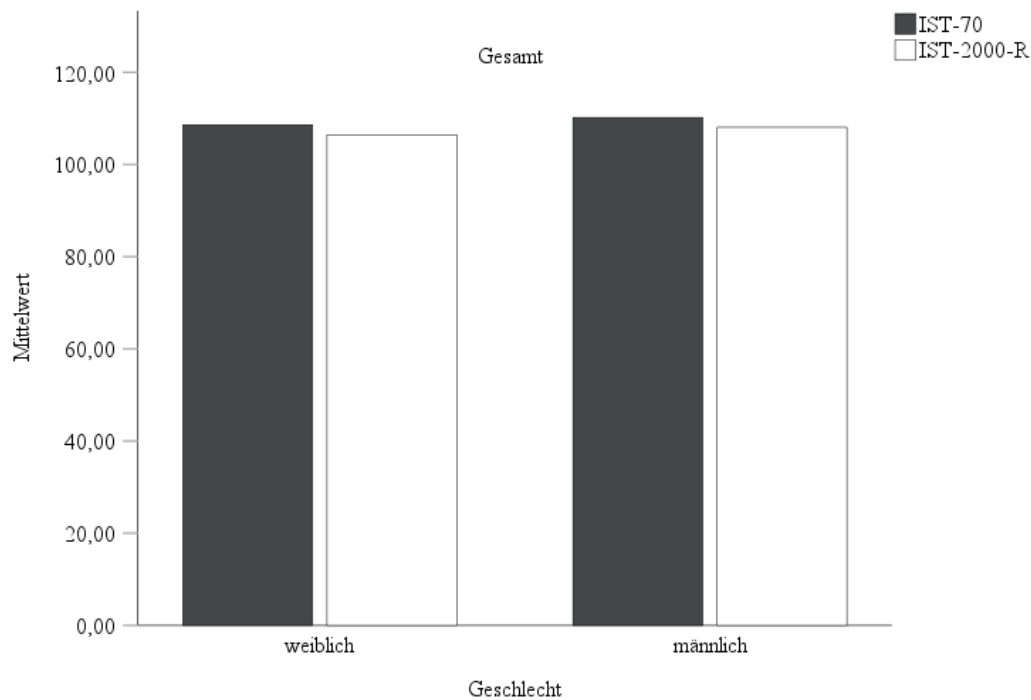
Abbildung 1

Durchschnittlicher IQ der Männer und Frauen in den verschiedenen Subtests









Anmerkung. $N = 152$, Frauen = 92, Männer = 60.

Unterschiede im psychometrischem g

Die Voraussetzungen für die Durchführung der Hauptkomponentenanalyse waren erfüllt. Das Kaiser-Meyer-Olkin (KMO) Maß zur Bestimmung der Stichprobeneignung lag für beide Testversionen über dem Mindestschwellenwert von .5 (IST-70: .73; IST-2000-R: .78). Für die Subtests des IST-70 ergab sich ein R^2 von .43 (95% KI [0.32; 0.54]), während für den IST-2000-R ein R^2 von .45 (95% KI [0.34; 0.56]) berechnet wurde. Da sich die Konfidenzintervalle der beiden R^2 -Werte überschneiden und der p -Wert für den Unterschied zwischen den beiden R^2 -Werten bei .980 lag, ist der Unterschied nicht signifikant. Die g -Ladungen der einzelnen Subtests sind in Tabelle 3 dargestellt.

Tabelle 3

G-Ladungen der einzelnen Subtests im IST-70 und IST-2000-R

	IST-70	IST-2000R
Satzergänzungen	.48	.55
Analogien	.68	.69
Gemeinsamkeiten	.64	.65
Zahlenreihen	.73	.74
Figurenauswahl	.70	.68
Würfelaufgaben	.68	.70

Beziehung zwischen g-Ladungen und IQ-Veränderungen

Die Pearson-Korrelation ergab einen signifikant positiven Zusammenhang zwischen den Cohen's d -Werten der Subtests und den g-Ladungen der Subtests des IST-70 ($r = .990, p < .001$) sowie des IST-2000-R ($r = .997, p < .001$). Die Interkorrelationen sind in Tabelle 4 dargestellt.

Tabelle 4

Korrelationsmatrix der g-Ladungen und der Cohen's d-Werte

	1	2	3
1. Cohen's d -Werte	1		
2. g-Ladungen IST-70	.990*	1	
3. g-Ladungen IST-2000-R	.997*	.979*	1

Anmerkung. * $p < .01$ (einseitig)

Diskussion

Die Ergebnisse zeigten eine zeitliche Stabilität der Stärke der positiven Kupplung über einen Zeitraum von 30 Jahren innerhalb einer deutschsprachigen Stichprobe. Es gab demnach keine Hinweise darauf, dass sich das psychometrische g im Laufe der Zeit verändert hat, obwohl in vier von sechs Subtests Zuwächse festgestellt wurden. Diese Stabilität des psychometrischen g deutet darauf hin, dass die allgemeinen kognitiven Fähigkeiten unverändert geblieben sind, während sich spezifische Fähigkeiten weiterentwickelt haben.

In der vorliegenden Studie wurde ein differenziertes, und domänenspezifisches Muster von Veränderungen in den IQ-Werten mit unterschiedlich starken Effektgrößen festgestellt (Lazaridis et al., 2022). Die Daten zeigten einen positiven Flynn Effekt in den Subtests Analogien, Zahlenreihen, Figurenauswahl und Würfelaufgaben. Auch für die Gesamtskala konnte eine Leistungssteigerung zwischen der originalen Version (Amthauer, 1970) und der revidierten Version (Amthauer et al., 2007) festgestellt werden. Ein negativer Flynn Effekt wurde im Subtest Satzergänzungen beobachtet, während im Subtest Gemeinsamkeiten keine signifikanten Veränderungen in den IQ-Werten auftraten.

Während in der Mehrheit der bisherigen Studien ein negativer Zusammenhang zwischen dem psychometrischen g und dem Flynn Effekt postuliert wurde (Must et al., 2003; Pietschnig & Voracek, 2015; Te Nijenhuis, 2013; Woodley et al., 2013; nicht bei Colom et al., 2001), zeigte die vorliegende Studie einen positiven Zusammenhang zwischen den g-Ladungen der Subtests und dem Flynn Effekt. Das bedeutet: Je höher die g-Ladungen der Subtests, desto stärker ist

der Flynn Effekt. Zudem hatte das Geschlecht keinen Einfluss auf den Flynn Effekt (Dutton et al., 2016; Pietschnig et al., 2011).

Flynn Effekt

Im Subtest Zahlenreihen wurde zwischen 1970 und 2000 die größte signifikante Leistungssteigerung beobachtet, mit einem Anstieg von 3.25 IQ-Punkten pro Dekade und einer mittleren Effektstärke. Dieser Subtest erfasst Aspekte der fluiden Intelligenz im Bereich der numerischen Fähigkeiten, da hierfür kein Vorwissen erforderlich ist. Das bedeutet, dass sich in diesem Zeitraum weiterhin eine positive Entwicklung abzeichnet. Auch im Zeitraum von 2005 bis 2024 zeigten sich weiterhin Zuwächse in den numerischen Fähigkeiten, diese fielen jedoch mit 2.8 IQ-Punkten pro Dekade etwas geringer aus als in der vorliegenden Studie (Oberleiter, Fries, et al., 2024). Dies könnte darauf hindeuten, dass der Flynn Effekt in den numerischen Fähigkeiten im deutschsprachigen Raum in den letzten Jahrzehnten zwar fortbestand, sich jedoch leicht abgeschwächt hat.

Es ist jedoch zu beachten, dass einige Teilnehmer*innen im Subtest Zahlenreihen alle 20 Items korrekt beantwortet haben, was auf einen Deckeneffekt hindeutet. Ein Deckeneffekt tritt auf, wenn ein Subtest oder Test so einfach ist, dass eine beträchtliche Anzahl von Teilnehmer*innen die maximale Punktzahl erreicht (Wang et al., 2008). Dadurch sind die Items nicht mehr in der Lage, feine Unterschiede in höheren Leistungsbereichen zu erfassen, was zu einer ungenauen Messung der tatsächlichen Unterschiede in den kognitiven Fähigkeiten führt.

Der Deckeneffekt kann zu einer Überschätzung des Flynn Effekts führen, da die beobachteten Leistungssteigerungen eher die begrenzte Differenzierungsfähigkeit des Subtests in höheren IQ-Bereichen widerspiegeln als tatsächliche Verbesserungen der kognitiven Fähigkeiten der Teilnehmer*innen. Der Flynn Effekt im Subtest Zahlenreihen könnte daher durch diesen Effekt verzerrt worden sein. Folglich ist es möglich, dass der Flynn Effekt in diesem Bereich weniger stark ausgeprägt ist.

Im Subtest Figurenauswahl zeigte sich zwischen 1970 und 2000 ein Anstieg von 1.42 IQ-Punkten pro Dekade. Ähnlich verzeichnete der Subtest Würfelaufgaben eine Zunahme von 1.55 IQ-Punkten im gleichen Zeitraum. Der Subtest Analogien wies einen ähnlichen Anstieg von 1.46 IQ-Punkten pro Dekade auf. Alle diese Subtests zeigten kleine Effektstärken und lassen sich der fluiden Intelligenz zuordnen. Diese Ergebnisse deuten darauf hin, dass in der fluiden Intelligenz weiterhin positive Entwicklungen stattfinden, also ein positiver Flynn Effekt vorliegt. Die Zuwächse in diesen Subtests sind jedoch nicht so hoch, wie in der Metanalyse von

Pietschnig und Voracek (2015), deren Zuwächse von 1924 bis 2013 4.1 IQ-Punkte pro Dekade betrugen.

Neuere Forschungsergebnisse haben jedoch Hinweise darauf gefunden, dass die IQ-Werte in der fluiden Intelligenz zurückgehen (z. B. Dworak et al., 2023; Lazaridis et al., 2022; Oberleiter, Patzl, et al., 2024). Im Gegensatz zu den neueren Erkenntnissen zeigten die Ergebnisse dieser Studie hingegen in allen Bereichen der fluiden Intelligenz kleine bis moderate positive Flynn Effekte.

Es ist aber wichtig zu beachten, dass die in dieser Studie verwendeten Normen aus dem Jahr 2000 stammen. Es ist möglich, dass bei diesen älteren Normen noch ein Flynn Effekt sichtbar war, der sich in positiven Intelligenzzuwächsen widerspiegelt. Bei aktuellen Normen könnte dieser Effekt jedoch schwächer ausfallen oder sogar ganz ausbleiben, wie in neueren Studien gezeigt wurde (z. B. Dworak et al., 2023; Lazaridis et al., 2022; Oberleiter, Patzl, et al., 2024). Dworak et al. (2023) fanden in einer US-amerikanischen Studie zwischen 2006 und 2018 heraus, dass die durchschnittlichen Werte für fluide Fähigkeiten bei Teilnehmenden aus den letzten Jahren niedriger waren. Im Bereich des figuralen Denkens wurde ein negativer Flynn Effekt zwischen 2012 und 2022 festgestellt (Oberleiter, Patzl, et al., 2024). Lazaridis et al. (2022) berichteten, dass bei Matrizenaufgaben im Zeitraum von 1996 bis 2018, kaum oder gar kein Flynn Effekt nachweisbar war.

Diese Ergebnisse stehen im Widerspruch zu den Befunden der vorliegenden Studie, die positive Flynn Effekte in den Bereichen fluider Aufgabengruppen feststellte. Die Verwendung neuerer Normen könnte daher neue Trends oder eine Abschwächung beziehungsweise Umkehr des Flynn Effekts im Bereich der fluiden Intelligenz aufdecken. Angesichts dieser Entwicklung ist es entscheidend, weitere Forschung durchzuführen, um festzustellen, ob die in dieser Studie beobachteten positiven Effekte tatsächlich fortbestehen oder ob sich neue Muster in der Entwicklung der fluiden Intelligenz abzeichnen.

Im Subtest Gemeinsamkeiten wurden zwischen 1970 und 2000 keine signifikanten Veränderungen in den IQ-Werten beobachtet. Da dieser Test sowohl fluide als auch kristalline Komponenten umfasst, könnte die Stabilität der Ergebnisse darauf hinweisen, dass sich die entsprechenden kognitiven Fähigkeiten in diesem Bereich im Laufe der Zeit nicht verändert haben.

Im Gegensatz zu den fluiden Aufgabengruppen wurde im Subtest Satzergänzungen eine Leistungsminderung von -4.17 IQ-Punkten pro Dekade zwischen 1970 und 2000 festgestellt. Diese Abnahme deutet auf einen negativen Flynn Effekt in der kristallinen Intelligenz hin, da dieser Test die Fähigkeit misst, Sprache und Wissen anzuwenden. Während in der Metaanalyse

von Pietschnig und Voracek (2015) eine geringere Zunahme der kristallinen Intelligenz (2.1 IQ-Punkte pro Dekade) im Vergleich zur fluiden Intelligenz (4.1 IQ-Punkte pro Dekade) beobachtet wurde, zeigt die vorliegende Studie einen signifikanten Rückgang in diesem Bereich.

Einerseits könnte der Rückgang der IQ-Werte in dieser Aufgabengruppe auf eine Umkehrung des Flynn Effekts hindeuten, wie es bereits in einigen Ländern in unterschiedlichen kognitiven Domänen beobachtet wurde (z. B. Norwegen: Bratsberg & Rogeberg, 2018; USA: Dworak et al., 2023; Österreich: Oberleiter, Patzl, et al., 2024). Andererseits könnte eine mögliche Erklärung darin liegen, dass obsoleete Items, die veraltetes Wissen abfragen, den Flynn Effekt in diesem Subtest verfälscht haben. Solche Veränderungen in der Schwierigkeit der Items können dazu führen, dass der Flynn Effekt unterschätzt oder sogar als negativ wahrgenommen wird, wenn ältere Items als schwieriger gelten (Gonthier et al., 2021; Gonthier & Grégoire, 2022). Zum Beispiel wurde im IST-70 gefragt, was der wichtigste Bestandteil eines Fernsehers ist; damals war es die Bildröhre (Amthauer, 1970). Wenn veraltete Items aufgrund ihrer veränderten Relevanz in der heutigen Zeit schwieriger geworden sind, könnte dies auch in dieser Studie zu den beobachteten Unterschieden geführt haben, da normalerweise geringere Unterschiede im Bereich der kristallinen Intelligenz erwartet werden (z. B. Pietschnig & Voracek, 2015). Zudem wurden in der originalen Version des Tests zehn Items vollständig ersetzt, wobei diese im älteren Test generell stärker wissensbezogener waren als in der revidierten Version. Es ist auch denkbar, dass der Subtest Satzergänzungen in den beiden Versionen unterschiedliche Konstrukte misst. Dies würde bedeuten, dass sich die Eigenschaften des Subtests und die Messwerte im Laufe der Zeit verändert haben, was die Vergleichbarkeit der Ergebnisse zwischen den beiden Zeitpunkten beeinträchtigen könnte (Gonthier & Grégoire, 2022).

Die beobachteten Veränderungen des IQ in dieser Studie stehen im Einklang mit bisherigen Untersuchungen, die zeigten, dass die Zuwächse in der fluiden Intelligenz tendenziell stärker ausgeprägt sind als in der kristallinen Intelligenz (Nisbett et al., 2012; Pietschnig & Voracek, 2015). Diese Befunde unterstützen die Theorie einer domänenspezifischen Ausprägung des Flynn Effekt (Lazaridis et al., 2022; Pietschnig & Voracek, 2015; Te Nijenhuis & Van Der Flier, 2013), bei dem die Steigerungen in verschiedenen Bereichen der Intelligenz nicht gleichmäßig verteilt sind.

Die Ergebnisse der durchgeführten ANOVA mit Messwiederholung untermauern diese Annahme, da sie eine signifikante Interaktion zwischen dem Standardisierungsjahr und den Subtests aufzeigten. Dies deutet darauf hin, dass die Veränderungen in den IQ-Leistungen nicht über alle Subtests hinweg einheitlich sind. Die Ergebnisse legen somit nahe, dass die

beobachteten Veränderungen in den IQ-Werten bestimmte Bereiche der Intelligenz stärker betreffen.

Geschlechtsspezifischer Flynn Effekt

In der Forschung zum Flynn Effekt wurden häufig Daten von Männern verwendet, da viele Untersuchungen an Wehrpflichtigen durchgeführt wurden. Diese Vorgehensweise führte dazu, dass viele Untersuchungen zum Flynn Effekt auf Männer beschränkt sind, was die Ergebnisse beeinflussen könnte. Infolgedessen wird das Geschlechterverhältnis als mögliche Einflussgröße für die Veränderungen im Flynn Effekt diskutiert (Dutton et al., 2016).

In dieser Studie wurden zwar Unterschiede zwischen den Geschlechtern in den einzelnen Subtests festgestellt (siehe Abbildung 1), jedoch zeigte die durchgeführte ANOVA mit Messwiederholung keine signifikante Interaktion zwischen dem Standardisierungsjahr und dem Geschlecht. Dies bedeutet, dass das Ausmaß des Flynn Effekts unabhängig vom Geschlecht ist und die Veränderungen der IQ-Werte nicht durch das Geschlecht beeinflusst werden. Diese Ergebnisse stimmen mit den Befunden von Pietschnig et al. (2011) überein, die ebenfalls Unterschiede in den Leistungen der Geschlechter identifizierten, jedoch keinen Einfluss des Geschlechts auf den Flynn Effekt feststellten. Auch Dutton et al. (2016) kamen zu dem Schluss, dass das Geschlecht keine bedeutsame Rolle bei den Veränderungen der IQ-Werte spielt.

Das psychometrische g

In der vorliegenden Studie wurden keine Hinweise darauf gefunden, dass sich das psychometrische g im Zeitraum zwischen 1970 und 2000 verändert hat. Die Ergebnisse deuten darauf hin, dass dieses Maß für die allgemeinen kognitiven Fähigkeiten über drei Jahrzehnte hinweg stabil geblieben ist. Diese Befunde stehen im Gegensatz zu den Ergebnissen von Oberleiter, Fries, et al. (2024), die einen Rückgang der Stärke der positiven Kupplung zwischen 2005 bis 2024 feststellten, obwohl in sechs Intelligenzbereichen Leistungszuwächse verzeichnet wurden. Eine mögliche Erklärung für diese Diskrepanz könnte die Varianzeinschränkung in der vorliegenden Stichprobe sein, die aus besonders fähigen (High-Ability) Personen bestand. Diese eingeschränkte Varianz kann dazu führen, dass Effekte schwerer zu erkennen sind oder geringer ausfallen (Lakes, 2013). Laut Spearman's Law of Diminishing Returns (Spearman, 1927) nimmt bei höherer Intelligenz die Korrelation zwischen den Subtests ab. Das bedeutet, dass intelligentere Individuen eine geringere Interkorrelation zwischen den Subtests aufweisen als weniger intelligente Personen. Mit zunehmender Fähigkeit hat das psychometrische g also einen geringeren Einfluss (Blum & Holling, 2017).

Veränderungen im psychometrischem g könnten daher weniger aussagekräftig sein, da g in High-Ability Stichproben insgesamt weniger Erklärungswert besitzt.

Die in dieser Studie beobachtete Stabilität widerspricht der Hypothese, dass eine Zunahme spezifischer Fähigkeiten die Interkorrelationen, die das g -Faktor Modell stützen, abgeschwächt hätte (Pietschnig et al., 2023). Es gibt keine Evidenz dafür, dass sich die Stärke der positiven Kupplung über die Zeit verändert hat. Folglich wird die Theorie eines abnehmenden psychometrischen g (Oberleiter, Fries, et al., 2024; Oberleiter, Patzl, et al., 2024; Pietschnig et al., 2021, 2023) durch die Ergebnisse dieser Studie nicht gestützt.

Vielmehr haben sich die allgemeinen kognitiven Fähigkeiten als stabil erwiesen, während einzelne kognitive Domänen einen Flynn Effekt oder Anti-Flynn Effekt gezeigt haben. Dies könnte dennoch die Theorie der Fähigkeitsdifferenzierung stützen. Nach dieser Theorie könnten spezifische Fähigkeiten im Laufe der Zeit Veränderungen erfahren haben, während das übergeordnete Maß für allgemeine kognitive Fähigkeiten unverändert bleibt (Armstrong & Woodley, 2014; Pietschnig et al., 2023).

Ein Teil der Veränderungen in den IQ-Werten lässt sich möglicherweise durch diese Theorie erklären. Im Laufe der Zeit haben sich die Menschen zunehmend spezialisiert, was zu Verbesserungen in spezifischen Fähigkeiten geführt hat. Gleichzeitig sind die allgemeinen kognitiven Fähigkeiten weitgehend stabil geblieben. Dies bedeutet, dass die Interkorrelation zwischen den verschiedenen kognitiven Bereichen konstant geblieben ist, auch wenn in bestimmten Bereichen ein Flynn Effekt zu beobachten war.

Zusammenhang psychometrisches g und Flynn Effekt

In den meisten Studien wurde ein negativer Zusammenhang zwischen dem psychometrischem g und dem Flynn Effekt festgestellt (Must et al., 2003; Pietschnig & Voracek, 2015; Te Nijenhuis, 2013; Woodley et al., 2013). In der vorliegenden Untersuchung zeigte sich jedoch ein abweichendes Bild. Hier wiesen die g -Ladungen der Subtests im IST-70 (Amthauer, 1970) und im IST-2000-R (Amthauer et al., 2007) eine signifikant positive Beziehung zum Flynn Effekt auf und zwar mit großen Effektstärken. Diese positiven Assoziationen deuten darauf hin, dass Subtests mit höheren g -Ladungen stärkere Zuwächse im Flynn Effekt aufweisen.

Bisherige Studien zeigten tendenziell, dass Subtests mit höheren g -Ladungen einen geringeren Flynn Effekt aufweisen (Pietschnig & Voracek, 2015; Te Nijenhuis & Van Der Flier, 2013; Woodley et al., 2014). Im Gegensatz dazu zeigte die vorliegende Studie, dass der Subtest Zahlenreihen, der die höchste g -Ladung aufweist, auch den größten Flynn Effekt verzeichnet.

Frühere Untersuchungen haben häufig gezeigt, dass Bereiche, die die fluide Intelligenz messen, mit niedrigen g-Ladungen und einem stärkeren Flynn Effekt assoziiert sind (z. B. Pietschnig & Voracek, 2015). In dieser Studie jedoch weisen die fluiden Intelligenzbereiche sowohl die höchsten g-Ladungen als auch die größten Zuwächse auf. Zusätzlich wurde der größte Verlust der IQ-Werte in dem Subtest beobachtet, der die niedrigste g-Ladung aufweist. Dieses Ergebnis stimmt mit den Befunden von Woodley Of Menie et al. (2018) überein, die feststellten, dass die größten Verluste bei Subtests mit niedrigen g-Ladungen auftreten.

Eine mögliche Erklärung für den positiven Zusammenhang zwischen den g-Ladungen der Subtests und den IQ-Veränderungen könnte darin liegen, dass die Ergebnisse der Hauptkomponentenanalyse maßgeblich von der Auswahl und Art der einbezogenen Subtests beeinflusst werden (Jensen, 1998). In der vorliegenden Studie war die Auswahl der Subtests nicht gleichmäßig über verschiedene kognitive Fähigkeiten verteilt. Konkret wurden drei Subtests zur Messung verbaler Fähigkeiten, zwei zur Messung figuraler Fähigkeiten und einer zur Messung numerischer Fähigkeiten verwendet. Diese Ungleichverteilung könnte die Ergebnisse der Analyse beeinflusst haben (Jensen, 1998). Da die verbalen Subtests in dieser Studie die Mehrheit ausmachten, könnte dies zu einer Übergewichtung der verbalen Fähigkeiten in der Berechnung der g-Ladungen geführt haben. Dies könnte den positiven Zusammenhang zwischen den g-Ladungen und den IQ-Veränderungen erklären. Die unterschiedlichen Gewichtungen der Subtests beeinflussen die Berechnung des psychometrischen g und somit auch die g-Ladungen (Jensen, 1998). Die ungleiche Verteilung und Gewichtung in der Analyse könnten dazu geführt haben, dass Subtests mit höheren g-Ladungen größere Zuwächse in den IQ-Werten aufweisen, was den beobachteten positiven Zusammenhang erklärt.

Limitationen

Die vorliegende Studie weist mehrere Limitationen auf, die bei der Interpretation der Ergebnisse berücksichtigt werden sollten. Erstens ist die Repräsentativität der Stichprobe eingeschränkt, da die Teilnehmenden überwiegend Studierende waren. Dies führt zu einer gewissen Homogenität innerhalb der Stichprobe, insbesondere hinsichtlich des Bildungsabschlusses und des Alters. Die Mehrheit der Teilnehmenden hat die Matura beziehungsweise das Abitur als höchsten Bildungsabschluss angegeben und das Alter lag in einem ähnlichen Bereich. Diese Faktoren schränken die Generalisierbarkeit der Ergebnisse ein, da sie nicht auf andere Bevölkerungsgruppen übertragbar sind.

Zudem weist die Auswahl der Subtests in dieser Studie eine ungleiche Verteilung der kognitiven Fähigkeiten auf, wobei der Anteil der verbalen Subtests überproportional hoch ist. Diese Ungleichverteilung kann zu einer Verzerrung des psychometrischen g führen, da verbale Fähigkeiten einen größeren Einfluss auf die Berechnung des g -Faktors haben (Jensen, 1998). Infolgedessen könnte die Aussagekraft und die Vergleichbarkeit des psychometrischen g in dieser Studie eingeschränkt sein.

Weiters wird in dieser Querschnittstudie von einem linearen Zusammenhang zwischen den beiden Standardisierungsjahren ausgegangen, was impliziert, dass sich die Veränderungen der IQ-Werte gleichmäßig über die Zeit entwickeln. Frühere Studien (Bratsberg & Rogeberg, 2018; Pietschnig & Voracek, 2015) haben jedoch gezeigt, dass sowohl der positive als auch der negative Flynn Effekt nicht linear verlaufen. Das bedeutet, dass die Zunahmen oder Abnahmen in verschiedenen Zeiträumen unterschiedlich stark ausfallen können. Die angenommene Linearität in dieser Studie ist jedoch durch das Design bedingt und kann nicht beliebig verändert werden.

Conclusio

Die Ergebnisse dieser Studie zeigten, dass zwischen 1970 und 2000 bei vier von sechs Subtests sowie bei der Gesamtskala ein kleiner bis moderater positiver Flynn Effekt zu beobachten war. Dies deutet darauf hin, dass der Flynn Effekt in diesem Zeitraum weiterhin präsent ist. Zudem wurde in dieser Stichprobe festgestellt, dass das psychometrische g über einen Zeitraum von 30 Jahren stabil geblieben ist. Dies deutet darauf hin, dass die allgemeinen kognitiven Fähigkeiten in diesem Zeitraum konstant geblieben sind. Während in den meisten Studien ein negativer Zusammenhang zwischen dem psychometrischen g und dem Flynn Effekt festgestellt wurde (Must et al., 2003; Pietschnig & Voracek, 2015; Te Nijenhuis, 2013; Woodley et al., 2013), ergab die vorliegende Studie jedoch einen positiven Zusammenhang zwischen dem Flynn Effekt und den g -Ladungen.

Die in dieser Studie erhobene Stabilität des psychometrischen g stellt die Hypothese infrage, dass die zunehmende Spezialisierung in der Bevölkerung die positive Korrelation zwischen verschiedenen Intelligenzdomänen verringert hat. Dennoch könnte diese Stabilität eine Rolle bei der Erklärung des Flynn Effekts spielen. Während spezifische Fähigkeiten im Laufe der Zeit zunehmen, könnten die allgemeinen kognitiven Fähigkeiten unverändert bleiben (z. B. Armstrong & Woodley, 2014). Dies legt nahe, dass der Flynn Effekt eher Verbesserungen in spezifischen Fähigkeiten widerspiegelt als Veränderungen im psychometrischen g .

Für zukünftige Forschungsarbeiten wäre eine Untersuchung über einen längeren Zeitraum wichtig, um mögliche nicht lineare Zusammenhänge besser aufzudecken. Frühere Untersuchungen (Oberleiter, Patzl, et al., 2024; Pietschnig & Voracek, 2015) haben gezeigt, dass sowohl positive als auch negative Flynn Effekte nicht immer linear verlaufen. Längere Beobachtungszeiträume können dabei helfen, komplexe Muster und Schwankungen im Verlauf der IQ-Veränderungen zu identifizieren.

Literaturverzeichnis

- Amthauer, R. (1970). *Intelligenz-Struktur-Test: IST 70*. Göttingen: Hogrefe.
- Amthauer, R., Brocke, B., Liepmann, D., & Beauducel, A. (2007). *Intelligenz-Struktur-Test 2000 R* (2. erweiterte und überarbeitete Auflage). Göttingen: Hogrefe.
- Armstrong, E. L., & Woodley, M. A. (2014). The rule-dependence model explains the commonalities between the Flynn effect and IQ gains via retesting. *Learning and Individual Differences, 29*, 41–49. <https://doi.org/10.1016/j.lindif.2013.10.009>
- Blum, D., & Holling, H. (2017). Spearman's law of diminishing returns. A meta-analysis. *Intelligence, 65*, 60–66. <https://doi.org/10.1016/j.intell.2017.07.004>
- Brand, C. R., Freshwater, S., & Dockrell, W. B. (1989). Has there been a 'massive' rise in IQ levels in the west? Evidence from Scottish children. *The Irish Journal of Psychology, 10*(3), 388–393. <https://doi.org/10.1080/03033910.1989.10557756>
- Bratsberg, B., & Rogeberg, O. (2018). Flynn effect and its reversal are both environmentally caused. *Proceedings of the National Academy of Sciences, 115*(26), 6674–6678. <https://doi.org/10.1073/pnas.1718793115>
- Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies* (1. Aufl.). Cambridge University Press. <https://doi.org/10.1017/CBO9780511571312>
- Cattell, R. B. (1963). Theory of fluid and crystallized intelligence: A critical experiment. *Journal of Educational Psychology, 54*(1), Article 1. <https://doi.org/10.1037/h0046743>
- Clark, C. M., Lawlor-Savage, L., & Goghari, V. M. (2016). The Flynn effect: A quantitative commentary on modernity and human intelligence. *Measurement: Interdisciplinary Research and Perspectives, 14*(2), 39–53. <https://doi.org/10.1080/15366367.2016.1156910>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed). L. Erlbaum Associates.

- Colom, R., Juanespinosa, M., & Garcia, L. (2001). The secular increase in test scores is a “Jensen effect”. *Personality and Individual Differences*, 30(4), 553–559.
[https://doi.org/10.1016/S0191-8869\(00\)00054-4](https://doi.org/10.1016/S0191-8869(00)00054-4)
- Colom, R., Lluísfont, J., & Andrespueyo, A. (2005). The generational intelligence gains are caused by decreasing variance in the lower half of the distribution: Supporting evidence for the nutrition hypothesis. *Intelligence*, 33(1), 83–91. <https://doi.org/10.1016/j.intell.2004.07.010>
- Daley, T. C., Whaley, S. E., Sigman, M. D., Espinosa, M. P., & Neumann, C. (2003). IQ on the rise: The Flynn effect in rural Kenyan children. *Psychological Science*, 14(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., Hill, W. D., & Gale, C. R. (2021). Intelligence, health and death. *Nature Human Behaviour*, 5(4), 416–430. <https://doi.org/10.1038/s41562-021-01078-9>
- Dickens, W. T., & Flynn, J. R. (2001). Heritability estimates versus large environmental effects: The IQ paradox resolved. *Psychological Review*, 108(2), 346–369.
<https://doi.org/10.1037/0033-295X.108.2.346>
- 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, 163–169. <https://doi.org/10.1016/j.intell.2016.10.002>

- Dworak, E. M., Revelle, W., & Condon, D. M. (2023). Looking for Flynn effects in a recent online U.S. adult sample: Examining shifts within the SAPA project. *Intelligence*, 98, 101734. <https://doi.org/10.1016/j.intell.2023.101734>
- Fletcher, T. D. (2023). *Psychometric: Applied psychometric theory* (Version 2.4) [Software]. <https://doi.org/10.32614/CRAN.package.psychometric>
- Flynn, J. R. (1984). The mean IQ of Americans: Massive gains 1932 to 1978. *Psychological Bulletin*, 95(1), 29–51. <https://doi.org/10.1037/0033-2909.95.1.29>
- Flynn, J. R. (2007). *What is intelligence?: Beyond the Flynn effect* (1. Aufl.). Cambridge University Press. <https://doi.org/10.1017/CBO9780511605253>
- 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. (2018). Reflections about intelligence over 40 years. *Intelligence*, 70, 73–83. <https://doi.org/10.1016/j.intell.2018.06.007>
- Flynn, J. R. (2019). Secular changes in intelligence: The “Flynn effect”. In R. J. Sternberg (Hrsg.), *The Cambridge Handbook of Intelligence* (2. Aufl., S. 940–963). Cambridge University Press. <https://doi.org/10.1017/9781108770422.040>
- Flynn, J. R., & Shayer, M. (2018). IQ decline and Piaget: Does the rot start at the top? *Intelligence*, 66, 112–121. <https://doi.org/10.1016/j.intell.2017.11.010>
- Focosi Mazzei, R., Antonelli-Ponti, M., Almeida, S. D. S., Riul, T. R., & Da Silva, J. A. (2019). Intelligence, income and their relation to nutrition. *Psych*, 2(1), 1–10. <https://doi.org/10.3390/psych2010001>
- Gonthier, C., & Grégoire, J. (2022). Flynn effects are biased by differential item functioning over time: A test using overlapping items in Wechsler scales. *Intelligence*, 95, 101688. <https://doi.org/10.1016/j.intell.2022.101688>

- 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>
- Herrnstein, R. J., & Murray, C. A. (1994). *The bell curve: Intelligence and class structure in American life*. Free Press.
- 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* (1. Aufl.). Cambridge University Press. <https://doi.org/10.1017/CBO9780511781308>
- IBM SPSS Statistics. (2020). [Software] (Version 29). IBM Corporation. <https://www.ibm.com/products/spss-statistics>
- Jensen, A. R. (1998). *The g factor: The science of mental ability*. Praeger.
- Kaufman, A. S., Kaufman, J. C., Liu, X., & Johnson, C. K. (2009). How do educational attainment and gender relate to fluid intelligence, crystallized intelligence, and academic skills at ages 22-90 years? *Archives of Clinical Neuropsychology*, 24(2), 153–163. <https://doi.org/10.1093/arclin/acp015>
- 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>
- Kriegbaum, K., Becker, N., & Spinath, B. (2018). The relative importance of intelligence and motivation as predictors of school achievement: A meta-analysis. *Educational Research Review*, 25, 120–148. <https://doi.org/10.1016/j.edurev.2018.10.001>
- Lakes, K. D. (2013). Restricted sample variance reduces generalizability. *Psychological Assessment*, 25(2), 643–650. <https://doi.org/10.1037/a0030912>

- 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>
- 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. (2009). What has caused the Flynn effect? Secular increases in the development quotients of infants. *Intelligence*, 37(1), 16–24. <https://doi.org/10.1016/j.intell.2008.07.008>
- Lynn, R. (2011). *Dysgenics: Genetic deterioration in modern populations* (2., rev. ed). Ulster Inst. for Social Research.
- 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)
- 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>
- Nisbett, R. E., Aronson, J., Blair, C., Dickens, W., Flynn, J., Halpern, D. F., & Turkheimer, E. (2012). Intelligence: New findings and theoretical developments. *American Psychologist*, 67(2), 130–159. <https://doi.org/10.1037/a0026699>

- 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>
- Oberleiter, S., Fries, J., Dejardin, F., Heller, J., Schaible, C., Vetter, M., Voracek, M., & Pietschnig, J. (2024). Inconsistent Flynn effect patterns may be due to a decreasing positive manifold: Cohort-based measurement-invariant IQ test score changes from 2005 to 2024. *Intelligence*, 107, 101867. <https://doi.org/10.1016/j.intell.2024.101867>
- Oberleiter, S., Patzl, S., Fries, J., Diedrich, J., Voracek, M., & Pietschnig, J. (2024). Measurement-invariant fluid anti-Flynn effects in population—Representative German student samples (2012–2022). *Journal of Intelligence*, 12(1), 9. <https://doi.org/10.3390/jintelligence12010009>
- Pietschnig, J. (2016). The Flynn effect: Technology may be part of it, but is most certainly not all of it. *Measurement: Interdisciplinary Research and Perspectives*, 14(2), 70–73.
<https://doi.org/10.1080/15366367.2016.1171612>
- 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, 101544. <https://doi.org/10.1016/j.intell.2021.101544>
- Pietschnig, J., & Gittler, G. (2015). A reversal of the Flynn effect for spatial perception in German-speaking countries: Evidence from a cross-temporal IRT-based meta-analysis (1977–2014). *Intelligence*, 53, 145–153. <https://doi.org/10.1016/j.intell.2015.10.004>
- 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. (2015). One century of global IQ gains: A formal meta-analysis of the Flynn effect (1909–2013). *Perspectives on Psychological Science*, 10(3), 282–306.
<https://doi.org/10.1177/1745691615577701>
- 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(2), 267–283.
- Pokropek, A., Marks, G. N., & Borgonovi, F. (2022). How much do students' scores in PISA reflect general intelligence and how much do they reflect specific abilities? *Journal of Educational Psychology*, 114(5), 1121–1135. <https://doi.org/10.1037/edu0000687>
- R Core Team. (2024). *Language and environment for statistical computing* (Version 4.4.0) [Software]. <https://www.r-project.org/>
- Rehnberg, J., Huisman, M., Fors, S., Marseglia, A., & Kok, A. (2024). The Association between education and cognitive performance varies at different levels of cognitive performance: A quantile regression approach. *Gerontology*, 70(3), 318–326.
<https://doi.org/10.1159/000535717>
- Rodgers, J. L. (1998). A critique of the Flynn effect: Massive IQ gains, methodological artifacts, or both? *Intelligence*, 26(4), 337–356. [https://doi.org/10.1016/S0160-2896\(99\)00004-5](https://doi.org/10.1016/S0160-2896(99)00004-5)
- Runquist, E. A. (1936). Intelligence test scores and school marks of high school seniors in 1929 and 1933. *School and Society*, 43, 301–304.
- Schaie, K. W., & Strother, C. R. (1968). A cross-sequential study of age changes in cognitive behavior. *Psychological Bulletin*, 70(6, Pt.1), 671–680. <https://doi.org/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 (Hrsg.), *Contemporary intellectual assessment: Theories, tests and issues* (S. 73–163). The Guilford Press.
- Spearman, C. (1927). *The abilities of man*. Macmillan.
- Spearman, C. (2010). The proof and measurement of association between two things. *International Journal of Epidemiology*, 39(5), 1137–1150. <https://doi.org/10.1093/ije/dyq191>
- Storfer, M. (1999). Myopia, intelligence, and the expanding human neocortex: Behavioral influences and evolutionary implications. *International Journal of Neuroscience*, 98(3–4), 153–276. <https://doi.org/10.3109/00207459908997465>
- Strenze, T. (2007). Intelligence and socioeconomic success: A meta-analytic review of longitudinal research. *Intelligence*, 35(5), 401–426. <https://doi.org/10.1016/j.intell.2006.09.004>
- Te Nijenhuis, J. (2013). The Flynn effect, group differences, and g loadings. *Personality and Individual Differences*, 55(3), 224–228. <https://doi.org/10.1016/j.paid.2011.12.023>
- Te Nijenhuis, J., De Jong, M., 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., & 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>
- Teasdale, T. W., & Owen, D. R. (2005). A long-term rise and recent decline in intelligence test performance: The Flynn effect in reverse. *Personality and Individual Differences*, 39(4), 837–843. <https://doi.org/10.1016/j.paid.2005.01.029>
- Trahan, L. H., Stuebing, K. K., Fletcher, J. M., & Hiscock, M. (2014). The Flynn effect: A meta-analysis. *Psychological Bulletin*, 140(5), 1332–1360. <https://doi.org/10.1037/a0037173>
- Tuddenham, R. D. (1948). Soldier intelligence in World Wars I and II. *American Psychologist*, 3(2), 54–56. <https://doi.org/10.1037/h0054962>

- Wang, L., Zhang, Z., McArdle, J. J., & Salthouse, T. A. (2008). Investigating ceiling effects in longitudinal data analysis. *Multivariate Behavioral Research*, 43(3), 476–496.
<https://doi.org/10.1080/00273170802285941>
- Wichman, A. L., Rodgers, J. L., & MacCallum, R. C. (2006). A multilevel approach to the relationship between birth order and intelligence. *Personality and Social Psychology Bulletin*, 32(1), 117–127. <https://doi.org/10.1177/0146167205279581>
- 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., & Madison, G. (2013). Establishing an association between the Flynn effect and ability differentiation. *Personality and Individual Differences*, 55(4), 387–390.
<https://doi.org/10.1016/j.paid.2013.03.016>
- Woodley, M. A., Te Nijenhuis, J., & 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., Te Nijenhuis, J., Must, O., & Must, A. (2014). Controlling for increased guessing enhances the independence of the Flynn effect from g: The return of the Brand effect. *Intelligence*, 43, 27–34. <https://doi.org/10.1016/j.intell.2013.12.004>
- Woodley Of Menie, M. A., Peñaherrera-Aguirre, M., Fernandes, H. B. F., & Figueredo, A.-J. (2018). What causes the anti-Flynn effect? A data synthesis and analysis of predictors. *Evolutionary Behavioral Sciences*, 12(4), 276–295. <https://doi.org/10.1037/ebs0000106>
- Zajonc, R. B., & Mullaney, P. R. (1997). Birth order: Reconciling conflicting effects. *American Psychologist*, 52(7), 685–699. <https://doi.org/10.1037/0003-066X.52.7.685>
- Zumbo, B. D. (2007). Three generations of DIF analyses: Considering where it has been, where it is now, and where it is going. *Language Assessment Quarterly*, 4(2), 223–233.
<https://doi.org/10.1080/15434300701375832>

Abbildungsverzeichnis

Abbildung 1: Durchschnittlicher IQ der Männer und Frauen in den verschiedenen Subtests ...29

Tabellenverzeichnis

Tabelle 1: Deskriptive Statistiken der IQ-Werte für die Subtests und die Gesamtskala	27
Tabelle 2: Übersicht der Cohen's <i>d</i> -Werte und IQ-Veränderungen pro Dekade (DIQ)	28
Tabelle 3: <i>G</i> -Ladungen der einzelnen Subtests im IST-70 und IST-2000-R	32
Tabelle 4: Korrelationsmatrix der <i>g</i> -Ladungen und der Cohen's <i>d</i> -Werte	33

Anhang

Zusammenfassung

Der Flynn Effekt zeigt signifikante Zuwächse bei Intelligenztestergebnissen auf, die je nach Domäne, Alter und Land variieren. In einigen Ländern, insbesondere in skandinavischen Ländern, sowie in bestimmten Intelligenzbereichen wie der räumlichen Wahrnehmung, wurden jedoch auch Rückgänge der IQ-Werte beobachtet. Während Ursachen wie beispielsweise verbesserte Bildung oder Ernährung untersucht wurden, bieten diese Theorien keine vollständige Erklärung für die Veränderungen in den Testergebnissen. Eine alternative Erklärung könnte in der zunehmenden Spezialisierung der Bevölkerung liegen, die spezifische Intelligenzbereiche fördert, während allgemeine kognitive Fähigkeiten stagnieren oder sogar zurückgehen. In einer Querschnittsstudie mit 154 deutschsprachigen Personen ($M = 25.91$ Jahre, $SD = 9.53$, 92 weiblich, 60 männlich, 2 divers) wurden sechs Subtests der originalen Version (IST-70) und der revidierten Version (IST-2000-R) eines deutschsprachigen Intelligenztests untersucht. Die Ergebnisse zeigten einen positiven Flynn Effekt in den Subtests Analogien, Zahlenreihen, Figurenauswahl, Würfelaufgaben und in der Gesamtskala, einen negativen Flynn Effekt bei Satzergänzungen und keinen Effekt bei Gemeinsamkeiten. Das psychometrische g blieb über die 30 Jahre hinweg stabil, was darauf hindeutet, dass sich spezifische Fähigkeiten verbessert haben, während die allgemeinen kognitiven Fähigkeiten unverändert blieben. Überraschenderweise war der Zusammenhang zwischen dem Flynn Effekt und den g -Ladungen der Subtests signifikant positiv, im Gegensatz zu früheren Studien, die einen negativen Zusammenhang fanden.

Abstract

The Flynn effect describes significant increases in intelligence test scores, which vary according to domain, age and country. However, declines in IQ scores have also been observed in some countries, particularly in Scandinavian countries as well as in certain domains of intelligence such as spatial perception. While causes such as improved education or nutrition have been investigated, these theories do not provide a complete explanation for the changes in test scores. An alternative explanation could lie in the increasing specialization of the population, which enhances specific cognitive domains while general cognitive abilities stagnate or even decline. In a cross-sectional study of 154 German-speaking participants ($M = 25.91$ years, $SD = 9.53$; 92 female, 60 male, 2 diverse), six subtests from the original version (IST-70) and the revised version (IST-2000-R) of a German-language intelligence test were examined. The results showed a positive Flynn effect in the subtests Analogien, Zahlenreihen, Figurenauswahl, Würfelaufgaben and the overall scale, a negative Flynn effect in Satzergänzungen and no effect in Gemeinsamkeiten. Psychometric g remained stable over 30 years, suggesting that specific abilities improved, while general cognitive abilities did not change. Surprisingly, the relationship between the Flynn effect and the g -loadings of the subtests was significantly positive, in contrast to earlier studies that showed a negative relationship.

R-Code

```
+ install.packages('psychometric')
+ library('psychometric')

# Konfidenzintervalle berechnen
CI.Rsq(0.43165, 154, 6, level=0.95)
CI.Rsq(0.44999, 154, 6, level=0.95)

# Konfidenzintervalle
ci_r2_1_lower <- 0.3201395
ci_r2_1_upper <- 0.5431605
ci_r2_2_lower <- 0.3398092
ci_r2_2_upper <- 0.5601708

# Berechnung der Standardfehler
z_975 <- qnorm(0.975)
se_r2_1 <- (ci_r2_1_upper - ci_r2_1_lower) / (2 * z_975 / sqrt(n))
se_r2_2 <- (ci_r2_2_upper - ci_r2_2_lower) / (2 * z_975 / sqrt(n))

# R-Squared Werte
r2_1 <- 0.43165
r2_2 <- 0.44999

# Fisher-Z Transformation
fisher_z <- function(r2) {
  0.5 * log((1 + sqrt(r2)) / (1 - sqrt(r2)))
}

z1 <- fisher_z(r2_1)
z2 <- fisher_z(r2_2)

# Berechnung der Z-Differenz
z_diff <- abs(z1 - z2) / sqrt(se_r2_1^2 + se_r2_2^2)
```

```
# Berechnung des p-Werts
```

```
p_value <- 2 * (1 - pnorm(z_diff))
```

```
cat("p-Wert für den Unterschied zwischen den beiden R-Squared Werten:", p_value, "\n")
```