

# Intelligenz(tests) verstehen und missverstehen

Jörn R. Sparfeldt<sup>1</sup>, Nicolas Becker<sup>1</sup>, Samuel Greiff<sup>2</sup>, Martin Kersting<sup>3</sup>, Cornelius J. König<sup>1</sup>, Jonas W. B. Lang<sup>4</sup> und André Beauducel<sup>5</sup>

<sup>1</sup>Universität des Saarlandes, Saarbrücken, Deutschland

<sup>2</sup>Universität Luxemburg, Luxemburg

<sup>3</sup>Justus-Liebig-Universität Gießen, Deutschland

<sup>4</sup>Universität Gent, Belgien

<sup>5</sup>Rheinische-Friedrich-Wilhelms-Universität Bonn, Deutschland

**Zusammenfassung:** Die vorliegende Standortbestimmung zeigt die hohe wissenschaftliche Qualität der Intelligenzforschung und von Intelligenztests. Es werden aber auch mögliche Missverständnisse und Einseitigkeiten der Ergebnisrezeption und -interpretation thematisiert. Im Einzelnen werden (1) die hohe prognostische und kriterienbezogene Validität bei gleichzeitigen Vorbehalten wie teils niedriger Akzeptanz bzw. Augenscheinvalidität, (2) die Darstellung empirischer Befunde aus der Perspektive ausgewählter Theorien sowie (3) die Bedeutung von Umwelteinflüssen und hohen Erblichkeitskoeffizienten eingehender betrachtet. Für jeden dieser Bereiche wird verdeutlicht, dass vor allem Präzision bei der Rezeption und Darstellung von Forschungsergebnissen notwendig ist, um Einseitigkeiten, Missverständnisse und Instrumentalisierungen zu vermeiden. Der vorliegende Beitrag zeigt, dass einiges, was als Problem der Intelligenzforschung und von Intelligenztests kritisiert wird, letztendlich auf die dargestellten Missverständnisse zurückzuführen ist. Vor diesem Hintergrund wird der Unterschied zwischen der qualitativ hochwertigen Intelligenzforschung und Intelligenztestung einerseits sowie den Missverständnissen und Einseitigkeiten bei der Rezeption andererseits herausgearbeitet. Weiterhin werden berechnete Kritikpunkte an der Intelligenzforschung und an Intelligenztests sowie Forschungsdesiderata benannt.

**Schlüsselwörter:** Intelligenz, Intelligenztest, Missverständnisse, Rezeption

## Understanding and Misunderstanding Intelligence and Intelligence Tests

**Abstract:** The present overview shows the high scientific quality of intelligence research and the respective intelligence tests. It discusses possible misunderstandings and one-sided reception and interpretation of research results. More specifically, we consider (1) the contrast between high predictive and criterion-related validity, on the one hand, and low acceptance and face validity, on the other hand; (2) the presentation of results in the light of theoretical perspectives; and (3) the relevance of environmental effects and high heritability estimates. We demonstrate that maximum precision of the presentation and reception of research results is necessary for each of these three areas to avoid unbalanced perceptions, misunderstandings, and unilateral exploitation of results. The present contribution also shows that some problems of intelligence research and intelligence tests occur because of such misunderstandings. Against this background, we elaborate the contrast between the high quality of intelligence research and testing, on the one hand, and the misunderstandings and one-sided reception, on the other hand. Further, we address other points of criticism of intelligence research and intelligence tests as well as research desiderata.

**Keywords:** intelligence, intelligence test, misunderstanding, reception

Die Psychologie kann bei den Themen „Intelligenz“ und „Intelligenztests“ auf eine beeindruckende Erfolgsgeschichte zurückblicken, und zwar sowohl hinsichtlich des wissenschaftlichen Erkenntnisgewinns als auch hinsichtlich der praktischen Relevanz. Dies begründet die besondere Bedeutung der Themen in Forschung, Lehre und Praxis. Bezogen auf die *Forschung* gelten Begabungsunterschiede seit jeher als zentrale Erklärung interindividueller Leistungsunterschiede: Personen mit vergleichbarer Lerngeschichte zeigen mit Begabungsunterschieden

korrespondierende Leistungsunterschiede. Innerhalb der akademischen Psychologie gehört die systematische Erforschung dieser Leistungsunterschiede und der menschlichen Intelligenz seit jeher zu den zentralen Forschungsgebieten. Wie in allen Wissenschaftsbereichen stellten sich im Nachhinein manche der eingeschlagenen Wege als Sackgasse oder Irrweg heraus; nichtsdestotrotz kann das Forschungsfeld heute auf ein breites Spektrum an etablierten Theorien und Modellvorstellungen, auf vielfältige und vielfach replizierte Befunde sowie auf solide Intelli-

genztests mit sehr guten psychometrischen Kennwerten zur Beantwortung praktisch bedeutsamer diagnostischer Fragestellungen blicken. Bezogen auf die *Lehre* gehören folgerichtig „Intelligenz“ und „Intelligenztests“ zum Kanon eines Psychologiestudiums. In einschlägigen Lehrbüchern nehmen diese Themen einen zentralen Platz ein. Für viele Fragestellungen in der *Praxis* liefert ein Intelligenztestergebnis wertvolle Informationen. Intelligenztests gehören – bezogen auf die psychometrischen Gütekriterien – zu den besten Verfahren der psychologischen Diagnostik. Der Güte verbreiteter Messungen in den sog. exakten Naturwissenschaften und bspw. der medizinischen Diagnostik steht die Psychologie im Bereich der Intelligenzdiagnostik – trotz weiterer Optimierungsbedarfe (s. u.) – keinesfalls nach. Ein Verzicht auf den Einsatz von Intelligenztests würde mit einem erheblichen Qualitätsverlust psychologischer Diagnostik und der darauf aufbauenden Empfehlungen und Entscheidungen einhergehen.

Doch das Bild einer Erfolgsgeschichte wird an verschiedenen Stellen getrübt. Über die wünschenswerte und den Erkenntnisfortschritt vorantreibende Fachdiskussion hinaus sind die Themen „Intelligenz“ und „Intelligenztests“ immer wieder Inhalt von grundsätzlichen Debatten im Fach, aber auch von Diskussionen in der Öffentlichkeit. Dabei werden bestimmte etablierte Befunde der Intelligenzforschung hinterfragt – oft weniger aus fachlichen Gründen, sondern mehr aus einer Gemengelage aus gesellschaftspolitischen Überzeugungen und individueller Betroffenheit. Immer wieder wurden von einzelnen Personen aus Befunden der Intelligenzforschung, respektive aus einer (mangelhaften) Rezeption dieser Befunde gesellschaftspolitische Forderungen abgeleitet, die ihrerseits heftig umstritten waren (z. B. Herrnstein & Murray, 1994; Sarrazin, 2010). Die umstrittenen Schlussfolgerungen einzelner Personen haben die Intelligenzforschung als Ganzes sowie den Einsatz von Intelligenztests immer wieder in Verruf gebracht. Die teilweise wissenschaftlich problematischen Darstellungen und Diskussionen haben wir zum Anlass genommen, die vorliegende Übersicht über Möglichkeiten und Grenzen, aber auch über offene Forschungsfragen zu ausgewählten Problemstellungen sowie über (teils scheinbare) Widersprüche und Paradoxien der Intelligenzforschung zu verfassen. Dabei greifen wir auf Beispiele aus verschiedenen psychologischen Teilfächern zurück. Gleichmaßen wollen wir auch die Frage beleuchten, ob gesellschaftliche Kritik an Intelligenztests oder die politische Instrumentalisierung von Intelligenztestergebnissen substantiell mit der wissenschaftlichen Begründung und Qualität von Intelligenzmessungen oder eher mit deren gesellschaftlicher Relevanz assoziiert sind.

Eine umfassende Übersicht über Intelligenzdefinitionen sowie die Intelligenzstruktur wäre einen eigenen Bei-

trag wert. Innerhalb der scientific community besteht weitgehende Einigkeit über die zentralen Intelligenzaspekte „schlussfolgerndes Denken“, „Problemlösen“, „Verständnis komplexer Ideen“ und „Lernen aus Erfahrungen“ (Gottfredson, 1997). Gleichmaßen wird Intelligenz in der Forschung konsensuell als hierarchisch strukturiert verstanden – mit der allgemeinen Intelligenz als Dimension an der Spitze der Hierarchie und spezifischeren Dimensionen darunter (Gottfredson, 1997; Lubinski, 2004). Nichtsdestotrotz sind die genaue Form der Hierarchie (z. B. Carroll, 1993; Lang, Kersting & Beauducel, 2016; vgl. Rost, 2013), die (Kausal-)Beziehungen zwischen den hierarchischen Ebenen (z. B. Lang & Kell, 2020; Ree & Carretta, 2002), die Umsetzung spezifischer Vorstellungen zur Hierarchie in einzelnen Intelligenzbatterien (z. B. Dombrowski, Canivez & Watkins, 2018) oder die Bedeutung der verschiedenen Dimensionen hierarchischer Modelle als Korrelat oder Prädiktor bedeutsamer Außenvariablen (z. B. Beier, Kell & Lang, 2019; Roth et al., 2015) weiter Gegenstand aktueller Forschung.

## Vorbehalte und Kritikpunkte trotz hoher Validitätskoeffizienten

Aus wissenschaftlicher Sicht werden Intelligenztests als ausgesprochen hochwertige Messinstrumente geschätzt. Beispielsweise werden für Maße der allgemeinen Intelligenz etablierter Intelligenztests häufig Reliabilitätsschätzungen berichtet, die Werte von  $r_{tt} = .90$  erreichen und teilweise sogar übersteigen; dies gilt u. a. für split-half-Reliabilitätskoeffizienten und interne Konsistenzen sowie für Test-Retest-Reliabilitäten über übliche Zeiträume (vgl. z. B. Kersting, Althoff & Jäger, 2008; Liepmann, Beauducel, Brocke & Amthauer, 2007; Petermann & Petermann, 2008; vgl. Rost, 2013; Salgado et al., 2003; Scharfen, Peters & Holling, 2018). In Bezug auf eine Vielzahl an bedeutsamen Maßen des Bildungs- und Berufserfolgs finden sich in nationalen und internationalen Studien mindestens mittelhohe und teils sehr hohe Validitätskoeffizienten (z. B. Deary et al., 2007; Jensen, 1998; Lubinski, 2004; Rost, 2013; Schmidt & Hunter, 1998; Wilhelm & Schroeders, 2019). Weit verbreitet ist die Vermutung, die prädiktive Kraft der Intelligenztests sei auf proximale Kriterien beschränkt (z. B. Sternberg & Williams, 1997). Diese Vermutung ist unzutreffend; so zeigt z. B. die Metaanalyse von Strenze (2007), dass Intelligenzwerte von Kindern und Jugendlichen unter 18 Jahren auch nach über 11 Jahren substantiell mit Bildungs- und Karriereerfolg korrelieren ( $\rho \approx .50$ ; Überblick: Reeve & Bonaccio, 2011). Nichtsdestotrotz sind Vorbehalte und Kritik an In-

telligenztests verbreitet. Beispielhaft soll hier kurz auf die geringe Einsatzhäufigkeit sowie die teils geringe Akzeptanz von Intelligenztests, die geäußerte Kritik an Testaufgaben, Aspekte der Testfairness sowie am mit Intelligenztests (vermeintlich) verbundenen Weltbild aufgeführt werden.

In der Praxis bestehen hinsichtlich der Anwendung von Intelligenztests teils deutliche Vorbehalte. Ein besonders beeindruckendes Beispiel ist der Bereich der Personalauswahl. Die sehr hohe Treffsicherheit von Intelligenztest-basierten Vorhersagen des Ausbildungs-, Trainings- und Berufserfolgs konnte wiederholt in Metaanalysen mit nordamerikanischen (Schmidt & Hunter, 1998), europäischen (Salgado et al., 2003) und deutschen Daten (Hülshager, Maier & Stumpp, 2007; Kramer, 2009) gezeigt werden. Dessen ungeachtet werden Intelligenztests in der Personalauswahl seit jeher vergleichsweise zurückhaltend eingesetzt. Das gilt für den deutschsprachigen Bereich (Arnoneit et al., 2020) und darüber hinaus (USA, Kanada: Risavy et al., 2019). Häufigkeiten in Umfragen rangieren zwischen 5 % (Diekmann & König, 2015) und 35 % (Benit & Soellner, 2013), wobei zumeist nur erfasst wurde, ob Organisationen Intelligenztests in ihrem Anwendungsrepertoire haben, nicht aber, wie häufig sie die Verfahren pro Stelle nutzen.

Die Augenscheinvalidität ist ein maßgeblicher Faktor für die (geringe) soziale Akzeptanz diagnostischer Verfahren. Die erwähnte seltene Einsatzhäufigkeit von Intelligenztests in der Personalauswahl wird oft auch darauf zurückgeführt, dass diese Tests von Personalverantwortlichen sowie Bewerberinnen und Bewerbern nicht gut akzeptiert würden. Nach der Metaanalyse von Anderson, Salgado und Hülshager (2010) erzielen Intelligenztests im Gesamturteil ein positives Akzeptanzurteil, werden aber kritischer bewertet als Arbeitsproben, Interviews und Lebenslaufanalysen. In den Primärstudien wurde allerdings zumeist nicht kontrolliert, inwiefern die Befragten die zu bewertenden Verfahren aus eigener Erfahrung kennen. In der Regel wird das Verfahren lediglich kurz erläutert. Es handelt sich also nicht um das Akzeptanzurteil: Die Befragungen spiegeln das „Image“ der Auswahlverfahren wider, das unbestreitbar bedeutsam ist, da es die Entscheidung einer Person zur Bewerbung oder Verfahrensteilnahme beeinflussen kann. Allerdings korreliert das nach einer Kurzbeschreibung der Verfahren gemessene Image der Intelligenztests nur gering mit den Akzeptanzurteilen, die nach intensiverer Auseinandersetzung mit den Verfahren erfolgen (Marcus, 2003). Unserer Kenntnis nach fehlen Studien, bei denen Bewerbende in echten Bewerbungsverfahren im within-subject design die Akzeptanz von Intelligenztests im Vergleich zu anderen Auswahlverfahren beurteilen.

Entscheidend für den Einsatz von Intelligenztests sind aber nicht (nur) die tatsächliche Akzeptanz, die die Ver-

fahren bei den Bewerberinnen und Bewerbern finden, sondern (auch) die Vermutungen über diese Akzeptanz, die diejenigen anstellen, die in Organisationen die Personalauswahl gestalten und verantworten. Diesbezüglich schneiden Intelligenztests schlecht ab (Schuler et al., 2007). Darüber hinaus wird die Validität von Intelligenztests von Personalverantwortlichen unzutreffend eingeschätzt. So stimmen mehr als die Hälfte der unzutreffenden Aussage zu, dass Gewissenhaftigkeit ein besserer Prädiktor von beruflicher Leistung als Intelligenz ist (Fisher et al., 2021).

In deutlichem Gegensatz zum Personalbereich beinhaltet eine umfassende psychologische Diagnostik im Schul- und Bildungsbereich häufig eine Intelligenzdiagnostik (z. B. in der Schulpsychologie; vgl. Gold, Gawrilow & Hasselhorn, 2016). Jedoch wird der Intelligenz sowie einer Anwendung von Intelligenztests auch im Schul- und Bildungsbereich häufig Skepsis entgegengebracht – insbesondere durch Personen ohne wissenschaftlichen Psychologie-Hintergrund. In den großen Studien der Bildungsforschung werden in der Regel nur wenige Intelligenzsubtests und Intelligenztestitems eingesetzt (z. B. PISA 2000: 2 Subtests, vgl. Brunner, 2006; vgl. Saß, Kampa & Köller, 2017), was für eine überzeugende und breite Intelligenzmessung nicht ausreicht. In den USA wird die Intelligenzskepsis im Schulwesen mit einer Gemengelage verschiedener Gründe erklärt; dies sind unter anderem eine tradierte Vorstellung, der zufolge Intelligenz als wenig veränderbar verstanden und mit einem traditionellen Eliteverständnis in Verbindung gebracht wird, was in gewissem Widerspruch zu einem pädagogischen Veränderbarkeitsoptimismus steht, sowie der Überzeugung, dass Schulen und Unterricht wie eine *Tayloristisch* geführte Fabrik optimal funktionierten (z. B. Maranto & Wai, 2020).

Eine ähnlich lange Tradition hat die Kritik an Intelligenztestaufgaben. Diese ließen, so die Kritik, Augenscheinvalidität vermissen, das Aufgabenmaterial in nichtsprachlichen Intelligenztestaufgaben (z. B. figurale Matrizenaufgaben) wirke z. T. artifiziell, einfach oder veraltet. Trotz vermeintlich „einfacher“ Aufgaben sind die für eine richtige Lösungsfindung beispielsweise figuraler Matrizenaufgaben erforderlichen kognitiven Prozesse – wie induktive Prozesse beim Erkennen der zugrunde liegenden Regeln (*rule induction*), die koordinierte Bearbeitung von Zwischenzielen bis zur Aufgabenlösung (*goal management*) sowie allgemeine exekutive und visuell-räumliche Prozesse – anspruchsvoll (vgl. z. B. Carpenter, Just & Shell, 1990; Krieger et al., 2019; Kovacs & Conway, 2016); zudem weisen beispielsweise figurale Matrizentests sehr gute kriteriumsbezogene und prognostische Validitätskoeffizienten auf (z. B. Jensen, 1998). Gelegentlich wird zudem die Bildungsabhängigkeit von Intelligenztestaufgaben kritisiert. Dies betrifft insbesondere Aufgaben zur Mes-

sung der *Kristallinen Intelligenz*, womit nach Horn und Noll (1997) vor allem das mittels Akkulturation aufgebaute Wissen gemeint ist. In entsprechenden Intelligenzsubtests wird jedoch weniger Wissen in einem engen Spezialthema, sondern Wissen in verschiedenen Themen erfragt, wie es in Lerngelegenheiten in z.B. Schule und Alltagserfahrungen erworben werden kann. Intelligenteren Personen gelingt es bei vergleichbaren Lerngelegenheiten besser, sich einen entsprechenden Wissenskorporus anzueignen. Kristalline Subtests sind, was die prognostische und kriteriumsbezogene Validität angeht, ebenfalls als sehr gut einzuschätzen (z.B. Jensen, 1998). Einschränkend ist jedoch die Voraussetzung vergleichbarer Lerngelegenheiten zu berücksichtigen (vgl. Stern & Neubauer, 2016). An Intelligenztestaufgaben wird zudem gelegentlich kritisiert, dass im „wirklichen“ Leben bedeutsame Aspekte, wie das Lösen sich über die Zeit verändernder, komplexer Probleme, keine Rolle spielten (Dörner, 1976). Komplexes Problemlösen korreliert jedoch relativ eng mit Intelligenz: In einer Metaanalyse resultierte ein mittlerer Zusammenhang von  $M(g) = .43$  (Stadler et al., 2015); bei breiter Operationalisierung der Intelligenz und psychometrisch überzeugender Erfassung komplexen Problemlösens liegen die latenten Korrelationen teils noch höher ( $r = .85$  bei Kretschmar, Neubert, Wüstenberg & Greiff, 2016;  $r \approx .70$  bei Lotz, Sparfeldt & Greiff, 2016). Über Intelligenz hinaus sind zweifelsohne bspw. nicht-kognitive Personenmerkmale für Erfolg in unserer Gesellschaft relevant. Im Sinne einer differenzierten Diagnose sollten diese Variablen – bezogen auf die jeweilige diagnostische Fragestellung – mit weiteren Verfahren ergänzend erfasst werden.

Bedenken bestehen auch hinsichtlich der Testfairness: Werden Angehörige bestimmter Gruppen bei der Vergabe knapper Güter wie Studienplätzen benachteiligt? Dies wird u.a. für Frauen, Personen aus Kulturen, in denen die Tests nicht entwickelt wurden, Personen mit niedrigem sozioökonomischen Status (z.B. Strenze, 2007) sowie Personen, die über keine allgemeine oder spezifische Testerfahrung verfügen, diskutiert (Cohen, 2006; Scharfen et al., 2018). Die Debatte über die Fairness von Intelligenztestbasierten Entscheidungen verläuft häufig fruchtlos, weil in den Argumentationen der Begriff „Fairness“ mit jeweils unterschiedlicher Bedeutung genutzt wird. Gesellschaftspolitisch hohe Bedeutung haben Quotenmodelle. Die Forderung, bei Auswahlentscheidungen gesellschaftlich relevante Gruppen nach der Vorgabe von Quoten zu berücksichtigen, führt häufig zu der unsachlichen Auffassung, dass ein Intelligenztest, der in diesem Kontext eingesetzt wird, nur dann ein „guter“ Test ist, wenn sich anhand der Testergebnisse die gewünschte Quote erzielen lässt. Die wissenschaftliche Psychologie legt hingegen mehrheitlich einen anderen Fairnessbegriff an. Nach dem

wissenschaftlich zumeist genutzten Fairness-Modell von Cleary (1968) ist ein Test dann fair, wenn Personen mit gleichen Erfolgsaussichten auch gleiche Auswahlchancen aufgrund des Testergebnisses zeigen. Personen sollten dieser Auffassung zufolge also nach Leistung und gerade nicht nach Gruppenzugehörigkeit ausgewählt werden – eine dem Quotenmodell diametral entgegenstehende Position. Unfair ist ein Test in diesem Sinne dann, wenn die Vorhersagbarkeit von Kriterien (z.B. Studienleistungen) durch den Test für verschiedene Personengruppen (z.B. Frauen und Männer) ungleich ausfällt. Entsprechende Effekte der Unfairness bestehen, sind aber vergleichsweise klein. So zeigte sich beispielsweise in der Metaanalyse von Fischer, Schult und Hell (2013) ein nur geringer Vorhersagebias bei der Prognose von Studienleistungen von Frauen und Männern durch Studierfähigkeitstests. Durch den Test wurde die Studienleistung von Frauen auf einer Skala von 0 bis 4 um ca. 0,24 Punkte unterschätzt.

Die auf dem Quotenmodell basierende Fairnessauffassung läuft darauf hinaus, dass die Mittelwerte (sowie weitere Verteilungskennwerte) für definierte Subgruppen (z.B. Geschlecht, regionale Herkunft) gleich ausfallen. Die damit verbundene Bedeutung von Mittelwertunterschieden zwischen Gruppen ist auch vor dem Hintergrund der statistischen Auswertung zu problematisieren; Gruppenvergleiche anhand von einfachen Summenwerten, Faktorwerten oder latenten Mittelwerten können nämlich zu unterschiedlichen Ergebnissen führen. So zeigten Steinmayr, Beauducel und Spinath (2010) im Vergleich der geschlechtsspezifischen Intelligenzleistungen bei Frauen höhere Mittelwerte der latenten „verbalen Intelligenz“ und bei Männern höhere Mittelwerte in den ungewichteten Summenwerten der verbalen Intelligenz. Allerdings haben fast alle Studien zu Mittelwertunterschieden zwischen Gruppen mit ungewichteten Summenwerten gearbeitet und Auswirkungen unterschiedlicher Berechnungsmethoden nicht geprüft (z.B. Johnson & Bouchard, 2007; Lynn & Irwing, 2004; Rushton & Jensen, 2005). Auch wurden die seit langem verfügbaren Methoden zur Ermittlung der Äquivalenz von Messverfahren in verschiedenen Gruppen (Meredith, 1993) in den genannten Studien nicht betrachtet. Messäquivalenz bedeutet, dass beobachtete Variablen latente Konstrukte in verschiedenen Gruppen in gleicher Weise messen. Wenn Messäquivalenz gegeben ist, hängt die Beziehung zwischen Konstruktausprägungen und Antwortwahrscheinlichkeiten nicht von der Gruppenzugehörigkeit ab. Eine ausreichende Messäquivalenz zwischen Gruppen stellt eine zentrale Voraussetzung für eine angemessene Interpretation von Gruppenunterschieden dar; tatsächlich ist deren Nicht-Berücksichtigung ein Problem vieler bisheriger Studien zu Gruppenunterschieden (siehe z.B. Liu & Lynn, 2011). Zudem weisen unterschiedliche Gruppen häufig unterschiedliche Um-

weltbedingungen und Lernerfahrungen auf; es stellt damit stets eine Herausforderung dar, dokumentierte Gruppenunterschiede sorgfältig und angemessen zu interpretieren. Es ist unangemessen und schädlich, den großen prädiktiven Nutzen von Intelligenztestergebnissen zu diskreditieren, nur weil mitunter eine methodisch unausgereifte Ermittlung von Gruppenunterschieden instrumentalisiert wurde oder weil ein Mangel an Kenntnissen einer Kompensation gruppenspezifischer Benachteiligungen besteht.

Die Begriffe „Intelligenz“ und „Intelligenztest“ werden außerdem immer wieder als Sinnbild eines bestimmten Weltbildes aufgefasst, das die gesellschaftlichen Bedingungen von (Un-)Gleichheit ignorieren und sich allein auf messbare Merkmale des Individuums konzentrieren würde (z. B. Holzkamp, 1993). Das Intelligenzkonstrukt, das von einer Normalverteilung, Stabilität und einem gewissen genetischen Einfluss (s. u.) ausgeht, ist bei vielen unbeliebt. Das Negativimage von Intelligenztests dürfte auch mit ihrer Anwendung zu tun haben: Intelligenztests werden häufig in Auswahl-situationen genutzt, etwa bei der Vergabe von Ausbildungs-, Studien- oder Arbeitsplätzen. Erhält man ein begehrtes Gut wie den Wunsch Arbeitsplatz nicht, dient es dem Selbstwert, die verwendete Auswahlmethode abzuwerten. Es drängt sich der Eindruck auf, dass sich die gesellschaftspolitische Kritik zwar an Intelligenztests entzündet, tatsächlich aber eine Unzufriedenheit über die Knappheit begehrter Güter wie bestimmte Studienplätze zum Ausdruck bringt.

Um der erwähnten Kritik – insbesondere an einzelnen Validitätsaspekten und der Fairness von Intelligenztests – zu begegnen, empfiehlt sich ein Vergleich mit alternativen Verfahren, die beispielsweise in der Personalauswahl eingesetzt werden (z. B. Situational Judgment Tests, McDaniel, Hartman, Whetzel & Grubb, 2007). So sollten Schwächen der verschiedenen Verfahren bzw. Verfahrensklassen benannt und das für die Beantwortung der Fragestellung am besten geeignete Verfahren eingesetzt werden. Es sollte jedoch deutlich geworden sein, dass ein Verzicht auf eine Intelligenzdiagnostik als Teil einer umfassenden psychologischen Diagnostik für die Beantwortung vieler praktischer (z. B. schulbezogener und personalpsychologischer) Fragestellungen eine deutliche Qualitätsabnahme der Validität und Fairness der darauf aufbauenden Empfehlungen implizieren würde. Auch der Diskussion um das Weltbild, das nach Ansicht der Kritik mit Intelligenztests verbunden ist, sollte man nicht aus dem Weg gehen, sondern die Vor- und Nachteile der Möglichkeiten, Verteilungsprobleme wie der Vergabe von Arbeitsplätzen anzugehen, transparent offenlegen. Kritik ist berechtigt; aber man sollte nicht bei der Kritik stehen bleiben, sondern konkret aufzeigen, wie es besser gehen könnte.

## Darstellung empirischer Realitäten aus der Perspektive ausgewählter Theorien

Ein Problem, mit dem sich die Intelligenzforschung immer wieder konfrontiert sieht, ist die Art und Weise, wie empirische Befunde dargestellt und rezipiert werden. Dabei ergeben sich insbesondere zwei typische Muster.

Beim ersten Muster werden empirische Befunde mit theoretischen Überlegungen vermischt. Als Folge davon können Leserinnen und Leser nicht erkennen, zu welchem Anteil ein Befund auf empirischen Realitäten beruht und zu welchem Anteil ein Befund das Resultat der Anwendung einer spezifischen Theorie zur Intelligenz ist. Ein wichtiges Beispiel für dieses Muster ist der häufig rezipierte Befund, dass die allgemeine Intelligenz der wichtigste Prädiktor von praxisrelevanten Kriterien wie Schul- (Jensen, 1998), Berufs- (Ree & Carretta, 2002) und Karriereerfolg (Gottfredson, 1998) ist und spezifische Fähigkeiten – so diese Sichtweise – dem gegenüber vernachlässigt werden können. Was jedoch oft nicht offengelegt wird, ist die Tatsache, dass der Befund auf der theoretischen Vorannahme von Spearman (Lang & Kell, 2020; Lang, Kersting, Hülshager & Lang, 2010) basiert, der zufolge die gesamte gemeinsam (zwischen allgemeiner Intelligenz und spezifischeren Faktoren) erklärte Varianz der allgemeinen Intelligenz zugeschlagen wird. Aus intelligenztheoretischer Sicht benötigt die Entscheidung, die allgemeine Intelligenz als prioritär zu behandeln, eine überzeugende Begründung. Die Literatur kennt nämlich mehrere Intelligenzmodelle, die einen umgekehrten Kausalbezug annehmen (Überblick: Lang & Kell, 2020; Lang et al., 2016) oder argumentieren, dass es keine kausale Richtung geben sollte, wie etwa in Modellen mit geschachtelten Faktoren (Lang et al., 2010; Schult & Sparfeldt, 2016; Valerius & Sparfeldt, 2014, 2015; Wee, 2018) oder mit Hierarchien von Faktorenlösungen (Lang et al., 2016).

Abbildung 1 illustriert anhand einer Studie zur Vorhersage von Karriereerfolg über einen Zeitraum von 50 Jahren, dass Analysen mit Spearman's Vorannahme zu sehr anderen Ergebnissen kommen können als Analysen ohne die Vorannahme. Im Einklang mit Spearman's Vorannahme zeigt Abbildung 1 A Ergebnisse einer inkrementellen Validitätsanalyse, bei der die allgemeine Intelligenz zuerst aufgenommen wurde. Im Einklang mit anderen Befundübersichten (Beier, Kell & Lang, 2019; Kell & Lang, 2017; Mainert, Niepel, Murphy & Greiff, 2019) erklären die spezifischen Faktoren dann zwar mitunter relevante, aber in der Gesamtschau nur unwesentliche zusätzliche Varianz. Abbildung 1 B zeigt alternativ eine relative Bedeutungsanalyse (Braun, Converse & Oswald, 2019; Lang & Kell, 2020; Wee, 2018) ohne a-priori Annahme

zur Wichtigkeit der Prädiktoren (für alternative Ansätze ohne a-priori Annahmen siehe Eid, Krumm, Koch & Schulze, 2018; Zhang, Sun, Cao & Drasgow, 2020). Im Einklang mit anderen Befunden in der Literatur auf der Basis dieser Methode (Beier et al., 2019; Kell & Lang, 2017; Lang et al., 2010; Wee, 2018) ist die allgemeine Intelligenz ein bedeutsamer, aber nicht klar der wichtigste Prädiktor in Abbildung 1 B – obwohl die Analysen in Abbildung 1 A und 1 B auf den gleichen Daten basieren. Der Befund, dass die allgemeine Intelligenz nicht unbedingt der wichtigste Prädiktor vieler relevanter Kriterien ist, hat große Implikationen für die Praxis: Allgemeine Intelligenz ist in vielen Fällen sicherlich ein sinnvoller Prädiktor, aber nicht alternativlos. In der Praxis kann zwischen verschiedenen alternativen Prädiktoren gewählt werden, und weitere Gesichtspunkte wie Unterschiede zwischen gesellschaftlichen Gruppen auf dem Prädiktor oder die Akzeptanz des Messinstruments bei den Teilnehmenden könnten die Wahl mitbestimmen. Zum Beispiel hat eine Befragung von Studierenden ergeben, dass diese studienfachspezifische Fähigkeitstests (die in der Regel spezifischere Intelligenzdimensionen erfassen) einem allgemeinen Intelligenztest vorziehen (Stegt, Didi, Zimmerhofer & Seegers, 2018).

Bei dem zweiten problematischen Muster handelt es sich um selektive Darstellungen von Einzelaspekten. Dabei werden empirische Befunde hervorgehoben, die nicht repräsentativ für die Gesamtbefundlage sind. Ein wichtiges Beispiel ist die Nutzung von Messinstrumenten, die nur einen sehr spezifischen Teilaspekt der Intelligenz erfassen (z. B. figurale Intelligenz, teils erfasst mit nur einem oder wenigen Subtests), um daran anschließend weitreichende Schlussfolgerungen zu ziehen, die – häufig ungeprüft – auf breite Intelligenzgruppenunterschiede in verschiedenen Intelligenzfacetten generalisiert werden (z. B. Cronshaw, Hamilton, Onyura & Winston, 2006; Rushton & Skuy, 2000). Ein weiteres Beispiel für selektive Darstellungen von Einzelaspekten ist Kritik an Oberflächenmerkmalen oder spezifischen Aspekten eines Messinstruments (einzelne Formulierungen oder Items), die aber nicht unbedingt eine realistische Einschätzung der Gesamteffizienz des Messinstruments wiedergeben (z. B. Lorenz, 1988).

Derartige Muster zu erkennen, erfordert oft eine hohe Detailkenntnis der Literatur. Selektive Darstellungen sind besonders problematisch, wenn damit der Eindruck erweckt wird, dass die Intelligenzforschung immer wieder unterschiedliche Ergebnisse liefert. In Wahrheit nimmt die Intelligenzforschung innerhalb der Psychologie eine Sonderstellung ein, weil die meisten empirischen Befunde in der Intelligenzforschung immer wieder über verschiedene Datensätze hin repliziert worden sind. Abweichende Ergebnisse sind häufig das Ergebnis von divergierenden theoretischen und methodischen Überlegungen bei

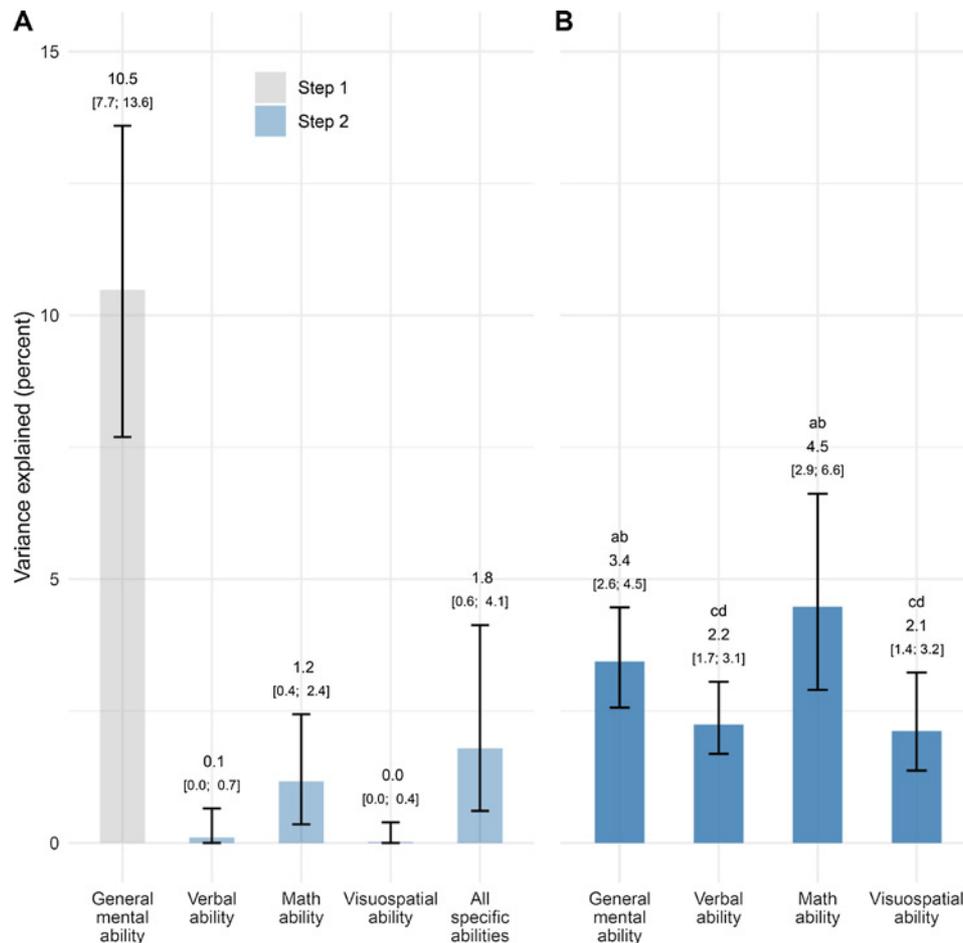
der Analyse der Daten zwischen verschiedenen Autorentams (siehe auch Abbildung 1).

## **Bedeutung von Umwelteinflüssen bei gleichzeitig hohen Erblichkeitskoeffizienten**

Bekanntermaßen spiegelt das Ergebnis einer Person in einem Intelligenztest auch frühere Lern- und Umwelterfahrungen wider. Betrachtet man Umwelteinflüsse auf Intelligenztestergebnisse, scheint eine konzeptuelle Unterscheidung bedeutsam zu sein: (a) eher kurzfristige und eng umgrenzte Steigerungen des Ergebnisses in einem spezifischen Intelligenz(sub)test, (b) Versuche, mittels umgrenzter Intelligenztrainings Intelligenzsteigerungen zu erzielen, (c) umweltbedingte Veränderungen der Intelligenz aufgrund umfangreicher Umweltveränderungen (z. B. Adoption, längerer Schulbesuch). An dieser Stelle muss aus den vielen Befunden zu vorübergehenden bzw. dauerhaften sowie intelligenzmindernden (z. B. Konsum spezifischer Substanzen, Kopfverletzungen) bzw. intelligenzsteigernden Umwelteinflüssen ausgewählt werden.

(a) Eher kurzfristige und eng umgrenzte Steigerungen des Ergebnisses: Kurzfristige Steigerungen der Ergebnisse in einem spezifischen Intelligenztest von kleiner bis mittlerer Effektgröße sind durch wiederholte Testbearbeitungen, intensives Üben von Intelligenztestaufgaben oder Coaching zu erzielen (z. B. Hausknecht, Day & Thomas, 2007). So wurde metaanalytisch ein mittlerer Test-Wiederholungseffekt bei Verwendung der identischen Testversion berichtet, der in der IQ-Skalierung einem Zuwachs von 6 bis 7 IQ-Punkten entspricht; ab der dritten Testwiederholung ergaben sich keine bedeutsamen Punktzuwächse mehr (Scharfen et al., 2018). Werden – über reine Testwiederholungseffekte hinaus – bspw. im Rahmen kurzer Video-Tutorials die figuralen Matrizenaufgaben zugrundeliegenden Regeln erläutert, zeigten sich im Vergleich zu einem Kontroll-Video substantiell bessere Testleistungen in figuralen Matrizentestaufgaben (z. B. Levacher et al., 2022; Schneider et al., 2020). Bei derartigen Testwertsteigerungen wird nicht von damit einhergehenden Steigerungen der Intelligenz und erst recht nicht von direkten Effekten auf Bildungs- und Berufserfolg ausgegangen; diskutiert werden allerdings Verzerrungen in Selektionsentscheidungen sowie Gegenmaßnahmen wie ausführliche Erläuterungen der zugrundeliegenden Regeln für alle Testpersonen (z. B. Levacher et al., 2022; Schneider et al., 2020).

(b) Versuche, mittels umgrenzter Intelligenztrainings Intelligenzsteigerungen zu erzielen: Im deutschen Sprach-



**Abbildung 1.** Inkrementelle Validitätsanalyse und relative Bedeutungsanalyse mit Intelligenz in der Schulzeit (Schuljahr 1959 – 1960) als unabhängige und Einkommen im Jahr 2011 – 2012 als abhängige Variable (aus Lang & Kell, 2020).

raum zeigten die theoretisch begründeten und breit evaluierten Denktrainings von Klauer (z. B. 1991) einer aktualisierten Metaanalyse (Klauer, 2014) zufolge mittelgroße Fördereffekte auf Testaufgaben fluider Intelligenz – bei mittelfristiger Stabilität der Fördereffekte – sowie Aspekte schulischen Lernens. Die vielfältigen Bemühungen, mithilfe umgrenzter Intelligenztrainings nachhaltige, substantielle sowie lebenspraktisch bedeutsame Intelligenzsteigerungen zu erzielen, sind vor dem Hintergrund der Literatur jedoch als in ihrer Gesamtheit nicht erfolgreich zu beurteilen (im Überblick z. B. Rost, 2013). So finden sich nach Ende der Intervention zwar manchmal bedeutsam höhere Intelligenztestleistungen im Vergleich zu nichttrainierten Kindern, diese Vorteile verschwinden mit der Zeit jedoch häufig wieder (z. B. Protzko, 2015).

(c) Umweltbedingte Veränderungen der Intelligenz aufgrund umfangreicher Umweltveränderungen: Für erhebliche Umwelteinflüsse auf die Intelligenzleistung sprechen auch Studien, die bspw. Adaptionen in intellektuell anregende und entwicklungsförderliche Umwelten oder Migration betrachtet haben (vgl. Sauce & Matzel, 2018).

Gleichermaßen hängt die Schulbesuchsdauer bedeutsam mit der Intelligenz zusammen (z. B. Stelzl, Merz, Ehlers & Remer, 1995; Ceci & Williams, 1997), wobei ein zusätzliches Schuljahr mit Intelligenzsteigerungen von bis zu 4 IQ-Punkten (vgl. Rost, 2013) bzw. – so eine aktuelle Metaanalyse – in Abhängigkeit vom Untersuchungsdesign zwischen im Mittel 1 und 5 IQ-Punkten (Ritchie & Tucker-Drob, 2018) einhergeht. Als Erklärung liegt nahe, dass in der Schule eine intelligenzförderliche, also intensive, längerfristige, dauerhafte und kontinuierliche sowie systematische und strukturierte Auseinandersetzung mit unterschiedlichen intellektuellen Anregungen und Problemstellungen in verschiedenen Inhaltsbereichen sowie ein breiter Wissensaufbau und Kompetenzzuwachs stattfindet (z. B. Rost, 2013). Zudem wurden größere Intelligenzzuwächse beim Besuch des intellektuell anregungsreicheren Gymnasiums als nicht-gymnasialer Schulformen berichtet (z. B. Guill, Lüdtke & Köller, 2017). Derartige Intelligenzzuwächse dürften den weiteren Bildungs- und Berufsweg nachhaltig positiv beeinflussen.

In der Zusammenschau der drei Ansätze und entsprechend unterschiedlicher Interpretationen als entweder Steigerung der Testwerte in einem spezifischen Intelligenzsubtest oder Steigerungen der Intelligenz im Sinne eines breiteren Intelligenzfaktors verdient die Abgrenzung von Intelligenztestleistung und Intelligenz, also die Unterscheidung von Mess- und Konstruktebene, Beachtung. Aus angewandter Perspektive bieten die dokumentierten Umwelteinflüsse vielversprechende psychologische Ansatzpunkte.

Ein ganz anderer Aspekt, der ebenfalls die Relevanz von Umwelteinflüssen betrifft, ergibt sich aus der Forschung zur Erbllichkeit von Intelligenztestwerten. Neben den erwähnten Umwelteinflüssen sind auch die Heritabilitätskoeffizienten für die allgemeine Intelligenz bedeutsam, die auf etwa 50 % erbliche Varianz hindeuten (Bouchard, 1997; Plomin & von Stumm, 2018; bis ca. 70 %, vgl. Rost, 2013). Konzeptuell liegt als Erklärung für einerseits bedeutsame Heritabilität und andererseits bedeutsame Umwelteinflüsse der Intelligenz nahe, neben den entsprechenden Haupteffekten insbesondere das Zusammenspiel – also Gen-Umwelt-Korrelationen sowie entsprechende Gen-Umwelt-Interaktionen – in den Blick zu nehmen (z. B. Sauce & Matzel, 2018). Heritabilitätskoeffizienten beschreiben nicht den Anteil kausaler Wirkung von Genen auf ein Merkmal, sondern den Anteil der Merkmalsvarianz, der auf erbliche Wirkungen zurückgeführt werden kann. So ist auch erklärbar, dass Heritabilitätskoeffizienten mit dem Lebensalter zunehmen (Plomin & Deary, 2015) und dass der sozioökonomische Status nur in einigen Ländern mit der Heritabilität von Intelligenztestergebnissen interagiert (Tucker-Drob & Bates, 2016). Die statistische Modellierung von Varianzanteilen, die durch Gene, Umwelt oder Gen-Umwelt-Interaktionen hervorgerufen werden, ist anspruchsvoll (Franić, Dolan, Borsboom & Boomsma, 2012), was einer leichtfertigen Interpretation derartiger Befunde im Rahmen von manchen gesellschaftspolitischen Diskursen entgegensteht. In diesem Zusammenhang wird auch vom Intelligenz-Paradox gesprochen: Intelligenz ist gleichzeitig erblich *und* kann stark von Umwelteinflüssen verändert werden (Dickens & Flynn, 2001; Loehlin, 2002). Dies kann durch die angesprochenen Interaktionseffekte erklärt werden, die sich über die Zeit „aufschaukeln“ können.

Während die populationsgenetische Intelligenzforschung seit den 1990er Jahren als vergleichsweise konsolidiert gelten kann, wurde eine molekulargenetische Untersuchung von Intelligenzunterschieden erst ab da möglich. In der Regel werden in molekulargenetischen Studien Einzelnukleotid-Polymorphismen mit Intelligenztestwerten assoziiert. Einzelnukleotid-Polymorphismen sind Abschnitte auf der DNA (Desoxyribonucleinacid, Träger der Erbinformation in den Chromosomen), auf dem

Menschen sich in einem einzelnen Basenpaar unterscheiden (siehe Montag, 2018). Zwischen 1990 und 2017 wurden allerdings keine replizierbaren Assoziationen zwischen molekulargenetischen Polymorphismen und Intelligenz festgestellt (Plomin & von Stumm, 2018). Dies wurde so interpretiert, dass sehr viele Polymorphismen sehr gering mit Intelligenz assoziiert sind. Erst eine Metaanalyse auf der Basis von über 70.000 Probanden konnte bedeutsame genom-weite Assoziationen mit Intelligenz aufdecken und in 18-genomweit signifikanten Regionen 3 % der Intelligenzvarianz erklären. Dabei waren eine sehr große Anzahl von Einzelnukleotid-Polymorphismen (> 300) mit Intelligenzunterschieden assoziiert (Sniekers et al., 2017). Die molekulargenetisch erklärte Intelligenzvarianz ist trotz der großen Zahl einzelner Prädiktoren deutlich geringer, als man vor dem Hintergrund einer populationsgenetischen Erbllichkeit von ca. 50 % (Plomin & von Stumm, 2018) für die Intelligenz erwarten könnte. Somit dürften künftig an noch größeren Stichproben weitere relevante Assoziationen molekulargenetischer Polymorphismen mit Intelligenz aufgezeigt werden. Auch wenn die molekulargenetisch erklärbare Intelligenzvarianz im Vergleich zur populationsgenetisch erklärbaren Intelligenzvarianz sehr gering ist, nimmt sie – genau wie die populationsgenetisch erklärbare Intelligenzvarianz – von der frühen Kindheit über die Adoleszenz bis zum Erwachsenenalter zu (Plomin & von Stumm, 2018), was mit dem genannten Zusammenspiel von Genen und Umwelt erklärt werden kann (z. B. indem Menschen sich im Laufe ihres Lebens Umwelten „suchen“, die zu ihrer genetisch prädisponierten Intelligenz passen). So wie das Zusammenspiel von Anlage- und Umweltfaktoren hohe Erbllichkeit *und* deutliche Veränderbarkeit von Intelligenz erklären kann, deuten auch die molekulargenetischen Befunde insgesamt auf eine komplexe Determiniertheit der Intelligenzvarianz hin.

## Schlussfolgerungen

Unser Überblick verdeutlicht die wissenschaftliche Qualität und Tragweite der Intelligenzforschung sowie den praktischen Nutzen der Anwendung von Intelligenztests. Dabei zeigt sich durchgehend, dass ein beachtlicher methodischer und inhaltlicher Kenntnisstand für eine angemessene Befund-Interpretation und für wissenschaftlich tragfähige Schlussfolgerungen essentiell ist. Eine angemessene Ergebnisinterpretation erfordert hohe Methodenkompetenzen und Fachkenntnisse. Aus Missverständnissen und oberflächlichen sowie einseitigen Befundrezptionen können Abwehrreaktionen und gesellschaftspolitische Instrumentalisierungen entstehen. Da Intelligenztest-

ergebnisse bei wichtigen Entscheidungen eine erhebliche Rolle spielen können, können vereinfachende Darstellungen der methodisch komplexen Ergebnisse erhebliche Bedenken hervorrufen. Diese Problematik, die sich insbesondere bei verkürzten Darstellungen und populärwissenschaftlichen Aufarbeitungen ergeben kann, ist jedoch nicht dem (multidimensionalen) Intelligenzkonstrukt, den Intelligenztests oder der Intelligenzforschung per se anzulasten. Vor diesem Hintergrund sollen die folgenden Aspekte nochmals betont werden, um Missverständnissen im Rahmen des gesellschaftlichen Diskurses entgegen zu wirken:

- Trotz hervorragender Reliabilitäts- sowie u. a. prognostischer und kriteriumsbezogener Validitätshinweise von (multidimensionalen) Intelligenztests sind diese nicht perfekt; eine (weitere) Optimierung stellt eine Herausforderung für Forschung und Testentwicklung dar. Darüber hinaus zeigten Interventionen deutliche Testwertveränderungen; diese können mit Verzerrungen in z. B. Selektionsentscheidungen einhergehen (z. B. Auswahl von Testpersonen mit „wahrem“ Wert knapp unterhalb und entsprechend gesteigertem Wert oberhalb der Selektionsgrenze).
- Auch wenn entsprechende prognostische und kriteriumsbezogene Validitätskoeffizienten vieler mehrdimensionaler Intelligenztests sehr hoch sind, sind auch diese nicht perfekt; es handelt sich also stets um einen probabilistischen Zusammenhang zwischen Intelligenz und Leistungskriterien.
- Unterschiede in *real-life* Kriterien sind in der Regel multikausal bedingt. Zweifelsohne ist Intelligenz ein (besonders) wichtiger, aber nicht der einzig wichtige Prädiktor für Kriterien wie Erfolg. Die relative Bedeutung der allgemeinen Intelligenz im Verhältnis zu spezielleren Intelligenzfaktoren ist keine absolute empirische Größe, sondern hängt erheblich von der gewählten Analyseverfahren ab.
- Hohe Intelligenz-Erblichkeitsschätzungen dürfen nicht darüber hinwegtäuschen, dass die Anteile der Umweltvarianz häufig in gleicher Größenordnung liegen und dass die Erblichkeitsschätzungen selbst auch von den Umweltbedingungen abhängen. Daher moderieren z. B. das Lebensalter und sozioökonomische Faktoren die Erblichkeitsschätzungen. Die molekulargenetisch erklärbare Intelligenzvarianz ist in bisherigen Metaanalysen deutlich geringer als die Erblichkeitsschätzungen. Diese Diskrepanz kann bisher nicht erklärt werden und deutet auf ein zukünftiges Forschungspotential hin.
- Mythen, Meinungen und Vorurteile gegenüber dem Intelligenzkonstrukt sowie gegenüber Intelligenztests gehen vielfach mit einer selektiven Darstellung und Rezeption der Forschungsergebnisse einher. Diesen sollte mit guter Forschung und entsprechend präzise darge-

stellten Forschungsbefunden entgegengetreten werden.

- Intelligenztests haben primär ein Akzeptanz- und kein Validitätsproblem. Diesem Akzeptanzproblem sollte durch Aufklärung begegnet werden. Vorbehalte gegenüber dekontextualisierten Intelligenztests lassen sich beispielsweise durch Informationen über die Vorteile dieser Abstraktion angehen. Alternativ kann man anstelle der Nutzung abstrakten Materials die Testaufgaben so gestalten, dass sie einen unmittelbaren, augenscheinlichen Bezug zur Wirklichkeit haben (z. B. Kersting, 2014). Schließlich sollten beispielsweise die Vor- und Nachteile verschiedener Auswahlinstrumente gegeneinander abgewogen werden. Es sollte verdeutlicht werden, dass ein Verzicht auf eine Intelligenzdiagnostik als Teil einer umfassenderen psychologischen Diagnostik eine deutliche Qualitätsminderung der Validität und Fairness sowie der darauf aufbauenden Empfehlungen impliziert.
- Interindividuelle Unterschiede sind in vielen Gebieten nicht ungewöhnlich (Sport, Musik, Schule); dokumentierte interindividuelle Intelligenz-Unterschiede erzeugen hingegen gelegentlich Unbehagen. Doch für viele Beratungsanliegen ist eine realistische Begabungseinschätzung wichtig, um Fehlentscheidungen zu minimieren. Auch wenn eine Gesellschaft, um erfolgreich zu sein, auf eine erfolgreiche Umsetzung von Intelligenzpotentialen in entsprechende Leistungen angewiesen ist, haben Intelligenz-Testergebnisse nichts mit dem Wert als Mensch zu tun. Zudem sind neben Intelligenz weitere Merkmale für relevante Leistungen bedeutsam.
- Mögliche Benachteiligungen von Personengruppen sind ein Grundproblem jedweder Auswahlverfahren, das oft auf gesellschaftliche Unterschiede zurückgeht und kein Spezifikum von Intelligenztests darstellt. Sofern das Angebot begehrter Güter (z. B. Medizinstudienplätze) knapper als die Nachfrage ist, muss geregelt werden, wie die Verteilung erfolgen soll. Aktuell lautet die Entscheidung häufig, dass die Vergabe unabhängig von der Gruppenzugehörigkeit nach Leistung entschieden werden soll. Dieses Vergabeprinzip wird mit Intelligenztests gut umgesetzt. Wer eine andere Verteilung – z. B. ein Quotenmodell – bevorzugt, sollte nicht die Intelligenztests, sondern die entsprechenden politischen Entscheidungen kritisieren.

Unser Überblick würdigt die beachtliche Qualität der Intelligenzforschung sowie der Intelligenztests und weist auf erhebliche Probleme, die durch Missverständnisse, undifferenzierte und einseitige Rezeptionen sowie gesellschaftspolitische Instrumentalisierungen von Ergebnissen entstehen können. Wir wollen mit unserem Beitrag zu einer Schärfung der Unterscheidung zwischen den

oben dargestellten Verständnis- und Darstellungsproblemen einerseits sowie der hohen Qualität der Intelligenzforschung und der Intelligenztests andererseits beitragen. Ein besseres Verstehen der Qualitäten und Grenzen der Intelligenzforschung und Intelligenztests sowie der Gründe für Missverständnisse und Instrumentalisierungen sollte zu einer weiterhin höchst erfolgreichen sowie verantwortungsvollen Forschung und Praxis beitragen.

## Literatur

- Anderson, N., Salgado, J. F. & Hülsheger, U. R. (2010). Applicant reactions in selection: Comprehensive meta-analysis into reaction generalization versus situational specificity. *International Journal of Selection and Assessment*, 18, 291–304.
- Arnoneit, C., Schuler, H. & Hell, B. (2020). Nutzung, Validität, Praktikabilität und Akzeptanz psychologischer Personalauswahlverfahren in Deutschland 1985, 1993, 2007, 2020. *Zeitschrift für Arbeits- und Organisationspsychologie*, 64, 67–82.
- Beier, M., Kell, H. & Lang, J. W. B. (2019). Commenting on the “Great Debate”: General abilities, specific abilities, and the tools of the trade. *Journal of Intelligence*, 7 (1), 5.
- Benit, N. & Soellner, R. (2013). Scientist-practitioner gap in Deutschland: Eine empirische Studie am Beispiel psychologischer Testverfahren. *Zeitschrift für Arbeits- und Organisationspsychologie*, 57, 145–153.
- Bouchard, T. J. (1997). IQ similarity in twins reared apart: findings and responses to critics. In R. J. Sternberg & E. L. Grigorenko (Eds.), *Intelligence: Heredity and Environment* (pp. 126–160). Cambridge University Press.
- Braun, M. T., Converse, P. D. & Oswald, F. L. (2019). The accuracy of dominance analysis as a metric to assess relative importance: The joint impact of sampling error variance and measurement unreliability. *Journal of Applied Psychology*, 104, 593–602.
- Brunner, M. (2006). *Mathematische Schülerleistung: Struktur, Schulformunterschiede und Validität*. Berlin: Humboldt-Universität. Verfügbar unter <https://edoc.hu-berlin.de/handle/18452/16132>
- Carpenter, P. A., Just, M. A. & Shell, P. (1990). What one intelligence test measures: A theoretical account of the processing in the Raven progressive matrices test. *Psychological Review*, 97, 404–431.
- Carroll, J. B. (1993). *Human cognitive abilities*. Cambridge University Press.
- Ceci, S. J. & Williams, W. M. (1997). Schooling, intelligence, and income. *American Psychologist*, 52, 1051–1058.
- Cleary, T. A. (1968). Test bias: Prediction of grades of Negro and white students in integrated colleges. *Journal of Educational Measurement*, 5, 115–124.
- Cohen, A. D. (2006). The coming of age of research on test-taking strategies. *Language Assessment Quarterly*, 3, 307–331.
- Cronshaw, S. F., Hamilton, L. K., Onyura, B. R. & Winston, A. S. (2006). Case for non-biased intelligence testing against Black Africans has not been made: A comment on Rushton, Skuy, and Bons (2004). *International Journal of Selection and Assessment*, 14, 278–287.
- Deary, I. J., Strand, S., Smith, P. & Fernandes, C. (2007). Intelligence and educational achievement. *Intelligence*, 35, 13–21.
- Dickens, W. T. & Flynn, J. R. (2001). Heritability estimates versus large environmental effects: The IQ paradox resolved. *Psychological Review*, 108, 346–369.
- Diekmann, J. & König, C. J. (2015). Personality testing in personnel selection: Love it? Leave it? Understand it! In I. Nikolaou & J. Oostrom (Eds.), *Employee recruitment, selection, and assessment: Contemporary issues for theory and practice* (pp. 117–135). Psychology Press.
- Dombrowski, S. C., Canivez, G. L. & Watkins, M. W. (2018). Factor structure of the 10 WISC-V primary subtests across four standardization age groups. *Contemporary School Psychology*, 22, 90–104.
- Dörner, D. (1976). *Problemlösen als Informationsverarbeitung*. Kohlhammer.
- Eid, M., Krumm, S., Koch, T. & Schulze, J. (2018). Bifactor models for predicting criteria by general and specific factors: Problems of nonidentifiability and alternative solutions. *Journal of Intelligence*, 6 (3), 42.
- Fischer, F., Schult, J. & Hell, B. (2013). Sex-specific differential prediction of college admission tests: A meta-analysis. *Journal of Educational Psychology*, 105, 478–488.
- Fisher, P. A., Risavy, S. D., Robie, C., König, C. J., Christiansen, N. D., Tett, R. P. & Simonet, D. V. (2021). Selection myths: A conceptual replication of HR professionals’ beliefs about effective human resource practices in the United States and Canada. *Journal of Personnel Psychology*, 20, 51–60.
- Franić, S., Dolan, C. V., Borsboom, D. & Boomsma, D. I. (2012). Structural equation modeling in genetics. In R. H. Hoyle (Ed.), *Handbook of structural equation modeling* (p. 617–635). Guilford Press.
- Gold, A., Gawrilow, C. & Hasselhorn, M. (2016). Grundlagen schulpsychologischer Diagnostik. In K. Seifried, S. Drewes & M. Hasselhorn (Hrsg.), *Handbuch Schulpsychologie* (2. Aufl., S. 117–127). Stuttgart: Kohlhammer.
- Gottfredson, L. S. (1997). Mainstream science on intelligence. An editorial with 52 signatories, history, and bibliography. *Intelligence*, 24, 13–23.
- Gottfredson, L. S. (1998). The general intelligence factor. *Scientific American Presents*, 9, 24–29.
- Guill, K., Lüdtke, O. & Köller, O. (2017). Academic tracking is related to gains in students’ intelligence over four years: Evidence from a propensity score matching study. *Learning and Instruction*, 47, 43–52.
- Hausknecht, J. P., Halpert, J. A., Di Paolo, N. T. & Moriarty Gerrard, M. O. (2007). Retesting in selection: A meta-analysis of coaching and practice effects for tests of cognitive ability. *Journal of Applied Psychology*, 92, 373–385.
- Herrnstein, R. & Murray, C. (1994). *The Bell Curve – Intelligence and Class Structure in America*. New York: Free Press.
- Holzkamp, K. (1993) *Lernen: Subjektwissenschaftliche Grundlegung*. Frankfurt/M.: Campus.
- Horn, J. L. & Noll, J. (1997). Human cognitive capabilities: Gf-Gc theory. In D. P. Flanagan, J. L. Genshaft & P. L. Harrison (Eds.), *Contemporary intellectual assessment. Theories, tests, and issues* (pp. 53–91). New York: Guilford Press.
- Hülsheger, U. R., Maier G. W. & Stumpp T. (2007) Validity of general mental ability for the prediction of job performance and training success in Germany: A meta-analysis. *International Journal of Selection and Assessment*, 15, 3–18.
- Jensen, A. R. (1998). The g factor and the design of education. In R. J. Sternberg & W. M. Williams (Eds.), *Intelligence, instruction, and assessment: Theory into practice* (pp. 111–131). Erlbaum.
- Johnson, W. & Bouchard, T. J. (2007). Sex differences in mental abilities: g masks the dimensions on which they lie. *Intelligence*, 35, 23–39.
- Kell, H. & Lang, J. W. B. (2017). Specific abilities in the workplace: More important than g? *Journal of Intelligence*, 5 (2), 13.
- Kersting, M. (2014). *Testmanual zum SMART – berufsbezogener Test zur kognitiven Kompetenz*. Mödling: Schuhfried.

- Kersting, M., Althoff, K. & Jäger, A. O. (2008). *WIT-2. Wilde-Intelligenz-Test 2*. Göttingen: Hogrefe.
- Klauer, K. J. (1991). *Denktraining für Kinder II*. Göttingen: Hogrefe.
- Klauer, K. J. (2014). Training des induktiven Denkens – Fortschreibung der Metaanalyse von 2008. *Zeitschrift für Pädagogische Psychologie*, 28, 5–19.
- Kovacs, K. & Conway, A. R. A. (2016). Process Overlap Theory: A Unified Account of the General Factor of Intelligence. *Psychological Inquiry*, 27, 151–177.
- Kramer, J. (2009). Allgemeine Intelligenz und beruflicher Erfolg in Deutschland: Vertiefende und weiterführende Metaanalysen. *Psychologische Rundschau*, 60, 82–98.
- Kretzschmar, A., Neubert, J. C., Wüstenberg, S. & Greiff, S. (2016). Construct validity of complex problem solving: A comprehensive view on different facets of intelligence and school grades. *Intelligence*, 54, 55–69.
- Krieger, F., Zimmer, H. D., Greiff, S., Spinath, F. M. & Becker, N. (2019). Why are difficult figural matrices hard to solve? The role of selective encoding and working memory capacity. *Intelligence*, 72, 35–48.
- 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, 1047–1061.
- Lang, J. W. B., Kersting, M. & Beauducel, A. (2016). Hierarchies of factor solutions in the intelligence domain: Applying methodology from personality psychology to gain insights into the nature of intelligence. *Learning and Individual Differences*, 47, 37–50.
- Lang, J. W. B., Kersting, M., Hülshager, U. R. & Lang, J. (2010). General mental ability, narrower cognitive abilities, and job performance: The perspective of the nested-factors model of cognitive abilities. *Personnel Psychology*, 63, 595–640.
- Levacher, J., Koch, M., Hissbach, J., Spinath, F. M. & Becker, N. (2022). You can play the game without knowing the rules – but you're better off knowing them: The influence of rule knowledge on figural matrices tests. *European Journal of Psychological Assessment*, 38, 15–23. <https://doi.org/10.1027/1015-5759/a000637>
- Liepmann, D., Beauducel, A., Brocke, B. & Amthauer, R. (2007). *Intelligenz-Struktur-Test 2000 R*. Göttingen: Hogrefe.
- Loehlin, J. C. (2002). The IQ paradox: Resolved? Still an open question. *Psychological Review*, 109, 754–758.
- Lorenz, A. L. (1988). Berufspraxis als kritischer Psychologe – Wandel und Beständigkeit. In G. Rexilius (Hrsg.), *Psychologie als Gesellschaftswissenschaft: Geschichte, Theorie und Praxis* (S. 116–131). Opladen: Westdeutscher Verlag.
- Liu, J. & Lynn, R. (2011). Factor structure and sex differences on the Wechsler Preschool and Primary Scale of Intelligence in China, Japan, and United States. *Personality and Individual Differences*, 50, 1222–1226.
- Lotz, C., Sparfeldt, J. R. & Greiff, S. (2016). Complex problem solving in educational contexts – Still something beyond a “good g”? *Intelligence*, 59, 127–138.
- Lubinski, D. (2004). Introduction to the Special Section on Cognitive Abilities: 100 Years After Spearman's (1904) “General Intelligence, Objectively Determined and Measured”. *Journal of Personality and Social Psychology*, 86, 96–111.
- Lynn, R. & Irving, P. (2004). Sex differences on the progressive matrices: A meta-analysis. *Intelligence*, 32, 481–498.
- Mainert, J., Niepel, C., Murphy, K. & Greiff, S. (2019). The incremental contribution of complex problem solving skills to the prediction of job level, job complexity, and salary. *Journal of Business and Psychology*, 34, 825–845.
- Maranto, R. & Wai, J. (2020). Why intelligence is missing from American education policy and practice, and what can be done about it. *Journal of Intelligence*, 8 (1), 2.
- Marcus, B. (2003). Attitudes toward personnel selection methods: A partial replication and extension in a German sample. *Applied Psychology: An International Review*, 52, 515–532.
- McDaniel, M. A., Hartman, N. S., Whetzel, D. L. & Grubb, W. L., III. (2007). Situational judgment tests, response instructions, and validity: A meta-analysis. *Personnel Psychology*, 60, 63–91.
- Meredith, W. (1993). Measurement invariance, factor analysis, and factorial invariance. *Psychometrika*, 58, 525–543.
- Montag, C. (2018). *Eine kurze Einführung in die Molekulare Psychologie. Band 1: Definitionen und molekulargenetische Grundbegriffe*. Wiesbaden: Springer.
- Petermann, F. & Petermann, U. (Hrsg.). (2008). *Hamburg-Wechsler-Intelligenztest für Kinder-IV*. Bern: Huber.
- Plomin, R. & Deary, I. J. (2015). Genetics and intelligence differences: Five special findings. *Molecular Psychiatry*, 20, 98–108.
- Plomin, R. & von Stumm, S. (2018). The new genetics of intelligence. *Nature Review Genetics*, 19, 148–159.
- Protzko, J. (2015). The environment in raising early intelligence: A meta-analysis of the fadeout effect. *Intelligence*, 53, 202–210.
- Ree, M. J. & Carretta, T. R. (2002). g2K. *Human Performance*, 15 (1–2), 3–23.
- Reeve, C. L. & Bonaccio, S. (2011). On the myth and the reality of the temporal validity degradation of general mental ability test scores. *Intelligence*, 39, 255–272.
- Risavy, S., Robie, C., Fisher, P. & König, C. J. (2019). Selection tool use: A focus on personality testing in Canada, the United States, and Germany. *Personnel Assessment and Decisions*, 5 (1), 62–72.
- Ritchie, S. J. & Tucker-Drob, E. M. (2018). How much does education improve intelligence? A meta-analysis. *Psychological Science*, 29, 1358–1369.
- Rost, D. H. (2013). *Handbuch Intelligenz*. Weinheim: Beltz.
- Roth, B., Becker, N., Romeyke, S., Schäfer, S., Domnick, F. & Spinath, F. M. (2015). Intelligence and school grades: A meta-analysis. *Intelligence*, 53, 118–137.
- Rushton, J. P. & Jensen, A. R. (2005). Thirty years of research on race differences in cognitive ability. *Psychology, Public Policy, and Law*, 11, 235–294.
- Rushton, J. P. & Skuy, M. (2000). Performance on Raven's Matrices by African and White university students in South Africa. *Intelligence*, 28, 251–265.
- Salgado, J. F., Anderson, N., Moscoso, S., Bertua, C., de Fruyt, F. & Rolland, J. P. (2003). A meta-analytic study of general mental ability validity for different occupations in the European community. *Journal of Applied Psychology*, 88, 1068–1081.
- Sarrazin, T. (2010). *Deutschland schafft sich ab. Wie wir unser Land aufs Spiel setzen*. München: DVA.
- Saß, S., Kampa, N. & Köller, O. (2017) The interplay of g and mathematical abilities in large-scale assessments across grades. *Intelligence*, 63, 33–44.
- Sauce, B. & Matzel, L. D. (2018). The paradox of intelligence: Heritability and malleability coexist in hidden gene-environment interplay. *Psychological Bulletin*, 144, 26–47.
- Scharfen, J., Peters, J. M. & Holling, H. (2018). Retest effects in cognitive ability tests: A meta-analysis. *Intelligence*, 67, 44–66.
- Schmidt, F. L. & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology. Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin*, 124, 262–274.
- Schneider, B., Becker, N., Krieger, F., Spinath, F. M. & Sparfeldt, J. R. (2020). Teaching the underlying rules of figural matrices in a short video increases test scores. *Intelligence*, 82, 101473.
- Schuler, H., Hell, B., Trapmann, S., Schaar, H. & Boramir, I. (2007). Die Nutzung psychologischer Verfahren der externen Personal-

- auswahl in deutschen Unternehmen: Ein Vergleich über 20 Jahre. *Zeitschrift für Personalpsychologie*, 6, 60–70.
- Schult, J. & Sparfeldt, J. R. (2016). Do non-g factors of cognitive ability tests align with specific academic achievements? A combined bifactor modeling approach. *Intelligence*, 59, 96–102.
- Sniekers, S., Stringer, S., Watanabe, K., Jansen, P. R., Coleman, J. R. I., Krapohl E. et al. (2017). Genome-wide association meta-analysis of 78,308 individuals identifies new loci and genes influencing human intelligence. *Nature Genetics*, 49, 1107–1112.
- Stadler, M., Becker, N., Gödker, M., Leutner, D. & Greiff, S. (2015). Complex problem solving and intelligence: A meta-analysis. *Intelligence*, 53, 92–101.
- Steinmayr, R., Beauducel, A. & Spinath, B. (2010). Do sex differences in a faceted model of fluid and crystallized intelligence depend on the method applied? *Intelligence*, 38, 101–110.
- Stegt, S. J., Didi, H.-J., Zimmerhofer, A. & Seegers, P. K. (2018). Akzeptanz von Auswahlverfahren zur Studienplatzvergabe. *Zeitschrift für Hochschulentwicklung*, 13 (4). 15–35.
- Stelzl, I., Merz, F., Ehlers, T. & Remer, H. (1995). The effect of schooling on the development of fluid and crystallized intelligence: A quasi-experimental study. *Intelligence*, 21, 279–296.
- Stern, E. & Neubauer, A. (2016). Intelligenz: Kein Mythos, sondern Realität. *Psychologische Rundschau*, 67, 1–13.
- Sternberg, R. J. & Williams, W. M. (1997). Does the Graduate Record Examination predict meaningful success in the graduate training of psychologists? A case study. *American Psychologist*, 52, 630–641.
- Strenze, T. (2007). Intelligence and socioeconomic success: A meta-analytic review of longitudinal research. *Intelligence*, 35, 401–426.
- Tucker-Drob, E. M. & Bates, T. C. (2016). Large cross-national differences in gene  $\times$  socioeconomic status interaction on intelligence. *Psychological Science*, 27, 138–149.
- Valerius, S. & Sparfeldt, J. R. (2014). Consistent g- as well as consistent verbal-, numerical- and figural-factors in nested factor models? Confirmatory factor analyses using three test batteries. *Intelligence*, 44, 120–133.
- Valerius, S. & Sparfeldt, J. R. (2015). Zusammenhänge allgemeiner und spezifischer Intelligenzfaktoren mit allgemeinen und spezifischen Schulleistungen im Nested-Factor-Modell. *Zeitschrift für Pädagogische Psychologie*, 29, 101–108.
- Wee, S. (2018). Aligning predictor-criterion bandwidths: Specific abilities as predictors of specific performance. *Journal of Intelligence*, 6 (3), 40.
- Wilhelm, O. & Schroeders, U. (2019). Intelligence. In R. J. Sternberg & J. Funke (Eds.), *The Psychology of Human Thought: An Introduction* (pp. 255–275). Heidelberg: Heidelberg University Publishing.
- Zhang, B., Sun, T., Cao, M. & Drasgow, F. (2020). Using bifactor models to examine the predictive validity of hierarchical constructs: Pros, cons, and solutions (First published online April 15, 2020). *Organizational Research Methods*. <https://doi.org/10.1177/109442812091552>.

#### Förderung

Open Access-Veröffentlichung ermöglicht durch die Universität des Saarlandes.

#### Prof. Dr. Jörn R. Sparfeldt

Bildungswissenschaften  
Universität des Saarlandes  
Campus A 5.4  
66123 Saarbrücken  
Deutschland  
[j.sparfeldt@mx.uni-saarland.de](mailto:j.sparfeldt@mx.uni-saarland.de)