

BEFKI GC-K

*Eine Kurzskala zur Messung
kristalliner Intelligenz*

BEFKI GC-K

*A Short Scale for the
Measurement of Crystallized
Intelligence*

*Stefan Schipolowski, Oliver Wilhelm, Ulrich Schroeders,
Anastassiya Kovaleva, Christoph J. Kemper und
Beatrice Rammstedt*

Zusammenfassung

In aktuellen Intelligenzstrukturmodellen gehört kristalline Intelligenz (g_c) zu den am besten etablierten Fähigkeitsfaktoren. Dabei spiegelt g_c die Einflüsse von Lernen und Akkulturation wider und umfasst somit alles Wissen, das Menschen im Laufe ihres Lebens erwerben und zum Problemlösen einsetzen. In diesem Beitrag beschreiben wir die Entwicklung einer Kurzskala zur Messung kristalliner Intelligenz mit fünfminütiger Bearbeitungszeit, die auf deklarativen Wissensfragen aus den Natur-, Geistes- und Sozialwissenschaften beruht. Aus einem umfangreichen Itempool wurde ein 32 Fragen umfassender Wissenstest zusammengestellt und einer bundesweit repräsentativen Stichprobe von 1.134 Erwachsenen vorgelegt. Anhand psychometrischer Kennwerte und der Beziehungen zu Kovariaten erfolgte eine Auswahl von 12 Items für die Kurzskala. Ein eindimensionales Messmodell für diese Itemauswahl wies eine gute Passung und eine hohe Reliabilität des latenten Faktors auf. In der Zielpopulation der erwachsenen deutschen Bevölkerung wurden keine substanziellen Boden- oder Deckeneffekte

Abstract

Crystallized intelligence (g_c) is a well-established cognitive ability factor that has been conceptualized as reflecting influences of learning, education, and acculturation. In this article, we describe the development of a short knowledge scale for the measurement of g_c in five minutes administration time using declarative knowledge items from the sciences, the humanities, and civics. Based on a large item pool we compiled a 32-item knowledge test that was subsequently presented to a nationally representative sample of 1,134 German adults. In the next step, this data were used to derive a short 12-item knowledge scale. A unidimensional measurement model had satisfactory model fit and showed high reliability of the latent factor. There were no substantial floor or ceiling effects in the adult German population. Similar to the full scale, the short scale correlated highly positively with education (ISCED-97) and socio-economic status (ISEI) and was meaningfully related to self-reported knowledge and the Big Five personality traits. Therefore, the short knowledge scale allows for



beobachtet. Übereinstimmend mit der Langversion zeigten sich für die Kurzskala hohe Beziehungen zum Bildungsabschluss (ISCED-97) und sozioökonomischen Status (ISEI) sowie erwartungskonforme Korrelationen mit selbstberichtetem Wissen und den fünf Hauptdimensionen der Persönlichkeit (Big Five). Die Kurzskala ermöglicht folglich eine effiziente, reliable und valide Erfassung kristalliner Intelligenz im Rahmen der Umfrageforschung.

an efficient and valid measurement of crystallized intelligence in survey research.

1 Einleitung¹

Die Messung psychologischer Merkmale wie kognitiver Fähigkeiten und Persönlichkeitseigenschaften gewinnt in den Sozial- und Wirtschaftswissenschaften zunehmend an Bedeutung (Rat für Sozial- und Wirtschaftsdaten 2010). So verweisen etwa Grabner und Stern (2010) auf das komplexe, bisher jedoch nicht hinreichend erforschte Zusammenspiel zwischen individuellen kognitiven Ressourcen und sozioökonomischen Variablen. Auf Personenseite kommt hier neben der dekontextualisierten Fähigkeit zum schlussfolgernden Denken bzw. dem Arbeitsgedächtnis insbesondere jenen Fähigkeiten eine zentrale Rolle zu, die auf Lernen, Erfahrung und Wissen beruhen und unter dem Begriff der kristallinen Intelligenz zusammengefasst werden. Wiederholt wurde jedoch darauf hingewiesen, dass bislang nur wenige für die Umfrageforschung geeignete, d.h. hinreichend kurze und effiziente Messinstrumente zur Erfassung psychologischer Merkmale vorliegen (Rammstedt/Kemper/Klein/Beierlein/Kovaleva 2012; Lang/Weiss/Stocker/von Rosenblatt 2007).

Der vorliegende Beitrag beschreibt die Entwicklung und Eigenschaften einer Kurzskala zur Erfassung kristalliner Intelligenz anhand deklarativer Wissensfragen in fünf Minuten Bearbeitungszeit. Die Datenbasis hierfür ist eine für die erwachsene Wohnbevölkerung Deutschlands repräsentative Stichprobe, die dementsprechend eine hohe Heterogenität bzgl. Alter und Bildung aufweist. Zuerst wird unter Rückgriff auf einflussreiche Intelligenztheorien das Konstrukt der kristallinen Intelligenz näher erläutert. Mit Blick auf die Validierung der Kurzskala folgt eine Darstellung zentraler Befunde zu den Beziehungen zwischen kristalliner Intelligenz, relevanten Personen- und Umweltmerkmalen sowie anderen psychologischen Konstrukten.

1 Während der Erstellung dieses Beitrags war Stefan Schipolowski Fellow der International Max Planck Research School „The Life Course: Evolutionary and Ontogenetic Dynamics (LIFE)“.

1.1 Kristalline Intelligenz in konsensualen Intelligenzstrukturtheorien

Wissen und sprachliche Fähigkeiten spielen seit Beginn der psychometrischen Intelligenzforschung im späten 19. Jahrhundert eine bedeutende Rolle in Intelligenztheorien. So schlug Hermann Ebbinghaus bereits 1897 ein Verfahren zur „Prüfung geistiger Fähigkeiten“ bei Schulkindern vor, das darauf beruhte, in kurzen Texten unvollständige Wörter sinnhaft zu vervollständigen und charakterisierte diese Anforderung explizit als „Intelligenzthätigkeit“ (Ebbinghaus 1897: 414). Für Charles Spearman (1938) hingegen stellte non-verbales schlussfolgerndes Denken den eigentlichen Kern allgemeiner Intelligenz, des sog. g -Faktors, dar. Nach Binet und Simon (1905) sowie Hebb (1942) gehörte Cattell (1943) zu den ersten Intelligenzforschern, die diese beiden Intelligenzaspekte aufgriffen und gegenüberstellten. Statt eines einzigen Generalfaktors der Intelligenz postulierte Cattell zwei bedeutende Faktoren, die er als fluide Intelligenz (g_f) und kristalline Intelligenz (g_c) bezeichnete. Letztere zeigt sich nach Cattell (1971) in Leistungen, bei denen zuvor erlernte Fertigkeiten und Wissen die entscheidende Rolle spielen. Typische Indikatoren für g_c bezeichnet Cattell als „schulische“ oder „akademische“ Tests, die auf die Inhalte formaler Bildung abzielen. Damit übereinstimmend wurden in den Studien in Cattells Labor (Cattell 1963; Horn/Cattell 1966; Cattell 1971) starke Ladungen von sprachnahen Aufgaben und Wissenstests auf dem Faktor g_c berichtet (vgl. etwa Horn 1965). Aus Cattells theoretischen Schriften geht hervor, dass g_c die Gesamtheit des Wissens umfasst, das Menschen im Laufe ihres Lebens erwerben und zum Problemlösen einsetzen. Aufgrund der mit steigendem Lebensalter zunehmenden Spezialisierung in Form unterschiedlicher Berufswahlen, verschiedenartiger Freizeitaktivitäten und Interessen würde eine umfassende Messung kristalliner Intelligenz im Erwachsenenalter demnach eine nahezu unendliche Vielfalt an Iteminhalten erfordern. Übereinstimmend betont auch Horn (1988) die Breite des g_c -Konstrukts und seine Nähe zum Generalfaktor der Intelligenz. In seiner Auflistung typischer Indikatoren spielen verbale Fähigkeiten eine prominente Rolle (z.B. „verbal knowledge“) sowie Wissen in einer Vielzahl von Bereichen („information about the humanities, social and physical sciences, business and culture in general“; Horn 1988: 659). Horn und Noll (1997: 69) beschreiben kristalline Intelligenz dementsprechend als „akkulturiertes Wissen“, das über Aufgaben gemessen wird, die „Tiefe und Breite des Wissens der dominanten Kultur“ widerspiegeln.

Als weiteres einflussreiches Intelligenzmodell soll Carrolls (1993) Drei-Stratum-Theorie erwähnt werden, die nach wie vor als Status Quo der Intelligenzforschung gilt. Das Modell basiert auf der Reanalyse von mehr als 460 Datensätzen zu

kognitiven Fähigkeitskonstrukten und beschreibt verschiedene Schichten („Strata“) mit Intelligenzfaktoren unterschiedlicher Breite. Zu den insgesamt acht Faktoren auf Stratum II gehören auch g_f und g_c . Ähnlich wie Cattell betont auch Carroll mit Blick auf den g_c -Faktor die Rolle von Erfahrung, Lernen und Akkulturation, verschiebt jedoch die Definition von g_c in Richtung sprachlicher Fähigkeiten wie Leseverstehen und Fremdsprachenkenntnissen (Carroll 1993: 626). Allerdings dokumentiert Carroll auch die hohen Ladungen von Wissenstests auf dem g_c -Faktor. Eine besondere Bedeutung kommt dem Stratum-I-Faktor „General Information“ zu, der Unterschiede im Erwerb von Wissen jenseits von Sprachkenntnissen widerspiegelt (vgl. KO; Carroll 1993: 590 bzw. 634). Dieser Wissensfaktor gehört zu jenen Faktoren, die in Carrolls Reanalysen häufig die höchste oder zweithöchste Ladung auf einem übergeordneten g_c -Faktor aufwiesen (vgl. auch Carroll 2003). In Übereinstimmung mit den theoretischen Vorarbeiten Cattells untermauern diese Befunde die Position, dass eine Operationalisierung kristalliner Intelligenz im Erwachsenenalter Wissen aus möglichst vielen unterschiedlichen Bereichen berücksichtigen sollte.

1.2 Alters- und Geschlechtsunterschiede in kristalliner Intelligenz

Das Konzept der kristallinen Intelligenz wurde auch von Seiten der Entwicklungspsychologie über die Lebensspanne aufgegriffen (vgl. etwa Baltes 1987; Baltes/Staudinger/Lindenberger 1999), die wesentliche Erkenntnisse zum Entwicklungsverlauf von g_c liefern konnte. Nach Baltes et al. (1999: 486ff) zeigt die kristalline Pragmatik der Kognition, die kulturell vermitteltes Wissen repräsentiert, einen deutlichen Anstieg im Kindes- und Jugendalter, bleibt im Verlauf des Erwachsenenalters weitgehend stabil und nimmt erst im sehr hohen Lebensalter ab. Auch Ackerman (2008) argumentiert in seinem Literaturüberblick, dass sich g_c im Sinne von Allgemeinwissen und lexikalischem Wissen durch hohe Stabilität im Erwachsenenalter auszeichnet. Ein anderes Bild ergibt sich jedoch für spezialisiertes, bereichsspezifisches Wissen im Sinne von Expertise. So konnten Ackerman und Kollegen in mehreren Studien zeigen, dass im Erwachsenenalter teilweise deutliche Leistungszuwächse im bereichsspezifischen Wissen zu beobachten sind (Ackerman 2000; Ackerman/Rolfhus 1999; zusammenfassend Ackerman/Beier 2004). Diese Befunde illustrieren, dass die Beziehung kristalliner Intelligenz zum Alter auch von den gemessenen Inhalten abhängig ist. Wird g_c über das im Kindes- und Jugendalter erworbene schulische Wissen gemessen, das unabhängig vom Lebensalter häufig abgerufen wird – wie es etwa für lexikalisches Wissen der Fall ist – sind kaum Veränderungen im Erwachsenenalter zu beobachten; wird g_c hingegen als Expertise operationalisiert,

die erst durch eine spezielle Ausbildung oder durch die Berufsausübung erworben wird, können auch im mittleren Erwachsenenalter substanzielle Zuwächse beobachtet werden.

Verschiedene Arbeiten aus der psychologischen Forschung haben sich mit Geschlechtsunterschieden in Wissensleistungen auseinandergesetzt. Hier sind insbesondere die Arbeiten von Lynn sowie aus der Arbeitsgruppe von Ackerman zu erwähnen. In Untersuchungen an Jugendlichen bzw. jungen Erwachsenen mit einer viele Wissensbereiche umfassenden Testbatterie wurde wiederholt ein Leistungsvorsprung im Allgemeinwissen zugunsten der Männer von etwa einer halben Standardabweichung gefunden ($d = 0.50$ bis 0.60 ; Lynn/Irwing/Cammock 2001; Lynn/Irwing 2002; Lynn/Wilberg/Margraf-Stiksrud 2004). Ähnliche Ergebnisse berichten Ackerman, Bowen, Beier und Kanfer (2001), die eine Stichprobe von 320 Studierenden mit einer breit angelegten Wissenstestbatterie zu 19 spezifischen Wissensbereichen untersuchten. In 14 der 19 Wissensbereiche erzielten Männer signifikant bessere Ergebnisse, insbesondere in den naturwissenschaftlich-technischen und einigen sozialwissenschaftlichen Bereichen. Für keinen Wissensbereich ergaben sich signifikant bessere Resultate für Frauen; in vier Bereichen fanden sich keine oder sehr geringe Unterschiede. Insgesamt ergab sich für einen zusammengesetzten Gesamtwert zum deklarativen Wissen ein Unterschied von $d = 0.68$ zugunsten der Männer. Einschränkend muss jedoch gesagt werden, dass diese geschlechtsbezogenen Disparitäten kulturspezifisch sind und nicht auf andere Kulturen übertragbar sein müssen.

1.3 Zusammenhänge kristalliner Intelligenz mit Bildung und Lernumwelt

Wie oben ausgeführt, ist die Abhängigkeit kristalliner Intelligenz von Lernen und Bildung das zentrale Definitionsmerkmal des Konstrukts. Aufgrund der Langfristigkeit des Wissenserwerbs und dessen Abhängigkeit von den zur Verfügung stehenden Lerngelegenheiten und -ressourcen sind folglich hohe Zusammenhänge zwischen g_c und der Qualität und Quantität formaler Bildung zu erwarten, wie sie in formalen Bildungsabschlüssen und darauf beruhenden Indizes zum Ausdruck kommen (Cliffordson/Gustafsson 2008; Ceci 1991). Ebenso kann eine substanzielle positive Korrelation mit solchen Indikatoren und Indizes angenommen werden, die für den Wissenserwerb bedeutsame Ressourcen erfassen (Rowe/Jacobson/van den Oord 1999). Dies betrifft Maße des sozioökonomischen Status ebenso wie Fragen nach der Ausstattung des Haushalts, etwa zur Anzahl der verfügbaren Bücher (Ehmke/Siegle 2005).

1.4 Zusammenhänge kristalliner Intelligenz mit Selbsteinschätzungen des Wissens und Persönlichkeitskonstrukten

Neben einer Testung deklarativen Wissens über Wissensfragen mit mehreren Antwortalternativen, die eindeutig als richtig oder falsch zu werten sind, können auch Selbstberichte herangezogen werden. Derartige Selbsteinschätzungen des eigenen Wissens spiegeln jedoch neben der tatsächlichen Fähigkeitsausprägung weitere Varianzquellen wider (z.B. *faking-good*), so dass Selbsteinschätzungen des Wissens und objektive Wissenstests als unterschiedliche, wenngleich korrelierte Konstrukte aufzufassen sind (Furnham/Dissou 2007). Die Korrelationen zwischen entsprechenden Messungen sind somit zwar substanziell, fallen aber selbst unter Verwendung von Ansätzen zur Korrektur von Antwortverzerrungen deutlich niedriger aus als zwischen Messungen desselben Konstrukts (Hülür/Wilhelm/Schipolowski 2011; Rolfhus/Ackerman 1996; Paulhus/Harms 2004).

Eine Vielzahl an Studien hat sich mit den Zusammenhängen zwischen kognitiven Fähigkeiten einerseits und Persönlichkeitseigenschaften im Sinne typischen Verhaltens (Cronbach 1949) andererseits befasst. Hervorzuheben ist die Metaanalyse von Ackerman und Heggestad (1997), die anhand von 135 Studien die Beziehungen zwischen Persönlichkeits- und Fähigkeitskonstrukten beleuchtet. Die Autoren beschränken sich auf der Fähigkeitsseite nicht auf die Betrachtung allgemeiner Intelligenz, sondern differenzieren zwischen zehn verschiedenen Fähigkeitskonstrukten, darunter kristalline Intelligenz und eine zusammengefasste Kategorie Wissen/Achievement. Bei der Kategorisierung der gemessenen Konstrukte orientieren sich die Autoren an den Charakterisierungen der Fähigkeitsfaktoren bei Carroll (1993). Kristalline Intelligenz umfasst demnach sprachliche Leistungen und Allgemeinwissen (vgl. Carroll 1993: 599); zur Kategorie Wissen/Achievement zählen neben spezifischen Wissenstests auch Fachleistungstests etwa im Schulfach Biologie (vgl. Carroll 1993: 513). Die Persönlichkeitsskalen der Studien wurden anhand verschiedener in der Forschung etablierter Systeme klassifiziert, darunter das Modell der fünf Hauptdimensionen der Persönlichkeit (Big Five; für ausführliche Beschreibungen dieser Dimensionen siehe Ostendorf/Angleitner 2004). Die metaanalytische Betrachtung der korrelativen Zusammenhänge zwischen den Big Five-Faktoren und g_c bzw. Wissen/Achievement ergab jeweils die mit Abstand höchsten Beziehungen für Offenheit. Hier wurde eine minderungskorrigierte Korrelation von .30 (g_c) bzw. .28 (Wissen/Achievement) ermittelt. Eine ebenfalls positive, aber mit .11 deutlich niedrigere Korrelation zeigte sich zwischen g_c und Extraversion. Für Wissen/Achievement lag diese bei .05 und war nicht signifikant von null verschieden. Negative Zusammenhänge wurden hingegen mit Neurotizismus (Korrelationen

von $-.09$ mit g_c und $-.13$ mit Wissen/Achievement) und Gewissenhaftigkeit gefunden, wobei die Korrelation zwischen Gewissenhaftigkeit und g_c nicht signifikant war und für Wissen/Achievement mit $-.19$ zwar vom Betrag her höher ausfiel, diese Angabe jedoch nur auf einer einzigen Studie mit relativ geringer Fallzahl basiert. Die Korrelationen mit Verträglichkeit waren nicht signifikant von null verschieden.

Die vergleichsweise hohe Beziehung zwischen g_c bzw. Wissen/Achievement und Offenheit ist sowohl empirisch gut belegt (von Stumm/Ackerman 2012) als auch aus theoretischer Perspektive plausibel. So zielen Items zur Messung von Offenheit unter anderem auf intellektuelle Neugier, das Bestreben, neues Wissen zu erwerben und auf kulturelles Engagement (Ostendorf/Angleitner 2004). Folgerichtig beschreibt Ackerman (1996) neben „typischem intellektuellem Engagement (TIE)“ Offenheit als Persönlichkeitskonstrukt mit den höchsten Beziehungen zu g_c und geisteswissenschaftlichem Wissen. Übereinstimmend argumentieren Ziegler, Danay, Heene, Asendorpf und Bühner (2012), dass Personen mit hohen Offenheitswerten mehr Zeit in Lernen und Wissenserwerb investieren.

1.5 Zusammenfassung und erwartete Zusammenhänge

Aus theoretischen und diagnostischen Überlegungen heraus sollte eine Operationalisierung von g_c über deklaratives Wissen aus möglichst vielen verschiedenen Wissensbereichen erfolgen. Zu erwarten sind hohe positive Korrelationen zwischen kristalliner Intelligenz und der Qualität und Quantität formaler Bildung sowie mit Indikatoren, die für den Wissenserwerb bedeutsame Ressourcen erfassen, wie beispielsweise Maße des sozioökonomischen Status. Mit Blick auf Altersunterschiede ist in der Erwachsenenpopulation aufgrund der hohen Stabilität kristalliner Intelligenz von sehr geringen Effekten auszugehen, sofern die gemessenen Inhalte auf solche Wissensbestände abzielen, die in einer Vielzahl von Lernumwelten erworben werden können. In Anlehnung an die in der Literatur beschriebenen geschlechtsbezogenen Unterschiede in Wissensleistungen wird ein Vorsprung der Männer von etwa einer halben Standardabweichung bis zwei Drittel einer Standardabweichung erwartet. Des Weiteren ist von positiven Korrelationen von g_c mit selbsteingeschätztem Wissen sowie mit dem Persönlichkeitsfaktor Offenheit auszugehen.

2 Methode

2.1 Stichprobe

Als Grundgesamtheit für die Stichprobenziehung wurde die Wohnbevölkerung der Bundesrepublik Deutschland im Alter von 18 Jahren und älter definiert. Es wurden auch Personen mit Zuwanderungshintergrund berücksichtigt, sofern sie die deutschsprachigen Fragen und Aufgaben verstehen und auf Deutsch antworten konnten. Die Stichprobe wurde mithilfe des ADM-Stichprobensystems F2F der Arbeitsgemeinschaft deutscher Marktforschungsinstitute gezogen. Dabei handelt es sich um ein komplexes mehrstufiges Ziehungsverfahren, in dem zunächst Flächen, dann Privathaushalte und schließlich Zielpersonen innerhalb der Haushalte nach einem Zufallsverfahren ausgewählt werden (für Details vgl. von der Heyde 2009). Nach diesem Verfahren wurde eine Stichprobe von 1206 Personen realisiert, die an der Erhebung teilnahmen. Im Anschluss wurden auf Basis des Zensus von GESIS Fallgewichte erstellt, um Repräsentativität für die o.g. Grundgesamtheit mit Blick auf Region (Ost- bzw. Westdeutschland), Geschlecht, Bildung und Alter zu gewährleisten. Grundlage der Gewichtung war ein reduzierter Datensatz von 1.134 Fällen nach Ausschluss unbrauchbarer Datenpunkte sowie von Personen ohne deutsche Staatsbürgerschaft, um für die Gewichtung eine eindeutige Definition der Grundgesamtheit zu ermöglichen. Die gewichtete Stichprobe umfasst somit 1.134 Erwachsene (52,2% weiblich) im Alter von 18 bis 93 Jahren ($M = 52$ Jahre, $SD = 18$ Jahre) aus dem gesamten Bundesgebiet.

2.2 Messinstrumente

In einem umfangreichen Fragebogen wurden verschiedene soziodemographische Merkmale der Teilnehmerinnen und Teilnehmer erfasst. Hierzu zählten Geburtsjahr und -monat, Geschlecht, Familienstand, Staatsangehörigkeit, erreichter bzw. (bei Schülerinnen und Schülern) angestrebter allgemeinbildender Schulabschluss, beruflicher Ausbildungsabschluss, Erwerbsstatus, berufliche Stellung und Haushaltsnettoeinkommen. Die Erfassung dieser Merkmale orientierte sich an den Demographischen Standards des Statistischen Bundesamtes (2010). Ergänzend wurde anhand von sieben Kategorien die Anzahl der Bücher im Elternhaus erfragt, das Geburtsland sowie die berufliche Stellung der Eltern, als der/die Teilnehmende 15 Jahre alt war. Anhand zweier fünfstufiger Skalen wurden vom Interviewer die Deutschkenntnisse und die soziale Schichtzugehörigkeit der Teilnehmenden eingeschätzt.

Zur Messung kristalliner Intelligenz wurde auf den umfangreichen Itempool des BEFKI-Projekts (Berliner Test zur Erfassung Eluider und Kristalliner Intelligenz; Wilhelm/Schroeders/Schipolowski, in Vorbereitung; Wilhelm/Schipolowski 2010) zurückgegriffen, mit dem deklaratives Wissen in 16 verschiedenen Domänen erfasst werden kann. Im Einzelnen wird naturwissenschaftliches (Physik, Chemie, Biologie, Medizin, Geografie, Technologie), geisteswissenschaftliches (Literatur, Kunst, Musik, Religion, Philosophie) und sozialwissenschaftliches Wissen erfragt (Geschichte, Recht, Politik, Wirtschaft, Finanzen). Die Auswahl der Wissensbereiche orientierte sich an der empirisch begründeten Klassifikation von Ackerman (2000; Rolfhus/Ackerman 1999). Für die aktuelle Studie wurden insgesamt 32 Wissensitems anhand inhaltlicher und psychometrischer Kriterien aus dem Gesamtitempool ausgewählt. Konkret wurden zwei Items aus jedem der 16 Wissensbereiche eingesetzt, wobei eines der beiden Items zu jedem Bereich von geringer bis mittlerer Schwierigkeit war (entwickelt für Personen ohne Schulabschluss, mit Hauptschul- oder Mittlerem Schulabschluss), das andere von hoher Schwierigkeit (entwickelt für Personen, die über die Hochschulreife verfügen bzw. diese anstreben). Die psychometrische Eignung der Items wurde anhand von Vorinformationen aus verschiedenen Erhebungen sichergestellt (Schipolowski/Schroeders/Wilhelm 2008; Schroeders/Schipolowski/Wilhelm 2010; Schroeders/Schipolowski/Nelles/Wilhelm 2011). Ausschlaggebend war dabei neben den Informationen zur Itemschwierigkeit insbesondere eine hohe positive Trennschärfe in den genannten Voruntersuchungen. Alle Wissensitems waren ausschließlich textbasiert und hatten ein Multiple-Choice-Format mit vier Antwortalternativen, von denen genau eine die richtige Lösung darstellte.

Zusätzlich zur kristallinen Intelligenz wurden weitere psychologische Konstrukte erfasst. Der BFI-10 (Rammstedt/John 2007) ermöglichte die Erfassung der Big-Five-Persönlichkeitsdimensionen Neurotizismus (N), Extraversion (E), Offenheit (O), Verträglichkeit (V) und Gewissenhaftigkeit (G) anhand von jeweils zwei (N, E, O, G) bzw. drei Items (V). Die Items bestanden aus Aussagen, deren Zutreffen von den Teilnehmenden auf einer fünfstufigen Ratingskala eingeschätzt wurde. Mit dem VOC-T (Ziegler/Kemper/Rammstedt 2013) wurde zudem ein Maß für die Selbsteinschätzung des eigenen Wissens eingesetzt. Zu insgesamt 12 verschiedenen Begriffen aus den Natur-, Geistes- und Sozialwissenschaften sowie dem handwerklichen Bereich gaben die teilnehmenden Personen anhand einer siebenstufigen Ratingskala an, wie vertraut sie mit dem jeweiligen Begriff oder Konzept sind. Neben diesen real existierenden Begriffen enthält der VOC-T zusätzlich drei fiktive Begriffe.

Weitere eingesetzte Skalen dienten der Erfassung von Lebenszufriedenheit, politischer Partizipation, Werten, Kontrollüberzeugungen, Selbstwirksamkeitser-

wartungen, Impulsivität und dem Gesundheitszustand. Da diese Skalen nur in das Imputations-, nicht jedoch in das Analysemodell einbezogen wurden, werden sie hier nicht näher beschrieben.

2.3 Durchführung

Die Erhebung der Daten erfolgte im Zeitraum 2. Mai bis 23. Juni 2011 durch ein beauftragtes Erhebungsinstitut. Geschulte Interviewer suchten die Studienteilnehmerinnen und -teilnehmer zu vorab vereinbarten Terminen in ihren Wohnungen auf, um die Befragung bzw. Testung durchzuführen. Die soziodemographischen Angaben und die Antworten auf die Persönlichkeitsitems wurden vom Interviewer erfragt und in eine Eingabemaske am Notebook eingegeben (CAPI, computer assisted personal interview). Den g_c -Test bearbeiteten die Teilnehmenden selbstständig am Notebook (CASI, computer assisted self-interview). In beiden Fällen war die Abfolge der Fragen durch ein Skript vorgegeben, um einen standardisierten Ablauf zu gewährleisten. Beim g_c -Test wurden immer vier Fragen gleichzeitig auf dem Bildschirm dargestellt; um zur nächsten Bildschirmseite zu gelangen, musste die teilnehmende Person zunächst alle vier Fragen der aktuellen Seite beantworten (ggf. durch Raten). Für die Bearbeitung der 32 g_c -Items war ein Zeitlimit von 10 Minuten vorgegeben. Bei Erreichen des Zeitlimits brach der Wissenstest automatisch ab und es wurde mit dem nächsten Fragenkomplex fortgefahren. Die Dauer des gesamten Interviews betrug im Durchschnitt 43 Minuten ($SD = 13$).

2.4 Datenaufbereitung und Auswertungsverfahren

Zur Aufbereitung der gewichteten Stichprobendaten wurden im ersten Schritt anhand der vorliegenden demographischen Angaben verschiedene Indizes gebildet. Als Index zur formalen Bildung wurde die ISCED-97 (International Standard Classification of Education; UNESCO 1997) genutzt. Dabei wurden die Angaben zum höchsten erreichten Schulabschluss sowie zum Ausbildungsabschluss in einer ordinalskalierten Variable mit 6 Stufen zusammengeführt (vgl. Statistisches Bundesamt 2010: 79). Auf der niedrigsten Stufe (ISCED 1) befinden sich demnach Personen ohne Schul- und Ausbildungsabschluss, während die höchste Stufe (ISCED 6) mit Personen besetzt ist, die die allgemeine oder Fachhochschulreife besitzen und nach erfolgreichem Hoch- bzw. Fachhochschulabschluss zusätzlich einen weiterführenden akademischen Grad (Promotion) erlangt haben. Als Index des sozioökonomischen Status wurde der ISEI (International Socio-Economic Index of Occupational Status; Ganzeboom/De Graaf/Treiman 1992) verwendet. Der ISEI beruht auf Anga-

ben zur Berufstätigkeit und der Prämisse, dass diese Informationswert bezüglich Einkommen und Bildung besitzt, die wiederum „die Hauptquellen der Macht in modernen Gesellschaften“ sind (Ganzeboom et al. 1992: 9). Konkret werden einzelnen Berufen Statuswerte zugeordnet, die auf internationalen Daten zum Einkommen und zur Bildung beruhen, über die Ausübende dieser Berufstätigkeiten typischerweise verfügen. Auf dieser Basis ergibt sich eine Skala von 16 (niedriger Status; etwa Reinigungskräfte) bis 90 (hoher Status; Richter). Ein Weg zur Ermittlung der ISEI-Werte ist die Kodierung von Berufstätigkeiten nach der ISCO (International Standard Classification of Occupations; ILO 2007). Im vorliegenden Fall wurde ein weniger aufwändiges Vorgehen gewählt, das auf Angaben zur beruflichen Stellung beruht, von der nach Wolf (1995) ebenfalls auf den ISEI geschlossen werden kann (vgl. Statistisches Bundesamt 2010). Nach diesem Vorgehen ergibt sich für die aktuelle Studie eine vereinfachte ISEI-Skala mit 13 verschiedenen Ausprägungen. Die Bildung des ISEI wurde für die Teilnehmenden selbst sowie für deren Eltern durchgeführt; in letzterem Fall wird für die folgenden Analysen der höchste der beiden elterlichen ISEI-Werte (HISEI) verwendet. Das Haushaltsnettoeinkommen wurde auch direkt erfragt, wobei zum Teil Freitextantworten, überwiegend jedoch Angaben in Form von Einkommenskategorien vorlagen. Freitextantworten wurden auf die vorliegenden 24 Kategorien rekodiert, um ein einheitliches Antwortformat zu erhalten.

Für die Skalen zu den psychologischen Konstrukten wurden Summen- oder Mittelwerte so gebildet, wie dies von den Autoren der jeweiligen Instrumente vorgeschlagen wurde. Somit lag für die folgenden Analysen jeweils ein Wert für jede der fünf Big-Five-Dimensionen vor. Als Indikator des selbstberichteten Wissens wurde ein Gesamtwert über alle 12 Items des VOC-T genutzt, die sich auf real existierende Begriffe beziehen; die fiktiven Begriffe wurden nicht einbezogen².

Die Items zur Messung der kristallinen Intelligenz wurden zunächst in richtig versus falsch beantwortet rekodiert. Anschließend wurde ein Summenwert über alle 32 Items als Schätzer der Personenfähigkeit auf der Gesamtskala berechnet. Auf Basis der 32 Wissensitems der Gesamtskala erfolgte im nächsten Schritt die Itemselektion für die Kurzskala nach folgenden Kriterien:

- Um einen flexiblen Einsatz der Kurzskala in der Umfrageforschung zu ermöglichen, sollte deren Bearbeitungszeit bei 5 Minuten liegen. Dies entspricht der geschätzten Bearbeitungszeit von 12 Items.

2 Die fiktiven Begriffe des VOC-T, engl. *foils*, können zur Berechnung weiterer Kennwerte herangezogen werden, die jedoch im vorliegenden Beitrag nicht berücksichtigt werden.

- Zur bestmöglichen Erhaltung der inhaltlichen Breite der Wissensmessung sollte einerseits die Dreiteilung in natur-, geistes- und sozialwissenschaftliches Wissen beibehalten werden, andererseits sollten möglichst viele der 16 Wissensbereiche der Gesamtskala auch in der Kurzskaala enthalten sein.
- Um Boden- und Deckeneffekte zu minimieren, sollten die ausgewählten Items einen großen Schwierigkeitsbereich abdecken. Die relative Lösungshäufigkeit sollte jedoch stets oberhalb der Ratewahrscheinlichkeit von .25 liegen.
- Es wurde ein einfaktorielles Messmodell mit guter Modellpassung und Itemladungen (Trennschärfen) von .50 oder höher angestrebt. In keinem Fall sollten Ladungen $< .30$ auftreten.
- Die Kurzskaala sollte ähnliche Beziehungen zu Personen- und Umweltmerkmalen sowie anderen psychologischen Konstrukten aufweisen wie die Gesamtskaala.

Nach erfolgter Itemselektion wurde für die Kurzskaala mit 12 Items ebenfalls ein Summenwert berechnet.

Um Einschränkungen bei der Teststärke sowie Verzerrungen durch nicht zufällig fehlende Informationen³ zu minimieren, wurden fehlende Datenpunkte im Wissenstest sowie in allen Kovariaten imputiert (Lüdtke/Robitzsch/Trautwein/Köller 2007). Der Anteil fehlender Werte bei den 32 Wissensitems betrug im Mittel pro Item 5.8% ($SD = 7.0\%$, Spannweite 0.4% bis 20.8%); die vorliegenden Fallzahlen für die Kovariaten gehen aus der Ergebnistabelle hervor (vgl. Tabelle 4, Spalte N_{vi}). Speziell bei den Wissensitems, für die Datenpunkte fast ausschließlich aufgrund der Zeitbegrenzung fehlten, ermöglichte die Imputation eine Minimierung konstruktirrelevanter Varianz (etwa interindividuelle Unterschiede in mentaler Geschwindigkeit; Danthiir/Roberts/Schulze/Wilhelm 2004). Konkret wurde das Verfahren der multiplen Imputation mit 100 Replikationen genutzt (Graham/Olchowski/Gilreath 2007). Um die fehlenden Werte möglichst zuverlässig schätzen zu können, wurden alle verfügbaren Kovariaten, also Personen- und Umweltmerkmale sowie psychologische Konstrukte, in das Imputationsmodell einbezogen (Collins/Schafer/Kam 2001). Die im Ergebnisteil berichteten Statistiken und Koeffizienten sind Mittelwerte über alle Replikationen. Bei der Ermittlung von Standardfehlern wurde die Streuung zwischen den Replikationen berücksichtigt.

Zur Berechnung von Kovarianzen bzw. Korrelationen kamen Verfahren zum Einsatz, die dem Skalenniveau der einbezogenen Variablen Rechnung tragen.

3 Als „nicht zufällig fehlende Informationen“ sind solche Ausfälle zu betrachten, die von der Ausprägung der fraglichen Variable selbst und/oder anderen beobachteten Variablen abhängig sind; vgl. die Definition von *Missing Not At Random* (MNAR) und *Missing At Random* (MAR) bei Rubin (1976).

Tabelle 1 Schwierigkeiten und Trennschärfen der Kurzskaala-Items

Item-Nr.	Itembezeichner	Wissensbereich	Wissensdomäne	p	λ
1	med	Medizin	Naturw.	.88	.58
2	rel	Religion	Geistesw.	.69	.67
3	geo	Geografie	Naturw.	.78	.58
4	kun	Kunst	Geistesw.	.44	.64
5	bio	Biologie	Naturw.	.33	.45
6	pol	Politik	Sozialw.	.62	.52
7	phi	Philosophie	Geistesw.	.40	.51
8	phy	Physik	Naturw.	.44	.35
9	lit	Literatur	Geistesw.	.51	.56
10	fin	Finanzen	Sozialw.	.77	.58
11	rec	Recht	Sozialw.	.64	.39
12	ges	Geschichte	Sozialw.	.55	.46

Anmerkungen: *Naturw.:* Naturwissenschaften; *Geistesw.:* Geisteswissenschaften; *Sozialw.:* Sozialwissenschaften; *p:* relative Lösungshäufigkeit; *λ :* Itemladung im einfaktorischen Messmodell (vgl. Abbildung 1, Modell b).

Entsprechend der kategorialen (dichotomen) Natur der rekodierten Wissensitems wurden Messmodelle mittels konfirmatorischer Faktorenanalyse unter Verwendung des WLSMV-Schätzers in Mplus 6.1 (Muthén/Muthén 1998–2010) berechnet. Zur Beurteilung der Modellpassung wurden neben dem Chi-Quadrat-Wert und den Freiheitsgraden weitere gebräuchliche Fitindizes wie CFI (Comparative fit index), RMSEA (Root mean square error of approximation) und WRMR (Weighted root mean square residual) herangezogen. Nach Yu (2002) zeichnen sich Modelle mit kategorialen Daten und guter Passung durch folgende Werte aus: $CFI \geq .96$, $RMSEA \leq .05$ und $WRMR \leq .95$. Alle Analysen erfolgten unter Verwendung der Fallgewichte.

3 Ergebnisse

3.1 Itemstatistiken

Die Entwicklung der Kurzskaala erfolgte auf Basis des vorliegenden Datensatzes durch Itemselektion aus der 32 Items umfassenden Gesamtskaala nach den oben genannten Kriterien (vgl. Abschnitt 2.4). In Tabelle 1 sind die Schwierigkeiten, die Itemladungen (äquivalent zu Trennschärfen in der klassischen Testtheorie) und die

inhaltliche Zuordnung zu den drei breiten Wissensdomänen bzw. den einzelnen Wissensbereichen für die ausgewählten Items angegeben. Werden die entsprechenden Summenscores herangezogen, korreliert die Kurzskala mit der Gesamtskala zu $r = .91$.

3.2 Messmodelle und Reliabilität

Im Folgenden werden zwei konkurrierende Messmodelle für die Kurzskala gegenübergestellt und hinsichtlich ihrer Passung verglichen. Im dreifaktoriellen Modell werden drei Faktoren gemäß der drei breiten Wissensdomänen Natur-, Geistes- und Sozialwissenschaften spezifiziert. Auf jedem der drei Faktoren laden dabei nur die Items aus der entsprechenden Domäne. Da kristalline Intelligenz als übergeordnetes Fähigkeitskonstrukt Wissen aller Bereiche umfasst, ist von positiven Korrelationen zwischen den drei domänenspezifischen Faktoren auszugehen. Das einfaktorielles Messmodell repräsentiert kristalline Intelligenz hingegen mit einem einzigen Faktor, auf dem alle 12 Items der Kurzskala laden. Eine weitere Untergliederung in verschiedene Wissensdomänen wird nicht modelliert. Das einfaktorielles Modell stellt somit einen Spezialfall des komplexeren dreifaktoriellen Modells dar, weshalb zum Modellvergleich auch die Differenz der Chi-Quadrat-Werte formal auf Signifikanz getestet werden kann (Schulze 2004). Die beiden Messmodelle sind in Abbildung 1 dargestellt; die Passung der Modelle kann Tabelle 2 entnommen werden.

Das dreifaktorielle Modell (a) weist eine signifikant bessere Passung auf ($\Delta\chi^2(N = 1.134, 3) = 15.2, p = .002$), was sich auch an den anderen in Tabelle 2 ausgewiesenen Fitindizes zeigt. Der Unterschied ist allerdings gering; auch für das einfaktorielles Modell (b) ergibt sich eine zufriedenstellende Passung. Übereinstimmend mit dem geringen Unterschied in der Modellpassung zeigen sich im dreifaktoriellen Modell hohe messfehlerbereinigte Korrelationen zwischen den drei Wissensdomänen, die zwischen .80 und .96 liegen.

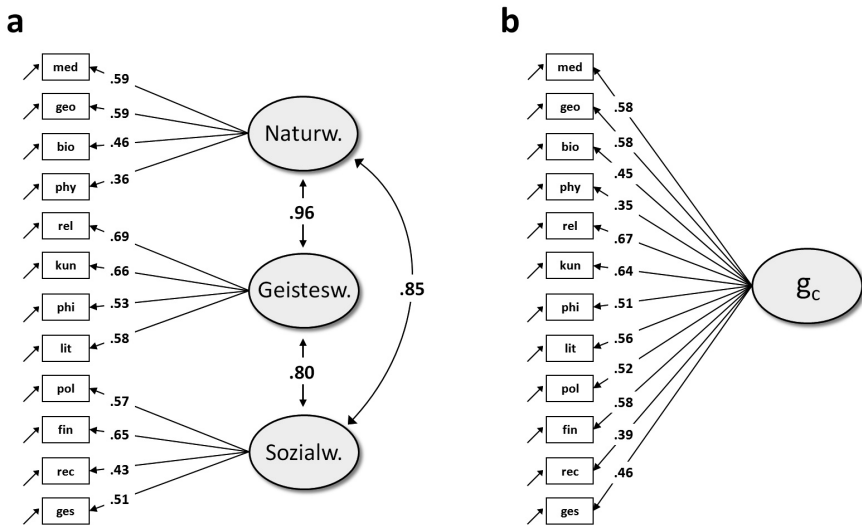
Für die folgenden Analysen wird das einfaktorielles Messmodell verwendet, da es eine zufriedenstellende Passung aufweist, die nur geringfügig schlechter ist als die Passung des dreifaktoriellen Modells. Insbesondere ist der inhaltliche Mehrwert des dreifaktoriellen Modells fraglich, da die drei breiten Wissensdomänen hoch zusammenhängen und in der Kurzskala nur durch jeweils vier Items repräsentiert werden, wodurch eine substanzielle Interpretation dieser spezifischeren Faktoren erschwert wird. Für das einfaktorielles Messmodell ergibt sich eine zufriedenstellende Reliabilität der latenten Variable von $\omega = .82$ (McDonald 1999) bzw. $\alpha = .81$ (Zumbo/Gadermann/Zeisser 2007). Die Reliabilität des manifesten Summenscores

Tabelle 2 Fitstatistiken konkurrierender Messmodelle

Modell	χ^2	df	CFI	RMSEA	WRMR
3 Faktoren	94.9	51	.973	.027	0.97
1 Faktor	110.8	54	.965	.030	1.06

Anmerkungen: χ^2 : Chi-Quadrat-Wert; df: Anzahl der Freiheitsgrade; CFI: Comparative fit index; RMSEA: Root mean square error of approximation; WRMR: Weighted root mean square residual.

Abbildung 1 Dreifaktorielles (a) und einfaktorielles Messmodell (b) der Kurzskala



Anmerkungen: Naturw.: Naturwissenschaften; Geistesw.: Geisteswissenschaften; Sozialw.: Sozialwissenschaften, g_c: kristalline Intelligenz; zu den Itembezeichnungen siehe Tabelle 1.

liegt bei .70 (Raykov/Dimitrov/Asparouhov 2010). Die entsprechenden Werte für die Gesamtskala mit 32 Items liegen mit .91 (ω/α) bzw. .84 (Skalenreliabilität nach Raykov et al. 2010) wegen der höheren Itemzahl erwartungsgemäß höher.

3.3 Überprüfung von Boden- und Deckeneffekten

Da bei Kurzskalen nur wenige Items eingesetzt werden, besteht im Vergleich zu umfangreicheren Skalen ein höheres Risiko von Boden- bzw. Deckeneffekten, d.h. mangelnder Differenzierungsfähigkeit innerhalb von Personengruppen mit sehr niedriger bzw. sehr hoher Fähigkeitsausprägung. Um zu überprüfen, ob solche

Effekte vorliegen, wird im Folgenden zunächst die Verteilung des Summenwerts der Kurzsкала in der Gesamtpopulation betrachtet (vgl. Abbildung 2). Zudem erfolgt eine Betrachtung zweier Subpopulationen mit geringer bzw. hoher Schulbildung (vgl. Abbildung 3). Die Subpopulation mit „geringer Schulbildung“ umfasst sowohl Personen ohne Schulabschluss als auch Personen mit Hauptschulabschluss (bzw. Äquivalent), die in der Regel nach der achten oder neunten Klasse die allgemeinbildende Schule verlassen haben. Ihr Anteil an der erwachsenen deutschen Wohnbevölkerung beträgt gemäß der vorliegenden Erhebung 45%. Personen mit „hoher Schulbildung“ im Sinne der hier vorgenommenen Analyse verfügen demgegenüber über eine fachgebundene oder allgemeine Hochschulreife bzw. Fachhochschulreife (26% der Gesamtpopulation), die typischerweise nach 12 oder 13 Jahren Schulbesuch erworben wurde.

Tabelle 3 gibt die Kennwerte der Verteilungen wieder. Für die Gesamtpopulation zeigen sich keine nennenswerten Boden- oder Deckeneffekte. Zwar liegt eine gering negative Schiefe vor, d.h. am oberen Ende der Fähigkeitskala ist die Differenzierungsfähigkeit minimal geringer. Dies ist jedoch praktisch kaum bedeutsam: Der Anteil der Personen, die alle Wissensitems der Kurzsкала richtig beantworten können, liegt in der Gesamtpopulation unter vier Prozent; im Mittel werden etwa sieben Items richtig gelöst. Ein differenzierteres Bild ergibt sich erwartungsgemäß für die zwei betrachteten Subpopulationen, die sich in der mittleren Lösungshäufigkeit deutlich unterscheiden. Während Personen mit „geringer Schulbildung“ im Mittel weniger als sechs Items richtig beantworten, liegt der Mittelwert in der Subpopulation mit „hoher Schulbildung“ etwas unter neun Items. Konkret beträgt die standardisierte Mittelwertdifferenz zwischen den beiden Subpopulationen $d = 1.12$ Standardabweichungseinheiten. In der relativ großen Gruppe der Personen ohne Schulabschluss bzw. mit Hauptschulabschluss liegt kein Bodeneffekt vor, hier liegt die Schiefe nahe null. Für die kleinere Gruppe der Personen mit Hochschulreife tritt erwartungsgemäß ein Deckeneffekt auf. Selbst in dieser sehr leistungsstarken Personengruppe liegt jedoch der Anteil derer, die alle Items richtig beantworten, unter 10 Prozent.

3.4 Zusammenhänge mit Kriterien

Ein wesentlicher Aspekt bei der Entwicklung und Beurteilung von Kurzsкаlen besteht in deren Beziehungen zu relevanten Personen- und Umweltmerkmalen sowie anderen psychologischen Konstrukten. Diese sollen einerseits im Einklang mit der Literatur stehen (Konstruktvalidierung). Andererseits sollen die für die Kurzsкала ermittelten Beziehungen möglichst den Befunden für die ungekürzte

Abbildung 2 Verteilung des Summenwerts der Kurzskala in der Gesamtpopulation

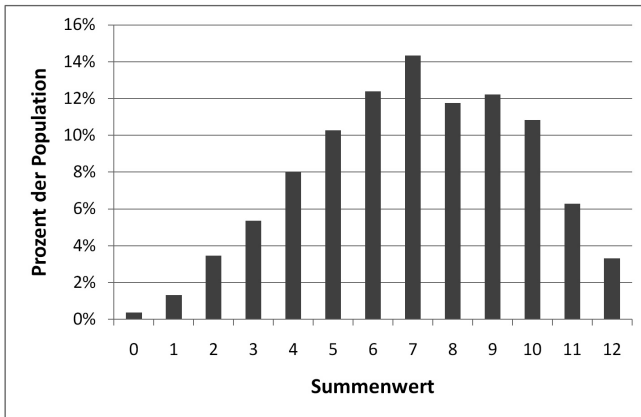


Abbildung 3 Verteilung des Summenwerts der Kurzskala in zwei Subpopulationen

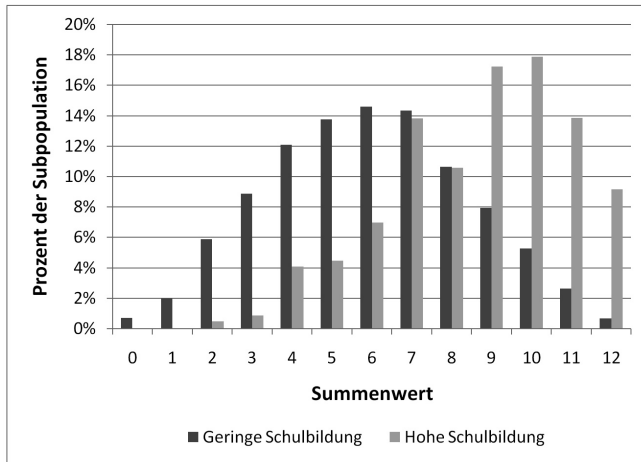


Tabelle 3 Kennwerte verschiedener Personenverteilungen für den Summenwert der Kurzskala

Population	<i>N</i>	<i>M</i>	<i>SD</i>	Schiefe	Exzess
Gesamtpopulation	1.134	7.04	2.66	-0.20	-0.64
„Geringe Schulbildung“	514	5.95	2.49	+0.06	-0.17
„Hohe Schulbildung“	290	8.66	2.27	-0.52	-0.38

Anmerkungen: *N*: Stichprobengröße; *M*: arithmetisches Mittel; *SD*: Standardabweichung.

Gesamtskala entsprechen, um sicherzustellen, dass die Itemselektion keine substantielle Minderung der Konstruktvalidität zur Folge hat (Widaman/Little/Preacher/Sawalani 2011). Im Folgenden werden daher die Korrelationen der Kurzskala mit verschiedenen Kovariaten näher betrachtet und den entsprechenden Korrelationen der Gesamtskala mit 32 Wissensitems gegenübergestellt (vgl. Tabelle 4). Zu den betrachteten Personenvariablen zählen Geschlecht, Alter, formale Bildung und sozioökonomischer Status der Teilnehmenden. Darüber hinaus werden an dieser Stelle auch Merkmale des Haushalts der Befragten (Haushaltsnettoeinkommen) sowie des elterlichen Haushalts analysiert (sozioökonomischer Status der Eltern, Anzahl der Bücher im elterlichen Haushalt zur Jugendzeit der Teilnehmenden), die als Indikatoren für Umfang und Reichhaltigkeit der früheren oder aktuellen Lernumwelt angesehen werden. Mit Blick auf psychologische Konstrukte werden die Beziehungen zu den fünf Hauptdimensionen der Persönlichkeit (Big Five) und zu selbstberichtetem Wissen untersucht. Bei der Interpretation der im folgenden dargestellten Beziehungen ist zu beachten, dass es sich um messfehlerbehaftete Korrelationen zwischen manifesten Variablen handelt. Zur Einordnung der Größe der Effekte kann eine Orientierung an Cohen (1988) erfolgen, der für Produkt-Moment-Korrelationen Werte um .10 als kleine Effekte, Werte um .30 als mittlere Effekte und Werte um .50 als große Effekte betrachtet. Wesentlicher ist jedoch der Vergleich der hier ermittelten Zusammenhänge mit den in der Einleitung dargestellten Erwartungen.

In Übereinstimmung mit der Literatur zeigen Männer etwas höhere Wissensleistungen als Frauen. Die standardisierte Mittelwertdifferenz beträgt $d = .30$ und ist somit inhaltlich bedeutsam, obgleich niedriger als von Ackerman et al. (2001) berichtet⁴. Wie erwartet wird in der hier untersuchten Erwachsenenpopulation kein bedeutsamer Alterseffekt beobachtet. Deklaratives Wissen weist eine hohe positive Korrelation mit dem ISCED-97 als Indikator formaler Bildung auf, der sowohl Schul- als auch Ausbildungsabschlüsse berücksichtigt: Für kein anderes hier untersuchtes Personen- oder Umweltmerkmal wurden höhere Korrelationen gefunden. Eine ebenfalls starke Beziehung zeigt sich zu dem auf der ISEI-Skala quantifizierten sozioökonomischen Status der Teilnehmenden. Auch für die anderen in Tabelle 4 genannten Personen- und Haushalts-

4 Für Männer und Frauen kann von Messinvarianz ausgegangen werden, d.h. die Kurzskala misst in beiden Gruppen dasselbe Konstrukt mit vergleichbarer Genauigkeit. Ein entsprechend restringiertes Multigruppenmodell unter Annahme strikter Messinvarianz zeigte sowohl für die Kurzskala mit 12 Items als auch für die Gesamtskala mit 32 Wissensitems eine befriedigende Passung (Werte für die Kurzskala: $\chi^2 = 203.7$, $df = 130$, RMSEA = .032, CFI = .953, WRMR = 1.548). Der latente Mittelwertunterschied betrug $0.32 SD$ zugunsten der Männer.

Tabelle 4 Korrelationen der Kurz- und Gesamtskala mit verschiedenen Personen- und Haushaltsmerkmalen sowie psychologischen Konstrukten

Variable	N_{vi}	Kurzskala		Gesamtskala	
		r	SE	r	SE
Geschlecht ¹	1.134	-.15	.03	-.15	.03
Alter	1.134	.01 ^{n.s.}	.03	.00 ^{n.s.}	.03
ISCED-97	1.091	.49	.03	.51	.03
ISEI	388	.44	.04	.45	.04
HISEI Eltern	1.082	.25	.03	.25	.03
Einkommen	638	.29	.04	.30	.04
Anzahl Bücher	1.101	.30	.03	.33	.03
selbstber. Wissen ²	1.134	.52	.03	.55	.03
Neurotizismus ³	1.104	-.10	.03	-.15	.03
Extraversion ³	1.104	.07 ^a	.04	.12	.04
Offenheit ³	1.104	.21	.03	.25	.03
Gewissenhaftigkeit ³	1.104	.07 ^a	.03	.09	.03
Verträglichkeit ³	1.104	-.02 ^{n.s.}	.03	-.02 ^{n.s.}	.03

Anmerkungen: $N = 1.134$. ¹ 0 = männlich, 1 = weiblich; ² selbstberichtetes Wissen (VOC-T Treffer); ³ Big Five-Dimensionen; ^a $p < .05$; ^{n.s.} nicht signifikant; N_{vi} : Fallzahl vor der Imputation; r : punkt-biseriale Korrelation (Geschlecht), polyseriale Korrelation (ISCED-97, Bücher), Produkt-Moment-Korrelation (alle anderen Variablen); SE : Standardfehler; ISCED-97: International Standard Classification of Education, Fassung 1997; ISEI: International Socio-Economic Index of Occupational Status; HISEI Eltern: Höchster ISEI-Wert der beiden Elternteile des Teilnehmenden. Sofern nicht anders gekennzeichnet, sind alle Korrelationen signifikant von null verschieden ($p < .01$).

merkmale liegen inhaltlich bedeutsame positive Korrelationen mit Wissen vor, die jedoch niedriger als die Zusammenhänge mit ISCED-97 und ISEI ausfallen.

Mit Blick auf die psychologischen Konstrukte liegt erwartungsgemäß eine vergleichsweise hohe Korrelation mit selbstberichtetem Wissen vor. Für die fünf Persönlichkeitsdimensionen ergibt sich ein differenziertes Ergebnismuster, das im Wesentlichen mit den in der Literatur berichteten Befunden übereinstimmt. Die Offenheitsdimension weist im Vergleich mit den anderen Big Five-Dimensionen die vom Betrag her höchste Korrelation mit Wissen auf. Auch die gering positive Korrelation mit Extraversion, die gering negative Beziehung zu Neurotizismus sowie die nicht signifikante Korrelation mit Verträglichkeit entsprechen den metaanalytisch gewonnenen Ergebnissen von Ackerman und Heggestad (1997). Eine Abweichung lässt sich allein für Gewissenhaftigkeit feststellen: Während Ackerman und Heggestad (1997) hier eine negative Korrelation mit Wissen bzw. eine nicht signifikante

Korrelation mit kristalliner Intelligenz berichten, wird in der aktuellen Analyse eine gering positive Beziehung zwischen Wissen und Gewissenhaftigkeit gefunden. Für alle berichteten Kovariaten ergeben sich im Vergleich von Kurz- und Gesamtskala sehr ähnliche Korrelationen.

4 Diskussion

Da der Umfrageforschung im deutschen Sprachraum bis dato nur wenige frei verfügbare und erprobte Instrumente zur Erfassung kognitiver Fähigkeiten vorliegen, war das Ziel dieses Beitrags die Entwicklung einer Kurzskaala zur Messung kristalliner Intelligenz (g_c) und die Prüfung ihrer psychometrischen Eigenschaften. Die hier beschriebene Kurzskaala BEFKI GC-K umfasst 12 Items, die in Übereinstimmung mit der Definition von g_c durch Cattell (1971) bzw. Carroll (1993) deklaratives Wissen aus ebensovielen Bereichen der Natur-, Geistes- und Sozialwissenschaften erfassen. Somit wird innerhalb von fünf Minuten Bearbeitungszeit ein möglichst breites Wissensspektrum berücksichtigt. Durch Einbezug von Items unterschiedlicher Schwierigkeit gelang es, trotz der vergleichsweise geringen Itemanzahl Boden- und Deckeneffekte in der erwachsenen deutschen Wohnbevölkerung fast vollständig zu vermeiden. Lediglich in der besonders leistungsstarken Subpopulation mit Hochschulabschluss bzw. Fachhochschulabschluss trat ein geringer Deckeneffekt auf. Weiterhin weist die Kurzskaala eine gute Reliabilität auf, welche gängige Standards für die Forschung erfüllt. Die zeitliche Stabilität der Testwerte konnte aufgrund der querschnittlichen Natur der Erhebung nicht geprüft werden. Daten einer mit Schülerinnen und Schülern der Mittelstufe durchgeführten Längsschnittuntersuchung mit Wissensitems aus dem BEFKI-Pool legen jedoch eine befriedigende zeitliche Stabilität nahe (Wilhelm et al., in Vorbereitung).

Sowohl für ein einfaktorielles Messmodell als auch für ein mehrdimensionales Modell, das drei korrelierte Faktoren gemäß der Unterscheidung zwischen Natur-, Geistes- und Sozialwissenschaften spezifiziert, ergab sich eine zufriedenstellende Passung bei statistisch signifikanter Überlegenheit des dreifaktoriellen Modells. Die Bevorzugung des eindimensionalen Modells im vorliegenden Beitrag beruht auf mehreren Argumenten: Erstens ist zu berücksichtigen, dass der χ^2 -Differenztest von der Stichprobengröße abhängig ist und daher in der vorliegenden, vergleichsweise großen Stichprobe auch inhaltlich unbedeutende Unterschiede als signifikant ausgewiesen werden. Zweitens ist das einfaktorielle Modell insofern theoretisch fundiert, als Wissen aus verschiedenen Bereichen in konsensualen Intelligenzstrukturtheorien einem gemeinsamen, bereichsübergreifenden g_c -

Faktor untergeordnet ist (Horn/Noll 1997; Carroll 1993, 2003; McGrew 2009). Dies wird auch empirisch durch die sehr hohen Korrelationen zwischen den Wissensfaktoren im dreidimensionalen Modell gestützt. Drittens ist zu berücksichtigen, dass die spezifischeren Faktoren des dreidimensionalen Modells mit nur jeweils vier Items extrem schmal operationalisiert sind. Eine inhaltliche Interpretation dieser Faktoren wäre fragwürdig, da die Itemstichprobe aufgrund dieser sehr geringen Itemzahl keine Repräsentativität für die jeweilige Wissensdomäne beanspruchen kann (Ackerman 2000). Ferner ist die im Vergleich zu einem Gesamtwert geringere Reliabilität von Subskalen zu berücksichtigen (Sinharay 2010). Im Extremfall kann dies dazu führen, dass der Gesamtwert einen besseren Prädiktor der Personenfähigkeit auf einer Subdimension darstellt als der entsprechende Subskalenwert (Sinharay/Haberman/Puhan 2007). Selbst bei vermeintlich ausreichender Reliabilität einer Subskala kann diese durch die übergeordnete Fähigkeit bedingt sein statt durch die spezifische Subdimension (Reise/Moore/Haviland 2010).

Die Validität der g_c -Kurzsкала wurde anhand der Korrelationen des Gesamtwertes mit verschiedenen Personen- und Umweltmerkmalen einerseits sowie ausgewählten psychologischen Konstrukten andererseits überprüft, wobei diese Beziehungen fast durchgängig mit den auf Basis der Fachliteratur formulierten Erwartungen übereinstimmten. Die Höhe der untersuchten Effekte bzw. Zusammenhänge lag für die Kurzsкала mit 12 Items nur geringfügig unter den entsprechenden Werten für die ungekürzte Wissensskala mit 32 Items; die Abweichungen lassen sich durch die etwas geringere Reliabilität der Kurzsкала erklären. Im Einzelnen wurde übereinstimmend mit den oben zitierten Befunden ein Geschlechtsunterschied im deklarativen Wissen zugunsten von Männern gefunden. Dies stellt keine Verzerrung und somit unfaire Messung dar, sondern repliziert den Befund von Ackerman et al. (2001), dass bei einem umfassenden Sampling von Items aus vielen verschiedenen Wissensbereichen in den meisten dieser Bereiche ein Wissensvorsprung der Männer zu beobachten ist. Dass der Mittelwertunterschied in der hier vorgestellten Skala weniger deutlich ausfällt, könnte auf kulturelle Unterschiede zurückzuführen sein oder darin begründet liegen, dass andere Studien teilweise deutlich größere Itemmengen eingesetzt haben, einschließlich sehr spezifischer Wissensfragen, die als berufsspezifische Expertise einzuordnen sind (vgl. den g_{kn} -Faktor im CHC-Modell; McGrew 2009). In der vorliegenden Kurzsкала wurde hingegen auf derartige Items zu hochspezifischem Wissen verzichtet, da sie nur innerhalb bestimmter Subpopulationen funktionieren. Dies ist auch ein Grund dafür, dass kein substanzieller Zusammenhang zwischen der g_c -Kurzsкала mit dem Alter festgestellt werden konnte. Mit Blick auf die hohe Stabilität kristalliner Intelligenz im Erwachsenenalter war ein kleiner oder nicht signifikanter Effekt erwartet

worden. Zwar werden in der Literatur Wissenszuwächse auch im jungen und mittleren Erwachsenenalter berichtet (Ackerman 2000), diese beziehen sich aber auf Expertise im Sinne von Wissen, das erst im Erwachsenenalter erworben werden kann, beispielsweise im Rahmen der Berufsausübung (Ackerman 1996). Zum anderen umfasst die untersuchte Population neben jungen Erwachsenen auch Erwachsene in sehr hohem Alter, für die in der Literatur biologisch bedingt abnehmende Leistungen auch in Tests kristalliner Fähigkeiten berichtet werden (Li 2003).

Erwartungskonform sind ebenfalls die substanziellen positiven Beziehungen der Kurzsкала mit Indikatoren der formalen Bildung und der Reichhaltigkeit der Lernumwelt, die im sozioökonomischen Status zum Ausdruck kommt. Sowohl ISCED-97 als auch ISEI der Teilnehmenden hängen mit dem über einen langen Zeitraum in einer Vielzahl von Bereichen erworbenen Wissen zusammen. Die vergleichsweise niedrigeren Korrelationen mit dem sozioökonomischen Status der Eltern, dem Haushaltsnettoeinkommen sowie der Anzahl der Bücher im elterlichen Haushalt zur Jugendzeit der Teilnehmenden sind insofern plausibel, als es sich dabei um eher indirekte bzw. inhaltlich weniger breite Indikatoren handelt, die begrenzte Aussagekraft bzgl. der Lernumwelt haben.

Mit Blick auf andere psychologische Konstrukte konnte gezeigt werden, dass eine relativ hohe positive Korrelation zwischen der Kurzsкала und selbstberichteten Wissen vorliegt. Dieser Befund überrascht nicht, da beide Skalen auf deklaratives Wissen abzielen und die Auswahl der erfassten Wissensbereiche in beiden Messverfahren auf der Systematik von Ackerman (2000) beruht. Dass kein perfekter Zusammenhang vorliegt, dürfte – neben der Tatsache, dass die berichteten Korrelationen messfehlerbehaftet sind – insbesondere an der unterschiedlichen Natur der Messverfahren liegen (Selbstauskünfte versus tatsächliches Wissen). Die Beziehungen der g_c -Kurzsкала zu den Big Five zeigten ein Korrelationsmuster, das mit den metaanalytischen Befunden von Ackerman und Heggstad (1997) weitgehend übereinstimmt. Die höchste Korrelation wies die g_c -Kurzsкала wie erwartet mit der Offenheitsdimension auf. Im Widerspruch zu den Ergebnissen von Ackerman und Heggstad (1997) steht lediglich die leicht positive Korrelation mit der Gewissenhaftigkeitsdimension. Hierbei ist jedoch zu bedenken, dass die diesbezügliche Datenbasis in der genannten Metaanalyse sehr klein war und ein gering positiver Zusammenhang zwischen Gewissenhaftigkeit und Wissensleistungen aus theoretischer Sicht erklärbar ist. Man denke etwa an die umfangreiche Literatur zur prädiktiven Validität von Gewissenhaftigkeit für den Berufserfolg (Barrick/Mount/Judge 2001). Befunde zu den Beziehungen der Kurzsкала mit anderen g_c -Indikatoren sowie mit g_f liegen bisher nur für Langformen aus Pilotierungs- und Normierungsstudien vor, die mit Schülerpopulationen durchgeführt wurden. Auf

der Ebene latenter Variablen konnte bei Schülerinnen und Schülern der achten bis zehnten Jahrgangsstufe eine hohe Korrelation zwischen deklarativem Wissen und Wortschatz ($\rho = .93$) und eine ebenfalls substanzielle Beziehung zwischen Wissen und schlussfolgerndem Denken ($\rho = .80$) gezeigt werden (Wilhelm et al., in Vorbereitung).

Bei der Interpretation der im vorliegenden Beitrag berichteten Befunde ist einschränkend zu berücksichtigen, dass die Itemselektion für die Kurzsкала und deren Evaluation auf derselben Erhebung basieren. Zudem lag für den Wissenstest eine relativ strenge Zeitbegrenzung vor, die zu fehlenden Werten führte, da nicht alle teilnehmenden Personen den Test in der vorgegebenen Zeit abschließen konnten. Diese aufgrund der Zeitbegrenzung fehlenden Werte als Falschantworten zu werten, hätte eine Verzerrung des Skalenwerts durch konstruktirrelevante Varianz – etwa Unterschiede in mentaler Geschwindigkeit – zur Folge gehabt. Um diese Verzerrung zu vermeiden, wurden daher fehlende Werte unter Berücksichtigung aller verfügbaren Informationen imputiert; dadurch ist die größtmögliche Vergleichbarkeit mit einer Durchführung ohne Zeitmangel gewährleistet. Eine Bearbeitungszeit von fünf Minuten für die 12 Items der Kurzsкала ist jedoch ausreichend, so dass eine Imputation fehlender Werte unter diesen Durchführungsbedingungen nicht erforderlich ist (stattdessen sollten ausgelassene Items wie Falschantworten ausgewertet werden). In zukünftigen Studien sollte geprüft werden, inwiefern die hier berichteten Befunde auf andere Stichproben und die veränderten Durchführungsbedingungen übertragbar sind.

5 Abschließende Bemerkungen

Die hier vorgestellte Kurzsкала BEFKI GC-K zur Erfassung kristalliner Intelligenz stellt unseres Wissens das erste frei verfügbare Verfahren zur g_c -Messung dar, das sich aufgrund seiner Kürze und psychometrischen Effizienz für den Einsatz in der Umfrageforschung eignet. Die Skala ist sowohl in technologiebasierten Testungen (computer assisted personal interview/self-interview/web-interview) als auch bei herkömmlichen Papier-Stift-Testungen (Selbstaussfüller) leicht anzuwenden; neben Einzeltestungen sind auch Gruppentestungen problemlos möglich. Einschränkend gilt hier, dass die Bewährung der Kurzsкала mit anderen Testmedien bislang nicht geprüft wurde. Mit Blick auf die umfangreiche Literatur zu Testmedienvergleichen (vgl. etwa Mead/Drasgow 1993; Schroeders/Wilhelm 2010) sind für die g_c -Skala bei ansonsten gleichen Durchführungsbedingungen jedoch keine nennenswerten Testmedieneffekte zu erwarten. Gegenüber gängigen Bildungsindikatoren wie

der ISCED-97 hat die Wissensskala den Vorteil, dass tatsächliches Wissen direkt erhoben wird, anstatt es aus Abschlüssen zu erschließen. Letzteres ist insbesondere deshalb problematisch, da Schul- und Ausbildungsabschlüsse einerseits zwischen verschiedenen Bundesländern oder gar Staaten schwer vergleichbar sind (Neumann/Nagy/Trautwein/Lüdtke 2009) und formale Abschlüsse andererseits dem zeitlichen Wandel unterliegen. Zudem sind Selbstauskünfte wie Angaben zu Bildungsabschlüssen oder zur Berufstätigkeit im Gegensatz zu objektiven Leistungstests leicht verfälschbar (Ziegler/MacCann/Roberts 2011). Ungeachtet der Vorzüge der hier vorgestellten g_c -Skala ist jedoch zu berücksichtigen, dass es sich um eine Kurzsкала handelt, die zur Wissensmessung in der sehr bildungsheterogenen erwachsenen Bevölkerung konstruiert wurde. Daher ist die inhaltliche Breite der Messung stark eingeschränkt und kann eine differenzierte Wissensdiagnostik nicht ersetzen. Stattdessen fokussiert die Kurzsкала auf „Wissen, von dem angenommen werden kann, dass es von einem großen Teil der Population geteilt wird“ (Ackerman 2003: 16). Mithilfe komplexer Testdesigns und Modellierungsmethoden ist es jedoch möglich, die Vorzüge von Kurzsкаlen mit den Vorteilen umfangreicher Messinstrumente zu kombinieren (Rhemtulla/Little 2012).

Die Kurzsкала kann für nicht-kommerzielle Forschungszwecke kostenfrei eingesetzt werden. Eine Druckvorlage ist über folgenden Link verfügbar: <http://befki.de/materialien> (Kennwort: trgw34785bns)

Literatur

- Ackerman, P. L., 1996: A theory of adult intellectual development: Process, personality, interests, and knowledge. *Intelligence* 22: 227-257.
- Ackerman, P. L., 2000: Domain-Specific Knowledge as the „Dark Matter“ of Adult Intelligence: Gf/Gc, Personality and Interest Correlates. *Journal of Gerontology: Psychological Sciences* 55B: 69-84.
- Ackerman, P. L., 2003: Cognitive ability and non-ability trait determinants of expertise. *Educational Researcher* 32 (8): 15-20.
- Ackerman, P. L., 2008: Knowledge and cognitive aging. S. 445-489 in: F. I. M. Craik und T. A. Salthouse (Hg.): *The handbook of aging and cognition* (3rd ed.). New York, NY, USA: Psychology Press.
- Ackerman, P. L. und M. E. Beier, 2004: Knowledge and Intelligence. S. 125-139 in: O. Wilhelm und R. Engle (Hg.): *Handbook of understanding and measuring intelligence*. Thousand Oaks, CA: Sage.
- Ackerman, P. L., K. R. Bowen, M. B. Beier und R. Kanfer, 2001: Determinants of individual differences and gender differences in knowledge. *Journal of Educational Psychology* 93: 797-825.
- Ackerman, P. L. und E. D. Heggstad, 1997: Intelligence, personality, and interests: Evidence for overlapping traits. *Psychological Bulletin* 121: 219-245.

- Ackerman, P. L. und E. L. Rolfhus, 1999: The locus of adult intelligence: Knowledge, abilities, and non-ability traits. *Psychology and Aging* 14: 314–330.
- Baltes, P. B., 1987: Theoretical propositions of life-span developmental psychology: on the dynamics between growth and decline. *Developmental Psychology* 23: 611–626.
- Baltes, P. B., U. M. Staudinger und U. Lindenberger, 1999: Lifespan Psychology: Theory and Application to Intellectual Functioning. *Annual Review of Psychology* 50: 471–507.
- Barrick, M. R., M. K. Mount und T. A. Judge, 2001: Personality and performance at the beginning of the new millennium: What do we know and where do we go next? *International Journal of Selection and Assessment* 9: 9–20.
- Binet, A. und T. Simon, 1905: Méthodes nouvelles pour le diagnostic du niveau intellectuel des anormaux. *L'année psychologique* 11: 191–244.
- Brennan, R. L., 2001: *Generalizability Theory*. New York: Springer.
- Carroll, J. B., 1993: *Human cognitive abilities: A survey of factor-analytic studies*. New York: Cambridge University Press.
- Carroll, J. B., 2003: The higher stratum structure of cognitive abilities: Current evidence supports g and about ten broad factors. S. 153–193 in: H. Nyborg (Hg.): *The scientific study of general intelligence: Tribute to Arthur R. Jensen*. New York: Pergamon.
- Cattell, R. B., 1943: The measurement of adult intelligence. *Psychological Bulletin* 40: 153–193.
- Cattell, R.B., 1963: Theory of fluid and crystallized intelligence: A critical experiment. *Journal of Educational Psychology* 54: 1–22.
- Cattell, R. B., 1971: *Abilities: Their Structure, Growth, and Action*. Boston, MA: Houghton Mifflin.
- Ceci, S. J., 1991: How much does schooling influence general intelligence and its cognitive components? A reassessment of the evidence. *Developmental Psychology* 27: 703–722.
- Cliffordson, C. und J.-E. Gustafsson, 2008: Effects of age and schooling on intellectual performance: Estimates obtained from analysis of continuous variation in age and length of schooling. *Intelligence* 36: 143–152.
- Cohen, J., 1988: *Statistical Power Analysis for the Behavioral Sciences*, 2nd ed. Hillsdale: Lawrence Erlbaum Associates.
- Collins, L. M., J. L. Schafer und C.-M. Kam, 2001: A comparison of inclusive and restrictive strategies in modern missing data procedures. *Psychological Methods* 6: 330–351.
- Cronbach, L. J., 1949: *Essentials of psychological testing*. New York: Harper.
- Danthiir, V., R. D. Roberts, R. Schulze und O. Wilhelm, 2004: Mental Speed: On Frameworks, Paradigms, and a Platform for the Future. S. 27–46 in: O. Wilhelm und R. W. Engle (Hg.): *Handbook of Understanding and Measuring Intelligence*. London: Sage.
- Ebbinghaus, H., 1897: Über eine neue Methode zur Prüfung geistiger Fähigkeiten und ihre Anwendung bei Schulkindern. *Zeitschrift für Psychologie und Physiologie der Sinnesorgane* 13: 401–459.
- Ehmke, T. und T. Siegle, 2005: ISEI, ISCED, HOMEPOS, ESCS. Indikatoren der sozialen Herkunft bei der Quantifizierung von sozialen Disparitäten. *Zeitschrift für Erziehungswissenschaft* 8: 521–539.
- Furnham, A. und G. Dissou, 2007: The relationship between self-estimated and test-derived scores of personality and intelligence. *Journal of Individual Differences* 28: 37–44.
- Ganzeboom, H. B. G., P. M. De Graaf und D. J. Treiman, 1992: A Standard International Socio-Economic Index of Occupational Status. *Social Science Research* 21: 1–56.
- Grabner, R. H. und E. Stern, 2010: Measuring Cognitive Ability. S. 753–768 in: *Rat für Sozial- und Wirtschaftsdaten* (Hg.): *Building on progress: Expanding the research infrastructure for the social, economic, and behavioral sciences*. Opladen: Budrich UniPress.

- Graham, J. W., A. E. Olchowski und T. D. Gilreath, 2007: How many imputations are really needed? Some Practical Clarifications of Multiple Imputation Theory. *Prevention Science* 8: 206-213.
- Hebb, D. O., 1942: The effect of early and late brain injury upon test scores and the nature of normal adult intelligence. *Proceedings of the American Philosophical Society* 85: 275-292.
- Horn, J. L., 1965: Fluid and crystallized intelligence: A factor analytic study of the structure among primary mental abilities. Dissertation, University of Illinois.
- Horn, J. L., 1988: Thinking about human abilities. S. 645-685 in: J. R. Nesselroade (Hg.): *Handbook of multivariate psychology*. New York: Academic Press.
- Horn, J. L., und R. B. Cattell, 1966: Refinement and test of the theory of fluid and crystallized general intelligences. *Journal of Educational Psychology* 57: 253-270.
- Horn, J. L. und J. Noll, 1997: Human cognitive capabilities: Gf-Gc theory. S. 53-91 in: D. P. Flanagan, J. L. Genshaft und P. L. Harrison (Hg.): *Contemporary intellectual assessment: Theories, tests and issues*. New York: Guilford.
- Hülür, G., O. Wilhelm und S. Schipolowski, 2011: Prediction of self-reported knowledge with over-claiming, fluid and crystallized intelligence and typical intellectual engagement. *Learning and Individual Differences* 21: 742-746.
- ILO, 2007: International Standard Classification of Occupations ISCO-08. <http://www.ilo.org/public/english/bureau/stat/isco/isco08/index.htm> (26.11.2012).
- Lang, F. R., D. Weiss, A. Stocker und B. von Rosenblatt, 2007: Assessing Cognitive Capacities in Computer-Assisted Survey Research: Two Ultra-Short Tests of Intellectual Ability in the German Socio-Economic Panel (SOEP). *Schmollers Jahrbuch* 127: 183-192.
- Li, S.-C., 2003: Biocultural orchestration of developmental plasticity across levels: The interplay of biology and culture in shaping the mind and behavior across the life span. *Psychological Bulletin* 129: 171-194.
- Lüdtke, O., A. Robitzsch, U. Trautwein und O. Köller, 2007: Umgang mit fehlenden Werten in der psychologischen Forschung: Probleme und Lösungen. *Psychologische Rundschau* 58: 103-117.
- Lynn, R. und P. Irwing, 2002: Sex differences in general knowledge, semantic knowledge and reasoning ability. *British Journal of Psychology* 93: 545-556.
- Lynn, R., P. Irwing und T. Cammock, 2001: Sex differences in general knowledge. *Intelligence* 30: 27-40.
- Lynn, R., S. Wilberg und J. Margraf-Stiksrud, 2004: Sex differences in general knowledge in German high school students. *Personality and Individual Differences* 37: 1643-1650.
- Mead, A. D. und F. Drasgow, 1993: Equivalence of computerized and paper-and-pencil cognitive ability tests: A meta-analysis. *Psychological Bulletin* 114: 449-458.
- McDonald, R. P., 1999: *Test theory: A unified treatment*. Mahwah, NJ: Erlbaum.
- McGrew, K. S., 2009: CHC theory and the human cognitive abilities project: Standing on the shoulders of the giants of psychometric intelligence research. *Intelligence* 37: 1-10.
- Muthén, L. K. und B. O. Muthén, 1998-2010: *Mplus User's Guide*. Sixth Edition. Los Angeles, CA: Muthén & Muthén.
- Neumann, M., G. Nagy, U. Trautwein und O. Lüdtke, 2009: Vergleichbarkeit von Abiturleistungen: Leistungs- und Bewertungsunterschiede zwischen Hamburger und Baden-Württemberger Abiturienten und die Rolle zentraler Abiturprüfungen. *Zeitschrift für Erziehungswissenschaft* 12: 691-714.
- Ostendorf, F. und A. Angleitner, 2004: NEO-Persönlichkeitsinventar nach Costa und McCrae, Revidierte Fassung (NEO-PI-R). Göttingen: Hogrefe.

- Paulhus, D. L. und P. D. Harms, 2004: Measuring cognitive ability with the overclaiming technique. *Intelligence* 32: 297-314.
- Rammstedt, B., C. J. Kemper, M. C. Klein, C. Beierlein und A. Kovaleva, 2012: Eine kurze Skala zur Messung der fünf Dimensionen der Persönlichkeit: Big-Five-Inventory-10 (BFI-10). *GESIS-Working Papers 2012|23*. Köln: GESIS.
- Rammstedt, B. und O. P. John, 2007: Measuring personality in one minute or less: A 10-item short version of the Big Five Inventory in English and German. *Journal of Research in Personality* 41: 203-212.
- Rat für Sozial- und Wirtschaftsdaten, 2010: Building on progress: Expanding the research infrastructure for the social, economic, and behavioral sciences. Opladen: Budrich Uni-Press.
- Raykov, T., D. M. Dimitrov und T. Asparouhov, 2010: Evaluation of scale reliability with binary measures using latent variable modeling. *Structural Equation Modeling: A Multidisciplinary Journal* 17: 265-279.
- Reise, S. P., T. M. Moore und M. G. Haviland, 2010: Bifactor Models and Rotations: Exploring the Extent to Which Multidimensional Data Yield Univocal Scale Scores. *Journal of Personality Assessment* 92: 544-559.
- Rhemtulla, M. und T. D. Little, 2012: Planned Missing Data Designs for Research in Cognitive Development. *Journal of Cognition and Development* 13: 425-438.
- Rolfhus, E. L. und P. L. Ackerman, 1999: Assessing individual differences in knowledge: Knowledge structures and traits. *Journal of Educational Psychology* 91: 511-526.
- Rowe, D. C., K. C. Jacobson und E. J. van den Oord, 1999: Genetic and environmental influences on vocabulary IQ: parental education level as moderator. *Child Development* 70: 1151-1162.
- Rubin, D. B., 1976: Inference and missing data. *Biometrika* 63: 581-592.
- Schipolowski, S., U. Schroeders und O. Wilhelm, 2008: BEFKI - Berlin Test of Fluid and Crystallized Intelligence. Präsentation auf dem XXIX International Congress of Psychology, Berlin.
- Schroeders, U. und O. Wilhelm, 2010: Testing reasoning ability with handheld computers, notebooks, and paper and pencil. *European Journal of Psychological Assessment* 26: 284-292.
- Schroeders, U., S. Schipolowski, C. Nelles und O. Wilhelm, 2011: Interest, domain specific knowledge, and fluid intelligence: Profile covariances and prediction of vocational training success. Präsentation auf dem 15th Biennial Meeting of the International Society for the Study of Individual Differences, London, UK.
- Schroeders, U., S. Schipolowski und O. Wilhelm, 2010: Berliner Test zur Erfassung fluider und kristalliner Intelligenz. Präsentation auf dem 47. Kongress der Deutschen Gesellschaft für Psychologie, Bremen.
- Schulze, R., 2004: Modeling Structures of Intelligence. S. 241-263 in O. Wilhelm und R. Engle (Hg.): *Handbook of understanding and measuring intelligence*. Thousand Oaks, CA: Sage.
- Sinharay, S., 2010: How Often Do Subscores Have Added Value? Results from Operational and Simulated Data. *Journal of Educational Measurement* 47: 150-174.
- Sinharay, S., S. Haberman und G. Puhon, 2007: Subscores Based on Classical Test Theory: To Report or Not to Report. *Educational Measurement: Issues and Practice* 26 (4): 21-28.
- Spearman, C. E., 1938: Measurement of intelligence. *Scientia* 64: 75-82.
- UNESCO, 1997: International Standard Classification of Education ISCED 1997. http://www.unesco.org/education/information/nfsunesco/doc/iscsed_1997.htm (20.11.2012).

- Widaman, K. F., T. D. Little, K. J. Preacher und G. M. Sawalani, 2011: On creating and using short forms of scales in secondary research. S. 39-61 in: K. H. Trzesniewski, M. B. Donnellan und R. E. Lucas (Hg.): *Secondary Data Analysis: An Introduction for Psychologists*. Washington, DC: American Psychological Association.
- Wilhelm, O. und S. Schipolowski, 2010: Intelligenzdiagnostik in der Pädagogischen Psychologie. In G. L. Huber (Hg.): *Enzyklopädie Erziehungswissenschaft Online*. Fachgebiet Pädagogische Psychologie. Weinheim/München: Juventa.
- Wilhelm, O., U. Schroeders und S. Schipolowski, in Vorbereitung: BEFKI – Berliner Test zur Erfassung fluider und kristalliner Intelligenz. Mittelstufenform. Göttingen: Hogrefe.
- Wolf, C., 1995: Sozio-ökonomischer Status und berufliches Prestige. *ZUMA-Nachrichten* 37: 102-136.
- Von der Heyde, C., 2009: Das ADM-Stichprobensystem für persönlich-mündliche Befragungen. http://www.adm-ev.de/fileadmin/user_upload/PDFS/Beschreibung-ADM-Stichproben-f2f_DE.pdf (30 KB) (20.11.2012)
- Von Stumm, S. und P. L. Ackerman, 2012: Investment and Intellect: A Review and Meta-Analysis. *Psychological Bulletin: Advance online publication*, doi: 10.1037/a0030746.
- Yu, C. Y., 2002: Evaluating cutoff criteria of model fit indices for latent variable models with binary and continuous outcomes. Doctoral dissertation, University of California, Los Angeles.
- Ziegler, M., E. Danay, M. Heene, J. Asendorpf und M. Bühner, 2012: Openness, fluid intelligence, and crystallized intelligence: Toward an integrative model. *Journal of Research in Personality* 46: 173-183.
- Ziegler, M., C. J. Kemper und B. Rammstedt, 2013: The Vocabulary and Overclaiming Test (VOC-T). *Journal of Individual Differences* 34: 32-40.
- Ziegler, M., C. MacCann und R. D. Roberts, 2011: *New Perspectives on Faking in Personality Assessment*. New York: Oxford University Press.
- Zumbo, B. D., A. M. Gadermann und C. Zeisser, 2007: Ordinal versions of coefficients alpha and theta for Likert rating scales. *Journal of Modern Applied Statistical Methods* 6: 21-29.

Anschrift des Autors	Stefan Schipolowski Humboldt-Universität zu Berlin Institut zur Qualitätsentwicklung im Bildungswesen Unter den Linden 6 10099 Berlin stefan.schipolowski@iqb.hu-berlin.de
Ko-Autor/-innen	Oliver Wilhelm Institut für Psychologie und Pädagogik Universität Ulm Ulrich Schroeders Institut zur Qualitätsentwicklung im Bildungswesen Humboldt-Universität zu Berlin Anastassiya Kovaleva Fakultät für Biologie, Universität Bielefeld Christoph J. Kemper Institut für medizinische und pharmazeutische Prüfungsfragen, Mainz Beatrice Rammstedt GESIS – Leibniz-Institut für Sozialwissenschaften, Mannheim