

gesis

Leibniz-Institut
für Sozialwissenschaften

Working Papers

2014|29

Eine kurze Skala zur Messung kristalliner Intelligenz

Die Kurzsкала g_c des Berliner Tests zur Erfassung
Fluiden und Kristallinen Intelligenz (BEFKI GC-K)

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

GESIS-Working Papers 2014|29

Eine kurze Skala zur Messung kristalliner Intelligenz

Die Kurzskala g_c des Berliner Tests zur Erfassung Fluider und Kristalliner Intelligenz (BEFKI GC-K)

Stefan Schipolowski¹, Oliver Wilhelm², Ulrich Schroeders¹, Anastassiya Kovaleva³, Christoph J. Kemper³ & Beatrice Rammstedt³

¹ Institut zur Qualitätsentwicklung im Bildungswesen, Humboldt-Universität zu Berlin

² Institut für Psychologie und Pädagogik, Universität Ulm

³ GESIS – Leibniz-Institut für Sozialwissenschaften

GESIS-Working Papers

GESIS – Leibniz-Institut für Sozialwissenschaften

Postfach 12 21 55

68072 Mannheim

Telefon: (0621) 1246 - 253

Telefax: (0621) 1246 - 500

E-Mail: constanze.beierlein@gesis.org

ISSN: 1869-0491 (Online)

Herausgeber,

Druck und Vertrieb:

GESIS – Leibniz-Institut für Sozialwissenschaften
Unter Sachsenhausen 6-8, 50667 Köln

1 Einleitung¹

Erhebungsinstrumente zur Erfassung psychologischer Merkmale wie Persönlichkeit, Risikobereitschaft, Lebenszufriedenheit, Optimismus oder Intelligenz werden immer häufiger in der Forschung eingesetzt. Neben der psychologischen Forschung, in der sie traditionell im Zentrum des Interesses stehen, werden psychologische Merkmale vermehrt auch in anderen Forschungsbereichen erhoben, zum Beispiel in den Sozialwissenschaften, den Wirtschaftswissenschaften und den Gesundheitswissenschaften. Die Forscherinnen und Forscher der verschiedenen Disziplinen erwarten von der Erfassung dieser Merkmale in Umfragen, Studien und Experimenten sowie deren Implementierung in ihre Modelle eine bessere Deskription und ein tieferes Verständnis wissenschaftlich und gesellschaftlich relevanter Prozesse und Phänomene.

Der Zusammenhang zwischen psychologischen Merkmalen und sozioökonomischen Erfolgsgrößen, die bei Sozialwissenschaftlern und Ökonomen im Fokus des Interesses stehen, gilt als gut gesichert (Gottfredson, 1997; Gottfredson & Deary, 2004; Schmidt & Hunter, 1998; Strenze, 2007). Diese Studien belegen, dass die kognitiven Fähigkeiten einer Person, insbesondere die allgemeine Intelligenz mit Erfolg in Schule, Studium, Ausbildung, Beruf und im Privatleben einhergehen. Mit der kognitiven Leistungsfähigkeit steigt beispielsweise das durchschnittliche Einkommen sowie die Position im Beruf. Kognitiv leistungsfähigere Personen lassen sich seltener scheiden und werden seltener delinquent oder arbeitslos. Neben der kognitiven Leistungsfähigkeit sind auch andere psychologische Merkmale dem sozioökonomischen Erfolg zuträglich, wie zum Beispiel Gewissenhaftigkeit und Optimismus (Barrick & Mount, 1991; Kemper, Kovaleva, Beierlein & Rammstedt, 2012).

Neben bedeutsamen Korrelationen mit sozioökonomischen Erfolgsgrößen wurden für viele psychologische Merkmale Assoziationen mit weiteren gesellschaftlich relevanten Verhaltensaspekten gefunden. So verbessern psychologische Merkmale die Vorhersage des Wahlverhaltens (Schumann & Schoen, 2005) und stehen in Zusammenhang mit Gesundheitsverhalten und dessen Folgen. Befunde aus der Fachliteratur zeigen, dass Merkmale wie Gewissenhaftigkeit und Optimismus die physische und die psychische Gesundheit, inklusive der Morbidität und der Mortalität beeinflussen können (Allison, Guichard, Fung & Gilain, 2003; Arthur & Graziano, 1996; Rasmussen, Scheier & Greenhouse, 2009). Schließlich stehen bestimmte psychologische Merkmale in Zusammenhang mit pathologischem oder delinquentem Verhalten wie beispielsweise Drogen- und Alkoholkonsum (Block, Block & Keyes, 1988; Block, Gjerde & Block, 1991; Gottfredson, 1997).

Aufgrund dieser Bedeutung psychologischer Merkmale für viele gesellschaftlich relevante Prozesse und Phänomene forderte kürzlich auch der Ökonomie-Nobelpreisträger James Heckman, dass zukünftige sozialwissenschaftliche Studien vermehrt validierte Persönlichkeitsskalen und Intelligenztests umfassen sollten (Borghans, Duckworth, Heckman & ter Weel, 2008). Dieser Standpunkt wird auch von anderen Forscherinnen und Forschern (Goldberg, 2005; Rammstedt, 2010) und Institutionen (Rat für Sozial- und Wirtschaftsdaten, 2010) vertreten. Seit einigen Jahren sind diesbezüglich deutliche Bestrebungen erkennbar. Im Sozio-ökonomischen Panel (SOEP) wurden 2004 erstmals Items zu Risikoaversion und 2005 zu Persönlichkeit (Big Five), Kontrollüberzeugungen und Reziprozität eingesetzt, 2006 auch Proxy-Variablen für bestimmte kognitive Fähigkeiten. Darüber hinaus wurden Skalen zur Erfassung psychologischer Merkmale in weitere Untersuchungen aufgenommen, zum Beispiel International Social Survey Programme (ISSP), Household, Income and Labour Dynamics in Australia (HILDA), UK House-

¹ Wesentliche Teile dieses Beitrags sind bereits andernorts veröffentlicht als Schipolowski, S., Wilhelm, O., Schroeders, U., Kovaleva, A., Kemper, C. J. & Rammstedt, B. (2013). BEFKI GC-K: Eine Kurzsкала zur Messung kristalliner Intelligenz. *Methoden, Daten, Analysen*, 7(2), 153-181.

hold Longitudinal Study (UKHLS) und DNB Household Survey (DHS). Der Bedarf an Verfahren zur Operationalisierung psychologischer Merkmale ist demnach gegeben und wird in den kommenden Jahren voraussichtlich steigen.

Forscherinnen und Forscher, die entsprechende Merkmale in ihrer Untersuchung erfassen möchten, stehen allerdings derzeit meist vor dem Problem, für ihre Zwecke geeignete Erhebungsinstrumente zu finden. Im Rahmen der psychologischen Forschung wurden zahlreiche Messinstrumente entwickelt (Leibniz-Zentrum für Psychologische Information und Dokumentation, 2011). Diese sind auch in den meisten Fällen sorgfältig im Hinblick auf ihre psychometrische Güte geprüft. Das heißt, durch empirische Untersuchungen wurde ermittelt, inwiefern die Erhebungsinstrumente verlässliche Messungen erlauben und theoretische Erwartungen hinsichtlich der Zusammenhänge mit weiteren Variablen zu treffen. Allerdings sind die Erhebungsinstrumente aus Gründen der Messgenauigkeit häufig umfangreich. Im Gegensatz zur psychologischen (Einzelfall-)Diagnostik, in der eine möglichst breite Messung verschiedener Aspekte eines Merkmales notwendig ist und die stark von der Dauer abhängigen Kosten der Erhebung eine relativ untergeordnete Rolle spielen, ist die Erhebungsdauer in den meisten sozialwissenschaftlichen Untersuchungen ein kritischer Faktor. Hier spielt die Anzahl der Items bzw. die Länge von Erhebungsinstrumenten eine entscheidende Rolle. Da psychologische Merkmale zwar nicht im Zentrum des Interesses stehen, aber aufgrund ihrer vielfältigen Beziehungen zu sozioökonomischen Erfolgsgrößen und anderen bedeutsamen Verhaltensaspekten dennoch mit erhoben werden sollen, werden oft auf der Grundlage etablierter psychologischer Skalen und Tests gekürzte Ad-Hoc-Instrumente erstellt. Der Einsatz solcher Ad-Hoc-Instrumente kann in manchen Fällen einen Mehrwert für eine bestimmte Untersuchung darstellen, bringt allerdings entscheidende Nachteile mit sich. So liegen für die gekürzten Fassungen der Instrumente in der Regel keine oder nur wenig belastbare Validitätsbelege auf Basis vergleichsweise kleiner Stichproben vor; die für das ungekürzte Erhebungsinstrument vorliegenden Befunde sind nicht übertragbar (Stanton, Sinar, Balzer & Smith, 2002). Eine Prüfung der Generalisierbarkeit von Kennwerten auf verschiedene Populationen fehlt häufig ebenso wie Prüfungen der Messinvarianz für unterschiedliche Testmedien. Schließlich sind Befunde zwischen verschiedenen Erhebungen nur eingeschränkt vergleichbar, wenn je nach Studie ausgehend vom Originalinstrument unterschiedliche Kürzungsstrategien verfolgt werden.

Die Bereitstellung von Beratungs- und Serviceleistungen für die Sozialwissenschaften gehört zu den Kernaufgaben von GESIS. Dies beinhaltet auch die Entwicklung und Verbreitung standardisierter Erhebungsinstrumente, die hinsichtlich ihrer psychometrischen Güte umfassend geprüft wurden. Bisher liegen für die Erfassung von für die sozialwissenschaftliche Forschung relevanten psychologischen Merkmalen nur wenige etablierte und disziplinübergreifend akzeptierte Erhebungsinstrumente vor, die eine ökonomische und effiziente Messung erlauben. Dies gilt insbesondere für die Messung kognitiver Fähigkeiten. Ziel des vorliegenden Arbeitsberichts ist es daher, die Entwicklung und Erprobung eines effizienten Erhebungsinstruments zur Messung der kristallinen Intelligenz zu beschreiben. Für eine ausführlichere Darstellung verweisen wir auf Schipolowski, Wilhelm, Schroeders, Kovaleva, Kemper & Rammstedt (2013).

2 Skalenkonzept

2.1 Theoretischer Hintergrund

Den theoretischen Hintergrund der Skala bildet die Definition kristalliner Intelligenz in etablierten Intelligenztheorien, insbesondere der Theorie fluider und kristalliner Intelligenz von Cattell und Horn (Horn, 1965, 1988; Cattell, 1971) sowie der Drei-Stratum-Theorie von Carroll (1993).

Nach Binet und Simon (1905) sowie Hebb (1942) gehörte Cattell (1943) zu den ersten Intelligenzforschern, die in ihren theoretischen Überlegungen neben dekontextualisierten Denkleistungen auch wissensbasierte Fähigkeiten berücksichtigten. Statt eines einzigen, übergeordneten 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 von Cattell und Kollegen starke Ladungen von sprachnahen Aufgaben und Wissenstests auf dem Faktor g_c berichtet (Cattell, 1963; Horn & Cattell, 1966; Cattell, 1971; Horn, 1965). Konzeptuell soll g_c diesen Autoren zufolge die Gesamtheit des Wissens umfassen, das Menschen im Laufe ihres Lebens erwerben und zum Problemlösen einsetzen. Diese Definition kommt auch in neueren Arbeiten zur Anwendung; so beschreiben Horn und Noll (1997, S. 69) kristalline Intelligenz als „akkulturiertes Wissen“, das über Aufgaben gemessen werden sollte, die „Tiefe und Breite des Wissens der dominanten Kultur“ widerspiegeln.

Als weiteres einflussreiches Intelligenzmodell ist Carrolls (1993) Drei-Stratum-Theorie zu nennen, die als wichtiger Bezugspunkt in der Intelligenzforschung gilt. Das Modell basiert auf der Reanalyse von 461 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 den Fokus in Richtung sprachlicher Fähigkeiten wie Leseverstehen und Fremdsprachenkenntnissen (Carroll, 1993, S. 626). Nichtsdestotrotz dokumentiert Carroll auch hohe Zusammenhänge von Wissenstests mit dem g_c -Faktor. Eine besondere Bedeutung kommt dabei dem Stratum-I-Faktor „General Information“ zu, der Unterschiede im Erwerb von Wissen jenseits von Sprachkenntnissen widerspiegelt (vgl. K0; Carroll, 1993, S. 590 und 634) und zu den Faktoren gehört, die in Carrolls Reanalysen häufig die höchste oder zweithöchste Ladung auf dem übergeordneten g_c -Faktor aufwiesen (vgl. auch Carroll, 2003).

2.2 Aufbau

Die Kurzskala BEFKI GC-K besteht aus insgesamt 12 Items zu ebenso vielen Wissensbereichen. Konkret wird Wissen aus den Naturwissenschaften (Biologie, Physik, Geografie, Medizin), den Geisteswissenschaften (Kunst, Literatur, Religion, Philosophie) und den Sozialwissenschaften erfasst (Politik, Geschichte, Finanzen, Recht). Alle Items haben ein Multiple-Choice-Format mit vier Antwortalternativen, von denen genau eine richtig ist. In [Tabelle 1](#) sind für die 12 Items Schwierigkeit und Trennschärfe (genauer: die Itemladung im einfaktoriellen Messmodell, vgl. [Abschnitt 5.2](#)) sowie die inhaltliche Zuordnung zu den drei breiten Wissensdomänen beziehungsweise den einzelnen Wissensbereichen angegeben. Die Kennwerte basieren auf einer repräsentativen Erhebung mit Erwachsenen zwischen 18 und 93 Jahren aus ganz Deutschland (vgl. Stichprobenbeschreibung in [Abschnitt 4](#)).

Tabelle 1: Schwierigkeiten und Trennschärfen der Kurzskala-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 einfaktoriellen Messmodell (vgl. [Abschnitt 5.2](#)).

2.3 Auswertung

Im ersten Schritt werden die Items anhand eines Lösungsschlüssels in richtig beantwortet (1) versus nicht richtig beantwortet (0) rekodiert. Der Lösungsschlüssel ist in der o. g. Druckvorlage enthalten. Items, bei denen die teilnehmende Person keine Antwort angekreuzt hat oder mehrdeutig geantwortet hat (z. B. durch Ankreuzen mehrerer Antwortoptionen) werden als nicht richtig beantwortet gewertet. Anschließend erfolgt die Bildung eines Gesamtwertes g_c durch Aufsummieren über alle 12 rekodierten Items. Eine getrennte Berechnung mehrerer Summenwerte für die drei breiten Wissensdomänen Natur-, Geistes- und Sozialwissenschaften wird nicht empfohlen, da wegen der geringen Itemzahl mit Boden- oder Deckeneffekten zu rechnen ist und zudem auf Basis von nur vier Items pro Wissensdomäne keine substantielle inhaltliche Interpretation der Werte möglich wäre (vgl. die ausführliche Diskussion konkurrierender Messmodelle in Schipolowski et al., 2013).

2.4 Items

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

3 Durchführung

Die Kurzsкала BEFKI GC-K wurde als zeiteffizientes Messinstrument zur Erfassung kristalliner Intelligenz entwickelt und eignet sich für Anwendungsfelder, in denen eine Proxy-Variablen für die kognitive Leistungsfähigkeit benötigt wird und mit geringem Zeitaufwand erhoben werden soll. Für die 12 Items der Kurzsкала ist eine Bearbeitungszeit von insgesamt fünf Minuten vorzusehen. Bei der Entwicklung und Erprobung der Skala wurde als Zielgruppe die Allgemeinbevölkerung ab 18 Jahren definiert, mit der Einschränkung, dass ausreichende Deutschkenntnisse zum Verständnis der Items gegeben sein müssen.

Die hier berichteten Befunde zu den Eigenschaften der Skala beziehen sich auf eine Administration im CASI-Modus (Computer Assisted Self-Interview). Messinvarianzprüfungen mit der Kurzsкала zur Vergleichbarkeit mit anderen Erhebungsmodalitäten stehen noch aus. Aufgrund der einschlägigen Literatur zu Testmedienvergleichen (etwa Mead & Drasgow, 1993; vgl. auch Schroeders & Wilhelm, 2010) ist jedoch davon auszugehen, dass beim Einsatz als Papierform (Selbstaussfüller), im CAPI-Modus (Computer Assisted Personal Interview) oder als Computer Assisted Web Interview (CAWI) bei ansonsten gleichen Durchführungsbedingungen keine nennenswerten Testmedieneffekte auftreten. Bei Online-Testungen sollte die Nutzung von Hilfsmitteln wie Internet-Suchmaschinen ausgeschlossen werden (Schroeders, Wilhelm & Schipolowski, 2010).

4 Entwicklung und Validierung

Die theoretischen Vorarbeiten Cattells und Horns sowie die in Carroll (1993, 2003) berichteten Befunde legen nahe, dass eine Operationalisierung kristalliner Intelligenz im Erwachsenenalter Wissen aus möglichst vielen unterschiedlichen Bereichen berücksichtigen sollte. Die Entwicklung der Wissensitems erfolgte im Rahmen des BEFKI-Projekts (Berliner Test zur Erfassung Fluiden und Kristalliner Intelligenz; Wilhelm, Schroeders & Schipolowski, in Vorbereitung; Wilhelm & Schipolowski, 2010), in dessen Verlauf ein umfangreicher Itempool aufgebaut wurde, der die Messung deklarativen Wissens in 16 verschiedenen Domänen ermöglicht. 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). Dabei orientierte sich die Auswahl der Wissensbereiche an der empirisch begründeten Klassifikation von Ackerman (2000; Rolfhus & Ackerman, 1999). Zur Entwicklung der Kurzskaala wurden zuerst aus diesem Itempool insgesamt 32 Wissensitems anhand inhaltlicher und psychometrischer Kriterien ausgewählt. Konkret wurden zwei Items aus jedem der 16 Wissensbereiche gewählt, wobei eines der beiden Items 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).

Die 32 ausgewählten Wissensitems wurden im nächsten Schritt einer für die erwachsene Wohnbevölkerung der Bundesrepublik Deutschland repräsentativen Stichprobe zur Bearbeitung vorgelegt. Diese umfasst auch Personen mit Zuwanderungshintergrund, sofern sie die deutschsprachigen Fragen und Aufgaben verstehen und auf Deutsch beantworten konnten. Die Ziehung erfolgte mithilfe des ADM-Stichprobensystems F2F und resultierte in einer Stichprobe von 1206 Personen, 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 1134 Fällen nach Ausschluss unbrauchbarer Datenpunkte sowie von Personen ohne deutsche Staatsbürgerschaft, um für die Gewichtung die Grundgesamtheit eindeutig zu definieren. Die gewichtete Stichprobe umfasst somit 1134 Erwachsene (52,2% weiblich) im Alter von 18 bis 93 Jahren ($M = 52$ Jahre, $SD = 18$ Jahre) aus dem gesamten Bundesgebiet.

Neben den Wissensitems wurden Indikatoren weiterer psychologischer Konstrukte vorgegeben, darunter der BFI-10 (Rammstedt & John, 2007) zur Erfassung der Big Five. 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 Ratingskala an, wie vertraut sie mit dem jeweiligen Begriff oder Konzept sind. Des Weiteren wurden verschiedene soziodemographische Merkmale der Teilnehmerinnen und Teilnehmer erfasst, darunter Alter, Geschlecht, Schulabschluss, beruflicher Ausbildungsabschluss, berufliche Stellung und Haushaltsnettoeinkommen. Ergänzend wurde die Anzahl der Bücher im Elternhaus erfragt sowie die berufliche Stellung der Eltern, als der/die Teilnehmende 15 Jahre alt war.

Die Erhebung der Daten erfolgte durch geschulte Interviewerinnen und Interviewer, welche die Studienteilnehmerinnen und -teilnehmer zu vorab vereinbarten Terminen in ihren Wohnungen aufsuchten. Nach Abfrage der soziodemographischen Angaben und der Persönlichkeitsitems durch den Interviewer bearbeiteten die Testpersonen den g_c -Test selbstständig am Notebook (CASI, Computer Assisted

Self-Interview). Dabei 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.

Im Zuge der Aufbereitung der gewichteten Stichprobendaten wurden im ersten Schritt anhand der vorliegenden demographischen Angaben verschiedene Indizes gebildet, darunter die ISCED-97 (International Standard Classification of Education; UNESCO, 1997) als Index zur formalen Bildung und der ISEI (International Socio-Economic Index of Occupational Status; Ganzeboom, De Graaf & Treiman, 1992) als Index des sozioökonomischen Status. Die Bildung des ISEI wurde für die Teilnehmenden selbst sowie für deren Eltern durchgeführt; in letzterem Fall wurde für die weiteren Analysen der höchste der beiden elterlichen ISEI-Werte (HISEI) verwendet. Das Haushaltsnettoeinkommen wurde auch direkt erfragt und liegt im aufbereiteten Datensatz als Variable mit 24 Einkommenskategorien vor. Für die Skalen zu den psychologischen Konstrukten wurden Summen- oder Mittelwerte so gebildet, wie von den Autoren der jeweiligen Instrumente vorgeschlagen. 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 Items zur Messung der kristallinen Intelligenz wurden zunächst in richtig beantwortet versus nicht richtig beantwortet rekodiert, anschließend wurde ein Summenwert über alle 32 Items berechnet.

Um Einschränkungen bei der Teststärke sowie Verzerrungen durch nicht zufällig fehlende Informationen 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 2](#), Spalte N_i). Speziell bei den Wissensitems, für die Datenpunkte fast ausschließlich aufgrund der Zeitbegrenzung – also nicht zufällig – fehlten, ermöglichte die Imputation eine Minimierung konstruktirrelevanter Varianz (etwa interindividuelle Unterschiede in mentaler Geschwindigkeit; Danthiir, Roberts, Schulze & Wilhelm, 2004). Konkret wurden fehlende Daten mittels multipler Imputation mit 100 Replikationen geschätzt. Das bedeutet, dass die berichteten Statistiken und Koeffizienten Mittelwerte über alle Replikationen sind. Bei der Ermittlung von Standardfehlern wurde die Streuung zwischen den Replikationen berücksichtigt.

Auf Basis der oben beschriebenen Datenerhebung mit 32 Items wurde im nächsten Schritt eine Itemselektion für die g_c-Kurzsкала nach folgenden Kriterien vorgenommen:

- Um einen flexiblen Einsatz der Kurzsкала in der Umfrageforschung zu ermöglichen, sollte deren Bearbeitungszeit bei 5 Minuten liegen. Dies entspricht der geschätzten Bearbeitungszeit von 12 Items.
- 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 Gesamtsкала auch in der Kurzsкала 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.
- Ein einfaktorielles Messmodell sollte gute Modellpassung und Itemladungen (Trennschärfen) von .50 oder höher ausweisen. In keinem Fall sollten Ladungen < .30 auftreten.
- Die Kurzsкала sollte ähnliche Beziehungen zu Personen- und Umweltmerkmalen sowie anderen psychologischen Konstrukten aufweisen wie die Gesamtsкала.

Nach erfolgter Itemselektion wurde für die Kurzsкала mit 12 Items ebenfalls ein Summenwert berechnet. Werden die entsprechenden Summenwerte herangezogen, korreliert die Kurzsкала mit 12 Items mit der 32 Items umfassenden Gesamtsкала zu $r = .91$.

5 Gütekriterien

5.1 Reliabilität

Zur Bestimmung der Reliabilität wurde neben dem bekannteren, aber weniger geeigneten Cronbach α auch der Koeffizient ω von McDonald (1999, S. 90) berechnet, dessen Höhe analog zu α interpretiert werden kann. Für ein einfaktorielles Messmodell ergab sich eine zufriedenstellende Reliabilität der latenten Variable von $\omega = .82$ bzw. $\alpha = .81$ (Zumbo, Gadermann & Zeisser, 2007). Die Reliabilität des manifesten Summenscores lag bei $\rho = .70$ (Skalenreliabilität nach Raykov, Dimitrov & Asparouhov, 2010).

5.2 Validität

Faktorielle Validität

Ein einfaktorielles Messmodell der Kurzskala, in dem kristalline Intelligenz durch einen einzigen Faktor g_c repräsentiert wird, auf dem alle 12 Items laden, zeigte gemessen an Standards aus der Literatur für Modelle mit binären Indikatoren (Yu, 2002) eine gute Passung: $\Delta\chi^2(N = 1134, 54) = 110.8$, Comparative fit index (CFI) = .965, Root mean square error of approximation (RMSEA) = .030, Weighted root mean square residual (WRMR) = 1.06. Die Ladungen der Items auf dem g_c -Faktor können [Tabelle 1](#) entnommen werden.

Konstruktvalidität

Ein wesentlicher Aspekt bei der Entwicklung und Beurteilung von Kurzskalen besteht in deren Beziehungen zu relevanten Personen- und Umweltmerkmalen sowie anderen psychologischen Konstrukten. Diese sollen einerseits im Einklang mit theoretischen Erwartungen stehen (Konstruktvalidierung). Andererseits sollen die für die Kurzskala ermittelten Beziehungen möglichst den Befunden für die ungekürzte Gesamtskala entsprechen, um sicherzustellen, dass die Itemselektion keine substantielle Minderung oder Veränderung 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 2](#)). Zu den betrachteten Personenvariablen zählten Geschlecht, Alter, formale Bildung und sozioökonomischer Status der Teilnehmenden. Darüber hinaus wurden 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 wurden 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 theoretischen Erwartungen. Aufgrund der hohen Stabilität kristalliner Intelligenz im Erwachsenenalter (Baltes, Staudinger & Lindenberger, 1999) war in der hier untersuchten Erwachsenenpopulation von sehr geringen Alterseffekten auszugehen. Es wurde jedoch ein bedeutsamer Geschlechtsunterschied erwartet; in vergleichbaren Studien mit umfangreichen Wissenstestbatterien wird ein Wissensvorsprung zugunsten der Männer von etwa .50 bis .65 Standardabweichungseinheiten berichtet (Ackerman, Bowen, Beier & Kanfer, 2001; Lynn, Irwing & Cammock, 2001; Lynn & Irwing, 2002). Vor

dem Hintergrund der Definition kristalliner Intelligenz als „akkulturiertes Wissen“ war zudem von hohen positiven Korrelationen zwischen kristalliner Intelligenz und der Qualität und Quantität formaler Bildung auszugehen (Cliffordson & Gustafsson, 2008) sowie von substantiellen Assoziationen mit Indikatoren, die für den Wissenserwerb bedeutsame Ressourcen erfassen, wie beispielsweise Maße des sozioökonomischen Status (Rowe, Jacobson & van den Oord, 1999). Mit Blick auf weitere untersuchte Kovariate wurde eine bedeutsame positive Korrelation von g_c mit selbsteingeschätztem Wissen erwartet. Da Selbsteinschätzungen des Wissens jedoch neben der tatsächlichen Fähigkeitsausprägung weitere Varianzquellen widerspiegeln (z.B. faking-good), ist davon auszugehen, dass diese Korrelation hoch, aber nicht perfekt sein würde (Hülür, Wilhelm & Schipolowski, 2011). Mit Blick auf die Big Five werden in der Literatur positive Korrelationen zwischen g_c bzw. Wissensleistungen und dem Faktor Offenheit berichtet (Ackerman & Heggestad, 1997; Ziegler, Danay, Heene, Asendorpf & Bühner, 2012). Ackerman und Heggestad (1997) berichten in ihrer Metaanalyse der Beziehungen zwischen Fähigkeits- und Persönlichkeitskonstrukten zudem einen gering positiven Zusammenhang zwischen g_c und Extraversion sowie einen gering negativen Zusammenhang mit Neurotizismus. Die Korrelationen von g_c mit Gewissenhaftigkeit und Verträglichkeit waren nicht signifikant von null verschieden.

Tabelle 2: Korrelationen der Kurz- und Gesamtskala mit verschiedenen Personen- und Haushaltsmerkmalen sowie psychologischen Konstrukten

Variable	N_{vi}	Kurzskala		Gesamtskala	
		r	SE	r	SE
Geschlecht ¹	1134	-.15	.03	-.15	.03
Alter	1134	.01 ^{n.s.}	.03	.00 ^{n.s.}	.03
ISCED-97	1091	.49	.03	.51	.03
ISEI	388	.44	.04	.45	.04
HISEI Eltern	1082	.25	.03	.25	.03
Einkommen	638	.29	.04	.30	.04
Anzahl Bücher	1101	.30	.03	.33	.03
selbstberichtetes Wissen ²	1134	.52	.03	.55	.03
Neurotizismus ³	1104	-.10	.03	-.15	.03
Extraversion ³	1104	.07 ^a	.04	.12	.04
Offenheit ³	1104	.21	.03	.25	.03
Gewissenhaftigkeit ³	1104	.07 ^a	.03	.09	.03
Verträglichkeit ³	1104	-.02 ^{n.s.}	.03	-.02 ^{n.s.}	.03

Anmerkungen: $N = 1134$. ¹ 0 = männlich, 1 = weiblich; ² 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 ($p < .01$).

Männer zeigten etwas höhere Wissensleistungen als Frauen: Die standardisierte Mittelwertdifferenz betrug $d = .30$ und ist somit inhaltlich bedeutsam, obgleich niedriger als in den oben zitierten Arbeiten. Wie erwartet wurde in der hier untersuchten Erwachsenenpopulation kein bedeutsamer Alterseffekt beobachtet. Deklaratives Wissen wies 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 zeigte sich zu dem auf der ISEI-Skala quantifizierten sozioökonomischen Status der Teilnehmenden. Auch für die anderen Personen- und Haushaltsmerkmale lagen inhaltlich bedeutsame positive Korrelationen mit Wissen vor (siehe [Tabelle 2](#)).

Mit selbstberichtetem Wissen ergab sich erwartungsgemäß eine vergleichsweise hohe Korrelation. Für die fünf Persönlichkeitsdimensionen zeigte sich ein differenziertes Ergebnismuster, das im Wesentlichen mit den in der Literatur berichteten Befunden übereinstimmt. Die Offenheitsdimension wies 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 ließ sich lediglich für die Skala Gewissenhaftigkeit feststellen: Während Ackerman und Heggestad (1997) hier eine nicht signifikante Korrelation mit kristalliner Intelligenz berichten, wurde in der aktuellen Analyse eine gering positive Beziehung zwischen Wissen und Gewissenhaftigkeit gefunden. Für Kurz- und Gesamtskala fielen die Korrelationen zu den Kovariaten ähnlich aus.

5.3 Referenzwerte

Kennwerte der Personenverteilungen für die hier untersuchte Gesamtpopulation (vgl. die Stichprobenbeschreibung in [Abschnitt 4](#)) und verschiedene Subpopulationen sind in [Tabelle 3](#) aufgeführt. 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%. Die Subpopulation mit „mittlerer Schulbildung“ umfasst Personen mit Mittlerer Reife (bzw. Äquivalent), die im Regelfall nach 10 Jahren Beschulung erworben wurde. Ihr Anteil an der Gesamtpopulation beträgt 29%. Personen mit „hoher Schulbildung“ im Sinne der hier vorgenommenen Kategorisierung verfügen über eine fachgebundene oder allgemeine Hochschulreife bzw. Fachhochschulreife (26% der Gesamtpopulation), die typischerweise nach 12 oder 13 Jahren Schulbesuch erworben wurde. Sowohl für die Gesamtpopulation als auch für alle drei Bildungsgruppen werden zusätzlich Kennwerte für Männer und Frauen berichtet. Auf eine Betrachtung verschiedener Altersgruppen wird hingegen verzichtet, da in der untersuchten Erwachsenenpopulation für die Kurzskala keine bedeutsamen Alterseffekte gefunden wurden (vgl. [Abschnitt Validitätsbelege, 5.2](#))

Tabelle 3: Kennwerte verschiedener Personenverteilungen für den Summenwert der Kurzska

Population	<i>N</i>	<i>M</i>	<i>SD</i>	Schiefe	Exzess
Gesamtpopulation	1134	7.04	2.66	-0.20	-0.64
männlich	542	7.44	2.61	-0.29	-0.47
weiblich	592	6.66	2.64	-0.13	-0.76
„Geringe Schulbildung“	514	5.95	2.49	+0.06	-0.17
männlich	240	6.37	2.48	+0.08	-0.34
weiblich	274	5.60	2.45	+0.04	-0.12
„Mittlere Schulbildung“	330	7.30	2.43	-0.32	-0.77
männlich	143	7.71	2.37	-0.53	-0.33
weiblich	187	6.99	2.44	-0.21	-0.98
„Hohe Schulbildung“	290	8.66	2.27	-0.52	-0.38
männlich	159	8.84	2.29	-0.73	+0.37
weiblich	131	8.44	2.23	-0.33	-1.04

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

6 Kurzfassung

Diagnostische Zielsetzung

Die Kurzskaala BEFKI GC-K wurde entwickelt um eine ökonomische Messung der kristallinen Intelligenz in sozialwissenschaftlichen Untersuchungen zu ermöglichen.

Aufbau

Die BEFKI GC-K erfasst Wissen aus 12 Wissensbereichen mit jeweils einem Item. Konkret wird Wissen aus den Naturwissenschaften, den Geisteswissenschaften und den Sozialwissenschaften erfasst. Alle Items haben ein Multiple-Choice-Format mit vier Antwortalternativen, von denen genau eine richtig ist.

Grundlagen und Konstruktion

Ziel der Skalenkonstruktion war die Messung der kristallinen Intelligenz (g_c). Den theoretischen Hintergrund der Skala bildet die Definition kristalliner Intelligenz in etablierten Intelligenzstrukturtheorien, insbesondere der Theorie fluider und kristalliner Intelligenz von Cattell und Horn (Horn, 1965, 1988; Cattell, 1971) sowie der Drei-Stratum-Theorie von Carroll (1993). Demnach kann g_c als „akkulturiertes Wissen“ beschrieben und über Aufgaben gemessen werden, die „Tiefe und Breite des Wissens der dominanten Kultur“ erfassen (Horn & Noll, 1997, S. 69). Zur Entwicklung der Kurzskaala BEFKI GC-K wurden aus dem umfangreichen Itempool des BEFKI-Projekts (Berliner Test zur Erfassung Fluider und Kristalliner Intelligenz; Wilhelm, Schroeders & Schipolowski, im Druck; Wilhelm & Schipolowski, 2010) zuerst 32 und schließlich 12 Items zum deklarativen Faktenwissen anhand inhaltlicher und psychometrischer Kriterien ausgewählt. Anschließend wurden die Reliabilität einschließlich konkurrierender Messmodelle und verschiedene Aspekte der Konstruktvalidität überprüft.

Gütekriterien

Reliabilität Zur Bestimmung der Reliabilität wurde neben dem bekannteren, aber weniger geeigneten Cronbach α auch der Koeffizient ω von McDonald (1999, S. 90) berechnet, dessen Höhe analog zu α interpretiert werden kann. Für ein einfaktorielles Modellmodell ergab sich eine zufriedenstellende Reliabilität der latenten Variable von $\omega = .82$ bzw. $\alpha = .81$ (Zumbo, Gadermann & Zeisser, 2007). Die Reliabilität des manifesten Summenscores lag bei $\rho = .70$ (Skalenreliabilität nach Raykov, Dimitrov & Asparouhov, 2010).

Validität Ein einfaktorielles Messmodell der Kurzskaala, in dem kristalline Intelligenz durch einen einzigen Faktor g_c repräsentiert wird, auf dem alle 12 Items laden, zeigte eine gute Passung. Zur Bestimmung der Konstruktvalidität wurden die Korrelationen der Kurzskaala mit verschiedenen Kovariaten näher betrachtet. Eine starke Beziehung zeigte sich zu dem auf der ISEI-Skala quantifizierten sozio-ökonomischen Status der Teilnehmenden. Mit selbstberichtetem Wissen ergab sich erwartungsgemäß ebenfalls eine substanzielle hohe Korrelation. Für die fünf Persönlichkeitsdimensionen (Big5) zeigte sich ein differenziertes Ergebnismuster, das im Wesentlichen mit den in der Literatur berichteten Befunden übereinstimmt. Eine Abweichung ließ sich lediglich für die Skala Gewissenhaftigkeit feststellen, da in der aktuellen Analyse eine geringe positive Beziehung zwischen Wissen und Gewissenhaftigkeit gefunden wurde. Die durchgeführten Analysen zeigen, dass die BEFKI GC-K eine valide ökonomische Erfassung der kristallinen Intelligenz erlaubt.

7 Bewertung

Die hier beschriebene Kurzskala BEFKI GC-K ist ein frei verfügbares Verfahren zur Messung kristalliner Intelligenz, das sich aufgrund seiner Kürze und psychometrischen Effizienz insbesondere für den Einsatz in der Umfrageforschung eignet. Die Kurzskala umfasst 12 Items, die in Übereinstimmung mit der Definition von g_c durch Cattell (1971) bzw. Carroll (1993) deklaratives Wissen aus ebenso vielen Bereichen der Natur-, Geistes- und Sozialwissenschaften erfassen. Somit wird innerhalb von fünf Minuten Bearbeitungszeit ein möglichst breites Wissensspektrum berücksichtigt. Des Weiteren weist die Kurzskala eine gute Reliabilität auf. Ihre Validität wurde anhand der Korrelationen des Summenwertes mit verschiedenen Personen- und Umweltmerkmalen einerseits sowie mit ausgewählten psychologischen Konstrukten andererseits überprüft. Die gefundenen Beziehungen stimmen nahezu vollständig mit den auf Basis der Fachliteratur formulierten Erwartungen überein. Die Höhe der untersuchten Effekte bzw. Zusammenhänge lag zudem für die Kurzskala mit 12 Items nur minimal unter den entsprechenden Werten für die ungekürzte Wissensskala mit 32 Items.

Bei der Interpretation der Befunde ist einschränkend zu berücksichtigen, dass die Itemselektion für die Kurzskala und deren Evaluation auf derselben Erhebung basieren. Zudem führte die relativ strenge Zeitbegrenzung in der hier beschriebenen Erhebung zu fehlenden Werten, die in der Folge imputiert wurden, um die größtmögliche Vergleichbarkeit mit einer Durchführung ohne Zeitmangel zu gewährleisten. Für die Kurzskala gilt jedoch, dass eine Bearbeitungszeit von fünf Minuten ausreichend ist, so dass eine Imputation fehlender Werte unter diesen Durchführungsbedingungen nicht erforderlich ist (stattdessen sollten ausgelassene Items wie nicht richtige Antworten ausgewertet werden).

Gegenüber gängigen Bildungsindikatoren wie der ISCED-97 hat die Wissensskala den Vorteil, dass tatsächliches Wissen direkt erhoben wird, anstatt dieses aus Abschlüssen zu erschließen. Letzteres ist insbesondere deshalb problematisch, da Schul- und Ausbildungsabschlüsse zwischen verschiedenen Bundesländern oder gar Staaten dem zeitlichen Wandel unterliegen und schwer vergleichbar sind (Neumann, Nagy, Trautwein & Lüdtke, 2009). 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 Kurzskala 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 Kurzskala auf „Wissen, von dem angenommen werden kann, dass es von einem großen Teil der Population geteilt wird“ (Ackerman, 2003, S. 16). Mithilfe komplexer Testdesigns und Modellierungsmethoden ist es jedoch möglich, die Vorzüge von Kurzskalen mit den Vorteilen umfangreicher Messinstrumente zu kombinieren (vgl. hierzu die Vorschläge von Rhemtulla & Little, 2012, sowie Graham, Taylor, Olchowski & Cumsille, 2006, zu planned missing data designs). In zukünftigen Studien sollte untersucht werden, inwieweit die g_c -Kurzskala gegenüber anderen in der Umfrageforschung häufig eingesetzten Instrumenten und Indikatoren inkrementell valide bei der Vorhersage bedeutsamer Outcomes ist.

8 Literatur

- Ackerman, P. L. (2000). Domain-specific knowledge as the „dark matter“ of adult intelligence: G_f/G_c , 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. & Heggestad, E. D. (1997). Intelligence, personality, and interests: Evidence for overlapping traits. *Psychological Bulletin*, *121*, 219-245.
- Ackerman, P. L., Bowen, K. R., Beier, M. B., & Kanfer, R. (2001). Determinants of individual differences and gender differences in knowledge. *Journal of Educational Psychology*, *93*, 797-825.
- Allison, P. J., Guichard, C., Fung, K., & Gilain, L. (2003). Dispositional optimism predicts survival status 1 year after diagnosis in head and neck cancer patients. *Journal of Clinical Oncology*, *21*, 543-548.
- Arthur, W. Jr. & Graziano, W. G. (1996). The five-factor model, conscientiousness, and driving accident involvement. *Journal of Personality*, *64*, 593-618.
- Baltes, P. B., Staudinger, U. M., & Lindenberger, U. (1999). Lifespan Psychology: Theory and Application to Intellectual Functioning. *Annual Review of Psychology*, *50*, 471-507.
- Barrick, M. R. & Mount, M. K. (1991). The big five personality dimensions and job performance: a metaanalysis. *Personnel Psychology*, *44*, 1-26.
- Binet, A. & Simon, T. (1905). Méthodes nouvelles pour le diagnostic du niveau intellectuel des anormaux. *L'année psychologique*, *11*, 191-244.
- Block, J., Block, J. H., & Keyes, S. (1988). Longitudinally foretelling drug usage in adolescence: Early childhood personality and environmental precursors. *Child Development*, *59*, 336-355.
- Block, J., Gjerde, P. F., & Block, J. H. (1991). Personality antecedents of depressive tendencies in 18-year-olds: A prospective study. *Journal of Personality and Social Psychology*, *60*, 726-738.
- Borghans, L., Duckworth, A. L., Heckman, J. J., & ter Weel, B. (2008). *The Economics and Psychology of Personality Traits*. IZA Discussion Paper No. 3333. Bonn: IZA.
- 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. In H. Nyborg (Hg.), *The scientific study of general intelligence: Tribute to Arthur R. Jensen* (S. 153-193). 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.

- Cliffordson, C. & Gustafsson, J.-E. (2008). Effects of age and schooling on intellectual performance: Estimates obtained from analysis of continuous variation in age and length of schooling. *Intelligence*, 36, 143-152.
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*, 2nd ed. Hillsdale: Lawrence Erlbaum Associates.
- Danthiir, V., Roberts, R. D., Schulze, R., & Wilhelm, O. (2004). Mental Speed: On Frameworks, Paradigms, and a Platform for the Future. In O. Wilhelm & R. W. Engle (Eds.), *Handbook of understanding and measuring intelligence* (pp. 27-46). London: Sage.
- Ganzeboom, H. B. G., De Graaf, P. M., & Treiman, D. J. (1992). A Standard International Socio-Economic Index of Occupational Status. *Social Science Research*, 21, 1-56.
- Goldberg, L. R. (2005). *Why Personality Measures Should Be Included in Epidemiological Surveys: A brief Commentary and a Reading List*. (Oregon Research Institute, Ed.).
- Gottfredson, L. S. (1997). Why g matters: The complexity of everyday life. *Intelligence*, 24, 79-132.
- Gottfredson, L. S. & Deary, I. J. (2004). Intelligence predicts health and longevity, but why? *Current Directions in Psychological Science*, 13, 1-4.
- Graham, J. W., Taylor, B. J., Olchowski, A. E., & Cumsille, P. E. (2006). Planned missing data designs in psychological research. *Psychological Methods*, 11, 323-343.
- 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. In J. R. Nesselroade (Ed.), *Handbook of multivariate psychology* (pp. 645-685). New York: Academic Press.
- Horn, J. L., & Cattell, R. B. (1966). Refinement and test of the theory of fluid and crystallized general intelligences. *Journal of Educational Psychology*, 57, 253-270.
- 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.
- Hülür, G., Wilhelm, O., & Schipolowski, S. (2011). Prediction of self-reported knowledge with overclaiming, fluid and crystallized intelligence and typical intellectual engagement. *Learning and Individual Differences*, 21, 742-746.
- Kemper, C. J., Kovaleva, A., Beierlein, C. & Rammstedt, B. (2012). *Entwicklung und Validierung einer ultrakurzen Operationalisierung des Konstrukts Optimismus-Pessimismus*. Manuskript eingereicht zur Publikation.
- Leibniz-Zentrum für Psychologische Information und Dokumentation (ZPID) (2011). *Verzeichnis Testverfahren. Kurznamen. Langnamen. Autoren. Testrezensionen (18., aktualisierte Auflage)*. Trier: ZPID. Online verfügbar unter: http://www.zpid.de/pub/tests/verz_teil1.pdf

- Lüdtke, O., Robitzsch, A., Trautwein, U., & Köller, O. (2007). Umgang mit fehlenden Werten in der psychologischen Forschung: Probleme und Lösungen. *Psychologische Rundschau*, *58*, 103-117.
- Lynn, R., & Irwing, P. (2002). Sex differences in general knowledge, semantic knowledge and reasoning ability. *British Journal of Psychology*, *93*, 545-556.
- Lynn, R., Irwing, P., & Cammock, T. (2001). Sex differences in general knowledge. *Intelligence*, *30*, 27-40.
- McDonald, R. P. (1999). *Test theory: A unified treatment*. Mahwah, NJ: Erlbaum.
- Mead, A. D. & Drasgow, F. (1993). Equivalence of computerized and paper-and-pencil cognitive ability tests: A meta-analysis. *Psychological Bulletin*, *114*, 449-458.
- Neumann, M., Nagy, G., Trautwein, U. & Lüdtke, O. (2009). Vergleichbarkeit von Abiturleistungen: Leistungs- und Bewertungsunterschiede zwischen Hamburger und Baden-Württemberger Abiturienten und die Rolle zentraler Abiturprüfungen. *Zeitschrift für Erziehungswissenschaft*, *12*, 691-714.
- Rammstedt, B. (2010). Subjective indicators. In Rat für Sozial- und Wirtschaftsdaten (Hrsg.), *Building on progress: Expanding the research infrastructure for the social, economic, and behavioral sciences* (pp. 813-824). Opladen: Budrich UniPress.
- Rammstedt, B. & John, O. P. (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.
- Rasmussen, H. N., Scheier, M. F., & Greenhouse, J. B. (2009). Optimism and physical health: A metaanalytic review. *Annals of Behavioral Medicine*, *37*, 239-256.
- Rat für Sozial- und Wirtschaftsdaten (2010). *Building on progress: Expanding the research infrastructure for the social, economic, and behavioral sciences*. Opladen: Budrich UniPress.
- Raykov, T., Dimitrov, D. M., & Asparouhov, T. (2010). Evaluation of scale reliability with binary measures using latent variable modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, *17*, 265-279.
- Rhemtulla, M. & Little, T. D. (2012). Planned Missing Data Designs for Research in Cognitive Development. *Journal of Cognition and Development*, *13*, 425-438.
- Rolfhus, E. L. & Ackerman, P. L. (1999). Assessing individual differences in knowledge: Knowledge structures and traits. *Journal of Educational Psychology*, *91*, 511-526.
- Rowe, D. C., Jacobson, K. C., & van den Oord, E. J. (1999). Genetic and environmental influences on vocabulary IQ: parental education level as moderator. *Child Development*, *70*, 1151-1162.
- Schipolowski, S., Schroeders, U., & Wilhelm, O. (2008). *BEFKI - Berlin Test of Fluid and Crystallized Intelligence*. Präsentation auf dem XXIX International Congress of Psychology, Berlin.
- Schipolowski, S., Wilhelm, O., Schroeders, U., Kovaleva, A., Kemper, C. J. & Rammstedt, B. (2013). BEFKI GC-K: Eine Kurzsкала zur Messung kristalliner Intelligenz. *Methoden, Daten, Analysen*, *7*(2), 153-181.

- Schroeders, U., Schipolowski, S., Nelles, C., & Wilhelm, O. (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., Schipolowski, S. & Wilhelm, O. (2010). *Berliner Test zur Erfassung fluider und kristalliner Intelligenz*. Präsentation auf dem 47. Kongress der Deutschen Gesellschaft für Psychologie, Bremen.
- Schroeders, U. & Wilhelm, O. (2010). Testing reasoning ability with handheld computers, notebooks, and paper and pencil. *European Journal of Psychological Assessment*, 26, 284–292.
- Schroeders, U., Wilhelm, O., & Schipolowski, S. (2010). Internet-based ability testing: Problems and opportunities. In S. D. Gosling & J. A. Johnson (Eds.), *Advanced methods for behavioral research on the Internet* (pp. 131–148). Washington, DC: American Psychological Association.
- 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.
- Schumann, S. & Schoen, H. (2005). *Persönlichkeit: Eine vergessene Größe der empirischen Sozialforschung*. Wiesbaden: VS Verlag.
- Stanton, J. M., Sinar, E. F., Balzer, W. K., & Smith, P. C. (2002). Issues and strategies for reducing the length of self-report scales. *Personnel Psychology*, 55, 167–194.
- Strenze, T. (2007). Intelligence and socioeconomic success: A meta-analytic review of longitudinal research. *Intelligence*, 35, 401–426.
- UNESCO (1997). *International Standard Classification of Education ISCED 1997*. Available online: http://www.unesco.org/education/information/nfsunesco/doc/isced_1997.htm
- Widaman, K. F., Little, T. D., Preacher, K. J., & Sawalani, G. M. (2011). On creating and using short forms of scales in secondary research. In K. H. Trzesniewski, M. B. Donnellan, & R. E. Lucas (Eds.), *Secondary Data Analysis: An Introduction for Psychologists* (pp. 39–61). Washington, DC: American Psychological Association.
- Wilhelm, O. & Schipolowski, S. (2010). Intelligenzdiagnostik in der Pädagogischen Psychologie. In G. L. Huber (Hg.), *Enzyklopädie Erziehungswissenschaft Online. Fachgebiet Pädagogische Psychologie*. Weinheim/München: Juventa.
- Wilhelm, O., Schroeders, U. & Schipolowski, S. (im Druck). *Berliner Test zur Erfassung fluider und kristalliner Intelligenz für die 8. bis 10. Jahrgangsstufe (BEFKI 8-10)*. Göttingen: Hogrefe.
- 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., Danay, E., Heene, M., Asendorpf, J., & Bühner, M. (2012). Openness, fluid intelligence, and crystallized intelligence: Toward an integrative model. *Journal of Research in Personality*, 46, 173–183.

- Ziegler, M., Kemper, C. J., & Rammstedt, B. (2013). The Vocabulary and Overclaiming Test (VOC-T). *Journal of Individual Differences, 34*, 32-40.
- Ziegler, M., MacCann, C., & Roberts, R. D. (2011). *New Perspectives on Faking in Personality Assessment*. New York: Oxford University Press.
- Zumbo, B. D., Gadermann, A. M., & Zeisser, C. (2007). Ordinal versions of coefficients alpha and theta for Likert rating scales. *Journal of Modern Applied Statistical Methods, 6*, 21-29.