

NEUERE TESTVERFAHREN

Kaufman, A. S., Kaufman, N. L. (2015). **KABC-II. Kaufman Assessment Battery for Children – II.** Deutschsprachige Fassung von P. Melchers und M. Melchers. Frankfurt: Pearson. Test komplett 1.389,- €, Auswertungsprogramm 273,70, Verbrauchsmaterial pro Durchführung 3,64 €.

Theoretische Grundlagen und Testaufbau

Vor knapp 25 Jahren erschien die deutsche Ausgabe der Kaufman-Assessment Battery for Children (K-ABC; Melchers u. Preuss, 1991), deren neuropsychologische Fundierung den Anspruch erhob, neue Wege in der kulturfairen Intelligenzdiagnostik von Kindern zu eröffnen. In den letzten Jahren geriet die K-ABC gegenüber zwischenzeitlich erschienenen Neuauflagen anderer Verfahren ins Hintertreffen. Jetzt stellt sich für viele Testanwender wiederum die Frage, ob die KABC-II wiederum innovative Akzente in der Kinderdiagnostik setzen kann, wobei das amerikanische Original der K-ABC-II bereits seit elf Jahren in Gebrauch ist (Kaufman u. Kaufman, 2004).

Die KABC-II besteht nunmehr aus insgesamt 18 Untertests. Bis auf die Dreijährigen, bei denen lediglich ein Gesamtwert ermittelt wird, lassen sich aus je zwei bis vier Untertests Skalenwerte bilden, die nun wahlweise auf der Grundlage der neuropsychologischen Konzeption Lurias oder der CHC-Theorie nach Cattell, Horn und Carroll (Schneider u. McGrew, 2012) interpretiert werden können.

Ursprünglich basiert die K-ABC auf Lurias neuropsychologischen Überlegungen zur Einteilung und Rolle verschiedener Hirnfunktionen, die bei der Bewältigung komplexer intellektueller Leistungen zusammenwirken (Luria, 1970). Grundlegend bei der Testkonstruktion war die Unterscheidung zwischen sequentieller und simultaner Verarbeitung, wobei es den Testautoren nicht darauf ankam, einzelne Intelligenzfaktoren möglichst unabhängig voneinander zu erfassen. KABC-II erweitert nun den Bezug zu Lurias Überlegungen um Aufgaben zur Lernfähigkeit und, ab 7;0 Jahre, auch zur Planungsfähigkeit. Aus den nunmehr vier Skalen, die verschiedenen Funktionen des neuropsychologischen Modells Lurias zugeordnet werden, ergibt sich ein Gesamtmaß, das als Intellektueller Verarbeitungsindex (IVI) bezeichnet wird. Separat davon kann auch ein Kennwert für Wissen, bestehend aus zwei Untertests, ermittelt werden, der die Fertigkeitenskala der K-ABC ablöst, allerdings keine Aufgaben zum Lesen und Rechnen enthält.

Den zweiten theoretischen Bezugsrahmen bildet die Cattell-Horn-Carroll-Theorie der Intelligenzmessung. Dazu werden mit der KABC-II bis zu fünf breite Fähigkeiten gemessen, nämlich Kurzzeitgedächtnis Gsm, visuelle Verarbeitung Gv, fluide Intelligenz Gf, Langzeitgedächtnis Glr sowie Kristalline Intelligenz Gc, die zu einem Gesamtindex namens Fluid-Kristallin-Index (FKI) zusammengefasst werden. Dieser ist in der Durchführung identisch mit dem IVI, be-

inhaltet aber zusätzlich den Kennwert für Wissen, der im Rahmen der CHC-Theorie Kristalline Fähigkeiten Gc genannt wird (vgl. Tab. 1).

Generell empfehlen die Autoren, bei der Testinterpretation den FKI zu benutzen, es sei denn, sprachliche Defizite und kulturelle Abweichungen (z. B. bei Kindern mit Migrationshintergrund) lassen es wichtig erscheinen, einen kulturfairen Gesamtwert zu bilden, der die kristallinen Fähigkeiten von der Intelligenzberechnung ausschließt. Für Kinder, die der deutschen Sprache nicht ausreichend mächtig sind, gibt es zusätzlich die Möglichkeit, einen sprachfreien Index (SFI) als Globalmaß der Intelligenz zu bilden, der sich aus Aufgaben zusammensetzt, die sprachfrei durchgeführt werden können. Der SFI empfiehlt sich vor allem bei Kindern mit gravierenden Sprach- und Hörproblemen.

Tabelle 1: Gegenüberstellung der Funktionen nach Luria und der Faktoren nach der CHC-Theorie, die mit der KABC-II erfasst werden können

Alter	Bezeichnung Luria-Modell	Bezeichnung CHC-Modell	Bezeichnung KABC-II
4;0-18;11	Sequentielle Verarbeitung	Kurzzeitgedächtnis Gsm	Sequentiell/Kurzzeitged. Gsm
4;0-18;11	Simultane Verarbeitung	Visuelle Verarbeitung Gv	Simultan/Visuelle Verarbeitung Gv
7;0-18;11	Planung	Fluide Intelligenz Gf	Planung/Fluide Fähigkeiten Gf
4;0-18;11	Lernfähigkeit	Langzeitspeicher und -erinnerung Glr	Lernen/Langzeitspeicher und -erinnerung Glr
4;0-18;11	(Wissen)	Kristalline Fähigkeiten Gc	Wissen/Kristalline Fähigkeiten Gc
3;0-18;11	Intellektueller Verarbeitungsindex (IVI)	Fluid-Kristallin-Index (FKI)	

Je nach Alter des Kindes und zugrunde gelegtem Modell werden zwischen fünf und zehn Untertests zur Berechnung des Gesamtwerts herangezogen. Hinzu kommen eine Reihe von Ergänzungsuntertests. Diese dienen der Erweiterung der diagnostischen Möglichkeiten, können aber auch Kernuntertests ersetzen, wenn aufgrund von Verweigerung oder anderer besonderer Umstände ein Kernuntertest entfallen muss. Das Zusammenstellen von Untertests nach Gutdünken zu einer eigenen Kerntestbatterie und deren Auswertung anhand der vorhandenen Normwerttabellen ist nicht zulässig. Einige Untertests sind in bestimmten Altersabschnitten Kernuntertests, während sie zu anderen Zeitpunkten als Ergänzungsuntertests fungieren. Zwei Untertests werden je nach Alter des Kindes verschiedenen Skalen zugeordnet (s. Tab. 2, folgende Seite). Die Namen der Skalen beinhalten sowohl die Bezeichnung nach dem Luria-Modell als auch den Namen des jeweiligen CHC-Faktors (s. Tab. 1).

Acht Untertests wurden mit geringfügiger Modifikation aus der K-ABC übernommen. Meist wurden Aufgaben niedriger und hoher Itemschwierigkeit hinzugefügt. Acht Untertests der K-ABC sind weggefallen, wofür zehn Untertests neu hinzu kamen. Nicht mehr enthalten sind die Untertests Zauberfenster, Räumliches Gedächtnis, Fotoserie (wurde durch Geschichten ergänzen abgelöst), Gesichter & Orte (ersetzt durch Wort- und Sachwissen), Rechnen, Lesen/Buchstabieren sowie Lesen/Verstehen. Den Wegfall der Aufgaben zum Lesen und Rechnen begründen die Autoren damit,

dass diese Fertigkeiten mithilfe eigens dafür entwickelter Schulleistungstests überprüft werden können.

Tabelle 2: Testaufbau und Durchführungsdauer der KABC-II

Alter in Jahren	3 ^a	4	5	6	7-12	13-18
Sequentiell/Kurzzeitgedächtnis (Gsm)						
Zahlen nachsprechen	◦ ^b	• ^c	•	•	•	•
Wortreihe	•	•	•	•	•	•
Handbewegungen	+ ^d	◦+	◦+	◦+	◦+	◦+
Simultan/visuelle Verarbeitung (Gv)						
Bausteine zählen			◦+	◦+	◦+	•+
Konzeptbildung	•+	•+	•+	•+		
Wiedererkennen von Gesichtern	•+	•+	◦+			
Muster ergänzen ^e			•	•		
Rover				•	•	•
Geschichten ergänzen ^e				◦		
Dreiecke	•+	•+	•+	•+	•+	•+
Gestaltschließen	◦	◦	◦	◦	◦	◦
Planung/fluide Fähigkeiten (Gf)						
Muster ergänzen ^e			+	+	•+	•+
Geschichten ergänzen ^e				+	•+	•+
Lernen/Langzeitspeicher und -erinnerung (Glr)						
Atlantis	•	•	•	•	•	•
Symbole		•	•	•	•	•
Atlantis – Abruf nach Intervall			◦	◦	◦	◦
Symbole – Abruf nach Intervall			◦	◦	◦	◦
Wissen/kristalline Fähigkeiten (Gc) (nur FKI)						
Wortschatz	•	•	•	•	◦	◦
Rätsel	•	•	•	•	•	•
Wort- und Sachwissen	◦	◦	◦	◦	•	•
Testdauer in Minuten (lt. Manual)						
Luria-Modell (IVI)	25-30	30-35	35-40	45-50	55-60	50-55
CHC-Modell (FKI)	30-35	40-45	45-50	55-60	70-75	65-70
Sprachfreie Skala (SFI)	20	20	30	30	40	40

Anmerkungen: ^{a)} Bei den Dreijährigen werden keine Skalenwerte ermittelt, sondern es wird nur ein Gesamtindex gebildet; ^{b)} ◦ Ergänzungsuntertest; ^{c)} • Kernuntertest; ^{d)} + Untertest der sprachfreien Skala SFI; ^{e)} Die Untertests *Muster ergänzen* und *Geschichten ergänzen* sind bei den Kindern bis 6;11 Jahren der Skala Simultan/G_v zugeordnet, ab 7;0 Jahren der Skala Planung/G_f.

Im Folgenden werden die einzelnen Untertests beschrieben. Hinweise zu den Altersbereichen der Skalen siehe Tabelle 1. Der sprachfreie Index setzt sich je nach Altersgruppe aus vier oder fünf Untertests zusammen, wobei dazu teilweise Untertests

herangezogen werden, die in der normalen Durchführung nur als Ergänzungsuntertests zum Einsatz kommen. In eckigen Klammern sind jeweils die CHC-Einzel-faktoren (Stratum-I) aufgeführt, die laut deutschem Manual mithilfe der Aufgaben erfasst werden können. Andere Autoren finden teilweise auch andere Zuordnung (vgl. Flanagan, Alfonso, Ortiz, 2012).

Zahlen nachsprechen: Vorgesprochene Zahlenreihen müssen in derselben Reihenfolge wiederholt werden [Gedächtnisspanne].

Wortreihe: Es werden Wortreihen vorgesprochen. Auf die zugehörigen Symbole muss anschließend auf einer Vorlage in der richtigen Reihenfolge getippt werden. Bei Aufgaben mit höherem Schwierigkeitsgrad ist zwischen Hören und Zeigen eine Interferenz-aufgabe (Farben benennen) zu absolvieren [Arbeitsgedächtnis].

Handbewegungen: Vorgemachte einfache Handpositionen müssen vom Kind in der richtigen Sequenz wiederholt werden [Gedächtnisspanne, visuelles Gedächtnis].

Bausteine zählen: Auf Abbildungen sind Bausteine in verschiedener Anordnung zu sehen, wobei in der perspektivischen Darstellung Steine zum Teil verdeckt sind. Diese müssen demnach beim Zählen erschlossen werden. Es ist jeweils die richtige Anzahl von Bausteinen anzugeben [Visualisierung, mathematische Leistungen].

Konzeptbildung: Aus einem Satz von vier oder fünf Bildern soll dasjenige herausgefunden werden, das nicht zu den übrigen passt [Visualisierung, induktives Denken].

Wiedererkennen von Gesichtern: Zunächst wird das Gesicht einer oder mehrerer Personen gezeigt. Dieselben Personen sollen dann anschließend auf Gruppenfotos herausgefunden werden [visuelles Gedächtnis].

Rover: Auf einer schachbrettartig gegliederten Fläche mit Hindernissen soll der kürzeste Weg vom Start- zum Zielpunkt gefunden werden. Die Aufgaben sind in die Rahmenhandlung eines Hundes eingebettet, der auf dem kürzesten Weg zu seinem Knochen gelangen möchte [räumliche Erfassung, generelles sequenzielles Denken, mathematische Leistungen].

Dreiecke: Bei den einfacheren Aufgaben dieses Untertests müssten aus verschieden geformten und gefärbten Plastikteilen vorgegebene Muster nachgelegt werden. Bei den schwierigeren Aufgaben stehen als Material zweifarbige Dreiecke aus Moosgummi zur Verfügung [räumliche Zusammenhänge, Visualisierung]

Gestaltschließen: aus unvollständigen Abbildungen soll erschlossen werden, welches Objekt dargestellt ist. Das Kind soll sagen, welches Objekt es erkannt hat [Schnelligkeit visueller Erkennung].

Muster ergänzen: Eine Reihe von Mustern soll um ein fehlendes Muster ergänzt werden, das aus vier oder sechs Alternativen auszuwählen ist. Zur Lösung der Aufgaben ist es erforderlich, die Logik zu erkennen, die der jeweiligen Serie inhärent ist [induktives Denken, Visualisierung].

Geschichten ergänzen: Eine Abfolge von Bildern stellt eine unvollständige Handlungssequenz dar. Aus einer Reihe von Bildkarten müssen dann diejenigen Bilder ausgewählt und eingefügt werden, die die Geschichte richtig vervollständigen [induktives Denken, Visualisierung, generelles sequenzielles Denken, Allgemeinwissen].

Atlantis: Sukzessive werden dem Kind die Phantasienamen von 12 Wasserlebewesen vermittelt. Diese müssen anschließend nach Nennung durch den Untersucher aus einer größeren Ansammlung von Abbildungen herausgefunden werden [assoziatives Gedächtnis].

Atlantis – Abruf nach Intervall: 20 bis 25 Minuten nach der Durchführung von *Atlantis* wird ohne weiteren Lerndurchgang erneut abgefragt, welche Namen-Objekt-Zuordnungen das Kind sich aus der Erstdarbietung behalten hat [assoziatives Gedächtnis, Lernfähigkeit].

Symbole: Das Kind bekommt eine Art Symbolschrift präsentiert und lernt nach und nach die Wortbedeutung der einzelnen Zeichen, mit denen dann Sätze mit zunehmender Anzahl von Symbolen „gelesen“ werden sollen [assoziatives Gedächtnis].

Symbole – Abruf nach Intervall: Auch zum Untertest *Symbole* gibt es eine zeitverzögerte Abfrage der behaltene Symbolnamen, ohne dass zuvor eine Auffrischung des Gelernten erfolgt [assoziatives Gedächtnis, Lernfähigkeit].

Wortschatz: Abgebildete lebensweltbezogene Objekte sollen benannt werden [lexikalisches Wissen].

Rätsel: Anhand mehrerer vom Untersucher in Frageform genannter Merkmale soll ein Begriff erraten werden [lexikalisches Wissen, generelles sequentielles Denken, Sprachentwicklung].

Wort- und Sachwissen: Aus jeweils sechs dargebotenen Bildern soll dasjenige ausgewählt werden, das zu einem genannten Begriff (Objekt, Eigenschaft etc.) passt [lexikalisches Wissen, Allgemeinwissen].

Die Subskalen werden jeweils aus zwei Kernuntertests gebildet. Eine Ausnahme stellt *Simultan/Gv* dar, wo in manchen Altersgruppen bis zu vier Untertests zur Indexbildung beitragen. Zusätzlich zu den genannten Subskalen gibt es die Möglichkeit, aus den beiden Ergänzungsuntertests *Atlantis – Abruf nach Intervall* und *Symbole – Abruf nach Intervall* einen separaten Index „Abruf nach Intervall“ zu errechnen, für den eigene Normen vorliegen.

Das Material umfasst vier Aufstellordner und zwei Spiralblöcke mit Testaufgaben und Instruktionen sowie Bildkarten und Legeplättchen. Es gibt ein Protokollheft für alle Altersgruppen sowie ein PC-Auswerteprogramm. Das Manual ist umfangreich (338 Seiten) und detailliert. Es enthält im Anhang die Normtabellen. Anfangs wurde der Test in einer Stofftasche geliefert, die den Inhalt nicht ausreichend schützte. Nach Kritik aus dem Kreis der Testbenutzer wurde diese zwischenzeitlich durch einen Pilotenkoffer ersetzt.

Testdurchführung, Auswertung und Interpretation

Die Durchführungsregeln unterscheiden sich in vielen Details von der Vorgängerversion, sodass auch Testleiter, die die K-ABC kennen, um eine sorgfältige Einarbeitung nicht umhin kommen. Außer einem geeigneten Arbeitsplatz sind keine besonderen räumlichen Anforderungen zu erfüllen. Der Test wird im Einzelsetting durchgeführt. Die Reihenfolge der Aufgaben ist festgelegt. Die Ergebnisse werden im Protokollheft

festgehalten, wo auch die Verrechnungen vorgenommen und die Ergebnisse dargestellt werden. Außerdem besteht dort die Möglichkeit, Verhaltensbeobachtungen einzutragen. Zu vielen Untertests gibt es Einführungs- und Lernaufgaben, die gewährleisten sollen, dass das Kind die Aufgabenstellung verstanden hat. Altersdifferenzierte Einstiegspunkte und Abbruchkriterien sowie Regeln für den Fall, dass die Einstiegs- und Lernaufgaben nicht bewältigt wurden, gewährleisten eine ökonomische Testdurchführung. Verrechnungshinweise gibt es auch für den Fall, dass in Untertests null Rohwertpunkte erreicht werden. Bei einigen Aufgaben ist exakte Zeitnahme erforderlich, um Darbietungs- und Bearbeitungszeiten zu kontrollieren. Ab 7;0 Jahren gibt es bei einigen Untertests Bonuspunkte für rasche Bearbeitung. Für Personen mit motorischen Handicaps liegen separate Normen ohne diese Bonusbepunktung vor. Bei einigen Aufgaben ist eine abgestufte Bewertung vorzunehmen, je nachdem; ob sie im ersten oder zweiten Lösungsversuch richtig bewältigt wurden.

Die Normwerte der Untertests werden in Wertpunkten ($M = 10$, $SD = 3$), die der Skalen als IQ-Werte ($M = 100$, $SD = 15$) angegeben. Zur Profilinterpretation werden ausschließlich die Subskalenindices herangezogen, wenn die Untertestergebnisse nicht zu große Unterschiede aufweisen. Es können sowohl normbezogene als auch individuelle Stärken und Schwächen unter Berücksichtigung von Signifikanzniveaus und kritischen Grundraten ermittelt werden.

Testentwicklung, Normierung und Testgütekriterien

Aufbauend auf die sorgfältige Entwicklung der amerikanischen Originalversion erfolgte eine Übersetzung und teilweise kulturelle Adaptierung des Verfahrens auf deutschsprachige Verhältnisse. Die Testnormierung erfolgte von April 2013 bis Februar 2014 an 1.745 Kindern in Deutschland, Österreich und der deutschsprachigen Schweiz, wobei Repräsentativität bezüglich Geschlecht, regionaler Verteilung, Siedlungsform, Migrationsanteil und Bildungsstand der Eltern angestrebt und weitgehend eingehalten wurde. Lediglich der Anteil von Probanden mit Eltern aus niedrigen Bildungsschichten war unterrepräsentiert. Zur Normierungsstudie selbst sind die Auskünfte eher spärlich. So erfährt man lediglich, dass die Durchführung in den Händen eines Marktforschungsunternehmens lag, das seinerseits Untersucher anlernte, deren beruflicher Hintergrund ungenannt bleibt. 1.628 Probanden wurden in Regelkindertagesstätten und Regelschulen gewonnen. Hinzu kam ein Anteil von 117 Kindern und Jugendlichen, die als klinische Stichprobe bezeichnet werden. Diese macht 6,7 % der Gesamtstichprobe aus und beinhaltet Kinder mit Entwicklungsstörungen und diversen psychischen Störungen, die sich nicht gleichmäßig über die Altersgruppen verteilen. Auch zur Gewinnung dieser Probandengruppen fehlen nähere Angaben.

Für die Gesamtstichprobe ergaben sich vielfach schlechtere Testleistungen bei niedrigem Bildungsniveau der Eltern und Migrationshintergrund, wobei das letztgenannte Merkmal im Manual zwar problematisiert, aber letztlich nicht definiert wird. Die Effekte fallen für die verschiedenen Skalen unterschiedlich groß aus (s. Tab. 3).

Tabelle 3: Effekte des elterlichen Bildungsgrads und des Migrationsstatus auf KABC-II-Ergebnisse

KABC-II Skalen	Bildungsstand der Eltern		Migration ^b	
	Differenz SA – HS ^a		Differenz Nein – Ja	
	3-6 Jahre	7-18 Jahre	3-6 Jahre	7-18 Jahre
Sequentiell/Gsm	10,8	15,6	5,6**	4,2***
Simultan/Gv	13,9	14,2	1,4	5,1***
Lernen/Glr	10,2	13,7	1,6	3,4***
Planung Gf	---	13,9	---	3,4***
Wissen/Gc	18,5	19,4	12,4***	8,5***
IVI	14,5	18,2	3,4*	5,0***
FKI	17,2	19,4	6,6***	6,1***
SFI	11,5	13,7	- 0,9 ^c	4,2***

Anmerkungen: ^a) SA = Studienabschluss, HS = Hauptschule mindestens 8 Jahre; ^b) Migrationshintergrund ja/nein unter Berücksichtigung des elterlichen Bildungsgrads ^c) hier schnitten die Kinder mit Migrationshintergrund geringfügig besser ab als die ohne * $p \leq .05$ ** $p \leq .01$ *** $p \leq .001$

Bei den Skalen sind Akademikerkinder den Kindern aus niedrigen Bildungsschichten im Schnitt um 1 bis 1,3 SD überlegen. Selbst wenn der Bildungsstand der Eltern kontrolliert wird, wirkt sich Migrationshintergrund ungünstig auf die Testergebnisse aus. Kulturfairness lässt sich demnach allenfalls für die sprachfreie Skala reklamieren.

Es wurden 47 Altersgruppen gebildet, bei den Drei- bis Fünfjährigen Gruppen à drei Monate pro Jahrgang, bei den Sechs- bis Vierzehnjährigen je drei Gruppen à vier Monate pro Jahrgang, und bei den Fünfzehn- bis Achtzehnjährigen erfolgte die Gruppeneinteilung in Halbjahresschritten. Die Normen gelten für Jungen und Mädchen gleichermaßen, auch wenn in der Normierungsstichprobe signifikante, aber geringfügige Geschlechtsunterschiede, zumeist zugunsten der Jungen (Simultan/Gv, Wissen/Gc und FKI), gefunden wurden. Für die amerikanische Fassung waren leichte, teilweise aber signifikante Unterschiede berichtet worden (Singer et al., 2012), wobei je nach Alter und Skala mal die Jungen und mal die Mädchen besser abschnitten.

Bei den Kindern im Alter von 3;0 bis 3;2 Jahren weisen drei der sieben Kernuntertests leichte Bodeneffekte auf, sodass bei einem Rohwert (RW) von 1 in allen Untertests minimal ein Gesamtwert von 57 (mehr als 3 SD) erreicht werden kann. Bei den Vierjährigen sind alle Kernuntertests frei von Bodeneffekten und es werden Skalenwerte bis in den untersten Messbereich hinein ausgewiesen. Leichte Bodeneffekte bestehen außerdem beim Untertest Muster ergänzen bei den Fünf- und Sechsjährigen sowie im Untertest Rover im Alter von 6;0 bis 6;2 Jahren. Ansonsten sind Kernuntertests und Skalen frei von Bodeneffekten. Bis zu einem Alter von 13 Jahren treten keine Deckeneffekte auf, ab dann zunehmend bei Ergänzungsuntertests. Kernuntertests sind nur in den höchsten Altersgruppen betroffen. Die Anstiege der Normwerte zwischen benachbarten Rohwertpunkten sind stetig, gelegentlich kommt es dabei zu Normwertsprüngen von 2/3 SD. Anstiege zwischen Altersgruppen betragen selten mehr als 1/3 SD.

Tabelle 4: Reliabilitäten der Testskalen

	3-6 Jahre	7-18 Jahre
Sequentiell/Gsm	.88 (.85 - .89)	.90 (.87 - .93)
Simultan/Gv	.93 (.88 - .95)	.95 (.94 - .97)
Lernen/Glr	.97 (.96 - .97)	.97 (.96 - .97)
Planung/Gf	---	.89 (.84 - .92)
Wissen/Gc	.94 (.92 - .95)	.95 (.94 - .96)
IVI	.95 (.94 - .96)	.97 (.96 - .98)
FKI	.97 (.96 - .97)	.98 (.97 - .98)
SFI	.92 (.90 - .92)	.95 (.93 - .96)

Anmerkung: Durchschnittswerte über alle Altersgruppen (in Klammern Range) berechnet mit Fisher's z-Transformation. Die Angaben im Manual beziehen sich auf gemittelte Werte der Altersnormgruppen.

Bei den Untertests erfolgte die Reliabilitätsberechnung anhand einer etwas unklar beschriebenen Prozedur, die der Split-Half-Methode angenähert ist. Die somit gewonnenen Kennwerte fallen überwiegend zufriedenstellend ($>.70$) bis hoch ($>.80$), teilweise sogar sehr hoch ($>.90$) aus. Bei den Kernuntertests treten keine Reliabilitätskoeffizienten $<.70$ auf, bei den Ergänzungsuntertests ist dies in einzelnen Altersgruppen der Fall. Für die Untertests Atlantis – Abruf nach Intervall und Symbole – Abruf nach Intervall wird ein gemeinsamer Reliabilitätskoeffizient angegeben, der in allen Altersgruppen Mittelwerte $>.90$ ergibt (s. Tab. 4).

Zur Retestreliabilität verweisen die Autoren auf die amerikanische Originalversion, von der eine zufriedenstellende bis hohe Mess-Stabilität über einen Zeitraum von durchschnittlich vier Wochen (12 bis 56 Tage) berichtet wird. Auf eine Überprüfung bei der deutschen Version haben die Autoren aus methodischen Gründen verzichtet. Sie nehmen an, dass die für die amerikanische Fassung berichteten Übungseffekte von 11,5 Standardwertpunkten für Lernen/ Glr und 7-10 Standardwertpunkten für Simultan/Gv und Planung/Gf vermutlich auch für die deutsche Fassung zutreffen, ebenso wie die geringeren Zuwächse bei Wissen/Gc (3-4 Punkte) und Sequentiell/Gsm (-0,8-2,2 Punkte). Zuwächse bei den Gesamtskalen IVI, FKI und SFI werden im Manual nicht referiert. In Anbetracht der teilweise deutlichen Übungseffekte empfehlen die Autoren, die KABC-II nicht häufiger als im Abstand von 12 Monaten zu wiederholen.

Viele Aufgabenformate der KABC-II sind in der Intelligenzdiagnostik etabliert und können als inhaltsvalide gelten. Für neu aufgenommene Untertests ist der Zusammenhang mit den zugrunde gelegten Intelligenzmodellen plausibel dargelegt. Die Subskala Planung/Gf wurde eher aus statistischen denn aus inhaltlichen Gründen zusammengestellt. Jede Subskala wird aus mindestens zwei verschiedenen Untertests gebildet, was im Rahmen der CHC-Theorie für ausreichend gehalten wird, um Stratum-II Faktoren angemessen zu erfassen (Flanagan et al., 2012).

Die Interkorrelationen der Kernuntertests fallen in den niedrigen bis mittleren Bereich und sind innerhalb der jeweiligen Subskalen durchweg gleich hoch oder höher

als die Zusammenhänge zwischen Aufgaben, die verschiedenen Skalen angehören. Alle Untertests korrelieren in mittlerer bis sehr hoher Ausprägung mit den jeweiligen Skalen, die ihrerseits auch hohe Zusammenhänge untereinander aufweisen. Planung/Gf hat von allen Skalen den höchsten Zusammenhang mit den Gesamtindices IVI und FKI. Die referierten konfirmatorischen Faktorenanalysen scheinen die Skalenstruktur der KABC-II für alle Altersgruppen zu bestätigen. Am höchsten laden Planung/Gf und Simultan/Gv auf dem Gesamtwert (IVI) mit Werten von .88 bis .99. Auch die Ladungen der Untertests auf den Fähigkeitsfaktoren, die die Subskalen repräsentieren, beträgt .50 und mehr. Allerdings lassen die im Manual enthaltenen Angaben keine umfassende Beurteilung der Faktorenanalysen zu.

Zur konkurrenten Validität werden Vergleichsuntersuchungen referiert, die mit Teilen der Normierungsstichprobe in verschiedenen Altersgruppen durchgeführt wurden und die überwiegend deutliche Zusammenhänge mit den Skalenwerten von WISC-IV (Petermann u. Petermann, 2011); IDS (Grob, Meyer, Hagmann-von-Arx, 2009); WPPSI-III (Petermann, 2009); SON-R 5½-17 (Tellegen u. Laros, 2005) und K-TIM (Melchers, Schürmann, Scholten, 2006) aufweisen. So beträgt beispielsweise der Zusammenhang zwischen FKI und dem WISC-IV-Gesamtwert .88. Ein Vergleich zwischen KABC-II und K-ABC bei Kindern im Alter von 8 bis 11 Jahren liefert ebenfalls deutliche statistische Zusammenhänge bei nur geringen Mittelwertdifferenzen im Gesamtwert. Die Korrelationen zwischen inhaltlich ähnlichen Untertests der beiden Verfahren fallen jedoch teilweise erstaunlich niedrig aus. Zur diskriminativen Validität wird lediglich von einer Studie mit der Kaseler Konzentrationsaufgabe (Krampen, 2007) berichtet, die für statistische Unabhängigkeit zwischen KABC-II und selektiver Aufmerksamkeit spricht. Zu Profilen klinischer Stichproben einschließlich Hochbegabung werden lediglich Ergebnisse genannt, die für die amerikanische Version vorliegen. Zur Kriteriums- und prognostischen Validität fehlen Angaben bzw. es wird cursorisch auf Befunde zur amerikanischen Version verwiesen.

Praktische Erprobung

Aufgrund der Vielzahl an Durchführungsregeln und zahlreicher Modifikationen gegenüber der K-ABC ist für alle Testanwender eine sorgfältige Einarbeitung erforderlich. Die Aufbewahrungs- und Transporttasche ermöglicht keine übersichtliche Anordnung des Materials und keinen materialschonenden Transport. Die Umstellung auf ein anderes Behältnis, die der Verlag zwischenzeitlich vorgenommen hat, ist daher prinzipiell zu begrüßen. Allerdings vermag der nunmehr zum Einsatz kommende Pilotenkoffer im praktischen Einsatz nicht zu überzeugen. Er gewährleistet zwar einen sicheren und rückschonenden Transport, ist aber platzintensiv und sollte von der Inneneinteilung her verbessert werden. Die meisten Aufgaben werden anhand der Aufstellordner dargeboten, was eine Sitzordnung „über Eck“ erfordert, damit die Zeigebewegungen des Kindes vom Untersucher auch gut beobachtet werden können. Die Skizze im Manual, die diese Sitzposition veranschaulichen soll, erscheint dem Rezensenten dazu allerdings ungeeignet.

Die Spiralheftungen der Aufstellordner biegen sich im Gebrauch auf, sodass ein reibungsloses Umblättern nicht durchgehend gewährleistet ist. Das erfordert häufiges Nachbiegen oder, falls man das versäumt hat, umständliches Wiedereinfädeln der Aufstellpappen. Während einer Testung kann es dadurch zu erheblicher Behinderung kommen. Die Registerblätter nutzen sich an den Griffstellen rasch ab und die Pappen, die statt des bisherigen Plastikeinbands die Aufstellordner stützen sollen, sind für Kinder, die kraftvoll auf die Abbildungen tippen, nicht stabil genug. Eine fehlerhafte Abbildung im Untertest Gestaltschließen wurde inzwischen ausgetauscht.

Die Handlungsanweisungen für den Testleiter sind klar beschrieben und die Instruktionen sind sprachlich einfach gehalten. Warum auf den Instruktionseiten der Aufstellordner auch ausführliche Hintergrundinformationen zu den einzelnen Untertests abgedruckt sind, ist nicht nachvollziehbar. Dadurch wird unnötiges Blättern bei jeder Testdurchführung erforderlich, was sich insbesondere bei Kindern, die konstante Aufmerksamkeit verlangen, ungünstig auswirkt. Die wichtigsten Durchführungsanweisungen hätten etwas prägnanter hervorgehoben werden können. Die Vorlage zum Untertest *Wortreihe* ist weder kratzfest noch abwaschbar oder praktisch in ihrer Handhabung. Bei regem Gebrauch bekommt der Softcover-Einband des Manuals auch bei vorsichtiger Handhabung rasch Gebrauchsspuren. Ein Hardcover-Einband oder ein separates spiralgebundenes Heft mit Testnormen wären bessere Alternativen.

Im Allgemeinen lässt sich das Verfahren mit Kindern aller Altersstufen gut durchführen, lediglich bei den Sechsjährigen gab es teilweise Durchhalteprobleme. Die Angaben zur Durchführungsdauer im Manual erwiesen sich in der eigenen Erprobung als realistisch. Atlantis ist als Einstiegsaufgabe und „Eisbrecher“ meist gut gewählt. Lediglich sehr schüchterne Kinder, die der Aufforderung auf die Abbildungen zu zeigen nicht nachkommen, lassen sich auf andere Untertests wie Wortschatz oder Dreiecke am Anfang unter Umständen besser ein. Die Aufgabenreihenfolge ermöglicht eine durchaus abwechslungsreiche Darbietung. Durch Zeitnahme, gestaffelte Bepunktung, Durchführungsunterschiede bei verschiedenen Altersgruppen und nicht zuletzt die differenzierten Kriterien für die Verhaltensbeobachtung stellt das Verfahren hohe, aber mit der nötigen Testroutine durchaus zu bewältigende Anforderungen an den Testleiter. Die Hinzunahme einfacher Aufgaben ermöglicht nun einen problemlosen Einstieg in den Untertest Dreiecke. Das Bildmaterial ist optisch klar gestaltet, aktualisiert und weist im Untertest Wort- und Sachwissen (Nachfolger von Gesichte & Orte) auch keine „Märchenlastigkeit“ mehr auf, erfordert allerdings intaktes Sehvermögen und gute Ausleuchtung des Arbeitsplatzes. Für die Anwendung bei sehbehinderten Kindern bietet KABC-II sich nicht unbedingt an. Die Anforderungen an die Handmotorik der Kinder sind bis auf Dreiecke und Rover gering. Bei entsprechenden motorischen Handicaps können, wo erforderlich, Normen ohne Bonuspunkte für rasche Bearbeitung herangezogen werden, die allerdings bei den älteren Probanden ausgeprägte Deckeneffekte aufweisen. Sprachlichen Problemen der Probanden wird generell durch sprachlich einfache und kurze Anweisungen und geringe Anforderungen an den sprachlichen Output Rechnung

getragen. Zusätzlich gibt es die sprachfreie Skala, jedoch ohne spezifische Instruktionen zur nonverbalen Durchführung.

Die Bewertung ist zumeist einfach. Der Protokollbogen bietet reichlich Platz für Eintragungen, ist vom Layout her aber gewöhnungsbedürftig. Die Auswertung anhand der Tabellen bereitet keine Schwierigkeiten. Alternativ liegt ein unkompliziert zu installierendes und handzuhabendes Auswertungsprogramm vor. Die Vorgaben, die bezüglich der Profilerinterpretation gemacht werden, sind methodisch angemessen. Die Ergebnisinterpretation setzt Kenntnis der theoretischen Grundlagen voraus. Dies trifft im Prinzip für alle (intelligenzdiagnostischen) Verfahren zu, ist aber wegen des expliziten Bezugs auf die neuropsychologische Theorie Lurias und das CHC-Modell hier besonders augenfällig.

Bewertung

Die KABC-II stellt eine gründliche Überarbeitung und Erweiterung der K-ABC dar. Das Verfahren basiert weiterhin auf dem Luria-Modell und fügt der simultanen und sukzessiven Verarbeitung noch Lernen und Planung als weitere Leistungsbereiche hinzu. Zugleich besteht die Möglichkeit, Befunde nach der CHC-Theorie zu interpretieren, auch wenn die Untertests in der Regel mehr als einem Intelligenzfaktor zuzuordnen sind. Dem Benutzer eröffnen sich damit zeitgemäße Interpretationsmöglichkeiten, und die Verknüpfung mit anderen Verfahren im Sinne eines Cross-Battery-Assessments (Flanagan et al., 2012) wird merklich erleichtert.

Begrüßenswert ist auch die Berücksichtigung von Lernen als spezifischer kognitiver Leistung und des verzögerten Abrufs, zwei Komponenten, die bislang trotz ihrer praktischen Relevanz in Intelligenztests kaum Berücksichtigung fanden. Kritisch anzumerken ist, dass im Untertest Symbole etliche Abbildungen die Symbolbedeutung anschaulich wiedergeben, sodass die Testleistung nicht nur das Lernen in der Testsituation wiedergibt, sondern auch durch bildungsabhängiges Vorwissen beeinflusst wird. Außerdem findet der verzögere Abruf bei der Berechnung des Index Lernen/Glr keine Berücksichtigung, obwohl der Recall als wesentlicher Bestandteil des CHC-Faktors Glr gilt. Die Unterscheidung zwischen simultaner/visueller Verarbeitung einerseits und Planung/fluiden Intelligenz andererseits ist von den Aufgabeninhalten her wenig eingängig und die Zuordnung der Aufgaben zu den beiden Indices erfolgte eher aus teststatistischen denn aus inhaltlichen Gründen.

Wissen/Kristalline Fähigkeiten Gc enthält vornehmlich Aufgaben zum Passivwortschatz und zur sprachlichen Konzeptbildung und kann damit inhaltlich nur bedingt mit dem Index Sprachverständnis der WISV-IV gleichgesetzt werden. Der Wegfall von Rechnen und Leseverständnis bei KABC-II ist zu bedauern, da es sich hierbei um orientierende Aufgaben handelte, die eingesetzt werden konnten, um eingehenderen Abklärungsbedarf zu ermitteln. Laut Manual erfassen zwar Rover und Bausteine zählen auch quantitative Fähigkeiten Gq, doch beschränken sich diese auf reine Zählfertigkeiten, die ab einem Alter von 6 bzw. 13 Jahren für die Beurteilung der Rechenfertigkeiten weitgehend ungeeignet sind.

Zu begrüßen ist, dass nunmehr normiertes Testen mit der KABC-II bis zum 19. Lebensjahr möglich ist. Bei der Interpretation der Ergebnisse ist allerdings zu berücksichtigen, dass Jugendliche im Alter von 16 bis 18 Jahren im Durchschnitt circa acht IQ-Punkte besser abschneiden als im K-TIM (Melchers et al., 2006), wobei sich die Differenz vornehmlich bei den kristallinen Leistungen ergibt.

Für die Diagnostik von Kindern und Jugendlichen mit (leichter) Intelligenzmindereung bietet sich die KABC-II mit ihren vielfältigen individuellen Adaptierungsmöglichkeiten und geringen Bodeneffekten an. Die Leistungsdifferenzierung im oberen Intelligenzbereich basiert bei den älteren Kindern nicht unwesentlich auf den Bonuspunkten für rasche Bearbeitung, vor allem bei Planung/Fluide Fähigkeiten Gf. Die Bonuspunkte machen am Gesamtergebnis bis zu 1½ SD aus, bei Planung/fluiden Intelligenz Gf bis zu 4½ SD und auch bei Simultan/visuelle Fähigkeiten Gv kann es zu erheblichen Unterschieden je nach Arbeitstempo kommen. Längst nicht alle hochleistenden Jugendlichen weisen hohes Arbeitstempo auf. Daher ist in diesen Fällen bei der Testinterpretation Vorsicht angesagt. Irreführend könnte außerdem sein, dass Probanden einerseits durch die Testinstruktionen aufgefordert werden, sich bei der Bearbeitung nicht zu beeilen und andererseits der Lösungszeit bei der Bewertung ein so hoher Stellenwert beigemessen wird. Der Wegfall der Altersgruppe 2;6 bis 2;11 Jahre fällt nicht ins Gewicht, da sich die K-ABC wegen großer Bodeneffekte und unausgewogener Aufgabenzusammenstellung für die Untersuchung dieser Altersgruppe ohnehin nicht empfahl.

Im Manual wird die K-ABC-II als kulturfares Verfahren bezeichnet. Allerdings wirken sich niedriger Sozialstatus und Migrationshintergrund bei der deutschen Normierung erheblich stärker auf die Testergebnisse aus als in amerikanischen Studien (Singer et al., 2012), wo nur halb so große Unterschiede berichtet werden. Dies sollte nicht vorschnell der deutschen Version des Tests angelastet werden, da die Ursachen auch in der Stichprobenauswahl oder in anderen gesellschaftlichen Bedingungen liegen können und in dieser Hinsicht Vergleiche mit anderen deutschsprachigen Verfahren fehlen. Die genannten Leistungsunterschiede sind jedoch so groß, dass Kulturfairness kaum glaubwürdig für das Verfahren als Ganzes reklamiert werden kann.

Der Test lässt sich im vorgesehenen Altersbereich im Allgemeinen gut durchführen. Die Instruktionen sind knapp gehalten, die Anleitungen für den Testleiter sind präzise und ermöglichen eine objektive Durchführung. Das Material weist einige gravierende Unzulänglichkeiten auf. Man wünscht sich, der Verlag hätte den Qualitätsstandard der alten K-ABC-Materialien beibehalten. Hier sollte im Interesse einer störungsfreien Testdurchführung dringend nachgebessert werden. Das Protokollheft ist ausführlich und bis auf einige gewöhnungsbedürftige Besonderheiten gut zu benutzen. Es enthält alle Verrechnungsoptionen. Die manuelle Testauswertung dauert circa zehn Minuten, die Auswertung mittels der empfehlenswerten Auswertungssoftware geht etwas rascher und minimiert Ablese- und Berechnungsfehler. Positiv hervorzuheben sind die methodisch klaren und fundierten Regeln zum Umgang mit Testergebnissen bei einem Rohwert von Null und zur Profilverinterpretation. Bezüglich der Validität von Testprofilen fehlen noch die empirischen Belege.

Die Normtabellen im Anhang des Manuals sind einigermaßen übersichtlich gestaltet und lassen sich zufriedenstellend ablesen. Das ansonsten ausführliche Handbuch liefert kaum Informationen zum Procedere der Testnormierung und auch die Angaben zur Konstruktvalidität sind nicht umfassend genug. Die Skalen sind weitgehend frei von Boden- und Deckeneffekten, die Normen sind stetig und liegen in ausreichend feiner Altersabstufung vor.

Die Reliabilität der Gesamtskalen ist gut bis sehr gut. Zu den Untertests werden insbesondere bei den Kernuntertests hohe bis sehr hohe Konsistenzmaße berichtet, die aber wegen verschiedener Berechnungsweisen mit den Reliabilitätswerten anderer Verfahren nur bedingt vergleichbar sind. Für die deutsche Fassung wären Retestergebnisse und Befunde zu Übungseffekten insbesondere für klinische Gruppen wünschenswert. Die Menge an Vergleichsstudien mit anderen Intelligenztestverfahren ist geradezu als vorbildlich zu bezeichnen. Dafür fehlen Befunde zur diskriminativen und Kriteriumsvalidität der Skalen und des Testprofils noch weitgehend.

Insgesamt kann KABC-II als gelungene Neufassung und Normierung dieses weit verbreiteten Intelligenztests angesehen werden. Inhaltliche und teststatistische Schwächen der Vorläuferversion wurden systematisch beseitigt, sodass dieses Verfahren nunmehr eine methodisch fundierte Intelligenzdiagnostik auf dem aktuellen wissenschaftlichen Stand ermöglicht. Unbefriedigend ist die Materialqualität, die zumindest teilweise für den täglichen Gebrauch unzulänglich und dem hohen Anschaffungspreis unwürdig ist. Falls es zu weiteren verlagsseitigen Nachbesserungen beim Material kommt, sollten davon auch diejenigen profitieren können, die das Verfahren bereits angeschafft haben und zum Teil durch ihre Kritik dem Verlag wichtige Hinweise zu dessen Verbesserung geliefert haben. Andernfalls ist Interessenten zu raten, mit der Anschaffung zu warten, bis die Materialqualität ausgereift ist.

Die theoretische Fundierung ist einerseits eine Stärke des Verfahrens, die in der Praxis allerdings nur dann zum Tragen kommen kann, wenn Testanwender über das erforderliche Grundwissen verfügen, um die Ergebnisse angemessen interpretieren zu können. Daher ist mit Nachdruck auf die Empfehlung der deutschen Testautoren im Manual hinzuweisen, dass nur diejenigen das Verfahren anwenden sollten, die sich im Rahmen eines Psychologiestudiums und/oder einer Weiterbildung zum Kinder- und Jugendlichenpsychotherapeuten das dazu erforderliche testtheoretische Wissen angeeignet und sich mit den theoretischen Grundlagen des Verfahrens hinreichend vertraut gemacht haben. Es ist zu hoffen, dass der Verlag beim Verkauf mithilft, diese in anderen Ländern selbstverständlichen Standards psychologischer Testung im Interesse der Klienten auch im deutschsprachigen Vertrieb zu beherzigen.

Literatur

- Flanagan, D. P., Alfonso, V. C., Ortiz, S. O. (2012). The Cross-Battery Assessment Approach. An overview, historic perspective and current directions. In D. P. Flanagan, P. L. Harrison (Hrsg.), *Contemporary intellectual assessment* (S. 459-483). New York: Guilford.
- Grob, A., Meyer, C. S., Hagemann-von-Arx, P. (2009). *Intelligence and Development Scales (IDS)*. Bern: Huber.
- Kaufman, A. S., Kaufman, N. L. (2004). *Kaufman Assessment Battery for Children – Second Edition*. Minneapolis: NCS Pearson.
- Krampe, G. (2007). *KKA: Kaseler Konzentrationsaufgabe für 3- bis 8-Jährige*. Göttingen: Hogrefe.
- Luria, A. R. (1970). The functional organization of the brain. *Scientific American*, 222, 66-78.
- Melchers, P., Preuß, U. (1991). *Kaufman Assessment Battery for Children (K-ABC)*. Deutsche Bearbeitung. Lisse: Swets & Zeitlinger.
- Melchers, P., Schürmann, S., Scholten S. (2006). *Kaufman Test zur Intelligenzmessung für Jugendliche und Erwachsene (K-TIM)*. Leiden: PITS.
- Petermann, F. (Hrsg.) (2009). *Wechsler Preschool and Primary Scale of Intelligence – III, deutsche Version (WPPSI-III)*. Frankfurt a. M.: Pearson Assessment.
- Petermann, F., Petermann, U. (2011). *Wechsler Intelligence Scale for Children – Fourth Edition (WISC-IV)*. Deutsche Bearbeitung. Frankfurt a. M.: Pearson Assessment.
- Schneider, W. J., McGrew, K. S. (2012). The Cattell-Horn-Carroll Model of Intelligence. In D. P. Flanagan, P. L. Harrison (Hrsg.), *Contemporary intellectual assessment* (S. 99-144). New York: Guilford.
- Singer, J. K., Lichtenberger, E. O., Kaufman, J. C., Kaufman, A. S., Kaufman, N. L. (2012). The Kaufman Assessment Battery for Children – Second Edition and the Kaufman Test of Educational Achievement – Second Edition. In D. P. Flanagan, P. L. Harrison (Hrsg.), *Contemporary intellectual assessment* (S. 269-296). New York: Guilford.
- Tellegen, P. J., Laros, J. A. (2005). *SON-R 5½-17. Non-verbaler Intelligenztest (3. korrigierte Aufl.)*. Göttingen: Hogrefe.

Dieter Irblich, Auel