

Psychometrische Eigenschaften der „Kaufman-Assessment Battery for Children" (K-ABC) bei 5- und 6-jährigen Kindern I: Reliabilität und Validität in einer klinischen Stichprobe¹

Gerolf Renner, Susanne Schmid, Dieter Irblich, Günter Krampen

Zusammenfassung: Bei 420 Kindern im Alter von 5 und 6 Jahren, die wegen Entwicklungs-, Verhaltens- oder emotionalen Störungen in einem Sozialpädiatrischen Zentrum vorgestellt wurden, wurde die „Kaufman-Assessment Battery for Children" (K-ABC) durchgeführt. Erfasst wurden ICD-10-Diagnosen, soziodemografische Variablen und bei einer Teilstichprobe Verhaltensratings (VBV 3-6 Elternfragebogen). Die Reliabilität der Skalen (Intellektuelle Fähigkeiten, Einzelheitliches Denken, Ganzheitliches Denken, Fertigkeiten) erwies sich mit Werten von .86 bis .92 als mittel bis hoch. Die Reliabilität der Subtests fällt nur teilweise befriedigend aus (.63-.92). Bei einer Teilstichprobe durchgeführte Wiederholungsmessungen zeigten eine hohe Stabilität der Skalen (.60- .79), bei den Subtests liegen die Stabilitäten zwischen .36 und .81. Zur differenziellen Validität für klinische Subgruppen sowie zur diskriminanten Validität ergaben sich weitgehend den Erwartungen entsprechende Befunde.

Schlüsselwörter: Intelligenzdiagnostik, Entwicklungsdiagnostik, Frühdiagnostik, Frühförderung, Reliabilität, Sozialpädiatrie, Validität, Vorschulalter; K-ABC; VBV 3 - 6; Profilanalyse

Psychometric properties of the "Kaufman-Assessment Battery for Children" (K-ABC) in 5- and 6-year old children I: Reliability and validity in a clinical sample

Summary: Reliability and validity of the "Kaufman-Assessment Battery for Children" (K-ABC; German version) were examined under routine conditions in a Community mental health Service for early assessment and early intervention (Social Pediatric Center). 420 children ages 5 to 6 with developmental, behavioural, and emotional disorders were tested. Sociodemographic data, ICD-10-diagnoses, and parental ratings of child behaviour (VBV 3-6 parent questionnaire) were collected. The reliability of the K-ABC scales (mental processing, sequential processing, simultaneous processing, achievement) ranged from .86 to .92. For several subtests reliability did not reach satisfactory values (ranging from .63 to .91). For a subgroup of the sample test-retest-data showed a high stability of K-ABC scales (.60 to .72). Retest-stability of subtests ranged from .36 to .81. Results demonstrate discriminative validity. Differential validity for clinical subgroups was supported empirically.

Keywords: Intelligence assessment, developmental assessment, preschool assessment, early intervention, preschool age, K-ABC, VBV 5-6, profile analysis

¹ Die Datenerhebung zu dieser Untersuchung wurde durch ein Forschungsstipendium der Deutschen Gesellschaft für Sozialpädiatrie und Jugendmedizin (DGSPJ) ermöglicht.

Die Erfassung des kognitiven Entwicklungsstandes mit mehrdimensionalen Intelligenztests ist ein wesentliches Element der Frühdiagnostik von Entwicklungsstörungen und trägt zur Differenzialdiagnostik sowie zur Behandlungsplanung bei (s. z.B. Daseking/Petermann, 2008; Renner/Irblich, 2009). Ein für diesen Altersbereich geeignetes individualdiagnostisches Verfahren ist die „Kaufman-Assessment Battery for Children“ (K-ABC; Melchers/Preuss, 1991, 2009), die häufig in Frühförderstellen (Grimm/Aktas 2002), Sozialpädiatrischen Zentren (Schmid/Weyrether/Weiß/Mickley, 2010), kinder- und jugendpsychiatrischen Einrichtungen (Bölte/Adam-Schwebe/Englert/Schmeck/Poustka, 2000) und Erziehungsberatungsstellen (Castello/Nestler, 2003) eingesetzt wird.

Die K-ABC soll einerseits grundlegende intellektuelle Fähigkeiten erfassen, wobei unter Bezug auf kognitions- und neuropsychologische Modellvorstellungen eine vornehmlich sequenzielle und eine vornehmlich simultane Informationsverarbeitung unterschieden und der Testkonstruktion zugrunde gelegt werden. Außerdem sollen eher kultur- und bildungsabhängige Fertigkeiten gemessen werden. Fähigkeiten und Fertigkeiten werden nicht zu einem globalen Intelligenzwert zusammengefasst, sondern separat sowie in ihrer Gegenüberstellung interpretiert.

Das Verfahren umfasst insgesamt 16 Untertests, von denen je nach Alter des Kindes bis zu 13 vorgegeben werden. Die „Skala intellektueller Fähigkeiten“ (SIF) soll grundlegende kognitive Leistungen erfassen und besteht aus zwei Unterskalen: Die „Skala einzelheitlichen Denkens“ (SED) dient der Erfassung serieller Kurzzeitgedächtnisleistungen, die „Skala ganzheitlichen Denkens“ (SGD) spricht vorrangig figurale, räumlich-integrative und abstrakt-logische Fähigkeiten an. Mit der „Fertigkeitenskala“ (FS) werden Kompetenzen wie

Wortschatz, sprachliches Verständnis und Rechnen abgebildet. Untertests, die keine aktivsprachlichen Leistungen verlangen, können zur „Nonverbalen Skala“ (NV) zusammengefasst werden, die sich inhaltlich weitgehend mit der SGD überschneidet. Bei den 5-Jährigen gehören die Subtests „Handbewegungen“, „Zahlennachsprechen“ und „Wortreihe“ zur SED, die Subtests „Gestaltschließen“, „Dreiecke“, „Räumliches Gedächtnis“ und „Bildhaftes Ergänzen“ zur SGD und die Subtests „Gesichter und Orte“, „Rechnen“ und „Rätsel“ zur FS. Bei den 6-Jährigen kommt der Subtest „Fotoserie“ (SGD) hinzu. Zur Standardauswertung der K-ABC werden im Testmanual neben der Bestimmung der Subtest- und Skalenwerte die Signifikanzbeurteilung und inhaltliche Interpretation von Profildifferenzen empfohlen.

Unter Bezug auf die Cattell-Horn-Carroll Intelligenztheorie (CHC-Theorie) als alternativen Interpretationsrahmen können die Untertests der K-ABC den Faktoren Visuelle Verarbeitung (Gv), Fluide Intelligenz (Gf), Kristalline Intelligenz (Gc) und Kurzzeitgedächtnis (Gsm) zugeordnet werden (Mickley/Renner, 2011).

Das amerikanische Original wurde schon vor einigen Jahren durch eine inhaltlich merklich veränderte Nachfolgeversion abgelöst („Kaufman Assessment Battery for Children, Second Edition“; Kaufman/Kaufman, 2004), deren deutsche Übertragung bisher nicht erfolgt ist. Trotz Kritik an möglicherweise veralteten Normen und nicht mehr zeitgemäßen Items (z. B. Horn, 2003) wird die deutschsprachige K-ABC weiterhin vom Testverlag vertrieben, in der Praxis eingesetzt und in der Fachliteratur empfohlen (z.B. Rindermann/ Kwiatkowski, 2010).

Eine Recherche in der deutschsprachigen psychologischen Datenbank PSYINDEX (Suchstichwort „K-ABC“; Stand: 8.1.2012) ergab

151 Treffer, die belegen, dass die K-ABC in vielfältigen Forschungs- und Anwendungskontexten zum Einsatz kommt. Allerdings befassten sich nur wenige Studien explizit mit den psychometrischen Eigenschaften des Verfahrens in klinischen Stichproben. Süß-Burghart (1994, 1995) berichtete über den Einsatz der K-ABC bei lern- und geistigbehinderten Kindern. Dabei zeigten sich ebenso wie in einer großen unveröffentlichten klinischen Studie (Düchting/Freiha/Zippel/Herkenrath, 2004) u. a. hohe Retest-Stabilitäten der K-ABC-Skalen. Die konvergente Validität zwischen der K-ABC und dem „Nonverbalen Intelligenztest“ (SON-R 2⁷²⁻⁷; Tellegen/Laros/Pe-termann, 2007) wurde in einer sozialpädiatrischen Stichprobe belegt (Renner/Rausch/ Krampen/Irblich, 2009). Zur Validität und Reliabilität der Profilanalyse finden sich bis heute weder im Testmanual noch in unabhängigen deutschsprachigen Studien Daten.

Dieser Mangel an Folgestudien war Anlass zur vorliegenden Untersuchung an einer klinisch-sozialpädiatrischen Stichprobe. Dabei stehen folgende Fragen im Mittelpunkt:

- 1) die Reliabilität und Retest-Stabilität der K-ABC-Skalen und Subtests,
- 2) die Reliabilität und Validität von Profildifferenzen der K-ABC-Skalen,
- 3) die differenzielle Validität für bedeutsame diagnostische Gruppen,
- 4) die diskriminative Validität der K-ABC gegenüber Verhaltensvariablen.

Methoden Stichprobe

Die Gesamtstichprobe umfasst 420 Kinder (144 5-Jährige, 276 6-Jährige), die zwischen 1994 und 2008 im SPZ kreuznacher diakonie, Bereichsstellen Idar-Oberstein und Simmern vorgestellt wurden. Sozialpädiatrische Zentren sind in Rheinland-Pfalz gleichzeitig Träger der Frühförderung. Alle Kinder wurden im

Rahmen der klinisch-psychologischen Routinediagnostik von zwei erfahrenen Diplom-Psychologen untersucht. Die „Fertigkeitenskala“ wurde nicht bei allen Kindern komplett durchgeführt, die entsprechende Teilstichprobe umfasst 78 5-Jährige und 140 6-Jährige.

Das Alter der Kinder lag zwischen 60 und 83 Monaten ($M = 6;1$ Jahre, $SD = 0;6$ Jahre). 297 der Kinder waren Jungen (70.7%) und 123 Mädchen (29.3%). Bei 88.1 % der Kinder war Deutsch die Muttersprache, 8.8% wuchsen zweisprachig auf (Deutsch und eine Fremdsprache), 3.1% sprachen ausschließlich eine Fremdsprache. 84.0% der Kinder besuchten einen Regelkindergarten, 3.6% einen Förderkindergarten, 5.2% eine Vor- oder Förderschule, 5.0% waren bereits eingeschult, nur 1.2% wurden ausschließlich in der Familie betreut. Bei 405 Kindern lagen Angaben zur Schwangerschaftsdauer vor: 1.2% waren sehr frühe Frühgeburten (24.-27. SSW), 3.1% frühe Frühgeburten (28.-33. SSW) und weitere 6.9 % späte Frühgeburten (34. - 36. SSW).

In einer bei allen Kindern vorab durchgeführten neuropädiatrischen Untersuchung wurden außerhalb des Kapitels V (Psychische und Verhaltensstörungen) der ICD-10 (Dilling/Mombour/Schmidt, 2010) am häufigsten Diagnosen gestellt aus den Gruppen H00-H95 (Krankheiten des Auges) mit 9.5%, P00-P96 (bestimmte Zustände, die ihren Ursprung in der Perinatalperiode haben) mit 7.4%, Q00-Q99 (angeborene Missbildungen, Deformitäten und Chromosomenanomalien) mit 6.9% , G00-G99 (Krankheiten des Nervensystems) mit 5.0% und E00-E90 (endokrine, Ernährungs- und Stoffwechselkrankheiten) mit 4.3%.

In der klinisch-psychologischen Diagnostik waren die folgenden Hauptdiagnosen nach ICD-10 am häufigsten vertreten:

199 | Fl 4 / 2012

- 38.4% Kombinierte Entwicklungsstörungen (F83)
 - 8.1 % Andere Entwicklungsstörungen (F88)
- 8.1 % Rezeptive Sprachentwicklungsstörungen (F80.2)
- 5.7% Störungen des Sozialverhaltens (F91.x),
 - 5.5% Hyperkinetische Störungen (F90.x)
- 5.2% Motorische Entwicklungsstörungen (F82)
 - 5.0% Artikulationsstörungen (F80.0)
 - 4.5% Intelligenzminderungen (F70, F71)

Erhebungsinstrumente

Neben der Intelligenztestung mit der K-ABC wurde bei 160 Kindern die Elternversion des „Verhaltensbeurteilungsbogens für Vorschulkinder“ (VBV-EL 3-6; Döpfner/Berner/Fleischmann/Schmidt, 1993) eingesetzt. Die diagnostischen Instrumente wurden nach dem in den Testmanualen beschriebenen Standardvorgehen durchgeführt und ausgewertet.

Ergebnisse

Die Datenauswertung erfolgte mit der Statistiksoftware SPSS 19. Aufgrund der hohen Interkorrelation (.92, $p < .001$) mit der „Skala Gesamtheitlichen Denkens“ (SGD) wird im Folgenden auf eine Darstellung von Ergebnissen zur „Nonverbalen Skala“ (N V) verzichtet.

Item- und Skalenparameter

Die Mittelwerte der IQ- und Subskalen lagen erwartungsgemäß deutlich unter denen der Normstichprobe (s. Tab. 1). Die Schwierigkeitsindizes (s. Tab. 2) deckten bei den meisten Untertests das ganze Schwierigkeitskontinuum ab. Testrohre von Null fanden sich bei den Untertests „*Handbewegungen*“, „*Dreiecke*“, „*Bildhaftes Ergänzen*“, „*Räumliches Gedächtnis*“ und „*Rätsel*“ in 0.2 bis 1.7 % der Fälle. Die Anfangsanforderungen von „*Fotoserie*“ erwiesen sich für viele 6-Jährige als zu schwierig: 23.4% lösten kein einziges Item.

Reliabilität

Für alle Untertests wurde die interne Konsistenz (Cronbachs α) berechnet (Tab. 2). Werte von $r_{ii} > .80$ in beiden Altersgruppen wurden nur bei „*Dreiecke*“, „*Räumliches Gedächtnis*“ und „*Rechnen*“ erreicht, die niedrigsten Reliabilitätsschätzungen ergaben sich für „*Bildhaftes Ergänzen*“. Die Zuverlässigkeit der K-ABC-Skalen, berechnet nach Formel 14.4 bei Lienert und Raatz (1995), beträgt .86 bis .92 (s. Tab. 1).

Tab. 1: Deskriptive Kennwerte und Reliabilität der K-ABC-Skalen

Skala	Mittelwert (SD)	Range	Reliabilität ^a	
			5-Jährige	6-Jährige
SIF	86.3 (12.1)	52-123	.89 (.90)	.89(.91)
SED	84.5 (13.0)	51-130	.86 (.89)	.87(.91)
SGD	88.0(14.9)	46-128	.90 (.94)	.90 (.93)
FS	82.1 (14.0)	55-131	.9 ² (.92)	.89 (.93)

^a Berechnet nach Formel 14.4 in Lienert & Raatz (1994) auf Basis der internen Konsistenzen

^b In Klammern Reliabilitätskennwerte der Normierungsstichprobe lt. Testmanual. SD = Standardabweichung.

Tab. 2: Reliabilität und Itemparameter der K-ABC-Subtests

Subtest	Reliabilität ^a		Spannweite der p.	
	5-Jährige	6-Jährige	5-Jährige	6-Jährige
Handbewegungen	.70 (.69)	■72 (77)	08-96	03-99
Gestaltschließen	75 (73)	73 (73)	09-99	.04- 1.0
Zahlennachsprechen	75 (-86)	74 (86)	00- 1.0	.00- 1.0
Dreiecke	.89 (.91)	91 (88)	.05-95	.00-97
Wortreihe	75 (.84)	.84 (.86)	02-96	.01- 1.0
Bildhaftes Ergänzen	63 (77)	69 (79)	09-80	.04-87
Räumliches Gedächtnis	.82 (.74)	82 (.76)	02-91	.05-89
Fotoserie	-	■75 ^c (.83)	-	.10-.54
Gesichter & Orte	■77(77)	80 (.79)	.00-99	00-99
Rechnen	87 (.90)	92 (-92)	.07-1.0	02- 1.0
Rätsel	.87(83)	78 (.84)	.04-.90	01-99

^a Interne Konsistenz

^b In Klammern Reliabilitäten der Normierungsstichprobe lt. Testmanual

^c Fälle mit einem Testrohwert von Null wurden ausgeschlossen

p = Itemschwierigkeit

Retest-Stabilität

Daten zu Wiederholungsmessungen in einem Retest-Intervall zwischen 6 bis 18 Monaten lagen für bis zu 50 Kinder vor (Tab. 3). Für die Skalenwerte zeigten sich durchgehend signifikante Stabilitätskoeffizienten zwischen .60 (FS) und .79 (SGD). Allerdings schließt dies erhebliche individuelle Veränderungen nicht aus: Die maximale Differenz zwischen den beiden Testungen lag für die SIF bei 26, für die SED bei 19, für die SGD bei 44 und für die FS bei 25 Standardwertpunkten. Auf Ebene der Subtests fielen die Ergebnisse bei teilweise geringen Stichprobengrößen sehr heterogen und nicht in allen Fällen statistisch signifikant aus. Dabei zeigten sich bei sieben der elf Untertests Koeffizienten $>.60$. Für die laut Manual interpretationsrelevanten Differenzen zwischen K-ABC-Skalenwerten ergaben sich folgende Stabilitäten: SED-SGD .49 (N=50, $p<.001$), SIF-FS .70 (N=21, $p<.001$), SED-FS .38 (N=21, n.s.) und SGD-FS .21 (N=21, n.s.).

Tab. 3: Retest-Stabilität der K-ABC-Skalen und Subtests

Skala/Subtest	N	r _{tt}
SIF	50	.74**
SED	50	.61**
SGD	50	.79"
FS	21	.60"
Handbewegungen	50	.60"
Gestaltschließen	50	.72"
Zahlennachsprechen	50	.81"
Dreiecke	50	.50"
Wortreihe	49	.61"
Bildhaftes Ergänzen	50	.36*
Räumliches Gedächtnis	50	.67"
Fotoserie	15	.39
Gesichter & Orte	21	■53*
Rechnen	44	.69"
Rätsel	43	.78"

N = Stichprobenumfang r_{tt} = Retest-Stabilität * $p<.05$

** $p<.01$

Differenzielle Validität

Die differenzielle Validität der K-ABC wurde in einfaktoriellen Varianzanalysen mit anschließenden β -postren'orz'-Mittelwertvergleichen nach Duncan für die häufigsten Hauptdiagnosen untersucht (s. Tab. 4). Dabei wurden nur Diagnosegruppen berücksichtigt, die mindestens 10 Fälle umfassten. Bei Analysen mit

der „Fertigkeitenskala“ war dies nur bei fünf Diagnosen der Fall.

- Bei der SIF ($F_{(10/369)} = 42.4, p < .001$, Effektstärke $w^2 = .52$) erreichten Kinder mit Intelligenzminderungen und Kombinierten Umschriebenen Entwicklungsstörungen die niedrigsten Werte, die sich jeweils signifikant von allen anderen Gruppen unterschieden. Die höchsten Werte zeigten Kinder mit Artikulationsstörungen und Kinder, bei denen keine ICD-10-Diagnose gestellt wurde (zu Details s. Tab. 4). Kinder mit emotionalen, motorischen und hyperkinetischen Störungen sowie mit Störungen des Sozialverhaltens erzielten Leistungen im Durchschnittsbereich.
- Ein im Wesentlichen vergleichbares Bild ergab sich bei der SED ($F_{(10/369)} = 24.3, p < .001$, Effektstärke $w^2 = .38$) und der SGD ($F_{(10/369)} = 30.6, p < .001$, Effektstärke $w^2 = .44$). Allerdings unterschieden sich bei der SED die Kombinierten

Umschriebenen Entwicklungsstörungen nicht signifikant von den Diagnosegruppen Rezeptive Sprachentwicklungsstörung und Expressive Sprachentwicklungsstörung (s. a. Tab. 4).

- Bei der FS ($F_{(10/369)} = 17.5, p < .001$, Effektstärke $w^2 = .30$) erreichten wiederum Kinder mit Intelligenzminderungen signifikant niedrigere Werte als die anderen Diagnosegruppen (zu Details s. Tab. 4).

Neben den Mittelwerten wurden auch die Differenzwerte zwischen den K-ABC-Skalen betrachtet (Tab. 5). Bei Einbeziehung der FS ergab sich wiederum eine Einschränkung auf nur fünf Diagnosegruppen.

- SED - SGD: Bei einem signifikanten Haupteffekt für die Diagnosegruppe ($F_{(10/369)} = 4.1, p < .001$, Effektstärke $w^2 = .07$) zeigten Kinder mit Expressiven und Rezeptiven

Tab. 4: Mittelwerte der K-ABC-Skalen in klinischen Diagnosegruppen

Diagnosegruppe	N'	SIF	SED	SGD	FS"
F70 Intelligenzminderung	19 (10)	64.6 _a	65.4 _a	64.2 _a	65.7 _a
F83 Komb. Entwicklungsstörungen	161 (92)	79.0 _b	78.6 _b	79.5 _b	77.1 _b
F88 Andere Entwicklungsstörungen	34 (17)	88.2 _c	87.9 _d	88.7 _c	89.8 _c
F80.2 Sprache rezeptiv	34 (20)	88.7 _c	80.9 _b	95.2 _{d,c}	73.0 _b
F80.1 Sprache expressiv	11	89.9 _{c,d}	82.3 _b	96.1 _{e,d}	
F82 Motor. Entwicklungsstörungen	22	91.4 _d	90.6 _d	93.1 _d	
F90 Hyperkinetische Störungen	23	94.7 _{d,e,f}	93.9 _{d,e}	96.0 _e	
F93 Emotionale Störungen	15	95.9 _e	91.3 _{d,e}		
F91 Störungen des Sozialverhaltens	24 (14)	96.7 _{e,f}	93.8 _{d,e}	98.3 _d	92.2 _c
F80.0 Artikulationsstörungen	21	99.6 _f	98.1 _e	101.3 _f	
Keine Diagnose	16	98.1 _e	97.8 _e	100.4 [^]	

Mittelwerte mit unterschiedlichem Subskript unterscheiden sich statistisch bedeutsam voneinander (Duncan-Test, $p < .05$). N = Stichprobenumfang.

^a In Klammern Stichprobenumfang für die Analyse der „Fertigkeitenskala“. ^b Diagnosegruppen mit einem N < 10 wurden nicht berücksichtigt.

Tab. 5: Mittelwerte von Profildifferenzen und Profilvariabilität in klinischen Diagnosegruppen

Diagnosegruppe	N'	SED-SCD	SIF-FS"	SED-FS"	SGD-FS"
F70 Intelligenzminderung	19 (10)	1.2	0.0 _d	1.8	-1.2 _a
F83 Komb. Entwicklungsstörungen	161 (92)		2.2 _c	1.5	3.0 _c
F88 Andere Entwicklungsstörungen	34(17)	" ^a -7 _b		-8.9	" ^a 2-5 _a
F802 Sprache rezeptiv	34 (20)	-14.2,	i6.i _b	7.2	23.6 _b
F80.1 Sprache expressiv	11	-13.8,			
F82 Motor. Entwicklungsstörungen	22	" ^a 2-5 _{b,c}			
F90 Hyperkinetische Störungen	23	-2 ^{-V}			
F93 Emotionale Störungen	15	-8.8 _c			
F91 Störungen des Sozialverhaltens	24 (14)	-4-3 _b	3-9,	1-3	7-8,
F80.o Artikulationsstörungen	21	" ^a 3-3 _b			
Keine Diagnose	16	-2.6 _h	b c		

Mittelwerte mit unterschiedlichem Subskript unterscheiden sich statistisch bedeutsam voneinander (Duncan-Test, $p < .05$). N = Stichprobenumfang.

' In Klammern Stichprobenumfang für die Analyse der „Fertigkeitenskala“. " Diagnosegruppen mit einem N < 10 wurden nicht berücksichtigt.

Sprachentwicklungsstörungen die größten Diskrepanzen (höhere Werte in der SGD). FS - SIF: Die Varianzanalyse fiel statistisch signifikant aus ($F_{m,w} = 8.5, p < .001$, Effektstärke $w^2 = .16$). Kinder mit Rezeptiven Sprachentwicklungsstörungen unterschieden sich dabei signifikant von den übrigen Diagnosegruppen und wiesen deutlich höhere Werte in der SIF auf. FS-SED: Die Varianzanalyse zeigte keinen statistisch signifikanten Effekt der Diagnosegruppe ($F_{(4/148)} = 1.2, p = .33$). FS-SGD: Bei einem signifikanten Haupteffekt ($F_{(4/148)} = 8.5, p < .001$, Effektstärke $w^2 = .24$) zeigten Post-hoc-Tests bei Kindern

mit Rezeptiven Sprachentwicklungsstörungen die größten Diskrepanzen mit besseren Leistungen in der SGD.

Diskriminative Validität

Zur Beurteilung der *diskriminativen Validität* wurde der Zusammenhang zwischen der K-ABC und dem VBV 3-6-Elternfragebogen untersucht. Nur bei der Skala „Hyperaktivität vs. Spielausdauer“ ergeben sich signifikante Korrelationen geringer bis mittlerer Effekt-

Tab. 6: Korrelationen der K-ABC-Skalen mit dem VBV 3-6-Elternfragebogen

Skala	Emotionale Kompetenzen	Oppositionell-aggressives Verhalten	Hyperaktivität vs. Spielausdauer	Emotionale Auffälligkeiten
SIF	.03	-.05	-.33**	-.05
SED	.02	.00	-.28**	-.06
SGD	.01	-.08	-.31**	-.05
FS	.00	.05	-.19	-.07

$p < .05$; ** $p < .01$; Stichprobenumfang = 160 (SIF, SED und SGD) bzw. 87 (FS)

stärke (Tab. 6). Zu den Differenzwerten zwischen den K-ABC-Skalen fanden sich keine signifikanten Zusammenhänge.

Fazit

Die K-ABC konnte in der klinisch-psychologischen Diagnostik bei 5-jährigen Kindern überwiegend problemlos durchgeführt werden. Ein beträchtlicher Anteil der 6-Jährigen war allerdings schon mit den Anfangsanforderungen des Subtests „Fotoserie“ überfordert, was aufgrund des auch in den Normtabellen der K-ABC erkennbaren Bodeneffektes in dieser Altersgruppe nicht überrascht. Für den Einsatz bei leistungsschwächeren 6-Jährigen sind daher Testverfahren zu bevorzugen, die in allen Untertests auch Items mit geringen Aufgabenschwierigkeiten anbieten (zu den aus Bodeneffekten resultierenden Problemen bei der Testinterpretation s. z.B. Renner, 2009).

Die ermittelten Reliabilitätskennwerte der K-ABC lagen in der klinischen Stichprobe nur geringfügig unter denen der Normierungsstichprobe. Für die Intelligenzskalen kann die Reliabilität nach der Einteilung von Flanagan, Ortiz, Alfonso und Mascolo (2006) als mittel bis hoch bezeichnet werden. Es zeigten sich zudem im Längsschnittverlauf hohe Stabilitäten (.60 bis .72), was jedoch auf individueller Ebene Spielraum für deutliche Veränderungen lässt. Für die SIF fällt dieser Wert niedriger aus als in der Untersuchung von Süß-Burghart (1995), die eine sehr hohe Stabilität von $r_{tt} = .93$ ergab. Diese Stichprobe zeigte gegenüber unserer Untersuchung einen deutlich niedrigeren IQ ($M=64.7$) mit stark eingeschränkter Varianz (zur höheren Stabilität von Intelligenzwerten im unteren Extrembereich s. z. B. Kamphaus, 2005). Auf Ebene der Subtests wurde die für klinische Interpretationen wünschenswerte Reliabilität von $r_{tt} > .80$ (Yousfi/Steyer, 2006) nur bei „Dreiecke“, „Räumliches Gedächtnis“ und „Rechnen“ in beiden Altersgruppen erreicht. Allerdings ergaben sich bei „Gestalt schließen“, „Zahlennachsprechen“ und „Rätsel“ hohe Retest-

Stabilitäten, die sogar das Niveau der Intelligenzskalen erreichten. Bei allen weiteren Subtests sollten spezifische inhaltliche Interpretationen stets durch ergänzende Verfahren abgesichert werden. Dies gilt insbesondere für „Bildhaftes Ergänzen“, bei dem sich sowohl für die interne Konsistenz als auch für die Retest-Stabilität die niedrigsten Werte fanden.

Daten zur differenziellen Validität zeigten -bei durchgehend großen Effekten in den Varianzanalysen - für Kinder mit Intelligenzminderungen und kombinierten Entwicklungsstörungen wie erwartet in allen Skalen die niedrigsten Testleistungen. Kinder mit Sprachentwicklungsstörungen erzielten in der SGD annähernd durchschnittliche Werte, während sie in der SED und der FS deutlich schlechter abschnitten. Für Diagnosegruppen, bei denen keine nennenswerten kognitiven Beeinträchtigungen erwartet wurden, zeigten sich homogene Profile im Durchschnittsbereich. Dieses Befundmuster wird den Erwartungen an ein Intelligenzdiagnostisches Verfahren gerecht.

Die Zusammenhänge zu Verhaltensvariablen (VBV-EL 3-6) fielen gering bis ganz unbedeutend aus. Die signifikanten Korrelationen zur Skala „Hyperaktivität vs. Spielauddauer“ konvergieren mit vorliegenden Befunden (Döpfner et al., 1993; Renner/Zenglein/Krampe 2004).

Als Besonderheit dieser Untersuchung wurden auch Reliabilität und Validität von Differenzwerten zwischen den K-ABC-Skalen betrachtet. Dabei ergaben sich - bei teilweise kleinen Stichproben - nur für die Diskrepanz zwischen SIF und FS befriedigende Retest-Stabilitäten. In den Analysen zur differenziellen Validität zeigten sich überwiegend signifikante Haupteffekte für die Diagnosegruppe, die Effektstärken fielen jedoch durchgehend geringer aus als für die Analysen auf Skalenebene. Deutliche Diskrepanzen zwischen SGD und SED sowie zwischen SGD und FS fanden sich bei Kindern mit Sprachentwicklungsstörungen, sodass bei Vorliegen entsprechender Ergebnisse

eine nähere Untersuchung der Sprachentwicklung erwogen werden sollte.

Bei der Bewertung der Befunde sind einige Einschränkungen zu berücksichtigen: Es handelte sich um eine selektierte Stichprobe mit einer eng umschriebenen regionalen Herkunft und einem eher geringen Bildungsstatus der Eltern. Es konnten ausschließlich Kinder berücksichtigt werden, die zu einer sozialpädiatrischen Abklärung überwiesen wurden und bei denen im SPZ eine psychologische Untersuchung für sinnvoll erachtet wurde. Der Einsatz der K-ABC erfolgte nicht aufgrund einer systematischen Erhebungsstrategie, sondern beruhte auf einzelfallbezogenen Entscheidungen der Untersucher. Die Daten wurden zudem nicht zeitgleich erhoben, sondern über einen langen Zeitraum in der klinischen Praxis gesammelt. Der Generalisierbarkeit der Ergebnisse sind damit Grenzen gesetzt.

Insgesamt konnten Reliabilität sowie differenzielle und diskriminative Validität der K-ABC auch unter den Routinebedingungen der klinisch-psychologischen Frühdiagnostik bestätigt werden. Bei den 6-Jährigen ergibt sich allerdings eine Einschränkung des Einsatzbereiches aufgrund des deutlichen Bodeneffektes im Subtest „Fotoserie“. Die vorliegenden Daten können jedoch die für den künftigen klinischen Einsatz der K-ABC hoch relevanten Fragen nach der aktuellen Gültigkeit der Normierung und den Auswirkungen der zunehmenden Veralterung der Testmaterialien nicht beantworten.

Gerolf Renner

Pädagogische Hochschule Ludwigsburg, Fakultät für
Sonderpädagogik **Susanne Schmid** Weissenstein-
Klinik St. Blasien **Dieter Irblich**

Sozialpädiatrisches Zentrum kreuznacher diakonie

Günter Krampen

Universität Trier, Fachbereich I - Psychologie,
Abteilung Klinische Psychologie und Psychotherapie

Korrespondenzanschrift: **Prof. Dr. Gerolf Renner**

Pädagogische Hochschule Ludwigsburg Fakultät für
Sonderpädagogik Pestalozzistr. 53 72762 Reutlingen

Literatur

Bölte, S., Adam-Schwebe, S., Englert, E., Schmeck, K. & Poustka, F. (2000): Zur Praxis der psychologischen Testdiagnostik in der deutschen Kinder- und Jugendpsychiatrie: Ergebnisse einer Umfrage. Zeitschrift für Kinder- und Jugendpsychiatrie und Psychotherapie, 283, 151-161

Castello, A. & Nestler, J. (2003): Praxis psychologischer Testdiagnostik an Erziehungsberatungsstellen Verhaltenstherapie & Psychosoziale Praxis, 35, 555-566

Daseking, M. & Petermann, F. (2008): Diagnostik kognitiver Leistungen im Vorschulalter. Monatsschrift Kinderheilkunde, 156, 685-694 **Dilling, H., Mombour, W. & Schmidt, M.H. (2010):** Internationale Klassifikation psychischer Störungen. ICD-10 Kapitel V (F). Klinisch-diagnostische Leitlinien (7., überarbeitete Auflage). Bern: Huber.

Döpfner, M., Berner, W., Fleischmann, T. & Schmidt, M. (1993): Verhaltensbeurteilungsbogen für Vorschulkinder (VBV 3-6). Weinheim: Beltz

Düchting, C, Freiha, T., Zippel, U. & Herkenrath, P. (2004): Kaufman-Assessment Battery for Children. Statistische Testauswertung von über 300 Testdurchführungen mit besonderem Augenmerk auf die prognostische Validität von Testwiederholungen. Vortrag gehalten auf der 8. Sozialpädiatrischen Forschungstagung, Köln, 19.6.2004

Flanagan, D. P., Ortiz, S.O., Alfonso, V.C. & Mascolo, J.T. (2006): The achievement test desk reference. A guide to learning disability identification (2nd ed.). Hoboken, NJ: Wiley & Sons

Grimm, H. & Aktas, M. (2002): Entwicklungstests im Vorschulalter: Beurteilung ihrer Nützlichkeit durch praktisch tätige Psychologen. Frühförderung interdisziplinär, 21, 163-177

Horn, R. (2003): Eine kritische Anmerkung zum K-ABC. Report Psychologie, 28, 189

Kamphaus, R.W. (2005): Clinical assessment of child and adolescent intelligence. New York: Springer

Kaufman, A.S. & Kaufman, N.L. (2004): Kaufman Assessment Battery for Children KABC-II. (2nd Edition). Circle Pines, MN: ACS

Lienert, G.A. & Raatz, U. (1994): Testaufbau und Testanalyse (5., überarbeitete Aufl.). Weinheim: Beltz

Melchers, P. & Preuss, U. (Hrsg.) (1991): Kaufman-Assessment Battery for Children. Deutschsprachige Fassung (K-ABC) (2. Aufl.). Amsterdam, NL: Swets & Zeitlinger

- Melchers, P. & Preuss, U. (Hrsg.) (2009):** Kaufman-Assessment Battery for Children. Deutschsprachige Fassung (K-ABC) (8, unveränd. Aufl.). Frankfurt: Pearson Assessment
- Mickley, M. & Renner, G. (2010):** Intelligenztheorie für die Praxis: Auswahl, Anwendung und Interpretation deutschsprachiger Testverfahren für Kinder und Jugendliche auf Grundlage der CHC-Theorie. Klinische Diagnostik und Evaluation, 3, 447-466
- Renner, G. (2009):** Testpsychologische Diagnostik. In D. Irblich & G. Renner (Hrsg.), Diagnostik in der Klinischen Kinderpsychologie. Die ersten sieben Lebensjahre (S. 73-85). Göttingen: Hogrefe
- Renner, G. & Irblich, D. (2009):** Intelligenzdiagnostik. In D. Irblich & G. Renner (Hrsg.), Diagnostik in der klinischen Kinderpsychologie. Die ersten sieben Lebensjahre (S. 136-151). Göttingen: Hogrefe
- Renner, G., Rausch, N., Krampen, G. & Irblich, D. (2009):** Der SON-R 2 VI-7 in der klinischen Praxis. Reliabilität, Validität und Erprobung einer Kurzform. Kindheit und Entwicklung, 18, 232-243
- Renner, G., Zenglein, C. & Krampen, G. (2004):** Zur Reliabilität und Validität der Elternfassung des „Verhaltensbeurteilungsbogen für Vorschulkinder“ (VBV-EL 3-6) in der klinischen Anwendungspraxis. Diagnostica, 50, 78-87
- Rindermann, H. & Kwiatkowski, V. (2010):** Intelligenzdiagnostik im Kindesalter. In C. Quaiser-Pohl & H. Rindermann (Hrsg.), Entwicklungsdiagnostik (S. 102-132). München: Reinhardt UTB
- Schmid, R.G., Weyrether, H., Weiß, C. & Mickley, M. (2010):** Testverfahren im Sozialpädiatrischen Zentrum. Zugriff am 2.1. 2012 <http://www.dgspj.de/media/QZTestverfahren.pdf>
- Süss-Burghart, Heinz (1994):** Die „Kaufman Assessment Battery for Children (K-ABC)“ und der „Hamburg-Wechsler Intelligenztest für Kinder (HAWIK-R)“ in der Diagnostik klinischer Populationen. Zeitschrift für Differentielle und Diagnostische Psychologie 15, 41-47
- Süss-Burghart, H. (1995):** Die Kaufman Assessment Battery for Children (K-ABC): Testergebnisse, Validität und Retestreliabilität bei mental retardierten Kindern. Frühförderung interdisziplinär, 14, 72-77
- Tellegen, P.J., Laros, J.A. & Petermann, F. (2007):** SON-R 2VI-7. Non-verbaler Intelligenztest. Göttingen: Hogrefe
- Yousfi, S. & Steyer, R. (2006):** Klassische Testtheorie. In F. Petermann & M. Eid (Hrsg.), Handbuch der Psychologischen Diagnostik (S. 288-303). Göttingen: Hogrefe.

