

Neuere Testverfahren

Praxis der Kinderpsychologie und Kinderpsychiatrie 58 (2009) 6, S. 467-476

urn:nbn:de:bsz-psydok-49081

Erstveröffentlichung bei:

Vandenhoeck & Ruprecht WISSENSWERTE SEIT 1735

<http://www.v-r.de/de/>

Nutzungsbedingungen

PsyDok gewährt ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Die Nutzung stellt keine Übertragung des Eigentumsrechts an diesem Dokument dar und gilt vorbehaltlich der folgenden Einschränkungen: Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen.

Mit dem Gebrauch von PsyDok und der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

Kontakt:

PsyDok

Saarländische Universitäts- und Landesbibliothek
Universität des Saarlandes,
Campus, Gebäude B 1 1, D-66123 Saarbrücken

E-Mail: psydok@sulb.uni-saarland.de
Internet: psydok.sulb.uni-saarland.de/

NEUERE TESTVERFAHREN

Ricken, G., Fritz, A., Schuck, K. D., Preuß, U. (Hrsg.) (2007). **Hannover-Wechsler-Intelligenztest für das Vorschulalter - III**. Bern: Huber, Testkoffer komplett: 980,- €, Verbrauchsmaterial pro Testung abhängig vom Alter des Kindes 3,- € bzw. 6,40 €, Testauswerteprogramm 249,- €.

Der Hannover-Wechsler-Intelligenztest für das Vorschulalter - III (HAWIVA-III) ist die deutsche Übersetzung, Adaptation und Normierung der Wechsler Preschool and Primary Scale of Intelligence - III (WPPSI-III; Wechsler, 2002). Sie löst in der Intelligenzdiagnostik bei Vorschulkindern den HAWIVA (Eggert, 1975) ab, der als sog. Experimentalversion jahrzehntlang in Gebrauch war, aber wegen der nicht mehr zeitgemäßen Testkonzeption und veralteter Normen in der klinischen Praxis weitgehend an Bedeutung verloren hatte.

Theoretischer Hintergrund

Der WPPSI-III liegt das Intelligenzkonzept von David Wechsler zugrunde, das Intelligenz als Zusammenwirken eines globalen mit mehreren spezifischen Faktoren versteht. Dazu wird eine Reihe sprachlicher und nichtsprachlicher Fähigkeiten und Fertigkeiten erfasst. Als eigenständiger Faktor neben sprachlichen und wahrnehmungsgebunden-logischen Anforderungen wird auch die Verarbeitungsgeschwindigkeit als wichtiger Intelligenzindikator berücksichtigt. Neu ist außerdem, dass neben dem Verbal-IQ auch eine „Allgemeine Sprachskala“ zur Erfassung des Aktiv- und Passivwortschatzes aufgenommen wurde. Obwohl weitgehend am WPPSI-III orientiert, weist der HAWIVA-III doch einige Änderungen gegenüber der amerikanischen Vorlage auf. So ist die obere Altersgrenze von 7;3 auf 6;11 reduziert, der Mosaik-Test wurde mit zusätzlichen Items im mittleren Schwierigkeitsbereich versehen und beim Untertest Figurenlegen werden Zeitbonuspunkte für kurze Lösungszeiten vergeben. Wegen kultureller Unterschiede zwischen den USA und dem deutschsprachigen Raum erfolgten punktuell Adaptationen der Abbildungen. Die Autoren der deutschen Übersetzung deuten an, dass sie nur einen geringen Spielraum für Modifikationen hatten.

Testaufbau, Material und Durchführung

Der HAWIVA-III ist für die Einzeluntersuchung von Kindern im Alter von 2;6 bis 6;11 Jahre vorgesehen. Kurz- oder Parallelformen sind nicht vorhanden, dafür be-

Der Autor dankt Dr. Gerolf Renner, Birkenfeld, und Dipl.-Psych. Susanne Schillig, Mainz, für ihre hilfreichen Hinweise.

steht durch die Unterscheidung zwischen obligatorischen und optionalen Untertests die Möglichkeit, sich flexibler den diagnostischen Erfordernissen anzupassen. Je nach Alter des Kindes kommen verschiedene Untertests zur Anwendung und es gibt altersabhängige Testeinstiege sowie Abbruchkriterien, wodurch der geringeren Ausdauer jüngerer Kinder Rechnung getragen wird.

Der Testkoffer ist für das darin enthaltene Material ausreichend dimensioniert. Der Test umfasst zwei verschiedene Protokollhefte (für die Altersgruppen 2;6-3;11 und 4;0-6;11), ein spiralgebundenes Manual (255 S.) mit allgemeinen und aufgabenspezifischen Durchführungs- und Auswertungshinweisen sowie ein weiteres mit Informationen zu Testentwicklung und Testgüte (216 S.). Das erforderliche Testmaterial (farbige Plastikwürfel und Puzzleteile) befindet sich in 2 Pappschachteln sowie in 2 Stimulusheften, die ebenfalls Spiralheftung aufweisen. Zusätzlich zu dem im Testkoffer vorhandenen Material benötigt man eine Stoppuhr und Bleistifte.

Bei vollständiger Durchführung umfasst der HAWIVA-III bei den 2;6 bis 3;11-Jährigen 5 Untertests und bei den 4;0- bis 6;11-Jährigen 14 Untertests. Bei der jüngeren Altersgruppe sind 4 Untertests obligatorisch und einer fakultativ, bei den älteren werden 7 Aufgabenarten dem Kernbereich zugeordnet, auf dem die Ermittlung des Verbal-, Handlungs- und des Gesamt-IQ beruht, und 7 weitere stehen für den fakultativen Einsatz zur Verfügung (s. Tab. 1). Diese können teilweise Kernuntertests ersetzen und werden für die Ermittlung weiterer Subskalen herangezogen.

Der Verbalteil setzt sich bei den 2;6- bis 3;11-Jährigen aus 2 Untertests zusammen: (1) *Passiver Wortschatz* (PW): Das Kind soll aus jeweils 4 Abbildungen diejenige auswählen, die der Untersucher nennt. Dabei handelt es sich um Körperteile, Tiere, Pflanzen, Gegenstände, Tätigkeiten, Objekteigenschaften und Präpositionen. (2) *Allgemeines Wissen* (AW): Es werden Fragen zum Alltagswissen des Kindes gestellt. Bei den ersten 4 Aufgaben soll das Kind ebenso wie bei PW auf diejenige von 4 Abbildungen zeigen, die als Antwort zu der Frage des Untersuchers passt. Bei allen weiteren Fragen ist die Antwort mündlich zu geben.

Der Handlungsteil wird in dieser Altersgruppe ebenfalls von 2 Untertests gebildet: (1) *Mosaik-Test* (MT): Der Testleiter baut mit einfarbigen (bei den schwierigeren Aufgaben auch mit zweifarbigen) Würfeln ein Muster vor, welches das Kind nachbauen soll. Nur bei den schwierigsten Aufgaben dienen Abbildungen als Vorlage. (2) *Figuren legen* (FL): Das Kind soll Puzzleteile zusammenlegen. Bei den Items 6 bis 14 werden Bonuspunkte für kurze Lösungszeiten vergeben. Die zu bildende Form wird vom Untersucher zuvor genannt. Verbal- und Handlungsteil bilden zusammen die Gesamtskala.

Als zusätzlicher Untertest fungiert der *Aktive Wortschatz* (AK), der durch das Benennen von Abbildungen (Lebewesen und Objekte) geprüft wird. Zusammen mit PW kann daraus die „Allgemeine Sprachskala“ (AS) gebildet werden, die nach Auffassung der Autoren das Sprachvermögen des Kindes beschreiben soll. Da jedoch grammatische, semantische und kommunikative Aspekte nicht berücksichtigt werden, entspricht die diagnostische Relevanz eher dem eines Wortschatztests (s. Renner u. Irlich, 2009).

Tabelle 1: Aufbau und Reliabilität des HAWIVA-III

	Itemanzahl	Testposition	Reliabilität ^{a,b}	Mittlere Reliabilität ^b
Altersbereich 2;6 bis 3;11				
Gesamt-IQ			0,89-0,94	0,93
<i>Verbalteil</i>			0,90-0,93	0,91
Passiver Wortschatz (PW) ^f	31	1	0,82-0,88	0,86
Allgemeines Wissen (AW)	28	3	0,82-0,86	0,84
<i>Handlungsteil</i>			0,74-0,92	0,87
Mosaik-Test (MT)	23	2	0,87-0,89	0,88
Figuren legen (FL)	14	4	0,48-0,86	0,77
<i>Allgemeine Sprachskala (AS)</i>			0,89-0,93	0,92
Aktiver Wortschatz (AK) ^d	26	5	0,85-0,91	0,87
Altersbereich 4;0 bis 6;11				
Gesamt-IQ			0,94-0,95	0,95
<i>Verbalteil</i>			0,91-0,94	0,93
Allgemeines Wissen (AW)	28	2	0,77-0,86	0,81
Begriffe erklären (BEL)	14	4	0,84-0,89	0,86
Begriffe erkennen (BEN)	15	7	0,79-0,89	0,86
Allgemeines Verständnis (AV) ^d	19	8	0,87-0,91	0,88
Gemeinsamkeiten finden (GF) ^d	19	11	0,87-0,91	0,90
<i>Handlungsteil</i>			0,87-0,92	0,90
Mosaik-Test (MT)	23	1	0,79-0,91	0,86
Matrizen-Test (MZ)	17	3	0,71-0,86	0,80
Klassen bilden (KB)	17	5	0,78-0,84	0,82
Bilder ergänzen (BE) ^d	31	10	0,69-0,93	0,86
Figuren legen (FL) ^d	14	13	0,53-0,85	0,73
<i>Verarbeitungsgeschwindigkeit^e</i>			0,76-0,90	0,84
Kodieren (KO)	59	8	0,76-0,81 ^c	0,79
Symbol-Suche (SS) ^d	50	6	0,56-0,87 ^c	0,75
<i>Allgemeine Sprachskala</i>			0,82-0,89	0,88
Passiver Wortschatz (PW) ^d	31	12	0,76-0,85	0,79
Aktiver Wortschatz (AK) ^d	26	14	0,75-0,85	0,81

Anmerkungen: ^{a)} Minimale und maximale Reliabilitätsschätzung in den Altersgruppen ^{b)} Wenn nicht anders angegeben Testhalbierungsreliabilität nach Guttman ^{c)} Retestreliabilität ^{d)} Optionaler Untertest ^{e)} zugehörige Untertests siehe Darstellung im Text ^{f)} Bildet zusammen mit AW die Allgemeine Sprachskala.

Bei den 4;0- bis 6;11-jährigen Kindern bilden 3 Aufgabenarten den Verbalteil. Dabei handelt es sich (1) um den oben beschriebenen Untertest AW; (2) *Begriffe erklären* (BEL): Vom Untersucher genannte Begriffe sollen von dem Kind verbal definiert werden. Die Qualität der Antwort wird anhand vorgegebener Kriterien mit 0, 1 oder 2 Punkten bewertet. Es gibt Vorgaben, wann der Untersucher nachzufragen hat. Dieser Untertest entspricht dem Wortschatztest im HAWIK-IV (Petermann u. Petermann,

2007); (3) *Begriffe erkennen* (BEN): Das Kind soll Begriffe anhand vorgelesener Fragen, ähnlich dem Untertest „Rätsel“ der K-ABC (Melchers u. Preuß, 1991) erraten.

Auch der Handlungsteil umfasst bei den älteren Kindern 3 Untertests: (1) MT (Beschreibung siehe oben); (2) *Matrizen-Test* (MZ): Konkrete Objekte oder geometrische Figuren sind in einer Matrice aus 2x2 oder 3x3 Objekten angeordnet. Ein fehlendes Element soll von dem Kind nach logischen Gesichtspunkten aus einer Reihe von 4 oder 5 Antwortalternativen durch Zeigen ausgewählt werden; (3) *Klassen bilden* (KB): Das Kind bekommt 2 Reihen mit Objektabbildungen gezeigt und soll aus jeder Reihe ein Objekt auswählen. Diese beiden sollen zu einer gemeinsamen Merkmals- oder Eigenschaftsklasse gehören, die das Kind selbst herausfinden muss.

Der Untertest *Kodieren* (KO) dient der Erfassung der Verarbeitungsgeschwindigkeit. Er gilt ebenfalls als Kerntest, ohne dem Verbal- oder dem Handlungsteil zugeordnet zu sein. Während 120 Sekunden soll das Kind mit einem Bleistift in geometrische Formen bestimmte Striche eintragen (entspricht dem Zahlen-Symbol-Tests im HAWIK-IV, Form A). Fakultativ kann durch die Untertests PW und AK auch bei den 4;0- bis 6;11-Jährigen die „Allgemeine Sprachskala“ als Kennwert der sprachlichen Fähigkeiten gebildet werden.

Der HAWIVA-III enthält 5 weitere Zusatztests für die älteren Kinder: (1) *Symbol-Suche* (SS): Das Kind soll im Arbeitsheft ankreuzen, ob ein bestimmtes Symbol sich in einer Reihe von 5 weiteren Symbolen wiederholt oder nicht. Es wird die Zahl der richtigen Antworten während 120 Sekunden Bearbeitungszeit ermittelt. Aus KO und SS kann ein Index für die Verarbeitungsgeschwindigkeit gebildet werden. (2) *Allgemeines Verständnis* (AV) enthält Fragen zum Verstehen von Alltagsgegebenheiten, z. B. warum man sich die Hände waschen oder „Danke“ sagen soll. Dieser Untertest kann einzelne Untertests aus dem Verbalteil ersetzen. (3) *Bilder ergänzen* (BE): Das Kind soll erkennen, welches wichtige Detail auf den vorgelegten Abbildungen fehlt. Der Untertest kann als mögliche Alternativaufgabe im Handlungsteil herangezogen werden. (4) *Gemeinsamkeiten finden* (GF): Die logische Gemeinsamkeit von 2 vorgelesenen Objektwörtern ist anzugeben. Auch diese Aufgabenart kann einen Untertest des Verbalteils ersetzen. (5) *Figuren legen* (FL), was als Kernuntertest bei den 2;6 bis 3;11-Jährigen zum Einsatz kommt, steht als optionaler Untertest auch für die ältere Probandengruppe zur Verfügung.

Testmaterial: Das Manual ist weitgehend übersichtlich gestaltet und enthält genaue Hinweise zu Darbietung, erlaubten Hilfen und zur Protokollierung. Die Durchführung wird für beide Altersgruppen getrennt beschrieben. Bei Untertests, die bei beiden Gruppen zum Einsatz kommen können, muss im Manual teilweise hin- und hergeblättert werden. In den Stimulusheften gibt es zwar Trennblätter, man vermisst aber Daumenregister für das rasche Auffinden der Untertestanfänge. Die Protokollhefte sind zweifarbig gestaltet, enthalten zahlreiche nützliche Durchführungshinweise und erleichtern somit die Testdurchführung. Das Aufgabenmaterial ist ansprechend und kindgerecht und lässt sich weitgehend gut handhaben.

Durchführung: Wie bei den meisten Intelligenztests ist eine Sitzposition vis à vis des Kindes vorgeschrieben. Bei den Untertests KO und SS ist die Bearbeitungszeit be-

grenzt und wird mittels Stoppuhr gemessen (Speed-Komponente). Bei dem Untertest FL gibt es bei einigen Aufgaben Bonuspunkte für rasche Bearbeitung. Im Verbalteil ist die Antwortzeit pro Aufgabe auf 30 Sekunden begrenzt. Teilweise gibt es vorgeschaltete Übungsaufgaben, anhand deren dem Kind die Anforderung vermittelt werden soll. Nicht alle davon sind glücklich gewählt. So ist die Formulierung bei Übungsstimulus 2 im Untertest BEN „Das ist ein Wort, das man sagt, wenn man geht“ unklar, da viele Kinder nicht erkennen, dass hier nach einer Verabschiedungsformel gefragt wird, die im Deutschen mit „Auf Wiedersehen“ auch streng genommen nicht aus nur einem Wort besteht. Außerdem sind die Instruktionen teilweise nicht altersgemäß, da Begriffe wie „Symbol“ (Untertest SS) oder „Querlinie“ (Untertest KO) vielen Kindern im Kindergartenalter nicht bekannt sind. Die Einstiegsaufgaben sind vom Schwierigkeitsgrad her zumeist angemessen. Das Abbruchkriterium von 5 falschen Antworten in Folge bei den Verbaltests zieht die Durchführung vor allen Dingen im Untertest BEL oft sehr in die Länge und wirkt sich dadurch nicht selten negativ auf die Motivation der Probanden aus. Die Durchlaufsdauer wird von den Autoren mit 20-60 Minuten bei den jüngeren und mit 30-120 Minuten bei den älteren Kindern angegeben. Die Maximalzahlen beziehen sich offensichtlich auf die Durchführung sämtlicher Untertests. Nach eigener Erfahrung sind für die Durchführung der Kernunters tests im Normalfall maximal 40 bzw. 60 Minuten zu veranschlagen. Eine sorgfältige Einarbeitung ist beim HAWIVA-III, wie bei anderen komplexen Intelligenztests auch, erforderlich.

Auswertung: Die Rohwerte werden anhand übersichtlicher Tabellen in Wertpunkte umgewandelt (Wertpunktskala $M = 10$, $SD = 3$). Die Normen sind in 3- bzw. 4-Monatsstufen unterteilt. Es gibt Angaben zur Wertpunktsommenschätzung bei fehlenden Untertests sowie Angaben zur Testalteräquivalenz. Für die Testskalen werden Standardwerte gebildet ($M = 100$, $SD = 15$). Signifikante Differenzen zwischen Subskalenwerten sind getrennt nach der Richtung der Unterschiede und für 5 verschiedene Intelligenzniveaus angegeben. Außerdem finden sich Hinweise zu kritischen Untertestdifferenzen.

Ein *computergestütztes Auswertungsprogramm* kann zusätzlich erworben werden (Lizenz jeweils für einen PC-Arbeitsplatz). Es erfordert Windows 2000 oder ein neueres Windows-Betriebssystem und lässt sich mühelos unter Windows XP und Vista installieren. Die Handhabung ist einfach, Auswertung und Darstellung der Ergebnisse werden erleichtert, die Auswertungsobjektivität wird erhöht und es besteht eine Zeitersparnis gegenüber der manuellen Bearbeitung, die mit der Differenziertheit der Datenanalyse zunimmt. Eingegeben werden Probandendaten und Rohwerte. Ablesbar sind u. a. Skalen- und Standardwerte, Prozentränge, Vertrauensintervalle (90 % und 95 %), die Signifikanz von Subskalendifferenzen sowie Altersäquivalenzwerte. Stichprobenartige Überprüfungen ergaben keine Abweichungen bei der Wertpunktermittlung von den Normtabellen des Manuals. Die Darstellung erfolgt in tabellarischer Form und in Diagrammen.

Das Auswertungsprogramm weist einige Schwachstellen auf: Die Ersetzungsregeln bei der Bildung der Subskalenwerte werden nicht zuverlässig berücksichtigt. So verändern sich die Skalenwerte für Verbal- und Handlungsskala, wenn zu den drei Kernunters tests ein optionaler Untertest hinzukommt, was zu erheblichen Auswer-

tungsfehlern führen kann. Optionale Untertests sollten daher nur dann eingegeben werden, wenn sie auf jeden Fall einen Kernuntertest ersetzen sollen. Bei nachträglicher Änderung der Probandendaten gehen bereits eingegebene Rohwerte verloren. Die Auswertung ist auf 4 Seiten verteilt, wobei 2 Blätter trotz verschiedener Bezeichnung identische Ausdrücke liefern. Es fehlt die Altersangabe, die Nennung der Rohwerte und die graphische Darstellung der Vertrauensintervalle. Beim Vergleich von 2 Datensätzen finden die Rohwerte ebenfalls keine Beachtung, sodass intraindividuelle Entwicklungsfortschritte bei Messwiederholung nicht unmittelbar sichtbar gemacht werden können. Datensätze werden in keinem gängigen Format ausgegeben, was den Import der Daten, die auf einem anderen PC angefallen sind, erschwert und ein erhebliches Hindernis für die Nutzung zu Analyse- und Forschungszwecken darstellt.

Testanalyse und Normierung

Die *Durchführungsobjektivität* ist durch genaue Vorgaben im Manual gesichert. Zweifel an der *Auswertungsobjektivität* ist bei den Untertests BEL, AV und GF angebracht, da einige im Manual vorgegebene Bewertungen unklar oder nicht nachzuvollziehen sind. Wenn das Kind z. B. im Untertest BEL auf die Frage „Was bedeutet Schaukeln“ die Antwort „Schaukeln“ gibt, bekommt es 2 Rohwertpunkte, obwohl Kinder bei Nichtwissen häufig den Stimulus einfach wiederholen. Im Untertest Allgemeines Verständnis werden treffendere Antworten bisweilen schlechter bewertet als ungenaue. Die Interpretationsobjektivität ergibt sich aus der eindeutigen Zuordnung der ermittelten Werte zu den Normdaten.

In einer Vorstudie wurde eine Übersetzung des WPPSI-III mit 479 Kindern erprobt. Dabei erwiesen sich einige Items als zu einfach und wurden daraufhin eliminiert. Die *Normierung* des HAWIVA-III erfolgte dann in den Jahren 2004 und 2005 an 960 Kindern in Deutschland und 362 Kindern in der deutschsprachigen Schweiz. Es wurde versucht, möglichst komplette Kindergartengruppen zu erfassen, was nach Meinung der Autoren Selektionseffekte vermeiden sollte. Dennoch überwog z. B. die Zahl der Jungen und derjenigen Kinder, die aus Mittelschichtfamilien stammen. Kinder, die einen Sprachheilkindergarten besuchten, wurden ebenso von der Normierung ausgeschlossen wie hochbegabte Kinder und solche, die bereits die Schule besuchten. Kinder, bei denen eine allgemeine Entwicklungsverzögerung oder Intelligenzminderung bekannt war, sind nach Auskunft der Testautoren in der Normierungsstichprobe ebenfalls nicht enthalten. Zur Testdurchführung wurden studentische Hilfskräfte herangezogen.

Für die *Berechnung der Normwerte* wurde die Normierungsstichprobe in Halbjahresschritte unterteilt. Die Stichprobengrößen schwanken erheblich von $N = 27$ bei den 2;6-2;11-jährigen bis zu $N = 242$ bei den 6;0-6;5-jährigen Kindern. Da die Rohwerte sich vielfach nicht normal verteilen, erfolgte die Wertpunktzuordnung aufgrund flächentransformierter Werte. Die Halbjahresgruppen wurden dann unter der Annahme kontinuierlicher Leistungszuwächse durch Interpolation in Normierungsgruppen à 3 Monate (bei den Kindern von 6;0-6;11 sind es 4 Monate) unter-

teilt. Die Normierung des MT beruht wegen nachträglicher Änderungen am Testmaterial auf lediglich 384 untersuchten Kindern aller Altersstufen (Gruppenstärken der Halbjahresstufen lediglich 16-65 Kinder).

Bei den Kindern unter 3 Jahren ergeben sich *Bodeneffekte* vor allen Dingen für den Handlungsteil. *Deckeneffekte* bestehen insbesondere bei Kindern ab einem Alter von 6;0 Jahren. Bei dem Untertest BEN zeichnet sich bereits bei den 4;6-Jährigen ein Deckeneffekt ab. Im Untertest MT erhalten die 6;8-6;11-Jährigen 2 Wertpunkte, egal, ob sie 0 oder 25 Rohwertpunkte erzielen, was mehr als 50% der maximal erreichbaren Rohwertpunkte bedeutet. Eine Erklärung für diese ungünstige Wertpunktverteilung wird nicht geliefert. Die Itemgradienten sind ansonsten zumeist gleichmäßig abgestuft. Für die Untertests MZ, KB und BEN ergeben sich im Altersbereich 5-6 Jahre so gut wie keine erkennbaren Mittelwertsanstiege. Angaben zu Itemschwierigkeiten fehlen im Manual.

Die *Reliabilitätsangaben* beruhen überwiegend auf der Testhalbierungsmethode. Bei den Untertests KO und SS wurde die Retestreliabilität ermittelt. Die Reliabilität des Gesamttests ist mit .89-.95 für die verschiedenen Altersgruppen als gut bis sehr gut zu bezeichnen. Auch die Subskalen sowie die Untertests des Verbalteils weisen zufriedenstellende bis gute Reliabilitäten auf (s. Tab. 1). Im Handlungsteil verfügt lediglich MT durchgängig bei allen Altersgruppen über eine gute Reliabilität. Die Reliabilität der Untertests FL und SS fällt in mehreren Altersgruppen unzureichend aus (FL .48 bei den 2;6-2;11-Jährigen und .53-.54 bei den 6-Jährigen, SS .56 bei den Kindern unter 5;6 Jahren). Eine Reteststudie, die mit z. T. sehr kleinen Stichproben von $N = 13$ bis 36 (je nach Alter und Untertest, getrennt für die 2;6-3;11-Jährigen und die 4;0-6;11-Jährigen) und ohne den MT bei den älteren Kindern durchgeführt wurde, ergab bei einem Zeitabstand von 1 bis 5 Monaten (mittlerer Wert $2\frac{1}{2}$ Monate) recht stabile Werte (gemittelter korrigierter Retestreliabilitätskoeffizient für den gesamten Altersbereich über alle Untertests .78, über die Skalen .85 und für den Gesamttest .89). Im Mittel erzielten die 2;6-3;11-Jährigen bei der Testwiederholung ca. 1 Wertpunkt bzw. 5,5 IQ-Punkte mehr als bei der Ersttestung. Bei den älteren Kindern wurden nur geringe und von der Richtung her uneinheitliche Veränderungen bei der Testwiederholung gefunden. Insgesamt liegen die Reliabilitätskoeffizienten beim HAWIVA-III etwas niedriger als bei der WPPSI-III.

Der HAWIVA-III enthält Aufgabenarten, die im Bereich der Intelligenzdiagnostik gut eingeführt sind, einen breiten Bereich kognitiver Fähigkeiten abdecken und die daher die *inhaltliche Validität* des Verfahrens begründen können. Ein Abgleich mit neueren Intelligenzkonzeptionen, z. B. der CHC-Theorie (McGrew, 2005), wird aber nicht hergestellt.

Zur Überprüfung der *Konstruktvalidität* wurden Interkorrelationen der Untertests und der Skalen des HAWIVA-III ermittelt. Die Untertests des Verbalteils und der Verarbeitungsgeschwindigkeit weisen erwartungsgemäß jeweils untereinander die höchsten Interkorrelationen auf. Dies trifft für den Handlungsteil nicht durchgängig in allen Altersgruppen zu. Insbesondere die optionalen Untertests FL und BE fallen hier aus dem Rahmen. Aber auch der Untertest KB korreliert bei den 4-6-Jährigen deutlich höher mit einigen Untertests des Verbalteils als mit denen des Handlungsteils, was für einen ernst zu

nehmenden Einfluss sprachlicher Kompetenzen auf die Ergebnisse in diesem Untertest spricht. In den exploratorischen Faktorenanalysen, die ohne Berücksichtigung des Mosaiktests durchgeführt wurden, bilden bei den 4-6-Jährigen zwar die Untertests MZ und KB einen gemeinsamen Faktor, doch gelingt es nicht, die Faktorenstruktur zu replizieren, wenn auch die optionalen Untertests einbezogen werden. Der Aussagewert der Faktorenanalyse ist allerdings ohnehin fraglich, da mit MT einer der 3 Kernuntertests von der Analyse ausgeschlossen blieb. Auch die „Allgemeine Sprachskala“ bildet zumindest in den oberen Altersgruppen keinen eigenständigen Faktor. Aufgrund der Interkorrelationen formulieren die Autoren Regeln, nach denen Kernuntertests durch optionale Tests ersetzt werden können. Inhaltliche Gesichtspunkte spielen dabei offenbar keine Rolle.

Zur Überprüfung der *Kriteriumsvalidität* des HAWIVA-III wurden Zusammenhänge zwischen den Untertests des Verbalteils mit anderen Intelligenzmaßen ermittelt. Bei einer Stichprobe von 48 Kindern ergab sich ein niedriger Zusammenhang von .16-.38 (Spearman Rangkorrelation) mit den Untertests 3, 4 und 5 des CFT 1 (Cattell, Weiß, Osterland, 1997) und von .73-.80 mit den Untertests der K-ABC-Fertigkeitenskala (Melchers u. Preuß, 1991). Die Autoren schließen daraus, dass der Verbalteil als Maß für kristalline Intelligenz und nicht für fluide Intelligenz angesehen werden kann. Angaben zur Kriteriumsvalidität des Handlungsteils und der Gesamtskala liegen nicht vor. Der Vergleich der HAWIVA-Ergebnisse mit dem Erzieherurteil über den globalen Entwicklungsstand der Kinder der Normstichprobe liefert für die Testskalen Rangkorrelationen von .18-.48. Für die Risikopunkte im Bielefelder Screening zur Früherkennung von Leserechtschreibschwierigkeiten (BISC; Jansen et al., 1999) wurden für 108 Kinder im Alter von 5;6-6;11 Jahren keine signifikanten Zusammenhänge mit dem HAWIVA-III Verbalteil gefunden. Der Strength and Difficulties Questionnaire (SDQ; Woerner et al., 2002) zeigt bei 70 Kindern im Alter von 4;0 bis 6;11 nur vereinzelt bedeutsame Korrelationen mit dem HAWIVA-III. Der Gesamtproblemwert steht in negativem Zusammenhang mit der Verarbeitungsgeschwindigkeit und der Summenwert für Verhaltensprobleme korreliert mit -.28 mit der Gesamtskala, was von den Autoren als Hinweis auf einen Zusammenhang zwischen Verhaltens- und Leistungsproblemen gewertet wird. Es werden keine Studien mit klinischen Stichproben berichtet. Stattdessen verweisen die Autoren auf die entsprechenden Validitätsstudien des WPPSI-III.

Bewertung

Verglichen mit dem HAWIVA von 1975 stellt der HAWIVA-III einen deutlichen Fortschritt hinsichtlich der Aufgabenzusammenstellung, der standardisierten Testdurchführung und der Normierung dar. Er ist ansprechend gestaltet, weitgehend benutzerfreundlich in der Anwendung und lässt sich mit Kindern des entsprechenden Alters im Allgemeinen gut durchführen. Was ihn gegenüber den Konkurrenzprodukten SON-R 2½-7 (Tellegen, Laros, Petermann, 2007) und K-ABC (Melchers u. Preuß, 1991) hervorhebt, ist die umfassendere Messung der verbalen Fähigkeiten und Fertigkeiten.

Allerdings gibt es in der Zusammenstellung der Untertests auch Schwachpunkte. So vermisst man Aufgaben zur Merkfähigkeit und zum rechnerischen Denken. In Bezug auf den Untertest KB gibt es begründete Zweifel, ob dieser Untertest wegen seiner starken verbalen Komponente nicht eher dem Verbalteil zugeordnet werden müsste, womit verbale Aspekte in der Zusammenstellung der Untertests aber überrepräsentiert wären. Die Aufnahme einer Subskala zur Verarbeitungsgeschwindigkeit ist zu begrüßen, auch wenn die Ergebnisse in den beiden Untertests durch graphomotorische Geschicklichkeit in beträchtlichem Maße konfundiert sein können. Die zahlreichen Untertests bedeuten für den Untersucher eine Erweiterung der diagnostischen Möglichkeiten. Bezüglich der Subskalensubstitution von Kern- durch optionale Tests fehlt aber eine inhaltliche Begründung.

Die Testanweisungen ermöglichen eine weitgehend standardisierte Durchführung, die empfohlene Sitzposition gegenüber dem Kind ist aber insbesondere bei unkooperativen Kindern nicht immer günstig, wenn diese ein höheres Maß aktiven Eingreifens seitens des Untersuchers erfordern. Die Begrenzung der Antwortzeit bei sprachlichen Aufgaben auf 30 Sekunden ist als Richtwert akzeptabel, gerade jüngere Kinder benötigen aber vor allen Dingen beim Untertest BEL bisweilen mehr Zeit.

Verbal-IQ und Subskala „Verarbeitungsgeschwindigkeit“ weisen zufriedenstellende bis sehr gute Werte zur Testhalbierungs- und Retestreliabilität sowie zur Konstruktvalidität auf. Dafür bestehen im Verbalteil erhebliche Schwächen in der Auswertungsobjektivität infolge unklarer, vereinzelt sogar unsinniger Bewertungskategorien. Im Handlungsteil bestehen erhebliche Schwächen vornehmlich in der Konstruktvalidität. Bei Kindern ab 4;0 Jahren ist die faktorielle Validität des Untertests MT nicht belegt und die beiden in der Faktorenanalyse verbliebenen Untertests MZ und KB bilden nur dann einen gemeinsamen Faktor in allen Altersgruppen, wenn die Analyse auf die Kernuntertests beschränkt bleibt. Die Kriteriumsvalidität des gesamten Verfahrens ist bisher noch unzureichend belegt. Die Normierungsstichprobe ist für die Kinder unter 3½ Jahren sehr gering, was zu besonders vorsichtiger Bewertung der entsprechenden Ergebnisse Anlass geben sollte. Dies trifft auch für den Untertest MT in allen Altersgruppen zu. Da Kinder mit bereits bekannten Entwicklungsproblemen von der Normierungsstichprobe ausgeschlossen blieben, ist die Brauchbarkeit der Normen hinsichtlich der Untersuchung von Kindern im unterdurchschnittlichen Intelligenzspektrum, wo Kliniker eines der Hauptanwendungsgebiete des Verfahrens sehen dürften, noch nicht ausreichend nachgewiesen.

Eine Stärke des Verfahrens liegt in der Erfassung der Verbalintelligenz. Ist eine Durchführung des „Verbalteils“ nicht möglich, was bei Kindern mit Intelligenzminderung bisweilen auch bei guter Kooperation in den anderen Testteilen der Fall ist, kann, um überhaupt zu einem standardisierten sprachlichen Eindruck zu gelangen, auf die „Allgemeine Sprachskala“ zurückgegriffen werden, auch wenn diese am ehesten als Wortschatzskala zu charakterisieren ist und daher komplexeres sprachliches Denken nicht abbildet. Gegenüber dem Handlungsteil mit seinen teilweise unklaren Testgüteeigenschaften sollte anderen Verfahren wie dem SON-R 2½-7 oder der K-ABC weiterhin der Vorrang gegeben werden. Wegen vorhandener Boden- und Deckeneffekte sind die Anwendungsmöglichkeiten des HAWIVA-III in den Extrembereichen einge-

schränkt. Bei Kindern über 6 Jahren sollte der HAWIVA-III nicht zur Untersuchung im Bereich hoher Begabung eingesetzt werden. Darauf weisen die Autoren auch selbst hin und empfehlen für diesen Fall die Anwendung des HAWIK-IV.

Aufgrund der unübersehbaren Schwachstellen in der Testgüte ist nicht nachzuvollziehen, warum der HAWIVA-III mit dem Attribut „Erfüllung höchster Qualitätskriterien“ im Katalog der Testzentrale (Hogrefe-Verlag, 2008, S. 46) beworben wird, ein Superlativ, der anderen, mit dem HAWIVA-III vergleichbaren Verfahren, an gleicher Stelle vorenthalten bleibt. Es scheint dringend geboten, dass weitere empirische Ergebnisse vorgelegt werden, die die Brauchbarkeit der deutschen Version des WPPSI-III belegen.

Das Auswertungsprogramm erhöht Auswertungsobjektivität und Testökonomie, ist aber dringend verbesserungsbedürftig.

Literatur

- Cattell, R. B., Weiß, R. H., Osterland, J. (1997). Grundintelligenztest Skala 1 (CFT 1) (5. revidierte Auflage). Göttingen: Hogrefe.
- Eggert, D. (Hrsg.) (1975). Hannover-Wechsler-Intelligenztest für das Vorschulalter. Deutsche Bearbeitung der Wechsler Preschool and Primary Scale of Intelligence. Bern: Huber.
- Hogrefe-Verlag (2008). Testkatalog 2008/2009. Göttingen Hogrefe.
- Jansen, H., Mannhaupt, G., Marx, H., Skowronek, H. (1999). Bielefelder Screening zur Früherkennung von Lese-Rechtschreibschwierigkeiten (BISC). Göttingen: Hogrefe.
- Lichtenberger, E. O., Kaufman, A. S. (2004). Essentials of WPPSI-III Assessment. Hoboken, NJ: John Wiley.
- McGrew, K. S. (2005). The Cattell-Horn-Carroll Theory of Cognitive Abilities. In D. P. Flanagan, P. L. Harrison (Hrsg.), Contemporary intellectual assessment (2. Aufl., S. 136-181). New York: Guilford Press.
- Melchers, P., Preuß, U. (1991). Kaufman-Assessment Battery for Children. Deutschsprachige Fassung. Amsterdam: Swets & Zeitlinger.
- Petermann, F., Petermann, U. (Hrsg.) (2007). HAWIK-IV. Hamburg-Wechsler-Intelligenztest für Kinder. Bern: Huber.
- 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.
- Tellegen, P. J., Laros, J. A., Petermann, F. (2007). SON-R 2½-7. Snijders-Oomen Non-verbaler Intelligenztest. Göttingen: Hogrefe.
- Wechsler, D. (2002). Wechsler Preschool and Primary Scale of Intelligence – III (WPPSI-III). San Antonio, TX: The Psychological Corporation.
- Woerner, W., Becker, A., Friedrich, C., Klasen, H., Goodman, R., Rothenberger, A. (2002). Normierung und Evaluation der deutschen Elternversion des Strengths and Difficulties Questionnaire (SDQ): Ergebnisse einer repräsentativen Felderhebung. Zeitschrift für Kinder- und Jugendpsychiatrie und Psychotherapie, 30, 105-112. Fragebogen siehe <http://www.sdqinfo.com/questionnaires/german/s1.pdf> (Zugriff am 8.03.2009).

Dieter Irblich, Auel