



WOGÉ

Testbesprechung zum Adaptiven Intelligenz Diagnostikum 2

Gerolf Renner, Hermann-Josef Baur
& Barbara Lischke

1. Testart

Beim AID 2¹ handelt es sich um einen Intelligenztest für die Einzeluntersuchung von Kindern und Jugendlichen, der den 1985 erschienenen AID ablöst.

2. Grundkonzept und theoretischer Hintergrund

Die Entwicklung des AID 2 wurzelt in einer Auseinandersetzung Kubingers (1983; s.a. Kubinger, 1998) mit dem HAWIK (Hamburg-Wechsler Intelligenztest für Kinder; Hardesty & Priester, 1956). In einer vor allem von teststatistischen Erwägungen geprägten Kritik kam Kubinger zu der Folgerung, dass der Ansatz des HAWIK bei allen Mängeln eine Weiterentwicklung verdiene und verwirklichte im AID (Kubinger & Wurst, 1985, 1988) bei weitgehender inhaltlicher Anlehnung an Wechslers Testkonzept modifizierte Aufgabenstellungen und Verrechnungsstrategien, die adaptive Testvorgabe und in statistischer Hinsicht den konsequenten Bezug auf die probabilistische Testtheorie. In der Tradition Wechslers steht das Bemühen, während einer Testuntersuchung ein breites Spektrum unterschiedlicher kognitiver Fähigkeiten abzudecken und dabei Beobachtungsmöglichkeiten über Verhalten und Lösungsstrategien des Kindes zu schaffen.

Intelligenz definieren Kubinger & Wurst (2000) als »Bündel aller kognitiven Voraussetzungen, die notwendig sind, um Wissen zu erwerben und Handlungskompetenzen zu entwickeln« (Manual, S. 30). Sie bekennen sich zu einem »pragmatischen Ansatz« der Intelligenzmessung und verzichten ganz auf eine Auseinandersetzung mit dem aktuellen Forschungsstand in der Intelligenzdiagnostik (zum Überblick s. z.B. Flanagan, Genshaft & Harrison, 1997).

Ziel des AID 2 sei es, »ziemlich viele (komplexe und basale) Fähigkeiten, die für ‚intelligentes‘ Verhalten verantwortlich scheinen, zu erfassen« (Manual, S. 14). Als Anwendungsbereiche nennt das Manual eine »förderungsorientierte Diagnostik« bei Lern-, Konzentrations- und Verhaltensstörungen und den Einsatz als Screening-Verfahren zur Erfassung von Teilleistungsstörungen.

Der Schwerpunkt der Testinterpretation liegt dabei auf den Leistungen in den einzelnen Subtests, ein Gesamt-IQ wird nicht bestimmt. »Ein Intelligenzquotient, definiert als der Durchschnitt aller geprüften Fähigkeiten, ist intelligenztheoretisch nicht mehr vertretbar und förderungsorientiert nicht zielführend« (Manual, S. 14). Anstelle eines dem IQ-Konzept impliziten Kompensationsmodells (eine Schwäche in einem Leistungsbereich kann durch Stärken in anderen Leistungsbereichen ausgeglichen werden) propagieren die Autoren ein »Defizitmodell« der Intelligenz: »Im Sinne des Gleichnisses, ‚jede Kette ist nur so stark wie ihr schwächstes Glied‘, interessiert dabei die minimale aller erbrachten Testleistungen« (Manual, S. 31). Als globales Intelligenzmaß wird die sog. Intelligenzquantität bestimmt, operationalisiert über die niedrigste Untertestleistung des Probanden. Wie die Autoren sehr zu Recht bemerken, aber leider nicht mit der gebotenen Konsequenz diskutieren, ist die Sinnhaftigkeit dieses

Maßes »an die Voraussetzung verknüpft, daß allen erfaßten Fähigkeiten dieselbe Bedeutung (in Bezug auf alltägliche berufliche bzw. schulische Leistungsanforderungen) beizumessen ist« (Manual, S. 31).

Im Hinblick auf die Diagnostik spezifischer Leistungsdefizite stellen Kubinger & Wurst (2000, S. 34ff.) – wiederum ohne sich auf aktuelle Forschungsergebnisse zu beziehen – ein Teilleistungsmodell vor, das der differenzialdiagnostischen Interpretation des AID 2 dienen soll. Dabei werden drei Ebenen unterschieden: a) grundlegende Formen der Informationsverarbeitung, nämlich Wahrnehmen (Informationsaufnahme), Merken (Informationsspeicherung) und Verarbeiten/Benutzen (Verarbeiten und in Beziehung setzen von Informationen/intermodale Transfers); b) an der Informationsverarbeitung beteiligte Wahrnehmungsmodalitäten (visuell, akustisch, propriozeptiv), außerdem Sprechen und Motorik; c) die Fähigkeitsdimensionen Gliederungs-, Differenzierungs- und Serialitätsfähigkeit und als besondere Form der visuellen Gliederungsfähigkeit die Raum-Lage-Orientierung. Die Darstellung der entsprechenden Testergebnisse kann in einem Diagramm zur Diagnostik von Teilleistungsstörungen erfolgen (Abb. 1; die Ziffern in den weißen Feldern zeigen an, welcher Subtest der jeweiligen Dimension zugeordnet werden kann), Interpretationsmöglichkeiten werden im Handbuch erläutert und an einem Fallbeispiel demonstriert.

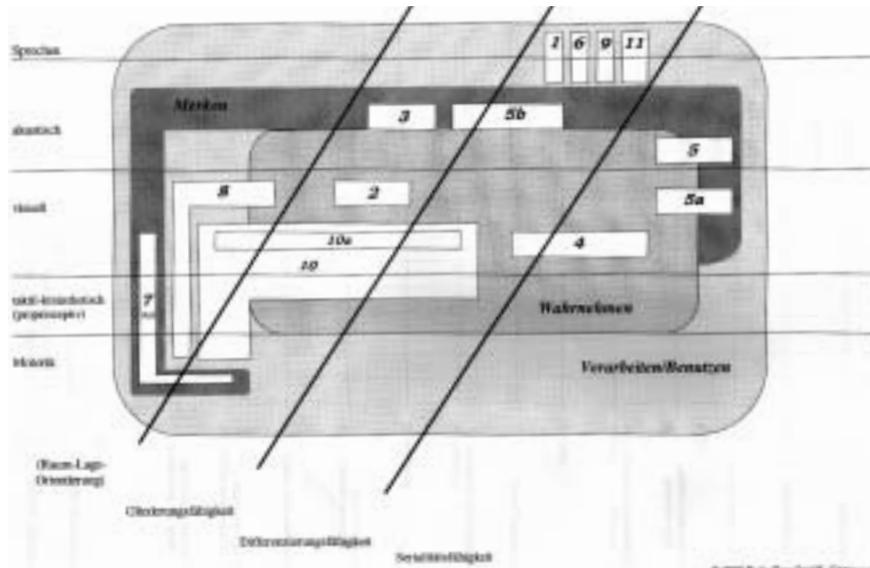


Abbildung 1: Diagramm zur Diagnose von Teilleistungsstörungen mit dem AID 2 (aus Kubinger & Wurst, 2000, S. 37; mit freundlicher Genehmigung der Beltz Test GmbH)

2. Testmaterial

Der Testkoffer enthält das 224-seitige Manual; 10 Protokollbogen mit je einem Arbeits- und Testblatt für den Untertest »Kodieren und Assoziieren« und den Zusatztest »Strukturieren-visumotorisch« und einem Beiblatt zur Beobachtung von Arbeitshaltungen sowie einem Diagramm zur Diagnostik von Teilleistungsstörungen; eine Schachtel mit neun Plastikwürfeln für den Untertest »Analysieren und Synthetisieren-abstrakt«; ein DIN A-6 Spiralheft mit den Vorlagen für den Untertest »Analysieren und Synthetisieren-abstrakt«; ein DIN A-5

Anmerkung:
¹ Im Folgenden wird die ursprüngliche Fassung des Verfahrens stets mit AID, die hier besprochene Überarbeitung mit AID 2 abgekürzt.

Spiralheft mit den Aufgaben für die Untertests »Realitätssicherheit«, »Angewandtes Rechnen« und »Kodieren und Assoziieren«; die Bildkarten des Untertests »Soziale und Sachliche Folgerichtigkeit«; einen Karton mit den Einzelteilen des Untertests »Antizipieren und Kombinieren-figural«; eine Bildkarte für den Untertest »Unmittelbares Reproduzieren-figural/abstrakt«; Auswertungsschablonen für den Untertest »Kodieren und Assoziieren« und den Zusatztest »Strukturieren-visumotorisch«; eine CD-ROM mit dem Computerauswerteprogramm AIDScore; ein ergänzendes »Informationsblatt zur Bestimmung eines globalen Maßes der Intelligenz (IQ)«. Für die Anwendung werden außerdem eine Stoppuhr, Papier und Bleistift benötigt.

3. Testgliederung

Bei vollständiger Durchführung umfasst der AID 2 14 Untertests, von denen drei als Zusatztests bezeichnet werden (Tabelle 1). Der Testbeschreibung stellen wir jeweils die Definition der überprüften Fähigkeitsdimensionen lt. Manual (S. 16-19) voran.

Untertest 1 Alltagswissen »soll die Fähigkeit prüfen, sich Sachkenntnisse über Inhalte anzueignen, die in der heutigen Gesellschaft alltäglich sind«. Fragen aus

verschiedenen Wissensgebieten sind mündlich zu beantworten (z.B. »Womit kannst du riechen?«, »Wer war Vincent van Gogh?«).

Untertest 2 Realitätssicherheit »soll prüfen, inwieweit die Wirklichkeit um Dinge des Alltags verstanden wird bzw. kontrolliert werden kann«. Auf einer Bildvorlage (Schwarz-weiß-Zeichnungen) soll das Kind ein bedeutendes fehlendes Element (z.B. den Henkel einer Teekanne; den Bauchschutzgürtel eines Gewichthebers) erkennen. Die Lösung durch Benennen oder Zeigen muss innerhalb von 15 Sekunden erfolgen.

Untertest 3 Angewandtes Rechnen »soll weitgehend unabhängig von schulischen Rechenfertigkeiten prüfen, inwieweit die Testperson bei der Problemlösung alltäglicher Aufgabenstellungen durch entsprechende Schlußfolgerungen die passenden Rechenoperationen anzuwenden imstande ist«. Textaufgaben (z.B. »Peter hat 3 Tafeln Schokolade. Eine schenkt er dir. Wie viele hat er dann noch?«; »Elisabeth ist 12 Jahre alt, ihr Vater ist 38 Jahre alt. In wieviel Jahren ist ihr Vater genau doppelt so alt wie sie?«) sind innerhalb einer Zeitgrenze von 30-60 Sekunden zu lösen. Kinder ab acht Jahren können bei zwölf, Kinder unter acht Jahren bei drei Aufgabengruppen zusätzlich mitlesen und dürfen Papier und Bleistift als Hilfsmittel benutzen.

Tabelle 1: Aufbau des Adaptiven Intelligenz Diagnostikums AID 2

Untertest	Vorgabe	Gesamtzahl Items	Items bei adaptiver Vorgabe	Kurzformen	Parallelformen	Sprachfreie Instruktion
1. Alltagswissen	Adaptiv	60	15 Items/3 Blöcke	Ja	Ja	Nein
2. Realitätssicherheit	Konventionell zeitbegrenzt	12 1 Übungsittem	---	Nein	Nein	Ja
3. Angewandtes Rechnen	Adaptiv zeitbegrenzt	60	15 Items/3 Blöcke	Ja	Ja	Nein
4. Soziale und Sachliche Folgerichtigkeit	Adaptiv zeitbegrenzt	18	6-7 Items/2 Blöcke, evtl. Zusatzaufgabe	Nein	Ja	Ja
5. Unmittelbares Reproduzieren – numerisch a) vorwärts b) rückwärts	Konventionell Abbruchkriterium	a) 24 b) 24	---	Nein	Nein	Nein
6. Synonyme Finden	Adaptiv	60	15 Items/3 Blöcke	Ja	Ja	Nein
7. Kodieren und Assoziieren a) Kodiermenge b) Assoziationen	Konventionell zeitbegrenzt	a) 104 b) 12	---	Nein	Nein	Ja
8. Antizipieren und Kombinieren – figural	Adaptiv zeitbegrenzt	11	5-7 Items/2-3 Blöcke	Nein	Nein	Ja
9. Funktionen Abstrahieren	Adaptiv	60	15 Items/3 Blöcke	Ja	Ja	Nein
10. Analysieren und Synthetisieren – abstrakt	Adaptiv zeitbegrenzt	18 2 Übungsittem	5-7 Items/2 Blöcke, evtl. Zusatzaufgabe	Nein	Ja	Ja
11. Soziales Erfassen und Sachliches Reflektieren	Adaptiv	60	15 Items/3 Blöcke	Ja	Ja	Nein
5a Unmittelbares Reproduzieren – figural/abstrakt	Konventionell Abbruchkriterium	14	---	Nein	Nein	Ja
5b Merken und Einprägen	Konventionell zeitbegrenzt	9 1 Übungsittem	---	Nein	Nein	Nein
10a Strukturieren-visumotorisch	Konventionell zeitbegrenzt	11 1 Beispielaufgabe	---	Nein	Nein	Ja

Untertest 4 Soziale und Sachliche Folgerichtigkeit soll die Fähigkeit erfassen »die Abfolge sozialen Geschehens bzw. alltäglicher Sachgegebenheiten zu verstehen und zu kontrollieren«. Serien von Bildern, die eine kleine Geschichte ergeben, werden in einer festgelegten falschen Reihenfolge ausgelegt und sollen vom Kind so geordnet werden, dass sich eine sinnvolle Anordnung ergibt (z.B. ein Junge baut einen Schneemann; der Ablauf eines Skisprunges vom Anlauf bis zur Siegerehrung). Für die Lösung steht ein Zeitrahmen von 30 bis 75 Sekunden zur Verfügung.

Untertest 5 Unmittelbares Reproduzieren-numerisch »soll die Kapazität der seriellen Informationsverarbeitung (im verbal-akustischen Bereich) messen«. Das Kind soll Ziffernfolgen, die vom Untersuchungsleiter vorgesprochen werden, unmittelbar in derselben oder in umgekehrter Reihenfolge wiedergeben. Die Ziffernfolgen umfassen zwei bis maximal neun Ziffern. Nach der ersten korrekten Lösung von einer bestimmten Länge wird gleich zur nächstschwierigeren Stufe übergegangen (bei maximal drei Versuchen). Erfasst werden sowohl der Umfang der längsten richtig reproduzierten Zahlenfolge, als auch die Anzahl der insgesamt vorgegebenen Items. Reproduzieren vorwärts und rückwärts werden getrennt ausgewertet.

Untertest 6 Synonyme Finden »soll das elementare Sprachverständnis prüfen, nämlich inwieweit die Testperson die Bedeutung sprachgebundener Begriffe erfaßt, bzw. über einen Wortschatz verfügt, der solche Begriffe alternativ ausdrücken läßt«. Das Kind soll zu einem mündlich dargebotenen Begriff ein anderes bedeutungsgleiches Wort finden (z.B. ‚kräftig‘ – ‚stark‘; ‚autonom‘ – ‚unabhängig; selbständig‘). Inhaltliche Erklärungen oder Oberbegriffe werden nicht als korrekte Lösungen akzeptiert.

Untertest 7 Kodieren und Assoziieren soll »zwei voneinander partiell unabhängige Fähigkeiten« erfassen: »Informationsverarbeitungsschnelligkeit und die Fähigkeit zum inzidentellen Lernen«. Auf einem Arbeitsblatt sind einer Folge von 104 anschaulichen Zeichen (12 einzelne Objekte, z.B. Apfel, Baum, Glocke) zugehörige geometrische Symbole (z.B. Dreieck, waagrechtlicher Strich, Pfeil) zuzuordnen. Das Kind wird aufgefordert, so schnell zu arbeiten wie es kann, ohne dabei Fehler zu machen. Gewertet werden die innerhalb von 120 Sekunden erzielten richtigen Lösungen. Zusätzlich soll anschließend aus dem Gedächtnis jedem der zwölf Objekte das korrekte Symbol zugeordnet werden, so dass neben der Kodiermenge auch die Gedächtnisleistung für die (inzidentell) gespeicherten Assoziationen ausgewertet werden kann.

Untertest 8 Antizipieren und Kombinieren-figural »soll schlußfolgerndes Denken in der Hinsicht prüfen, Teile eines (konkreten) Ganzen erkennen und dieses Ganze gestalten zu können«. Dem Kind werden Einzelteile einer Figur (z.B. Birne, Flugzeug) vorgelegt, die innerhalb einer Zeitgrenze zusammengesetzt werden sollen. Die Bewertung erfolgt bei neun der elf Items dreistufig, zwei Punkte werden für schnelle Lösungen vergeben.

Untertest 9 Funktionen Abstrahieren soll die Fähigkeit erfassen »durch Abstraktion zu einer Begriffsbildung zu

gelangen«. Das Kind wird nach dem Gemeinsamen von zwei Begriffen gefragt (»Was ist das Gemeinsame an Flugzeug und Vogel?«, »Was ist das Gemeinsame an Waage und Tachometer?«).

Untertest 10 Analysieren und Synthetisieren-abstrakt soll die Fähigkeit prüfen »komplexe (abstrakte) Gestalten durch eine geeignete Strukturierung reproduzieren zu können«. Mit Plastikwürfeln, die jeweils eine rote, eine weiße, eine diagonal geteilte rot-weiße, eine senkrecht geteilte rot-weiße und eine Seitenfläche mit drei weißen und einem roten Quadranten aufweisen (eine weitere blaue Würfelseite hat für die Lösung keine Bedeutung), sind innerhalb einer Zeitgrenze von 30-90 Sekunden vorgegebene Muster nachzulegen.

Untertest 11 Soziales Erfassen und Sachliches Reflektieren soll prüfen »inwieweit die Testperson Sachzusammenhänge der ‚gesellschaftlichen‘ Umwelt begreift bzw. inwieweit sie sozialisiert in dem Sinn ist, daß sie über sozial angepaßte Verhaltensweisen und gesellschaftliche Bedingungen Bescheid weiß«. Eine Reihe von Fragen nach sozialen und gesellschaftlichen Zusammenhängen ist zu beantworten (z.B. »Warum werden im Winter Rehe und Hirsche gefüttert?«). Bereits bei Erscheinen des Tests veraltet war das Item »Warum ist eine Flugkarte teurer als eine Bahnkarte für die gleiche Strecke?«

Zusatztest 5a Unmittelbares Reproduzieren-figural/abstrakt »soll die Kapazität der seriellen Informationsverarbeitung (im visumotorischen Bereich) messen«. Auf einer farbig gestalteten Bildtafel mit 49 abstrakten oder anschaulichen Objekten soll das Kind diejenigen Felder in der gleichen Reihenfolge antippen, die zuvor vom Versuchsleiter gezeigt wurden.

Zusatztest 5b Merken und Einprägen soll die Behaltenskapazität erfassen, »wie sie durch eine einmalige Wiederholung der Reizdarbietung erreichbar ist«. Das Kind soll neun Pseudowörter (Kombinationen aus einem Vokal und zwei Konsonanten; z.B. »sif«, »bax«), die der Untersuchungsleiter zweimal in verschiedener Reihenfolge vorspricht, unmittelbar reproduzieren. Gewertet wird die Anzahl richtig wiedergegebener Wörter der zweiten Darbietung.

Zusatztest 10a Strukturieren-visumotorisch »soll die Fähigkeit erfassen, komplexe (abstrakte) Gestalten in elementare Teilkomponenten zerlegen zu können«. Auf einem Arbeitsblatt sollen Muster durch Einzeichnen von Trennlinien in ihre (maximal neun) Einzelteile zerlegt werden. Die Formen der Einzelteile entsprechen den Würfelseiten, wie sie in Untertest 10 verwendet werden, die Muster sind jedoch schwarz-weiß dargestellt. Für die Bearbeitung der elf Aufgaben stehen zwei Minuten zur Verfügung. Dieser Test wird erst ab dem Alter von acht Jahren vorgegeben.

4. Testdurchführung

4.1 *Alter:* Der AID 2 ist für den Altersbereich von 6;0 bis 15;11 Jahren vorgesehen.

4.2 *Formen:* Der AID 2 wird stets im Einzelsetting ohne Begleitpersonen durchgeführt. Für die Untertests 1, 3, 6, 9 und 11 wurden aus dem Itempool jeweils 16 – sich inhaltlich teilweise überschneidende – Parallelformen

GEROLF RENNER, Dr.rer.nat., Studium der Psychologie in Göttingen, Chapel Hill (North Carolina, USA) und Trier. Psychologischer Psychotherapeut, Supervisor BDP, Ausbildung in Verhaltens- und Familientherapie. Arbeitsschwerpunkte: Psychologie in der Sozialpädiatrie, Entwicklungsdiagnostik, Verhaltens- und Familientherapie.

Anschrift:
Wilhelm-Dröschers-
Straße 17,
55765 Birkenfeld

HERMANN-JOSEF BAUR, Diplom-Psychologe, Psychologischer Psychotherapeut, Kinder- und Jugendlicher Psychotherapeut, Dipl. Ehe-Familien-Lebensberater (BAG). Psychotherapeutische Praxis und Co-Leitung des Instituts für integrative Förderung von Kindern und Jugendlichen (IFKJ).

Anschrift:
St. Annenstr. 18,
66606 St. Wendel

BARBARA LISCHKE, Diplom-Pädagogin. Arbeitsschwerpunkte: Entwicklungsdiagnostik als Mitarbeiterin in psychologischer Praxis. Legasthenie- und Dyskalkulie-therapie im Institut für integrative Förderung von Kindern und Jugendlichen (IFKJ), z.Zt. Ausbildung zur Lerntherapeutin.

Anschrift:
Spiemontstraße 28,
66606 St. Wendel

unterschiedlichen Schwierigkeitsgrades zusammengestellt, so dass gezielt eine Form gewählt werden kann, die dem in der Ersttestung gezeigten Fähigkeitsniveau des Kindes entspricht. Bei den Untertests 4 und 10 stehen 4 Parallelformen zur Verfügung. Bei den 12 Kurzformen der Untertests 1, 3, 6, 9 und 11 wird die Testdurchführung nach zwei statt nach drei Aufgabengruppen abgebrochen.

4.3 Handhabung: Die Durchführung erfolgt gemäß den detaillierten Beschreibungen des Manuals in festgelegter Reihenfolge möglichst in einem Durchgang. Die Anweisungen sind wörtlich vorzugeben, am besten auswendig zu lernen. Die differenzierten Durchführungsrichtlinien erfordern eine gründliche Einarbeitung, worauf die Autoren deutlich hinweisen.

Acht der elf Standardtests werden adaptiv im Sinne eines »branched-testing« durchgeführt. Bei den Untertests 1, 3, 6, 9 und 11 werden drei Aufgabenblöcke von je 5 Items vorgegeben. Nach Vorgabe des ersten Aufgabenblockes wird die Testung entsprechend der im Manual festgelegten Verzweigungsregeln fortgesetzt (z.B. wird ein 6-jähriges Kind, das im Untertest Alltagswissen mit Block 3 beginnt, bei 0-1 erzielten Punkten mit Block 2, bei 2-3 Punkten mit Block 10 und bei 4-5 Punkten mit Block 4 fortfahren). Eine weitere Verzweigungsregel führt schließlich zum dritten und letzten Aufgabenblock. Bei den Untertests 4, 8 und 10 findet dasselbe Prinzip in verkürzter Form Anwendung. Zur Umsetzung dieses Vorgehens ist das unmittelbare Bewerten der Antworten des Kindes und die fortlaufende Konsultation des Handbuchs im Hinblick auf die Verzweigungsregel nötig. Als Vorteile des adaptiven Testens nennen die Autoren die bessere Anpassung an das Fähigkeitsniveau des Probanden und somit eine differenziertere Bestimmung der gemessenen Fähigkeit. *Die von Jäger (1987) geäußerte Sorge, dass das adaptive Testen leistungshemmend wirken könnte, hat sich aufgrund eigener Erfahrungen (z.Zt. 66 Testungen an einer klinischen Stichprobe ohne Abbruch oder Verweigerung; Testleiterin B. Lischke) nicht bestätigt.*

Das Testmaterial ist eher wenig ansprechend gestaltet, einige Zeichnungen wirken nicht zeitgemäß. Die Plastikwürfel des Untertests 10 verrutschen leicht und sind nicht präzise gearbeitet, so dass beim Zusammenlegen irrelevante und irritierende Linien sichtbar werden. Der Protokollbogen bietet keine Erinnerungshilfen und sieht für die Protokollierung der Antworten keinen Raum vor.

Sprachfreie Instruktionen werden für die Untertests 2, 4, 7, 8, 10, 5a und 10a beschrieben. Dabei kommen vor allem pantomimische Demonstrationen zum Einsatz. Empirische Untersuchungen zur Qualität der sprachfreien Instruktion liegen noch nicht vor.

4.4 Zeit: Die Durchführungszeit liegt nach Angaben der Autoren zwischen 30 und 75 Minuten für die 11 Standardtests, für die drei Zusatztests werden weitere 10-15 Minuten veranschlagt. Die untere Grenze der Durchführungszeit ist sicherlich nur in seltenen Ausnahmefällen und bei Einsatz der Kurzformen zu erreichen. Empi-

rische Daten über die Bearbeitungsdauer in der Normstichprobe werden nicht angegeben. Nach eigenen Erfahrungen beträgt die Durchführungszeit für den Gesamttest bei relativ geringer Streuung ca. 90 Minuten.

5. Auswertung

5.1 Modus: Die Ermittlung der Testrohwerte ist standardisiert und erfolgt unter Zuhilfenahme von Lösungsbeispielen und -schablonen; die Ergebnisse werden im Protokollbogen eingetragen. Die weitere Auswertung erfolgt entweder anhand der Tabellen des Manuals oder – deutlich schneller und weniger fehleranfällig – mit dem Auswertungsprogramm AIDScore. Bei manueller Auswertung der adaptiv durchgeführten Tests werden die Fähigkeitsparameter für die jeweilige Altersgruppe und die durchgeführte Aufgabenkombination in den Tabellen abgelesen. In einer separaten Tabelle erfolgt anschließend die Transformation der Fähigkeitsparameter in T-Werte. Da nicht alle T-Werte tabelliert sind, muss ggf. interpoliert werden.

Bei Subtest 10a ist neben der Zahl der richtigen Lösungen zu berücksichtigen, welche Aufgaben überhaupt bearbeitet wurden. Von den theoretisch möglichen 2.048 Konstellationen sind im Manual 109 aufgeführt. *Eigene Erfahrungen zeigen, dass bei einem nicht unerheblichen Anteil der Kinder keine passende Zuordnung gefunden werden kann. Von 58 durchgeführten Tests ließen sich 14 (24%) nicht auswerten!*

Die Ergebnisse der elf Standardtests können im Protokollbogen in einer leider sehr eng dargestellten Graphik mit Hilfslinien für ausgewählte T-Werte als Leistungsprofil dargestellt werden. Als weitere Kennwerte werden auf Basis der elf Standardtests die Prozentränge für den Range der Intelligenz (Differenz zwischen maximalem und minimalem T-Wert), die Intelligenzquantität (minimaler T-Wert) und die zweitniedrigste Untertestleistung ermittelt. Für den niedrigsten tabellierten T-Wert von 19 beträgt der Prozentrang der Intelligenzquantität 1:1 (entsprechend einem IQ-Äquivalent von 67); niedrigere Intelligenzwerte können nicht abgebildet werden. Die ursprünglich nicht vorgesehene Berechnung eines Gesamt-IQ kann anhand einer im »Informationsblatt zur Bestimmung eines globalen Maßes der Intelligenz« angegebenen Berechnungsformel manuell erfolgen. *Ob sich der Testanwender für Intelligenzquantität oder Gesamt-IQ als Intelligenzmaß entscheidet, hat durchaus bedeutsame Konsequenzen: In unserer eigenen Stichprobe betrug die mittlere absolute Differenz der beiden Kennwerte 8 T-Punkte, mit einem Maximalwert von 35!*

Das Auswertungsprogramm AIDScore bietet neben der Ermittlung und Ausgabe des Testergebnisses die Möglichkeit der Probandenverwaltung, einschließlich einer Exportfunktion im dbase-Format. Zu beachten ist, dass bei Korrektur einer fehlerhaften Geburts- oder Testdatumsangabe nach Eingabe der Testwerte keine automatische Anpassung der Ergebnisse erfolgt.

5.2 Zeit: Die Ermittlung der Rohwerte läuft weitgehend parallel zur Testung und benötigt anschließend weitere 5-10 Minuten. Die Zeit für die Bestimmung der

Normwerte dürfte beträchtlich variieren, mit einem erheblichen Aufwand für wenig geübte Anwender, die auf eine manuelle Auswertung angewiesen sind. Bei zusätzlicher manueller Berechnung von Konfidenzintervallen dürften völlig unökonomische Auswertungszeiten erreicht werden. Der Einsatz von AIDScore wird die Normwertermittlung nach einer kurzen Einarbeitung deutlich reduzieren.

6. Gütekriterien

6.1 Objektivität: Aufgrund der überwiegend genauen und eindeutigen Beschreibungen des Manuals dürfte bei gründlicher Einarbeitung und regelmäßiger Testpraxis die *Durchführungsobjektivität* gewährleistet sein. Die allgemeine Testinstruktion gibt Hinweise zu äußeren Rahmenbedingungen, Transparenz, Beziehungsgestaltung und Motivierung des Kindes und trägt somit zu einer weiteren Homogenisierung der Testdurchführung bei. Testleitereffekte waren nach einer im Manual berichteten Studie beim AID vereinzelt nachweisbar, betroffen waren vor allem die Untertests 6 und 9.

Bei den Untertests 5 und 5b wäre eine klare Aussage, ob und wie die Aufmerksamkeit des Kindes während des Vorsprechens zu sichern ist, wünschenswert; z.B. kann sich über die gezielte Herstellung von Blickkontakt in Einzelfällen eine relevante Leistungssteigerung ergeben. Im Untertest 7 sollte explizit klargestellt werden, ob zeitaufwendige Bemühungen des Kindes, eine falsche oder ungenaue Lösung zu verbessern, zu ignorieren oder mit der Aufforderung zum schnellen Weiterarbeiten zu beantworten sind.

Empirische Befunde zur *Auswertungsobjektivität* werden im Manual nicht berichtet. Einen gewissen Ermessensspielraum des Testleiters sehen die Autoren bei den Untertests 6, 9, 11 und 10a; hinzuzufügen wäre auch der Untertest 7, bei dem Kriterien zur Bewertung von Schwächen in der zeichnerischen Umsetzung der Zeichen fehlen. Bei Subtest 10 vermissen wir einen Hinweis auf die Bewertung von Drehungen. Bei Untertest 10a sollte klargestellt werden, ob kleine Lösungsansätze schon ausreichen, um ein Item als bearbeitet zu werten. Umfangreiche Lösungstabellen für die sprachlichen Subtests oder ausführliche Darlegungen über die grundlegenden Bewertungsprinzipien, wie man sie (bei allen Mängeln, s. Renner & Fricke, 2001) aus dem HAWIK-III (Tewes, Rossmann & Schallberger, 1999) kennt, finden sich im AID 2 nicht. Nach eigenen Erfahrungen kann dies vor allem in den Untertests 6 und 11 zu Auswertungsunsicherheiten führen.

Bei Skalen mit Boden- und Deckeneffekten führen manuelle und computerisierte Auswertung nicht zu identischen Ergebnissen. Beispielsweise werden im Untertest 10 bei den 6-Jährigen für den minimalen Fähigkeitsparameter drei T-Werte (19, 27, 34) tabelliert. Dies spiegelt korrekt die Tatsache wider, dass bei einem Testrohwert von Null der T-Wert des Kindes 34 oder niedriger beträgt. AIDScore gibt an dieser Stelle ausschließlich den T-Wert 19 aus, was zu Fehlinterpretationen im Sinne einer Unterschätzung der kindlichen Fähigkeiten führen kann.

Die *Interpretationsobjektivität* ist für die quantitativen Befunde durch den Bezug auf gebräuchliche Normwerte gesichert. Verbale Beschreibungen der Skalenergebnisse werden nicht vorgeschlagen. Die statistische Beurteilung von Profildifferenzen beschränkt sich auf eine Faustregel für den paarweisen Vergleich von zwei Untertests (kritische Differenz der T-Werte von 15-18 für $\alpha = .05$ bei mittlerem Standardschätzfehler).

Zur beurteilerübergreifenden Beschreibung der einzelnen Untertests kann sich der Anwender auf die (teilweise vagen) Formulierungen der Testautoren stützen (s.o.). Als Grundlage für weitergehende inhaltliche Interpretationen dient das Diagramm zur Diagnostik von Teilleistungsstörungen (Abb. 1). Die Zuordnung der einzelnen Untertests zu den relevanten Dimensionen kann anhand des Diagramms eindeutig erfolgen. Für eine objektive Interpretation fehlen allerdings Regeln, welche Untertestwerte als auffällig zu gelten haben.

6.2 Skalierung: Die Skalenanalysen auf Grundlage der probabilistischen Testtheorie werden im Manual sehr ausführlich dargestellt. In einem ersten Analyseschritt nahmen die Autoren einen Vergleich von AID und AID 2 vor, der u.a. zur Elimination von veralteten Items führte. Im zweiten Analyseschritt wurden für sechs Untertests (1, 2, 3, 6, 9, 11) die Modellannahmen anhand verschiedener Aufteilungen der Stichprobe geprüft. Teilungskriterien waren a) Probanden mit hohen vs. Probanden mit niedrigen Testwerten, (»interne« Stichprobenaufteilung; mit Angaben zu Modelltests vor und nach dem Ausscheiden von Items mit unbefriedigenden Kennwerten), b) Geschlecht, c) Nationalität (deutsche und österreichische Kinder) und für den Untertest Realitätssicherheit d) Alter. Für die neu entwickelten Untertests 5a, 5b und 10a entfiel naturgemäß der Vergleich mit dem AID, die Modellprüfung erfolgte anhand der Teilungskriterien a-c (bei Untertest 10a zusätzlich d). Bei Subtest 5b ließ sich die Modellgültigkeit für die Gesamtstichprobe nicht zeigen. Der Untertest misst über den gesamten Altersbereich nicht eindimensional, erst eine Aufteilung der Stichprobe in Kinder bis/über 10 Jahre ergibt eine akzeptable Modellanpassung.

Bei der Bewertung der Modelle stützen sich die Autoren auf Likelihood-Tests nach Andersen. Im Hinblick auf durchaus häufige signifikante Modellabweichungen merken sie an, dass bei der bestehenden Stichprobengröße die Gefahr einer fehlerhaften Verwerfung der Modellgültigkeit (β -Fehler) unangemessen hoch sei. Im Fall signifikanter Likelihood-Tests wird daher zur weiteren Beurteilung stets die im Manual abgebildete graphische Modellüberprüfung herangezogen.

6.3 Reliabilität: Als Maß der Zuverlässigkeit werden – mit Ausnahme der Untertests 5 und 7 – die Standardschätzfehler der Fähigkeitsparameter mitgeteilt. Im Gegensatz zu Testverfahren, die auf der klassischen Testtheorie basieren, gibt es nicht einen für alle getesteten Personen als gleichermaßen gültig unterstellten Reliabilitätskennwert, sondern eine für die jeweilige Kombination von Rohwertsumme, durchgeführter

Aufgabenkombination und (teilweise) Alter des Kindes spezifische Schätzung der Messgenauigkeit.

Die Variationsbreite der Standardschätzfehler ist durchaus beträchtlich, z.B. zwischen .55 und 1.47 im Subtest Alltagswissen, wobei die Messgenauigkeit in den Extrembereichen generell merklich abnimmt. Negativ fällt der Subtest 10a mit einem maximalen Standardschätzfehler von 2.86 auf (lt. Manual, S. 22; die zugehörige Detailtabelle auf S. 220 enthält sogar noch höhere Werte). Die konkrete Bewertung der Messgenauigkeit ist im Einzelfall allerdings sehr schwierig, da das Manual keine tabellierten Konfidenzintervalle enthält. Ohne Anwendung des Computerprogramms müssen Konfidenzintervalle für die Fähigkeitsparameter manuell berechnet und anschließend in T-Werte transformiert werden. Die unteren Grenzen der Konfidenzintervalle bei leistungsschwachen Probanden fallen oft in den nicht tabellierten Bereich. Günstiger stellt sich die Berechnung mit AIDScore dar, das 95%-Konfidenzintervalle für neun der sechzehn Kennwerte graphisch ausgibt – leider nicht in voll befriedigender Darstellungsqualität. Grenzwerte, die unter $T=19$ bzw. über $T=81$ liegen, werden von AIDScore vereinfacht als $T=19$ bzw. $T=81$ dargestellt, so dass der Eindruck einer zu hohen Messgenauigkeit entstehen kann. Eine numerische Darstellung der Konfidenzintervalle kann nur umständlich und unübersichtlich über die Exportfunktion von AIDScore erreicht werden. Betrachtet man einige ausgewählte Konstellationen mit niedrigen Standardmessfehlern, beträgt die Spannbreite der 95%-Vertrauensintervalle 13-25 T-Werte, mit den ungünstigsten Werten für die Untertests 2, 4, 8 und 10. Bei höheren Standardmessfehlern können durchaus Intervalle von 30 T-Werten und mehr auftreten.

Befunde zur Testwiederholungszuverlässigkeit zeigen über einen Zeitraum von vier Wochen für sieben Kennwerte (Untertests 5 und 7, alle Zusatztests) eine Stabilität zwischen .57 (Merken und Einprägen) und .89 (Kodieren und Assoziieren – Kodiermenge).

Das Manual enthält zur Zuverlässigkeit der Kennwerte Range, Intelligenzquantität und zweitniedrigste Untertestleistung keinerlei Angaben. Im Handbuch des AID wurde für die Intelligenzquantität eine Reliabilität von .63 genannt (Kubinger & Wurst, 1988, S.30). Während die Autoren sonst gern auf Befunde zum AID zurückgreifen, wird diese für ein globales Intelligenzmaß bemerkenswert niedrige Kennziffer im AID 2 Manual nicht zitiert.

6.4 Validität: Die *inhaltliche Validität* des AID 2 ergibt sich durch den Bezug auf überwiegend gut eingeführte und bewährte Aufgabentypen, die auch in anderen Intelligenztests in ähnlicher Form verwendet werden. Bei der Entwicklung des AID wurden zusätzlich Expertenratings herangezogen.

Zur *konvergenten und diskriminanten Validität* (Zusammenhänge zwischen AID 2 und anderen gebräuchlichen Intelligenz- oder Leistungstests) werden keine neueren Daten berichtet. Das Manual referiert eine Vielzahl von Befunden, die mit dem AID erhoben wurden. Erwartungsgemäß fanden sich keine bedeutsa-

men Zusammenhänge zu eindeutig konstruktfernen Variablen (z.B. Schulangst, Erziehungsstil der Eltern). Auch die niedrigen bis mäßigen Korrelationen mit zahlreichen Leistungs- und Intelligenztests werden von den Autoren überwiegend im Sinne diskriminanter Validität interpretiert, da es sich ebenfalls um konstruktferne Variablen gehandelt habe. Positive Zusammenhänge hätten sich zwischen einzelnen AID Subtests und Konzentrations- und Gedächtnisvariablen, Informationsverarbeitungsgeschwindigkeit und Raumlage-Orientierung gezeigt – was allerdings in den im Manual dargestellten Korrelationstabellen nicht sehr prägnant in Erscheinung tritt. Für die neuen Zusatztests wurde in einer kleinen Stichprobe (37 9-11-Jährige) die Korrelation mit dem CFT20 (Weiß, 1987) bestimmt (.29 - .59).

Der *kriteriumsbezogenen Validität* lassen sich Daten über die mittleren T-Werte bei Hauptschülern und Gymnasiasten zuordnen, die überwiegend signifikante Unterschiede in der erwarteten Richtung zeigen. Bedeutsame Korrelationen (-.39 bis -.57) mit Schulnoten in den Fächern Deutsch und Mathematik werden nur für den AID berichtet, wobei die stärksten Zusammenhänge bei den Untertests 1, 3, 6 und 11 gefunden wurden.

Angesichts des Verzichts auf skalenaggregierende Auswertungen muss die Frage der *faktoriellen Validität* beim AID 2 nicht im Vordergrund stehen. Die im Manual knapp berichtete Hauptachsenfaktorenanalyse mit orthogonaler Varimax-Rotation aller 16 Testkennwerte ergab eine vierfaktorielle Lösung. Altersspezifische Lösungen werden nicht dargestellt, nach Angaben der Autoren hätten sich hinsichtlich Faktorenanzahl und Ladungsgrößen in vier Altersgruppen vergleichbare Ergebnisse gezeigt. Die Faktoren werden als »Informationsverarbeitung in der gesellschaftlichen Umwelt« (Ladungen über .50 für die Subtests 1, 3, 6, 9, 11; somit starke Übereinstimmung mit Wechslers Verbal-IQ); »Informationsverarbeitung neuer Inhalte« (hohe Ladungen nur für Untertest 7, Kodiermenge und Assoziationen); »Auffassungskapazität« (hohe Ladungen nur für Untertest 5, vorwärts und rückwärts) und »(Re)Produktionsfähigkeit durch Strukturierung« (Untertests 8 und 10, Zusatztest 10a; somit ein Faktor mit deutlicher visuell-konstruktiver Komponente) beschrieben. Mit Ladungen von maximal .43 und Komunalitäten zwischen .27 und .36 sind die Untertests 2 und 4 und die Zusatztests 5a und 5b nicht eindeutig zuordenbar.

Zur Validität der Kennwerte Range, Intelligenzquantität und zweitniedrigste Untertestleistung enthält das Manual keine Angaben.

6.5 Normierung: Normen werden für den Altersbereich von 6;0 bis 15;11 Jahren angeboten. Die individuellen Testwerte (Fähigkeitsparameter) wurden in T-Werte umgerechnet, bei fehlender Normalverteilung mittels Flächentransformationen. Die Normdaten des AID wurden zur Adjustierung der neuen Normwerte herangezogen, um Abweichungen vom bekannten Altersverlauf der Fähigkeitsparameter, die als »stichprobenbedingte Irregularitäten« (Manual, S. 68) eingeschätzt

Wir danken den Kollegen Torsten Fricke (Institut für Kinderneurologie und -epileptologie, Bremerhaven), Dieter Irblich und Susanne Schillig (Sozialpädiatrisches Zentrum der Kreuznacher Diakonie) für ihre konstruktiven Hinweise zu dieser Arbeit. Herrn Prof. Kubinger danken wir für seine hilfreiche Durchsicht des Manuskriptes.

wurden, zu erkennen und auszugleichen. In diesem Fall erfolgten Anpassungen durch logarithmische oder logistische Funktionen. Die Auswirkungen dieser Anpassung werden nur für den Subtest Alltagswissen näher dargestellt.

Die Altersabstände der acht Normgruppen betragen jeweils ein Jahr, die Normwerte für Intelligenzquantität, Range und zweitniedrigste Untertestleistung werden für die Gesamtstichprobe tabelliert. Für einzelne Altersjahrgänge werden in den Subtests 1, 3 und 10 geschlechtsspezifische Normen bereitgestellt.

Die Normierungsstichprobe umfasst 977 Kinder (495 Jungen/482 Mädchen) aus Österreich (67,6%) und Deutschland (alte Bundesländer 25%, neue Bundesländer 7,4%) bei weitgehend gleicher Geschlechterverteilung in den Altersjahrgängen. In den Altersgruppen sind zwischen 67 und 116 Kinder vertreten. Das Normierungsvorgehen wird als Kombination von Klumpenstichproben- und Quotenverfahren beschrieben. Ausgewählte Orte in fünf österreichischen und sechs deutschen Bundesländern und ausgewählte Schultypen wurden als Erhebungseinheiten festgelegt. Eine Gleichverteilung wurde in Bezug auf Alter, Geschlecht und Schultyp angestrebt. Als einfache demographische Variable wurde die Größe des Wohnorts (dichotomisiert in Gemeinden mit mehr bzw. weniger als 50.000 Einwohnern) kontrolliert. Die Sozialschicht wurde dokumentiert, aber nicht als Quotenvariable aufgenommen. Beim Vergleich der Stichprobenszusammensetzung mit schulstatistischen Daten sehen die Autoren selbst einen höheren Anteil an Hauptschülern und einen geringeren Anteil an Realschülern in der deutschen Stichprobe als kritisch. Bei der Darstellung irritiert, dass sich die schulstatistischen Angaben zur Grundgesamtheit in Deutschland zu 100% auffaddieren, obwohl Gesamtschulen und zahlreiche Sonderschulformen nicht berücksichtigt werden. Die Darstellung der besuchten Bildungseinrichtung erfolgt nicht nach Jahrgängen getrennt, so dass die Repräsentativität der Stichprobe in den einzelnen Normgruppen nicht nachvollzogen werden kann.

Die Testungen wurden von 30 Diplom-Psychologen und Studierenden der Psychologie während der Unterrichtszeit in den Schulen der Kinder durchgeführt. Die Testung erfolgte nur mit Einverständnis der Sorgeberechtigten. Wie oft dieses verweigert wurde, ist nicht dokumentiert. Angaben über den Anteil von drop-outs während der Untersuchung (z.B. bei mangelnder Kooperation des Kindes) fehlen.

Kritisch zu bewerten ist der im Vergleich mit anderen mehrdimensionalen Intelligenztests (z.B. HAWIK-III, Tewes et al., 1999) ungewöhnlich große Altersbereich der einzelnen Normgruppen. Bei den Subtests 1 und 11 unterscheiden sich die Normwerte von 6;11- und 7;0-jährigen Kindern mit identischen Fähigkeitsparametern um 7-9 T-Werte. Bei Mädchen liegen die entsprechenden Diskrepanzen im Angewandten Rechnen sogar bei ca. 10 T-Werten. Die Autoren sehen darin zwar keine wesentliche Problematik, verweisen aber auf die Möglichkeit der Interpolation bei manueller Auswertung.

Bei einer ganz auf die Untertestergebnisse bezogenen Interpretation von individuellen Stärken und Schwä-

chen ist eine gleichmäßige Ausschöpfung des tabellierten Wertebereiches von besonderer Bedeutung. Boden- und Deckeneffekte wurden im AID 2 allerdings nicht konsequent vermieden. Betroffen sind bei den 6-Jährigen die Untertests 5 (rückwärts), 7 (beide Kennwerte), 10, 5a und 5b. Im oberen Leistungsbereich ist bei älteren Kindern ebenfalls der Untertest 10 betroffen (max. T-Wert 59 bei den 15-Jährigen). Leichte Deckeneffekte zeigen die Untertests 2 und 7 (Assoziationen) mit einem max. T-Wert von 73 ab 12 bzw. 13 Jahren.

Bei der Auswertung des Untertests 10a ist neben der Anzahl der richtigen Lösungen zu berücksichtigen, welche Aufgaben jeweils bearbeitet wurden. Die dabei entstandenen Normtabellen enthalten z.T. bemerkenswerte Effekte (s. Tabelle 2): Erhebliche Unterschiede in der Testleistung werden nicht in den Normdaten abgebildet, sogar höhere Normwerte für schwächere Leistungen sind möglich. Ebenso kann der Normwert des Kindes absacken, wenn es eine weitere Aufgabe bearbeitet, aber nicht löst; z.B. erhält ein 10-jähriges Kind, das die ersten fünf Aufgaben korrekt bearbeitet, einen T-Wert von 76. Wird die nächste Aufgabe noch begonnen, aber nicht mehr vor der Zeitgrenze gelöst, fällt das Ergebnis um fast eine Standardabweichung niedriger aus!

Tabelle 2: AID 2 Untertest »Strukturieren-visumotorisch«: Bessere Testwerte bei schlechteren Leistungen – ausgewählte Beispiele

Normgruppe		Lösungsmuster ¹	Richtige Lösungen	T-Wert ²
15 Jahre	(1)	1 1 1 1 1 1 1 1 1 1 1 1	11	63
	(2)	1 1 1 / / / / / / / /	3	63
	(3)	1 1 1 1 0 0 1 1 / 0 /	6	49
	(4)	1 / 1 / / / / 1 / / /	3	71
10 Jahre	(1)	1 1 1 1 1 / / / / / /	5	76
	(2)	1 1 1 1 1 0 / / / / /	5	67
	(3)	1 1 / 1 / / / / / / /	3	81

Anmerkungen:
 1.) Korrekte Lösungen werden mit »1«, fehlerhafte Lösungen mit »0« und nicht bearbeitete Items mit »/« bezeichnet.
 2.) Übereinstimmende Ergebnisse von manueller und computerisierter Auswertung.

7. Abschließende Bewertung

Nach dem HAWIK-III (Tewes et al., 1999) liegt mit dem AID 2 ein weiterer mehrdimensionaler Intelligenztest in überarbeiteter Form und mit aktualisierten Normen vor. Der AID 2 ermöglicht bei niedrigen Verbrauchskosten einen breiten Überblick über verschiedene kognitive Fähigkeiten (sprachliche Intelligenz, visuell-konstruktive Leistungen, Gedächtnis u.a.m.) des untersuchten Kindes. Die Durchführung in der Einzelsituation begünstigt vielfältige Beobachtungen über Lösungsstrategien, Handlungsplanung, Aufmerksamkeitssteuerung, Motivation, Anstrengungsbereitschaft u.a.m. Positiv herauszuheben ist, dass die Autoren sich um inhaltliche Verbesserungen und Erweiterungen des Verfahrens bemühten, z.B. wurde aufgrund der empirisch fundierten Kritik von Hofmann und Ricken (1996) der Testkennwert »Lernquotient« aufgegeben. In teststatistischer Hinsicht beeindruckt die detaillierte Darstellung der Skaleneigenschaften auf Basis der probabilistischen Testtheorie. Der Testpraktiker wird das

teilweise Angebot von Parallel- und Kurzformen schätzen. Leider sind die Testbögen wenig benutzerfreundlich gestaltet, die manuelle Auswertung ist fehleranfällig. Das prinzipiell sehr hilfreiche Auswertungsprogramm AIDScore sollte durch eine direkte numerische Ausgabe von Konfidenzintervallen und eine bessere graphische Darstellung der Testergebnisse aufgewertet werden. Untere Extremwerte in Subtests mit Bodeneffekten sollten von AIDScore wie in den Tabellen des Manuals korrekt angegeben werden.

Im Hinblick auf die *Testgütekriterien* wären aktuelle empirische Befunde zur Durchführungs- und Auswertungsobjektivität wünschenswert. Ebenso fehlen neuere Studien zur Validität. Der Einwand Jägers (1987), dass der angestrebte Einsatzbereich des AID gemessen an den Validitätsnachweisen zu breit gewählt ist, bleibt weiterhin gültig und gilt insbesondere für die neu aufgenommenen Zusatztests. Die Angabe der personspezifischen Standardschätzfehler als Maß der Reliabilität erlaubt eine präzisere Abschätzung der Zuverlässigkeit des individuellen Testergebnisses, wodurch die (auch im AID 2 durchaus beträchtlichen) Messfehler in den Extrembereichen erkennbar werden. Die Reliabilität einzelner Untertests ist allerdings verbesserungsbedürftig. Wünschenswert wären ausführlichere Befunde zur zeitlichen Stabilität der Testergebnisse. Das Fehlen von Angaben zur Profilreliabilität überrascht bei einem Verfahren, das die Profilinterpretation ganz in den Vordergrund stellt. Äußerste Zurückhaltung ist bei der Interpretation von Subtest 10a anzuraten; wie oben beschrieben, besteht bei bestimmten Lösungskonstellationen ein ganz erhebliches Auswertungsproblem.

Als wesentliche inhaltliche Spezifika des AID 2 sind der Anspruch einer förderorientierten Teilleistungsdiagnostik und der Verzicht auf ein globales Intelligenzmaß zu nennen. Bei Fehlen jeglicher Validitätsangaben verwundert die Sicherheit, mit der Kubinger & Wurst (2000) den Kennwert Intelligenzquantität propagieren. In theoretischer Hinsicht wäre zu wünschen, dass die in ihrem Ausgangspunkt sicherlich gerechtfertigte Diskussion über die Sinnhaftigkeit eines Gesamt-IQ nicht allein auf der metaphorischen Ebene (»Jede Kette ist nur so stark wie ihr schwächstes Glied«) abgehandelt wird. Wie Rost (2002) vor kurzem in dieser Zeitschrift argumentiert hat, gibt es durchaus ernst zu nehmende Gründe und umfangreiche Forschungsarbeiten, die für die Verwendung eines globalen kognitiven Leistungsmaßes und eine besondere Sorgfalt und Zurückhaltung bei Profilinterpretationen sprechen.

Im Hinblick auf die Teilleistungsdiagnostik kann der AID 2 (wie andere mehrdimensionale Intelligenztests auch) den Anspruch eines Screening-Verfahrens durchaus erfüllen. Die detailliertere Erfassung spezifischer Teilleistungsschwächen ist mit dem AID 2 allerdings nicht leicht, weil die Ergebnisse der meisten Untertests von mehreren Fähigkeitsdimensionen bestimmt werden (die inhaltlich sinnvolle Trennung der Anforderungen von Zahlennachsprechen vorwärts und rückwärts stellt eine lobenswerte Ausnahme dar). Dies gilt nicht nur für die aus der Wechsler-Tradition stammenden

Untertests, sondern auch für die neuen Zusatztests (z.B. Vermischung von konkreten und abstrakten Symbolen im Untertest 5a; Vermischung von Kurz- und Langzeitgedächtniskomponenten im Untertest 5b). Eine förderorientierte Diagnostik ohne Bezug auf spezifische Störungsmodelle realisieren zu wollen, erscheint ohnehin wenig aussichtsreich. Bei Fehlen handlungsleitender Regeln zur Profilinterpretation bleibt der Brückenschlag vom AID 2 Testprofil zur klinischen Interpretation ein weitgehend unstandardisierter Prozess von unbekannter Zuverlässigkeit.

ZUSAMMENFASSUNG

Mit dem 2000 erschienenen AID 2 liegt eine an 977 Kindern aus Deutschland und Österreich neu normierte und erweiterte Überarbeitung des Adaptiven Intelligenz Diagnostikums AID vor. Der grundlegende Testaufbau ist unverändert, neu ist die Aufnahme der Zusatztests »Unmittelbares Reproduzieren-figural/abstrakt«, »Merken und Einprägen«, »Strukturieren-visumotorisch«. Als globales Intelligenzmaß wird standardmäßig die sog. Intelligenzquantität berechnet, die auf der niedrigsten Untertestleistung des Kindes basiert. Die inhaltlich vielfältigen Anforderungen des AID 2 und die individuelle Testadministration ermöglichen einen breiten Überblick über kognitive Leistungen und Arbeitsverhalten des Kindes. Für mehrere Untertests liegen Parallel- und Kurzformen vor.

Im Hinblick auf Testgütekriterien ist das Fehlen empirischer Befunde zur Objektivität und Validität der Zusatztests zu bemängeln. Die Validität des Kennwertes Intelligenzquantität ist nicht nachgewiesen. Die Reliabilität einzelner Untertests erscheint verbesserungsbedürftig. Die Auswertungsstrategie des Zusatztests »Strukturieren-visumotorisch« führt in Einzelfällen zu nicht nachvollziehbaren Ergebnissen. Der Altersabstand der Normgruppen ist mit einem Jahr für die jüngsten Kinder zu breit gewählt.

LITERATUR

- FLANAGAN, D.P., GENSHAFT, J.L. & HARRISON, P.L. (Eds.) (1997). *Contemporary Intellectual Assessment*. New York: Guilford Press.
- HARDESTY, F.P. & PRIESTER, H.J. (1956). *Hamburg-Wechsler Intelligenztest für Kinder*. Bern: Huber.
- HOFMANN, K. & RICKEN, G. (1996). Der Lernquotient im AID – ein Indikator für Lernfähigkeit? *Zeitschrift für Pädagogik und Psychologie Behinderteter*, 22, 123-130.
- JÄGER, R.S. (1987). Adaptives Intelligenz Diagnostikum. *Diagnostica*, 33, 81-84.
- KUBINGER, K.D. (Hrsg.). (1983). *Der HAWIK – Möglichkeiten und Grenzen seiner Anwendung*. Weinheim: Beltz.
- KUBINGER, K.D. (1998). Psychometric shortcomings of Wechsler's intelligence scales – results of the German WISC, conclusions for the WISC-R. *Review of Psychology*, 5, 3-12.
- KUBINGER, K.D. & WURST, E. (1985). *Adaptives Intelligenz Diagnostikum (AID) Manual*. Weinheim: Beltz.
- KUBINGER, K.D. & WURST, E. (1988). *Adaptives Intelligenz Diagnostikum (AID) 2., überarbeitete Auflage. Manual*. Weinheim: Beltz.
- KUBINGER, K.D. & WURST, E. (2000). *Adaptives Intelligenz Diagnostikum 2. Manual*. Weinheim: Beltz.
- RENNER, G. & FRICKE, T. (2001). Der Hamburg-Wechsler-Intelligenztest für Kinder – Dritte Auflage (HAWIK-III). *Report Psychologie*, 26, 460-477.
- ROST, D.H. (2002). Notwendige Klarstellungen. Zur Diskussion um Hochbegabung und Hochbegabte. *Report Psychologie*, 27, 624-634.
- TEWES, U., ROSSMANN, P. & SCHALLBERGER, U. (Hrsg.). (1999). *HAWIK-III. Hamburg-Wechsler-Intelligenztest für Kinder – dritte Auflage. Manual*. Bern: Huber.
- WEIß, H. (1987). *Grundintelligenztest CFT 20*. Göttingen: Hogrefe.

**Adaptives
Intelligenz
Diagnostikum 2
von
Klaus D. Kubinger
& Elisabeth Wurst
Göttingen:
Beltz Test, 2000
Testkoffer
komplett:
€ 648,00
Verbrauchsmaterial für
25 Anwendungen:
€ 30,00**