

# NEUERE TESTVERFAHREN

---

Grob, A., Meyer, C. S., Hagmann-von Arx, P. (2009). **Intelligence and Development Scales (IDS). Intelligenz- und Entwicklungsskalen für Kinder von 5-10 Jahren.** Bern: Hans Huber, Testkoffer komplett: 1.075,- € (inkl. Auswertungsprogramm und Verbrauchsmaterial für 25 Anwendungen), Verbrauchsmaterial je Testung: 7,44 €.

---

Mit den IDS legt das Autorenteam ein ambitioniertes Testverfahren vor, das neben der Erfassung kognitiver und sprachlicher Funktionen eine umfassende Beurteilung der kindlichen Entwicklung unter Einbeziehung motorischer Leistungen, sozial-emotionaler Kompetenzen und der Leistungsmotivation anstrebt und somit „eine umfangreiche Entwicklungsprofilanalyse“ (Manual, S. 9) ermöglichen will. Als Einsatzbereich der IDS nennen die Autoren das gesamte Spektrum der Entwicklungs- und Leistungsdiagnostik, die Schuleingangsdiagnostik und den klinischen Bereich, z. B. die Identifikation von Minder- und Hochbegabungen, die „Standortbestimmung der Intelligenz und Entwicklung“ (Manual, S. 12) bei der Diagnostik umschriebener und tiefgreifender Entwicklungsstörungen sowie die Diagnostik und Interventionsplanung bei emotionalen und Verhaltensstörungen in Erziehungs- und Familienberatung, Schulpsychologie, klinischer Kinderpsychologie, Kinderpsychiatrie und Pädiatrie.

## Theoretischer Hintergrund

Die Autoren beziehen sich in ihrer theoretischen Einführung allgemein auf eine „konstruktivistische“ Entwicklungsperspektive, die die kindliche Eigenaktivität bei der Aneignung neuer Erfahrungen betont. Die Auswahl der in den IDS repräsentierten Funktionsbereiche erfolgte unter Bezug auf Darstellungen in „prominenten Nachschlagewerken und Lehrbüchern der Entwicklungspsychologie“ (Manual, S. 16). Für die einzelnen Untertests referieren die Autoren jeweils theoretische Grundlagen und entwicklungspsychologische Befunde, wobei sie sich eher allgemein auf den Gegenstandsbereich (z. B. Aufmerksamkeit) und nicht auf die im jeweiligen Subtest realisierte Aufgabenstellung beziehen.

Bei der Zusammenstellung der Skalen des Bereiches *Kognition*, die zu einem Gesamt-IQ zusammengefasst werden, wurden nach Aussage der Autoren Wahrnehmung, Aufmerksamkeit, Gedächtnis und Sprache berücksichtigt. Die Autoren stellen keine Bezüge zwischen der Kognitionsskala und aktuellen Intelligenztheorien her. Nach ergänzenden Verlagsinformationen (Verlag Hans Huber, 2011a) sollten insbesondere elementare kognitive Funktionen erfasst werden, die es Kindern erlauben, sich neues Wissen anzueignen.

## Testaufbau, Material und Durchführung

Die IDS sind für den Altersbereich von 5;0 bis 10;11 Jahren vorgesehen und werden stets im Einzelsetting durchgeführt. Der Testkoffer enthält das Manual (212 Seiten), je 25 Sätze der erforderlichen Protokollbogen und Testhefte, zwei Aufstellordner zur Präsentation der Items aus den Bereichen *Kognitive Entwicklung* und *Allgemeine Entwicklung*, einen dazugehörigen Halter, der das Umfallen der Aufstellordner verhindern soll, eine Unterlage und eine Schachtel mit Kärtchen für den Untertest *Wahrnehmung Visuell*, ein Seil, einen Tennisball (Bereich *Motorik*), 25 Würfel, fünf Perlen und ein Bändchen (Bereich *Motorik*, *Denken Logisch-Mathematisch*), einen Satz Holzfiguren (*Sprache Rezeptiv*), jeweils vier drei- und rechteckige Holzklötze (*Denken Bildlich*), eine unbedruckte Unterlage (*Denken Bildlich*, *Denken Logisch-Mathematisch*), die Auswertungsschablonen für *Aufmerksamkeit Selektiv*, einen DIN-A6-Block mit Bleistift und das PC-Auswertungsprogramm für Windows-Betriebssysteme. Für die Testung wird außerdem eine Stoppuhr benötigt.

Bei vollständiger Durchführung umfassen die IDS 17 Untertests und zwei Ratingskalen, die sechs inhaltlichen Bereichen zugeordnet werden (s. a. Tab. 1). Nur die Untertests des Funktionsbereichs *Kognition* werden zu einem Gesamtwert (IQ) zusammengefasst. Bei den weiteren Bereichen wird „zugunsten einer differenzierten Profilanalyse“ (Manual, S. 17) auf Bildung eines Summenwertes verzichtet.

Zum Funktionsbereich *Kognitive Entwicklung* gehören sieben Untertests:

1. *Wahrnehmung Visuell* (7 Items). Dem Kind werden vier bis sieben Kärtchen mit unterschiedlich langen Linien (Längendifferenz zwischen 10 und 0,25 mm) vorgelegt, die in einer Reihe von der längsten bis zur kürzesten Linie angeordnet werden sollen. Teillösungen sind möglich, das Kind erhält einen Punkt für jede Karte, die an der richtigen Position liegt.
2. *Gedächtnis Auditiv* (24 Items). Dem Kind wird eine Geschichte vorgelesen (ca. 220 Wörter). Im Anschluss an Untertest 7 soll diese Geschichte nacherzählt werden. Es werden jeweils zwei Punkte vergeben, wenn einer der 24 Hauptgedanken der Geschichte frei erinnert wird. Anschließend stellt der Untersuchungsleiter zu den nicht erinnerten Punkten gezielte Fragen, für deren richtige Beantwortung jeweils ein Punkt vergeben wird.
3. *Aufmerksamkeit Selektiv*. Es handelt sich um die Variante eines Durchstreichtests. Auf einem Arbeitsblatt sind 225 Enten mit unterschiedlichen Merkmalen in neun Reihen angeordnet. Die Kinder sollen alle Enten markieren, die genau zwei orangefarbene Körperteile aufweisen (Füße oder Schnabelhälfen) und deren Schnabel nach rechts weist. Die Bearbeitungszeit je Zeile beträgt 15 Sekunden, wobei die Zeitmessung für jede Zeile getrennt erfolgt. Der Testrohwert ergibt sich aus der Summe der bearbeiteten Enten abzüglich der Fehler und Auslassungen.
4. *Gedächtnis Phonologisch* (12 Items). Folgen von zwei bis sieben Zahlen oder Buchstaben sollen unmittelbar nachgesprochen werden.

5. *Gedächtnis Räumlich-Visuell* (11 Items). Dem Kind werden ein bis sechs unterschiedliche geometrische Figuren präsentiert. Anschließend sollen diese auf einer Vorlage mit fünf bis zwölf Figuren gezeigt werden, wobei die Farbe der Formen mit einer Ausnahme von der Vorlage abweicht. Die Darbietungszeit beträgt 5-15 Sekunden.

*Tabelle 1:* Aufbau, Inhalte und Reliabilität der IDS

Skala	Inhaltliche Schwerpunkte <sup>a</sup>	Reliabilität $\alpha^b$ /Retest <sup>c</sup>
Kognitive Entwicklung (Intelligenz)		.92/.83
Wahrnehmung Visuell	Visuelle Diskriminationsfähigkeit (Längenunterscheidung)	.77/.62
Gedächtnis Auditiv	Verbales Langzeitgedächtnis, Sprachverständnis	.87/.81
Aufmerksamkeit Selektiv	Kurzfristige selektive visuelle Aufmerksamkeit (Speed-Test)	.96/.75
Gedächtnis Phonologisch	Kurzzeitgedächtnis in der akustischen Modalität	.73/.45
Gedächtnis Räumlich-Visuell	Figurales Kurzzeitgedächtnis	.68/.48
Denken Bildlich	Visuell-konstruktive Leistung	.70/.74
Denken Konzeptuell	Klassifikation, Erkennen von Gemeinsamkeiten. Allgemeine Entwicklung	.75/.52
Psychomotorik		
Grobmotorik	Grobmotorik & Körperkoordination	.57/.67
Feinmotorik	Fingerfeinmotorik	.60/.34
Visuomotorik	Visuomotorische und visuell-konstruktive Leistung	.71/.62
Sozial-Emotionale Kompetenz		
Emotionen Erkennen	Identifikation von mimischen Emotionsausdrücken	.59/.54
Emotionen Regulieren	Umgang mit negativen Gefühlszuständen	.79/.45
Soziale Situationen Verstehen	Verständnis für soziale Situationen und damit verbundene Gefühle	.74/.49
Sozial Kompetent Handeln	Soziale Kompetenz, Fähigkeit um Konfliktbewältigung	.63/.74
Mathematik		
Denken Logisch-Mathematisch	Rechnerisches Denken	.87/.88
Sprache		
Sprache Expressiv	Sprachlicher Ausdruck, Satzbildung	.81/.57
Sprache Rezeptiv	Sprachverständnis	.88/.72
Leistungsmotivation		
Durchhaltevermögen	Ausdauer in Leistungssituationen	.86/.71
Leistungsfreude	Freude an Bewältigung schwieriger Anforderungen	.76/.41

*Anmerkungen:* a) Angaben in Anlehnung an das Testmanual (zu den Einzelheiten s. dort), ergänzt durch eigene Einschätzungen. b)  $\alpha$  = interne Konsistenz. Diese Werte wurden nicht getrennt nach Altersgruppen berechnet, was bei alterskorrelierten Testleistungen zu Überschätzung der Reliabilität führen kann. c) Über einen Zeitraum von 15 Monaten

6. *Denken Bildlich* (10 Items). Auf Bildvorlagen präsentierte geometrische Figuren unterschiedlicher Komplexität sollen mit drei- und viereckigen Klötzchen nachgelegt werden. Die Lösung muss innerhalb von 30 bzw. 45 Sekunden erfolgen.
7. *Denken Konzeptuell* (12 Items). Es werden jeweils drei Bilder präsentiert, die etwas gemeinsam haben (z. B. drei Tiere, drei Lichtquellen). Aus fünf weiteren Bildern sind zwei auszuwählen, die zu diesem Konzept passen.

Der Bereich *Psychomotorik* wird durch drei Untertests erfasst:

8. *Grobmotorik* (3 Items). Die Kinder sollen über ein auf dem Boden liegendes Seil mit geschlossenen Augen balancieren, einen Tennisball fangen und werfen sowie beidbeinig seitlich über ein Seil hüpfen. Die Bewertung erfolgt unter Berücksichtigung unterschiedlicher Kategorien (z. B. „neben das Seil treten“ und „rudern mit den Armen“).
9. *Feinmotorik*. Es sollen schnellstmöglich je fünf runde und eckige Perlen abwechselnd aufgefädelt werden. Das Kind kann die Abfolge der Perlen vorab zehn Sekunden auf einer Bildvorlage betrachten. Die Bewertung erfolgt anhand von sechs Kriterien, die sich auf Greifen, Fädeln und Geschwindigkeit beziehen. Fehler in der Reihenfolge der Perlen führen zu Zeitzuschlägen.
10. *Visuomotorik* (2 Items). Es sollen auf zwei Vorlagen dargebotene geometrische Figuren frei abgezeichnet werden. Die Bewertung erfolgt nach jeweils acht Kriterien (z. B. abgerundete Ecken, Symmetrien nicht eingehalten).

Der Funktionsbereich *Sozial-Emotionale Kompetenz* ist mit vier Untertests vertreten:

11. *Emotionen Erkennen* (10 Items). Auf Fotos von Gesichtern sollen verschiedene Emotionen (Freude, Wut, Angst, Trauer, Überraschung) erkannt werden.
12. *Emotionen Regulieren* (3 Items). Im Subtest 11 werden bei drei Bildkarten, die Wut, Traurigkeit und Angst darstellen, zusätzlich Fragen nach Strategien zur Regulierung dieser Emotionen gestellt: „Was könnte dieses Kind tun, damit es nicht mehr wütend ist?“, „Was tust du, wenn du wütend bist, damit du nachher nicht mehr wütend bist?“, „Und was könntest du sonst noch tun, damit du nachher nicht mehr wütend bist?“. Die letzte Frage wird ausgelassen, wenn das Kind die ersten beiden nicht beantworten konnte. Bei der Bewertung werden drei Arten von Emotionsregulationsstrategien unterschieden: adaptive (z. B. „Vergessen“, „Problemorientiertes Handeln“; zwei Rohwertpunkte), maladaptive (z. B. „Aggressives Verhalten“, „Rückzug“; null Rohwertpunkte) und weitere (z. B. „soziale Unterstützung“, „Emotionskontrolle“; ein Rohwertpunkt).
13. *Soziale Situationen verstehen* (2 Items). Auf Bildkarten dargestellte soziale Situationen sollen verstanden und erklärt werden. Dabei werden sechs Situationsaspekte bewertet, die sich auf das Verständnis für die Ursachen bestimmter Verhaltensweisen und die Gefühle der beteiligten Personen beziehen. Korrekte Antworten werden mit zwei Punkten bewertet, wenn sie auf die pauschalen Fragen „Was ist hier los?“ und „Wie fühlen sich die Kinder auf dem Bild, und

warum fühlen sie sich so?“ genannt wurden. Ein Punkt wird gegeben, wenn ergänzende Kontrollfragen (z. B. „Warum weint das Mädchen?“) zu den noch nicht genannten Aspekten korrekt beantwortet werden.

14. *Sozial Kompetent Handeln* (6 Items). Die Kinder sollen angeben, wie sich in verschiedenen sozialen Situationen, die auf Bildkarten dargestellt sind, verhalten würden („Was würdest Du tun, wenn ein Kind Deine Sandburg kaputt macht?“). Bei der Bewertung wird zwischen direkten und indirekten prosozialen oder sozial kompetenten Verhaltensweisen unterschieden. Aktives Handeln wird mit zwei Punkten bewertet, ein Punkt wird vergeben, wenn das Kind eine andere Person um Unterstützung bittet.

Der Funktionsbereich *Mathematik* wird durch einen Untertest repräsentiert:

15. *Denken Logisch-Mathematisch* (18 Items). Es sollen unterschiedliche mathematische Aufgaben gelöst werden, die vom einfachen Abzählen bis zu einer komplexen Zahlenzerlegung reichen. Inhaltlich werden das Verständnis für natürliche Zahlen, logisches Denken, geometrische und arithmetische Fähigkeiten sowie die Verarbeitungskapazität angesprochen. Zwei Items gelten erst als gelöst, wenn auch eine korrekte Begründung für die richtige Antwort gegeben wird.

Zum Funktionsbereich *Sprache* gehören zwei Untertests:

16. *Sprache Expressiv* (12 Items). Unter Verwendung von ein bis drei vorgegebenen Wörtern soll ein sinnvoller und grammatikalisch richtiger Satz gebildet werden. Dialektsprache wird bis zum neunten Lebensjahr als korrekt akzeptiert. Bei der Bewertung werden grammatikalische Fehler, Auslassungen der vorgegebenen Wörter, unsinnige Sätze, die Bildung von mehr als einem Satz, eine bloße Aufzählung der relevanten Wörter und die Bildung von Sätzen, in denen die Wörter nicht in einem Zusammenhang stehen, berücksichtigt. Als Besonderheit bei fremdsprachigen Kindern wird eine Antwort mit einem  $\frac{1}{2}$  Punkt bewertet, wenn nur Fehler beim Genus vorliegen.
17. *Sprache Rezeptiv* (12 Items). Vorgesprochene Sätze sollen mit Holzfiguren nachgespielt werden. Die Bewertung erfolgt dreistufig: Der Inhalt wird vollständig richtig nachgespielt, der Inhalt wird richtig, aber in falscher Reihenfolge nachgespielt, der Inhalt wird fehlerhaft nachgespielt.

*Leistungsmotivation* wird anhand von zwei Ratingskalen bewertet:

18. *Durchhaltevermögen* (4 Items). Der Testleiter schätzt das Durchhaltevermögen ein (z. B. die Fähigkeit Bedürfnisse aufzuschieben, beharrliche Aufgabenbearbeitung auch bei schwierigen Anforderungen).
19. *Leistungsfreude* (4 Items). Der Testleiter bewertet die Leistungsfreude (z. B. Freude an der Bewältigung schwieriger Aufgaben, Stolz auf die eigene Leistung).

Die Vorgabe der Subtests erfolgt bei vollständiger Testdurchführung in der genannten Reihenfolge. Bei allen Subtests erfolgt der Testeinstieg bei Item 1. Bei einigen Subtests

werden bei älteren Kindern erst ein oder mehrere Folgeitems ausgelassen, wobei eine Umkehrregel vorsieht, dass bei einem Fehler in der zweiten vorgegebenen Aufgabe Items mit niedrigerer Aufgabenschwierigkeit doch noch durchgeführt werden. Bei acht Untertests ist ein altersunabhängiges Abbruchkriterium vorgesehen, das meist bei drei Fehlern in Folge liegt. Bei vier Untertests gibt es für Aufgabenbearbeitung eine Zeitgrenze, bei zwei weiteren Untertests betrifft dies nur einzelne Items.

Der Protokollbogen ist zwar umfangreich (16 Seiten), bietet aber nicht bei allen Untertests ausreichend Raum für die Protokollierung der Antworten. Erinnerungshilfen beschränken sich weitgehend auf die Einstiegs- und Umkehrregel. Eine grafische Darstellung der Testergebnisse ist möglich, es werden auch die kritischen Werte für Profilvergleiche angegeben.

Die Durchführung des Gesamttests dauert nach Angaben der Autoren ca. 90-120 Minuten, davon ca. 45 Minuten für die Bestimmung des Intelligenzwerts, jeweils ca. 10 Minuten für die Bereiche *Psychomotorik*, *Mathematik* und *Sprache* sowie ca. 15 Minuten für den Bereich *Sozial-Emotionale Kompetenz*. Empirische Daten über die Bearbeitungsdauer in der Normstichprobe werden nicht angegeben. Die Ermittlung der Testrohwerte erfolgt anhand der Hinweise und Lösungsbeispiele im Manual. In übersichtlichen Tabellen können die zugehörigen Standardwerte abgelesen oder schneller und komfortabler mit dem mitgelieferten Auswertungsprogramm ausgegeben werden.

## Testanalyse und Normierung

Die *Durchführungsobjektivität* wird durch eine detaillierte Aufgabenbeschreibung und allgemeine Hinweise zur Gestaltung der Testsituation überwiegend, aber nicht durchgehend gesichert. Im Untertest *Wahrnehmung Visuell* fehlt eine klare Vorgabe, wie die Stimuluskärtchen auszulegen sind (das Manual verlangt eine „willkürliche“ Reihenfolge). Bei *Gedächtnis Auditiv* soll die Geschichte bei Dialekt sprechenden Kindern im Dialekt erzählt werden, was dem Untersuchungsleiter einen Ermessensspielraum bei der konkreten Umsetzung einräumt und zudem nur im Rahmen der jeweiligen Dialektkenntnisse möglich ist.

Die Instruktionen erscheinen vereinzelt verbesserungsbedürftig. Bei *Sprache Expressiv* wird 5- bis 8-Jährigen der Begriff „Satz“ mit den Worten „Ein Satz ist eine ganz kurze Geschichte“ erklärt, was nahe legt, dass auch mehrere Sätze gebildet werden dürfen. Beim Balancieren soll das Kind die Augen geschlossen halten, Verstöße werden bei der Bewertung jedoch explizit nicht sanktioniert, so dass Kinder, die sich an die Anweisung halten, benachteiligt werden.

Die *Auswertungsobjektivität* ist an einigen Stellen nicht gewährleistet: Im Untertest *Grobmotorik* sind die Kriterien bei der Bewertung des Fangens und Werfens nicht hinreichend spezifiziert. Beim Untertest *Feinmotorik* wird nicht exakt definiert, ab wann genau das Kind das Band „zu weit hinten“ hält. Ein Bewertungskriterium in diesem Untertest bezieht sich darauf, dass das Kind einen Würfel statt einer Perle

greift (oder umgekehrt). Dabei wird nicht klar beschrieben, wann dieser Fehler als erfüllt gilt. Durch die Verwendung einer unnötig hohen und engen Schachtel, aus der das Kind die Perlen zu entnehmen hat, wird der Untersuchungsleiter zudem in der Regel nicht erkennen können, ob das Kind direkt das richtige oder erst ein falsches Objekt gegriffen und wieder losgelassen hat. Bei *Sprache Expressiv* wäre es wichtig zu wissen, ob bei den älteren Kindern schon ein einzelner Dialektausdruck eine Antwort ungültig macht, oder ob dies nur bei durchgehend dialektaler Sprache gilt. Ebenso ist unklar, ob schon eine dialektal geprägte Aussprache oder erst die Verwendung spezifischer Dialektausdrücke zu sanktionieren ist. Zudem wird nicht deutlich, welche Kinder als „fremdsprachig“ gelten. Rein fremdsprachig aufwachsende Kinder können kaum gemeint sein, da der Untertest ohne Deutschkenntnisse auch ansatzweise nicht zu bewältigen ist. Im gleichen Untertest wären zumindest bei einzelnen Aufgaben detailliertere Hinweise erforderlich, welche Kriterien die Darstellung des Kindes erfüllen muss, z. B. gibt es keine Erklärung, ob das Schnuppern eines Hundes bereits bei Berührung des zu beschnuppernden Objekts hinreichend dargestellt ist oder ob zusätzlich eine Bewegung der Hundeschnauze oder ein entsprechendes Geräusch erforderlich ist. Empirische Daten zur Auswertungsobjektivität liegen nicht vor.

Die *Interpretationsobjektivität* ist durch den Bezug auf Normwerte gesichert. Für die inhaltliche Interpretation werden ausführliche weitere Hinweise gegeben, dabei wird auch auf mögliche Interpretationsfehler und die Notwendigkeit, die Testergebnisse im Kontext weiterer diagnostischer Daten zu interpretieren, hingewiesen.

Als *Reliabilitätskennwert* wird die interne Konsistenz angegeben. Außerdem finden sich Daten zur Retestreliabilität über ein Testintervall von 12-18 Monaten für eine kleine Stichprobe Schweizer Kinder ( $N = 33$ ). Bei der Bestimmung der internen Konsistenz erfolgten die Berechnungen nicht getrennt nach Altersgruppen. Damit ist eine Überschätzung der Reliabilität zu befürchten (vgl. Reynolds, 1989). Selbst mit dieser Einschränkung fallen die Werte z. T. unbefriedigend aus (bei 13 Subtests  $\alpha < .80$ , bei 5 Subtests  $\alpha < .70$ ). Die Retestreliabilität erreicht für zwei Subskalen (*Gedächtnis Auditiv*, *Denken Logisch-Mathematisch*) und den Gesamt-IQ Werte über .80 (s. Tab. 1). Für die beiden Ratingskalen liegen keine Angaben zur Beurteilerübereinstimmung vor.

Die Bewertung der *inhaltlichen Validität* fällt recht unterschiedlich aus. Einzelne Subtests beruhen auf gut eingeführten Aufgabentypen, die in der Intelligenzdiagnostik in ähnlicher Form seit Jahrzehnten als verlässliche Indikatoren der kognitiven Leistungsfähigkeit angesehen werden. Kritisch anzumerken ist, dass die von den Autoren gewählten Skalenbezeichnungen zum Teil unzureichend durch die Testaufgaben repräsentiert werden. Exemplarisch sei auf den Untertest *Wahrnehmung Visuell* verwiesen, der nur eine spezifische Anforderung – Längenvergleich als visuell-perzeptive Leistung – berücksichtigt. Dieser Untertest erlaubt daher keine Aussage über das viel umfassendere Konstrukt „Visuelle Wahrnehmung“. *Gedächtnis Auditiv* ist nicht eine rein auditive Gedächtnisprüfung, sondern verlangt das Er-

fassen eines Textes und die Beantwortung von z. T. komplexen Verständnisfragen. *Gedächtnis Phonologisch* berührt phonologische Aspekte der Sprache (z. B. Lautunterscheidung) allenfalls am Rand.

Zur *konkurrenten Validität* werden Korrelationen mit dem „Hamburg-Wechsler-Intelligenztest für Kinder – IV“ (HAWIK-IV; Petermann u. Petermann, 2007) angegeben (N = 172). Die IQ-Werte beider Verfahren korrelieren mit .69 substantiell. Auf Ebene der HAWIK-IV Indexskalen zeigt sich der niedrigste Zusammenhang zum IDS-IQ bei *Verarbeitungsgeschwindigkeit* ( $r = .34$ ) und der höchste beim *Wahrnehmungsgesunden Logischen Denken* ( $r = .62$ ). Leider werden keine Daten zu Niveauunterschieden in den beiden Verfahren berichtet.

In den Bereichen *Psychomotorik*, *Sozial-Emotionale Kompetenz* und *Leistungsmotivation* finden sich teilweise signifikante, aber niedrige Korrelationen zum Gesamt-IQ des HAWIK-IV ( $r \leq .23$ ). Die entsprechenden Korrelationen mit *Denken Logisch-Mathematisch* ( $r = .53$ ), *Sprache Rezeptiv* ( $r = .42$ ) und *Sprache Expressiv* fallen erwartungsgemäß höher aus.

Weiter berichtet das Manual Daten zum Zusammenhang zwischen den IDS und dem „Salzburger Lese- und Rechtschreibtest“ (SLRT; Landerl, Wimmer u. Moser, 2006) und dem „Schweizer Rechentest“ (SR 1-3; Lobeck u. Frei, 1987). Da die Ergebnisse der Schulleistungstests in einer nicht beschriebenen Prozedur in eine Schulnotenskala (1-6) umgewandelt wurden, ist die Aussagekraft der so erhaltenen Korrelationen ganz unklar, sie werden daher hier nicht berichtet. Der Untertest *Fotoalbum* des Wiener Entwicklungstest (WET; Kastner-Koller u. Deimann, 2002) zeigt signifikante Korrelationen zu dem inhaltlich sehr ähnlichen Subtest *Emotionen Erkennen* ( $r = .50$ ) und zu *Soziale Situationen Verstehen* ( $r = .35$ ). Zur weiteren Validierung des Funktionsbereichs *Sozial-Emotionale Kompetenzen* wurden drei Fremdratingverfahren eingesetzt. Das mit der „Coping with Emotional Situations Scale“ (CESS; Eisenberg, Fabes, Nyman, Bernzweig, Pinuelas, 1994) und der „Parent Rating Scale of Social Competence“ (PRSSC; Denham, 2005) erhobene Elternurteil zeigt überwiegend signifikante, aber niedrige Zusammenhänge zu drei Skalen des IDS-Funktionsbereiches (max. Varianzaufklärung 3 %; keine signifikanten Zusammenhänge mit *Emotionen Erkennen*). Im Lehrerurteil, erfasst mit der „Lehrereinschätzliste für Sozial- und Lernverhalten“ (LSL; Petermann u. Petermann, 2006), ergeben sich signifikante, aber wiederum nicht hohe Korrelationen zu *Soziale Situationen Verstehen* und *Sozial Kompetent Handeln* ( $r = .18$  bzw.  $.27$ ).

Zur *faktoriellen Validität* wurde eine Hauptkomponentenanalyse mit Varimaxrotation durchgeführt. Berichtet wird eine vierfaktorielle Lösung (42,1 % Varianzaufklärung), wobei das Kriterium für die Bestimmung der Faktorenzahl nicht genannt wird und die Angemessenheit dieser Lösung somit nicht beurteilt werden kann. Die Faktoren werden als Kognition, Sozial-Emotionale Kompetenz, Leistungsmotivation und Psychomotorik interpretiert. Gegenüber der Testgliederung nach Funktionsbereichen ergeben sich einige Abweichungen: Die beiden sprachlichen Subtests und *Denken Logisch-Mathematisch* zeigen die höchsten Ladungen auf dem Faktor Kognition. Auf-

*merksamkeit Selektiv* lädt mit .39 am höchsten auf dem Faktor Motorik, *Sozial Kompetent Handeln* auf dem Faktor Leistungsmotivation. Wie sich die Faktorenstruktur in den unterschiedlichen Altersgruppen darstellt, wird nicht berichtet.

Zur *differentiellen Validität* werden verschiedene Daten berichtet. Kinder, die nach Angaben ihrer Eltern in einem Intelligenztest ein weit überdurchschnittliches Ergebnis erzielt hatten und im Manual als hochbegabt bezeichnet werden, erreichten einen mittleren IQ von 117.6, während der Mittelwert von Kindern, die eine Schweizer Kleinklasse (vergleichbar mit dem Bildungsgang für Lernbehinderte) besuchten, bei 78.1 lag. Angaben zum IQ-Niveau in weiteren Schulformen finden sich nicht. Fremdsprachige Kinder aus der Normstichprobe zeigten signifikant schlechtere Leistungen als eine Kontrollgruppe in den Untertests *Sprache Expressiv*, *Sprache Rezeptiv*, *Gedächtnis Auditiv* und *Feinmotorik*. Daten zur *prognostischen Validität* werden nicht berichtet.

Kinder, bei denen eine hyperkinetische Störung nach ICD-10 diagnostiziert wurde (N = 27), unterschieden sich signifikant im allgemeinen kognitiven Leistungsniveau und in den Untertests *Aufmerksamkeit Selektiv*, *Gedächtnis Phonologisch*, *Gedächtnis Räumlich-Visuell*, *Denken Logisch-Mathematisch*, *Sprache Expressiv* und in den Skalen zur *Leistungsmotivation* von einer Kontrollgruppe. Bei 38 Kinder mit der Diagnose eines Asperger-Syndroms fanden sich signifikante Mittelwertsunterschiede für *Aufmerksamkeit Selektiv*, *Gedächtnis Auditiv*, *Grobmotorik*, *Emotionen Regulieren*, *Soziale Situationen Verstehen*, *Sozial Kompetent Handeln* und *Durchhaltevermögen*. Die Mittelwerte der Asperger-Autisten im Bereich *Sozial-Emotionale Kompetenzen* lagen zwischen 1.2 und 2.2 Wertpunkten unter denen der Kontrollgruppe. In den kognitiven Skalen zeigten sich z. T. stärkere Mittelwertsunterschiede (*Aufmerksamkeit Selektiv*, *Gedächtnis Auditiv*). Als weitere klinische Gruppe wurden 57 Kinder mit aggressiven Verhaltensauffälligkeiten mit einer Kontrollgruppe verglichen. Signifikant niedrigere Werte bei den aggressiven Kindern fanden sich in 15 Skalen und im Gesamt-IQ.

Die Signifikanztestungen erfolgten übrigens bei allen Gruppenvergleichen durch 19 paarweise t-Tests, ohne dass das Signifikanzniveau für multiple Vergleiche adjustiert wurde. Effektstärken werden nicht angegeben. Es finden sich auch keine statistischen Prüfungen zur Frage, ob *zwischen* den klinischen Gruppen Leistungsunterschiede bestehen, die für die Testinterpretation bei differentialdiagnostischen Fragestellungen relevant wären.

Die *Normierung* erfolgte von August 2007 bis Juli 2008. Die Teilnehmer wurden über verschiedene Bildungseinrichtungen rekrutiert, indem Informationsbriefe verteilt wurden. Eltern, die eine Einverständniserklärung abgaben, wurden anschließend zur Terminvereinbarung kontaktiert. Inwieweit es durch dieses Vorgehen zu Selektionseffekten kam, kann anhand der spärlichen Angaben im Manual nicht beurteilt werden. Die Stichprobe umfasst 1.330 Kinder (662 Jungen und 668 Mädchen), davon 344 aus Deutschland (Raum Bremen, Jena und Freiburg), 239 aus Österreich (Raum Salzburg und Wien) und 747 aus der deutschsprachigen Schweiz. Die Testungen wurden von Schulpsychologen sowie Masterstudierenden und Diplomanden der Psychologie durchgeführt.

Standardnormen werden für den Altersbereich von 5;0 bis 10;11 Jahren angeboten. Für die Subtests wird die gebräuchliche Wertpunktskala ( $M = 10$ ,  $SD = 3$ , Wertebereich  $\pm 3$  SD) verwendet. Für die Gesamtleistung der kognitiven Untertests stehen IQ-Werte ( $M = 100$ ,  $SD = 15$ , Wertebereich  $\pm 3$  SD) und Prozentränge zur Verfügung. 90%-Vertrauensintervalle werden nur für den IQ angegeben. Die Altersabstände der zwölf Normgruppen betragen jeweils sechs Monate, wodurch eine gute Altersdifferenzierung gegeben ist.

Die Beurteilung von Leistungsprofilen auf Ebene der Funktionsbereiche erfolgt nach den Vorgaben des Manuals in zwei Schritten: Zuerst wird der Mittelwert der Wertpunkte pro Funktionsbereich gebildet. Im zweiten Schritt wird ein Wertpunktband um diesen Mittelwert konstruiert, indem die Standardabweichung der Wertpunktsumme des Funktionsbereiches addiert bzw. subtrahiert wird. Ein Profil wird dann als homogen bezeichnet, wenn alle Subtestergebnisse innerhalb des Wertpunktbandes liegen. Analog wird bei der Beurteilung intraindividuelle Stärken und Schwächen in der Gesamtentwicklung vorgegangen, also beim Vergleich des Mittelwerts eines Funktionsbereiches mit dem Mittelwert aller Funktionsbereiche. Diese Methode berücksichtigt nicht die Reliabilität der Untertests, verzichtet auf eine zufallskritische Absicherung der Profildifferenzen und entspricht damit nicht den notwendigen statistischen Standards. Eigene Nachberechnungen nach der Methode von Silverstein (1982), die allerdings nur auf den nicht nach Altersjahrgängen differenzierten Reliabilitätskennwerten beruhen können, zeigen z. B. für den Entwicklungsbereich Psychomotorik, dass auf dem 90%-Fehlerniveau erst Abweichungen von 3.06 bis 3.37 Wertpunkten als statistisch bedeutsam zu beurteilen wären – eine deutliche Diskrepanz gegenüber dem im Manual angegebenen Wert von 2.0 Wertpunkten.

Angaben über die Häufigkeit von Profildifferenzen in der Normstichprobe finden sich im Manual nicht. Statistisch signifikante Diskrepanzen können auch in Normalpopulationen häufig auftreten und sind allein kein Hinweis auf zugrundeliegende Pathologie.

Kritische Bodeneffekte finden sich in den jüngsten Altersgruppen bei mehreren Untertests. Bei einem Rohwert von Eins erhält z. B. ein 5;3-jähriges Kind sechs Wertpunkte in *Denken Konzeptuell* und acht Wertpunkte in *Sprache Expressiv*. Deckeneffekte zeigen sich z. T. schon bei Fünfjährigen, z. B. können im Subtest *Emotionen Erkennen* maximal 15 Wertpunkte erreicht werden. Boden- und Deckeneffekte sind für die klinische Anwendung und Interpretation relevant, weil sie dazu führen können, dass die Leistungsfähigkeit des Kindes über- oder unterschätzt wird. Zudem können sich bei unterschiedlicher Ausprägung von Boden- und Deckeneffekten Profildifferenzen ergeben, die nicht in der Begabungsstruktur des Kindes begründet sind. In den Normtabellen des IDS wird auf die stärksten Boden- und Deckeneffekte hingewiesen, die Implikationen für die Profilanalyse werden nicht thematisiert. Minimale Rohwertveränderungen können bei mehreren Untertests zum Anstieg der Skalenwerte um zwei Wertpunkte führen (ungünstige Itemgradienten; s. Bracken, 1988).

## Bewertung

Eine umfassende Diagnostik psychischer Funktionen erfolgt in der klinischen Praxis häufig durch eine auf die jeweilige Fragestellung zugeschnittene Kombination verschiedener Testverfahren, deren Ergebnisse neben Anamnese, Exploration und Verhaltensbeobachtung in ein klinisches Urteil integriert werden. Allerdings können bei diesem Vorgehen intraindividuelle Leistungsdiskrepanzen auch auf Unterschiede der Normierungen (Stichprobenzusammensetzungen, Erhebungszeitpunkte) zurückzuführen sein und sind daher nicht eindeutig interpretierbar. Der potentielle Vorteil der IDS besteht daher in der besseren Vergleichbarkeit der Subtestergebnisse auf Basis einer gemeinsamen Normierung. Außerdem liegen bei Einsatz unterschiedlicher Testverfahren selten Daten zu deren Skaleninterkorrelationen, zur gemeinsamen Faktorenstruktur oder zu Testbefunden in klinischen Gruppen vor. *Diese Vorteile kommen allerdings nur zum tragen, wenn ein mehrdimensionales Testverfahren insgesamt und für jeden der überprüften Bereiche die erforderlichen testmethodischen und testpraktischen Standards erfüllt*, was bei den IDS nur z. T. der Fall ist. Dies soll im Folgenden erst für die einzelnen Entwicklungsbereiche und anschließend für einige übergreifende Themen ausgeführt werden.

Die Anforderungen im Bereich *Kognitive Entwicklung* reichen von der Überprüfung der kurzzeitigen visuellen selektiven Aufmerksamkeit bis zu einer komplexen sprachlichen Verständnis- und Gedächtnisleistung. Bei einer nachträglichen Einordnung der Subtests in das Intelligenzmodell der CHC-Theorie (McGrew, 2005; s. a. Mickley u. Renner, 2010) zeigt sich der deutlichste Bezug zum Faktor Visuelle Verarbeitung (Gv; Subtests *Wahrnehmung Visuell*, *Gedächtnis Räumlich-Visuell* und *Denken Bildlich*). Weitere CHC-Faktoren sind nur mit einem Subtest vertreten: Kurzzeitgedächtnis (Gsm; *Gedächtnis Phonologisch*) und Fluide Intelligenz (Gf; *Denken Konzeptuell*). *Gedächtnis Auditiv* überprüft inhaltlich Sprachverständnis und Langzeitgedächtnis und ist damit innerhalb des CHC-Modells bei Bezügen zu Kristallinen Intelligenz (Gc) und zum Faktor Langzeitspeicherung und Abruf (Glr) nicht eindeutig interpretierbar. Bei *Aufmerksamkeit Selektiv* könnte allenfalls ein Bezug zu Verarbeitungsgeschwindigkeit (Gs) gesehen werden. Zur Bewertung der Kristallinen Intelligenz (Gc) können auch die Subtests des Bereiches Sprache herangezogen werden. Die Untertests *Wahrnehmung Visuell* und *Gedächtnis Auditiv* erweitern das diagnostische Repertoire, da sich vergleichbare Anforderungen in anderen mehrdimensionalen Intelligenztestbatterien für diesen Altersbereich nicht finden.

Mit der substantiellen Korrelation zum HAWIK-IV ist die konvergente Validität für den Gesamt-IQ belegt. Der Untertest *Denken Konzeptuell* sollte nicht – wie im Testmanual impliziert – als sprachunabhängig betrachtet werden, da verbale Vermittlungsprozesse bei der Bearbeitung durchaus eine Rolle spielen können. Wesentliche Vorteile der Skala *Kognitive Entwicklung* gegenüber etablierten mehrdimensionalen Intelligenztests lassen sich nicht erkennen.

Die Untertests des Bereichs *Psychomotorik* zeichnen sich durch nicht befriedigende Reliabilitätswerte aus, auch Verbesserungen in der Durchführungs- und Auswertungsobjektivität sind zu wünschen. Eine erste Einschätzung motorischer Fähigkeiten, die ggf. durch spezifische weitere Untersuchungsverfahren abgesichert werden muss, wird jedoch möglich sein. Die Ergebnisse im *Untertest Feinmotorik* könnten aufgrund der spezifischen Aufgabenstellung auch durch exekutive Leistungen mitbestimmt werden.

Da standardisierte Verfahren zur Erfassung *sozial-emotionaler Kompetenzen* in Deutschland kaum zur Verfügung stehen, ist die Berücksichtigung dieses Entwicklungsbereiches in den IDS vermutlich für viele potentielle Testanwender ein sehr attraktives Merkmal. Die berichteten Daten zu den Gütekriterien sind allerdings ernüchternd. Die Reliabilität erreicht keine befriedigenden Werte, die konvergente Validität zu Eltern- und Lehrerurteilen fällt zu niedrig aus. Nur bei der differenziellen Validität für klinische Gruppen deuten sich einige stimmige Tendenzen an. Generell ist die Frage zu stellen, ob Testaufgaben ohne Zeitdruck und ohne emotionale Involvierung die komplexe und dynamische Wirklichkeit sozialer Situationen adäquat abbilden können.

Auch inhaltliche Bedenken sind angebracht: Im Untertest *Emotionen Regulieren* sollen allein aufgrund eines Gefühlzustandes ohne Berücksichtigung des Kontextes Regulationsstrategien entwickelt werden, die dann als mehr oder weniger adäquat bewertet werden. Dieses Vorgehen verkennt, dass die Regulation von Gefühlszuständen ja nicht nur von der erlebten Emotion, sondern auch von ihrer Entstehung und den Handlungsmöglichkeiten abhängt. Es ist ein Unterschied, ob ein Kind bei der Frage nach der Regulation eines traurigen Gefühls an ein kleines Missgeschick oder den Tod eines Elternteils denkt. Auch im Untertest *Sozial Kompetent Handeln* erscheint die Bewertung zu schematisch. Generell wird nur aktives Handeln mit zwei Punkten honoriert. So genannte indirekte Strategien (Hilfe holen, aus dem Feld gehen) werden pauschal und ohne Berücksichtigung des Kontextes dagegen nur mit einem Punkt bewertet. Ob ein 5-Jähriger wirklich gut beraten ist, sich bei Bild C (ein aggressiv wirkender Jugendlicher) auf eine verbale Auseinandersetzung einzulassen, kann bezweifelt werden.

Der Entwicklungsbereich *Mathematik* wird durch einen Untertest mit befriedigender Reliabilität erfasst, der erste Hinweise auf eine mögliche Rechenstörung liefern kann. Auch im Bereich *Sprache* erreichen die Subtests befriedigende Reliabilitäten. Es ist zu begrüßen, dass auch das Sprachverständnis explizit überprüft wird. Auf Unklarheiten bei Instruktion und Auswertung wurde bereits hingewiesen. Im Sinne der Testfairness gegenüber Dialekt sprechenden Bevölkerungsgruppen wäre ein Nachweis zu wünschen, dass die in den Auswertungsrichtlinien verlangte Verwendung der Hochsprache im Subtest *Sprache Expressiv* die Validität der Testergebnisse nicht gefährdet.

Im Gegensatz zu den anderen Entwicklungsbereichen wird die *Leistungsmotivation* anhand des Verhaltens in der Testsituation bewertet. Leider fehlen Angaben zur Beurteilerübereinstimmung. Es ist außerdem noch unklar, inwieweit diese Einschätzungen über die Testsituation hinaus generalisiert werden können.

Die Belege für die Validität des Verfahrens fallen je nach Entwicklungsbereich recht heterogen aus. Wünschenswert wären ergänzende Angaben zur prognostischen Validität und nachvollziehbar ausgewertete Daten zum Zusammenhang mit Schulleistungen.

Für leistungsschwache 5-Jährige sind die IDS wegen der bestehenden Bodeneffekte weniger geeignet, z. T. gilt dies auch für die 6-Jährigen. Obwohl die IDS für einen breiten Einsatz in der klinischen Diagnostik vorgesehen sind, finden sich im Manual keine Hinweise zur Durchführung des Verfahrens bei behinderten Kindern.

Mehrfach fallen Mängel bei statistischen Auswertungen auf, die sich z. B. in unzureichenden Angaben zu den Faktorenanalysen, fehlender Anpassung des Signifikanzniveaus bei multiplen Prüfungen und den nicht nachvollziehbaren Transformationen der Ergebnisse von Schulleistungstests zeigen. Testwertdiskrepanzen werden nicht durch Signifikanzprüfungen abgesichert, außerdem fehlen Daten über ihr Auftreten in der Normierungsstichprobe. Auf dieser Grundlage sollten keine weitreichenden inhaltlichen Profilverinterpretationen vorgenommen werden. Die Autoren liefern über einige plausible Vermutungen hinaus keine empirischen Belege, dass bestimmte Profilkonstellationen im IDS *zusätzlich* zur normorientierten Testauswertung zu einer verbesserten Diagnosestellung und Behandlungsplanung beitragen.

Zu wünschen wären auch ergonomische Verbesserungen. Die z. T. kuriosen Stellungnahmen auf der IDS-Internetseite („Damit die Kärtchen [des Untertests *Visuell Selektiv*] nicht durcheinander kommen, empfiehlt es sich, den Tennisball in die Schachtel der Kärtchen zu legen“; Verlag Hans Huber, 2011b) lassen leider befürchten, dass beim Verlag kein Bewusstsein dafür besteht, dass ein funktionales Testmaterial bei einem Verfahren dieser Preiskategorie selbstverständlich sein sollte.

Die Schwächen der IDS sind insofern besonders bedauerlich, als ein Verfahren zur Überprüfung vielfältiger kognitiver und nicht-kognitiver Funktionen den Bedürfnissen der klinisch-diagnostischen Praxis sehr entgegenkäme. Einige der genannten Mängel ließen sich durch Nachberechnungen schnell beheben. Der Verlag wäre gut beraten, nach Altersgruppen getrennt berechnete Reliabilitäten, eine statistisch adäquate Methode zur Profilanalyse, Daten zur Häufigkeit von Profildifferenzen in der Normstichprobe, nachvollziehbare Berechnungen zum Zusammenhang mit Schulleistungstests und Niveauevergleiche mit dem HAWIK-IV nachzureichen.

## Literatur

- Bracken, B. A. (1988). Ten psychometric reasons why similar tests produce dissimilar results. *Journal of School Psychology*, 26, 155-166.
- Eisenberg, N., Fabes, R. A., Nyman, M., Bernzweig, J., Pinuelas, A. (1994). The relation of emotionality and regulation to preschoolers' anger related reactions. *Child Development*, 65, 1352-1366.
- Denham, S. A. (2005). Parent rating scale of social competence. Unveröffentlichtes Manuskript, National Institute for Child Health and Human Development at Bethesda, MD, USA.

- Landerl, K., Wimmer, H., Moser, E. (2006). Salzburger Lese- und Rechtschreibtest (SLRT) (2., korr. u. akt. Aufl.). Bern: Huber.
- Lobeck, A., Frei, M. (1987). Schweizer Rechentest 1.-3. Klasse (SRT 1-3). Göttingen: Hogrefe.
- 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.
- 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.
- Petermann, U., Petermann, F. (2006). Lehrereinschätzliste für Sozial- und Lernverhalten (LSL). Göttingen: Hogrefe.
- Reynolds, C. R. (1989). Measurement and statistical problems in neuropsychological assessment of children. In C. R. Reynolds, E. Fletcher-Janzen (Hrsg.), *Handbook of Clinical Child Neuropsychology* (S. 147-166). New York: Plenum Press.
- Silverstein, A.B. (1982). Pattern Analysis as Simultaneous Statistical Inference. *Journal of Consulting and Clinical Psychology*, 50, 234-240.
- Verlag Hans Huber (2011a). Fragen zur Auswertung. Zugriff am 06.05.2011 <http://www.intelligence-and-development-scales.com/de/praxis/haeufig-gestellte-fragen/fragen-zur-auswertung.html>
- Verlag Hans Huber (2011b). Fragen zum Material. Zugriff am 06.05.2011 <http://www.intelligence-and-development-scales.com/de/praxis/haeufig-gestellte-fragen/fragen-zum-material.html>

Gerolf Renner, Birkenfeld