

Kurt A. Heller:(Herausgeber)

Begabungsdiagnostik in der Schul- und Erziehungsberatung

Mit Beiträgen von

Ernst A. Hany, München

Kurt A. Heller, München

Christoph Perleth, München

Wolfgang Sierwald, München

Verlag Hans Huber
Bern Göttingen Toronto

Universitäts-
Bibliothek
München

45398542

Die Deutsche Bibliothek – CIP-Einheitsaufnahme

Begabungsdiagnostik in der Schul- und Erziehungsberatung /
hrsg. von Kurt A. Heller. Mit Beitr. von Ernst A. Hany ... –
1. Aufl. – Bern ; Göttingen ; Toronto : Huber, 1991
(Psychologie-Lehrbuch)
ISBN 3-456-81964-1
NE: Heller, Kurt [Hrsg.]; Hany, Ernst A.



1. Auflage 1991
© Verlag Hans Huber, Bern
Gesamtherstellung: Allgäuer Zeitungsverlag GmbH, Kempten/Allgäu
Printed in Germany

Inhaltsverzeichnis

1.	Einführung in den Gegenstandsbereich der Begabungsdiagnostik (K.A. Heller)	11
1.1	Historische Aspekte	11
1.2	Zur Krise der psychologischen Diagnostik und ihre Überwindung . .	13
1.3	Begabungsdiagnostische Funktionen und Entscheidungsstrategien . .	15
1.4	Psychologische Theorien über Intelligenz und Begabung	18
1.4.1	Ältere Intelligenzkonzepte	19
1.4.2	Neuere Begabungstheorien	19
1.4.3	Faktorentheorien	22
1.4.4	Kognitionspsychologische Begabungskonzepte	32
2.	Testtheoretische Konzepte der Begabungsdiagnostik (Ch. Perleth & W. Sierwald)	37
2.0	Statistische Grundbegriffe	38
2.1	Theoretische Grundlagen psychologischer Messungen	41
2.1.1	Was ist eine Messung?	41
2.1.2	Meßmodelle	45
2.1.3	Meßtheorie und (psychologische) Testtheorien	45
2.2	Klassische Testtheorie und Testgütekriterien	48
2.2.1	Überblick	48
2.2.2	Grundlegende Begriffe und Definitionen der Klassischen Testtheorie	50
2.2.3	Grundlegende Eigenschaften der gemessenen, der wahren und der Fehlerwerte	52
2.2.4	Reliabilität	54
2.2.4.1	Mathematische Definitionen der Reliabilität	54
2.2.4.2	Praktische Verfahren der Reliabilitätsbestimmung	56
2.2.4.3	Interpretation von Reliabilitätskoeffizienten	58
2.2.4.4	Die Reliabilität eines einzelnen Ergebnisses	61
2.2.5	Validität	63
2.2.5.1	Inhaltliche Validität	64
2.2.5.2	Kriteriumsvalidität	65
2.2.5.3	Konstruktvalidität	67
2.2.6	Testkonstruktion	70
2.2.6.1	Ziele und Phasen der Testkonstruktion innerhalb der Klassischen Testtheorie	70
2.2.6.2	Itemanalyse und Itemselektion im Sinne der Klassischen Test- theorie	73
2.2.7	Kritik an der Klassischen Testtheorie	75
2.3	Probabilistische Testmodelle	78
2.3.1	Grundannahmen probabilistischer Testmodelle	78
2.3.2	Grundlegende Annahmen und Eigenschaften des Rasch-Modells . .	79

2.3.3	Weitere probabilistische Testmodelle	83
2.4	Das Lerntestkonzept	84
2.5	Normierung	88
3.	Informationsquellen und Meßinstrumente (K.A. Heller & Ch. Perleth)	94
3.1	Verhaltensbeobachtung und Ratingverfahren	94
3.1.1	Beobachtungsverfahren	94
3.1.2	Rating- oder Schätzverfahren	98
3.2	Psychodiagnostisches Gespräch (Anamnese und Exploration)	102
3.2.1	Zur Formulierung der Fragen	104
3.2.2	Planung und Durchführung psychodiagnostischer Einzelfallgespräche	105
3.2.3	Zur Verarbeitung gesprächsdiagnostischer Informationen	106
3.3	Tests zur Erfassung kognitiver Fähigkeiten	108
3.3.1	Tests zur Erfassung der Grund- oder allgemeinen Intelligenz	110
3.3.1.1	Individualverfahren	110
3.3.1.2	Gruppenverfahren	141
3.3.2	Tests zur Erfassung differentieller kognitiver Fähigkeiten	153
3.3.3	Spezielle Fähigkeitstests	173
3.3.4	Kreativitätstests	177
3.4	Tests zur Erfassung begabungsrelevanter nichtkognitiver Persönlich- keitsmerkmale	182
3.4.1	Konzentrationstests	182
3.4.2	Motivationstests	190
3.4.3	Interessentests	193
3.4.4	Weitere Verfahren zur Erfassung nichtkognitiver Persönlichkeits- merkmale	201
3.5	Skalen zur Erfassung des sozialen Lernumfeldes	208
4.	Diagnose, Prognose und Entscheidung	213
4.1	<i>Schuleignungsprognostik</i> (K.A. Heller)	213
4.1.1	Schuleignungsbestimmung als Funktion der Schullaufbahnberatung	214
4.1.2	Zur multikausalen Bedingungsstruktur von Schulleistung und Schulerfolg	215
4.1.3	Methodenprobleme der Schuleignungsprognose	219
4.1.3.1	Definition des Kriteriums	219
4.1.3.2	Auswahl und Erfassung der Prädiktoren	220
4.1.3.3	Erstellung der Prognose	221
4.1.4	Entscheidungsstrategien	222
4.1.5	Prognosekriterium und Beurteilungsmaßstäbe für die Eignungsfest- stellung in der Schullaufbahnberatung	225

4.2	<i>(Begabungs-)Diagnostik bei Schülern mit Lernbehinderungen</i> (Ch. Perleth)	236
4.2.1	Lernbehinderung und (Minder-)Begabung	237
4.2.1.1	Traditionelle Definitionen von Lernbehinderung	237
4.2.1.2	Moderne Sichtweisen von Lernbehinderung	240
4.2.2	Aufnahmediagnostik (Lernbehindertenschule)	244
4.2.3	Zur Verwendung von (Begabungs-)Tests im Bereich sonderpädagogischer Diagnostik	260
4.2.4	Spezielle förderdiagnostische Maßnahmen	268
4.3	<i>Hochbegabungsdiagnostik</i> (K.A. Heller)	277
4.3.1	Theoretische Vorbemerkungen	277
4.3.2	Funktionen der Hochbegabungsdiagnostik	279
4.3.3	Methodenprobleme	281
4.3.3.1	Meßprobleme	281
4.3.3.2	Status- oder prozeßdiagnostische Tests in der Hochbegabungsdiagnostik	285
4.3.3.3	Entscheidungsstrategien in der Hochbegabungsdiagnostik	286
4.3.4	Hochbegabungsdiagnose als hypothesengeleiteter Prozeß	289
5.	Der diagnostische Prozeß (E.A. Hany)	292
5.1	Grundlagen des psychodiagnostischen Prozesses	292
5.1.1	Definition	292
5.1.2	Komponenten des diagnostischen Prozesses	293
5.1.3	Kontext psychodiagnostischen Handelns	295
5.1.4	Gütekriterien	297
5.1.5	Zusammenfassung	298
5.2	Präskriptive/normative Modelle des diagnostischen Prozesses	298
5.2.1	Das pädagogisch-psychologische Handlungsmodell von Krapp	299
5.2.2	Ablaufmodell der pädagogisch-psychologischen Begutachtung	301
5.2.3	Diagnostische Algorithmen	303
5.2.4	Diagnostische Kompetenzen	305
5.2.5	Zusammenfassung	306
5.3	Untersuchungen zur menschlichen Informationsverarbeitung	306
5.3.1	Subjektive Wahrscheinlichkeit	307
5.3.2	Subjektive Informationsintegration	310
5.3.3	Subjektive Entscheidungsstrategien	314
5.3.4	Subjektives Hypothesentesten	316
5.3.5	Zusammenfassung	317
5.4	Diagnostik als Hypothesenprüfen	318
5.4.1	Diagnose und psychologische Gesetze	319
5.4.2	Hypothesen	320
5.4.3	Operationalisierung der Hypothesen	322
5.4.4	Hypothesenentscheidung	323

5.4.5	Zusammenfassung	328
5.5	Randbedingungen des diagnostischen Prozesses	329
5.5.1	Rechtliche Aspekte	329
5.5.2	Ethische Aspekte	330
5.5.3	Institutionelle Aspekte der Diagnostik	331
5.5.4	Verwertungszusammenhang	332
5.5.5	Zusammenfassung	333
6.	Psychologische Begutachtung (E.A. Hany)	334
6.1	Begriffsklärungen	334
6.1.1	Begutachtung, Gutachten, gutachterliche Stellungnahme	334
6.1.2	Begutachtung und diagnostischer Prozeß	335
6.1.3	Arten von Gutachten	336
6.1.4	Voraussetzungen der Begutachtung	338
6.1.5	Funktionen des Gutachten	339
6.1.6	Rezeption von Gutachten	340
6.2	Struktur des Begutachtungsprozesses	342
6.2.1	Etablierung des Beratungskontaktes des Gutachtens	342
6.2.2	Festlegung der diagnostischen Zielsetzung	343
6.2.3	Datenerhebung und -auswertung	346
6.2.4	Formulierung der Befunde	349
6.2.5	Maßnahmenplanung	353
6.2.6	Kommunikation der Ergebnisse	354
6.3	Aufbau des Gutachtens	354
6.3.1	Identifikationsangaben	355
6.3.2	Ausgangssituation der Begutachtung	356
6.3.3	Eigene Untersuchung	356
6.3.4	Gesamtbefund, Stellungnahme	356
6.3.5	Fehlende Informationen	358
6.3.6	Stellungnahme, Maßnahmenempfehlung	358
6.4	Gestaltung des Gutachtentextes	359
6.5	Beispiele für Begutachtungen	362
6.5.1	Fall aus der schulpyschologischen Praxis	362
6.5.2	Fall aus der Beratungslehrerpraxis	373
7.	Literaturverzeichnis	386
8.	Personenregister	413
9.	Sachregister	419
10.	Testregister	426

3. Informationsquellen und Meßinstrumente

Kurt A. Heller & Christoph Perleth

Zuverlässige und differenzierte Begabungsdiagnosen sind auf unterschiedliche Informationsquellen angewiesen. Diese lassen sich in Anlehnung an Cattell (1973) in Life-Daten (L-Daten), Questionnaire-Daten (Q-Daten) und Testdaten (T-Daten) einteilen.

Mit **L-Daten** werden hier alle diagnostischen Informationen zusammengefaßt, die über die Begabung von Kindern und Jugendlichen in realen Lebenssituationen Auskunft geben können. Einschlägige Erfassungsinstrumente sind die Verhaltensbeobachtung und sog. Ratingverfahren, d.h. (halb-)standardisierte Methoden der Fremd- oder Selbstbeurteilung.

Q-Daten stammen von standardisierten Fragebögen (Questionnaires) oder auch von nichtstandardisierten psychodiagnostischen Gesprächen, die sowohl Explorations- als auch Anamnesedaten liefern. Meßtheoretisch wären jene standardisierten Fragebogenskalen zur Erfassung nichtkognitiver Persönlichkeitsmerkmale, deren Bearbeitung auf Selbstratings basiert, allerdings eher der L-Informationsquelle zuzuordnen. Dies trifft für die meisten sog. subjektiven (Mittenecker, 1982), aber standardisierten Persönlichkeitstests (z.B. Angst- und Motivationsskalen) zu, die nicht mit den projektiven Persönlichkeitsverfahren, z.B. Thematischer Apperzeptions-Test (TAT), Rorschach-Test (RT) oder Satzergänzungstests (vgl. Kornadt & Zumkley, 1982; Spitznagel, 1982) verwechselt werden dürfen. Soweit erstere für die **Begabungsdiagnostik** in der Schul- und Erziehungsberatung relevant sind, werden sie in Abschnitt 3.4 unten besprochen.

T-Daten betreffen hier allgemeine vs. differentielle Intelligenz- oder (kognitive) Fähigkeitstests sowie Kreativitätstests und denkprozeßdiagnostische Verfahren, wie sie in Abschnitt 3.3 ausführlicher zur Darstellung kommen.

3.1 Verhaltensbeobachtung und Ratingverfahren

3.1.1 Beobachtungsverfahren

Verhaltensbeobachtungen können vielfältige Informationen liefern. In der Begabungsdiagnostik sollen damit nicht nur Aussagen über das Leistungsverhalten einer Person und somit Aufschlüsse über das Zusammenspiel von kognitiven und nichtkognitiven Persönlichkeitsmerkmalen, sondern auch Einblicke in den situationalen und sozialen Bedingungskontext gewonnen werden. Für die Analyse sozialer Interaktionen, z.B. im Erziehungs- oder Unterrichtsgeschehen, ist die Verhaltensbeobachtung ohnehin die Methode der Wahl. Auf diese wird man deshalb kaum in einer diagnostischen Untersuchung verzichten können.

Die **Beobachtungsmethode** (observational technique) ist durch drei Kriterien, die auch als Prozeßabschnitte aufgefaßt werden können, charakterisiert:

- (1) **Beobachtungsakt**, d.h. "absichtliche, aufmerksam-selektive Art des Wahrnehmens, die ganz bestimmte Aspekte auf Kosten der Bestimmtheit von anderen beachtet" (Graumann, 1973, S. 15). **Beobachtung** in diesem Sinne meint also nicht bloße Wahrnehmung, sondern auf bestimmte Verhaltensaspekte oder Handlungsausschnitte fixierte Aufmerksamkeit, was prinzipiell sowohl für Selbst- als auch für Fremdwahrnehmungen gilt.
- (2) **Beschreibung oder Protokollierung** der Beobachtungsdaten, womit ein Kernproblem der Beobachtungsmethode angesprochen ist. Nach Graumann (1964, S. 90ff.) sind folgende **Beschreibungsmodi** möglich, wobei jedoch nur die beiden ersten Stufen wissenschaftlich zulässig sind:

- Das **verbale Niveau**. Hiermit ist eine "reine" Beschreibung prozessualer Abläufe gemeint. Dieser Verbalisierungsstil ist arm an Adverbien, womit Qualifikationen weitgehend vermieden werden sollen. Beispielhaft wären etwa folgende Protokollsätze: "Proband A faßt wiederholt (mehrmals, fortlaufend, viermal innerhalb von fünf Minuten) nach dem Arm der Testleiterin" - "Probandin B stellt während der Testdurchführung immerzu Fragen, auch nach wiederholter Aufgabeninstruktion" - "Proband C kaut an seinem Bleistift. Er wischt sich öfters den Schweiß von der Stirn, wobei sein Gesicht rot anläuft. Er spricht trotz Redeverbots mit dem Banknachbarn..."
- Das **adverbiale Niveau**. Hiermit sind nun bereits erste Qualifikationen der Handlungsabläufe erfaßt, wie folgende Beispiele verdeutlichen mögen: "Proband A greift zitternd nach dem Arm der Testleiterin" - "Probandin B stellt **unangefochten** durch die Zurechtweisung des Testleiters fortgesetzt Fragen" - Einen Schritt weiter gehen etwa die folgenden Beschreibungen: "Proband C kaut **verlegen** an seinem Bleistift. Er sitzt mit **fiebrigem (gerötetem, angespanntem Gesichtsausdruck)** über seiner Testaufgabe. Er neigt **Hilfe suchend (schielend, heimlich)** den Kopf zum Nachbarn..."

Während für die Fälle A und B relativ eindeutige Qualifikationen in den Beschreibungen gegeben werden, dürfte im dritten Beispielfall zwischen verschiedenen Beobachtern bzw. Protokollauswertern ein Konsens zunehmend erschwert werden, insofern hier bereits erste Schlußfolgerungen auf die den betr. Handlungen zugrunde liegenden Motive oder Ursachen in die Beschreibung miteinfließen. Sind Schweißperlen, Gesichtsrötung usw. Anzeichen für starke Konzentration, Unsicherheit, Ängstlichkeit oder eine fiebrige Erkältung des Schülers? Deutet das Kopfnicken die Absicht zu mogeln an oder ist es Ausdruck intensiven Nachdenkens? Eine Entscheidung hierüber sollte nicht schon in der Protokollierung, sondern erst in der (späteren) Interpretationsphase erfolgen.

- Das **adjektivische Niveau**. Genau betrachtet, handelt es sich bei diesem Modus nicht mehr um eine Beschreibung, sondern bereits um eine Bewer-

tung des beobachteten Verhaltens. Beschreibung und Beurteilung (Interpretation) sind aber wegen der Kontrollierbarkeit der Urteilsfindung im diagnostischen Prozeß nicht zu vermengen. Deshalb müssen Adjektive - und erst recht Substantive (siehe unten) - in der Protokollierungsphase vermieden werden. Beispiele für "adjektivische" Verwendungen: "Proband A ist ein furchtsamer (ängstlicher ...) Schüler" - "Probandin B ist pedantisch (nörglerisch ...)" - "Proband C ist aufgeregt (unsicher, leistungsschwach ...)" versus "erkältet..." Was eigentlich geschehen ist, erfährt man aus solchen Protokollsätzen nur undeutlich oder überhaupt nicht mehr. Nicht wenige Protokolle über Verhaltensbeobachtungen sind voll von "Vor-"Urteilen dieser Art, deren Vermeidung offensichtlich - auch geübten Beobachtern - schwer fällt. Verfrühte Deutungsversuche gefährden aber eine gesicherte diagnostische Urteilbildung und sind deshalb zu unterlassen.

- Das **substantivische** Niveau. Auf dieser Ebene münden Verhaltens-"Beschreibungen" vollends in bloße Klassifikationen, wie sie für Typologien kennzeichnend sind. Protokollsätze wie "Bei Proband A handelt es sich um einen Phobiker" oder "Probandin B ist eine Pedantin (Querulantin ...)" oder "Proband C ist ein nervöser Charakter. Er zeigt deutliche Züge von Leistungsinsuffizienz" vermitteln kaum noch konkrete oder genauere Vorstellungen über das Gemeinte, womit Validitätskontrollen praktisch unmöglich werden. Die Gefahr, daß es hierbei zu fehlerhaften Diagnosen kommt, ist deshalb sehr groß.

Es sollte deutlich geworden sein, daß wissenschaftlich vertretbare Beschreibungen konkreter Verhaltensbeobachtungen nur auf dem verbalen oder adverbialen Niveau legitim sind. Darüber hinaus gehende Deutungen gehören in die folgende Phase der Informationsverarbeitung.

- (3) In der **Interpretationsphase**, wo der legitime Ort für diagnostische Urteilsbildungen ist, werden alle Beobachtungsdaten berücksichtigt. Diese sind im Hinblick auf den Untersuchungsanlaß bzw. die Untersuchungshypothesen nach den üblichen Prinzipien der Befunddeutung zu ordnen (z. B. nach Konkordanz oder Inkonsistenz der Einzelbelege/Protokollsätze, Hypothesenbezogenheit, temporären vs. überdauernden Verhaltensmerkmalen) und erforderlichenfalls zu gewichten. So verstärken gleichsinnige Aussagen über identische Verhaltensmerkmale in unterschiedlichen Beobachtungssituationen die Annahme überdauernder Verhaltenseigenschaften oder auch von Dispositionen wie Leistungsmotivation, Interesse, Konzentrationsfähigkeit, Denkkompetenzen, kreative vs. intellektuelle Fähigkeiten; ausführlichere Informationen hierzu finden sich in den nachstehenden Buchkapiteln.

Bei der Beobachtung und Beurteilung menschlichen Verhaltens sind spezifische **Fehlerquellen** zu beachten. Als Erzfeind objektiver Beobachtungen bezeichnen Medley & Mitzel (1965) persönliche Voreingenommenheit (Vorlieben vs. Abneigungen), womit eine Verzerrungstendenz im Sinne des bereits 1920 von Thorndike erwähnten Hofeffekts gemeint ist: die vom Gesamteindruck einer z.B. als sym-

pathisch vs. unsympathisch erscheinenden Persönlichkeit hervorgerufene positive oder negative Ausstrahlung auf einzelne Verhaltensmerkmale - deshalb die Bezeichnung **Hof- oder Halo-Effekt**; engl. halo = (Licht-)Hof. Analog spricht man in bezug auf soziale Gruppen von **sozialen Vorurteilen**. In der Auswirkung verwandt und deshalb oft verwechselt mit dem Halo-Effekt ist der sog. **logische Fehler** (logical error), der auf impliziten Persönlichkeitstheorien (Hofer, 1970) basiert. Wirksam in der Beobachtung (Wahrnehmungsselektion) und der Beurteilung ("Charakter-Logik") werden hier zumeist unbewusste (implizite) Annahmen darüber, was ein bestimmtes Persönlichkeitsbild ausmacht, z.B. einer "ehrgeizigen", "intellektuellen" oder "kreativen" Persönlichkeit. Weitere Fehlerquellen sind **Projektionsfehler** (Freud), **Ähnlichkeitsfehler** (Guilford) vs. **Kontrastfehler** (Murray); ausführlicher vgl. Langhorst (1984, S. 208ff.) oder Hasemann (1983, S. 463ff.). Zur Vermeidung solcher Verzerrungstendenzen empfehlen sich für potentielle Beurteiler Trainingsseminare zur Verhaltensbeobachtung, da die Kenntnis möglicher Fehler nicht ohne weiteres schon deren Vermeidung garantiert.

In der Literatur werden verschiedene Formen der Verhaltensbeobachtung aufgeführt. Im Hinblick auf begabungsdagnostische Untersuchungen kommen vor allem folgende Varianten in Betracht:

- **Gelegenheitsbeobachtungen**. Diese Form ist etwa begleitend zu Testuntersuchungen, während der psychodiagnostischen Gesprächsführung oder bei nicht geplanten Ereignissen (event sampling) indiziert. Da solche Beobachtungen mehr oder weniger zufällig, d.h. nicht systematisch geplant, stattfinden, stellen sich besondere Anforderungen sowohl an die Wahrnehmungsfähigkeit als auch an die Protokollierung, die hierbei gewöhnlich frei - also nicht an vorgegebene Beschreibungskategorien gebunden - erfolgt. Neben den genannten Fehlerquellen wird deshalb die phänomengetreue Beschreibung sehr oft zum Hauptproblem. Andererseits würde ein Verzicht auf diese Form der Verhaltensbeobachtung nicht selten mit dem Verlust von diagnostisch bedeutsamen Informationen einhergehen.
- **Situationsbezogene freie Verhaltensbeobachtungen** können nicht nur diese Gefahr verringern, sondern auch die begabungsdagnostische Informationsgrundlage vervollständigen, z.B. Testdaten, Schulzensuren u.ä. Die Vorzüge situationsbezogener Verhaltensberichte (anecdotal records) wurden von Tismer (1976) in Anlehnung an Shertzer & Stone (1971) detailliert dargestellt. In die gleiche Kategorie wären **Schülerbeobachtungsbögen** einzuordnen, wozu Tismer ebenfalls praktische Beispiele liefert.
- **Systematische Verhaltensbeobachtungen**. Diese sind sowohl hinsichtlich der Datenerhebung (Beobachtung) als auch in bezug auf die Protokollierungsphase weitgehend standardisiert: Die zu beobachtenden Verhaltensbereiche und Kontexte (z.B. Schülermitarbeit im Englischunterricht, Problemlöseverhalten bei schwierigen Mathematikaufgaben, Kausalattributionsverhalten in bezug auf Lernerfolg vs. -mißerfolg) sowie die Beschreibungskategorien werden von vorneherein festgelegt. Planung und Kontrolle des gesamten Beobachtungsprozesses sind somit kennzeichnend für systematische Verhaltensbeobachtungen. Im Rahmen der Begabungsdagnostik wird diese Form der Verhaltensbeob-

achtung, die vor allem zu wissenschaftlichen Forschungszwecken eingesetzt wird, dann bedeutsam, wenn andere standardisierte Meßverfahren (z.B. Intelligenztests) für einzelne Probandengruppen nicht verfügbar sind. Eine solche Indikation kann sich etwa bei ausländischen Probanden, speziellen Behinderungen oder auch bei bestimmten klinischen Fällen ergeben.

3.1.2 Rating- oder Schätzverfahren

Die bisher besprochenen Techniken der Verhaltensbeobachtung gehören zum Basisinventar jedes diagnostisch tätigen Beratungslehrers, „Schulpsychologen oder Erziehungsberaters. Die Auswertung umfangreicher Beobachtungsprotokolle ist jedoch sehr zeitaufwendig und birgt auch manche Gefahren in bezug auf die (Auswertungs-)Objektivität in sich. Ökonomischer sind sog. **Ratingverfahren**. Diese beinhalten praktisch nur den ersten, und letzten Prozeßabschnitt der oben beschriebenen Verhaltensbeobachtung, d.h. beim Beurteilungs- oder Ratingverfahren wird auf eine Beschreibung des beobachteten Verhaltens verzichtet und stattdessen sofort eine Beurteilung oder **Einschätzung** (rating) vorgenommen. Manche Autoren sprechen deshalb in diesem Zusammenhang von "indirekter" Beobachtungsmethode. Diese Bezeichnung trifft jedoch nur insofern zu, als auch beim Ratingverfahren die Beobachtungsdaten zumindest kurzfristig im Gedächtnis festgehalten werden müssen, bevor eine Einschätzung oder Beurteilung erfolgen kann.

Eine Unterscheidung von Beobachtungs- und Beurteilungsverfahren läßt sich überzeugender mit Hilfe des Inferenz-Kriteriums vornehmen. **Inferenz** (Schlußfolgerung) meint hier den Abstraktionsgrad der verwendeten Kategorien. Demnach sind reine Beschreibungskategorien "niedrig inferent", Beurteilungskategorien oder Ratings dagegen "hoch inferent".

Obwohl hoch-inferente Ratingverfahren (z.B. Schulnoten als Leistungsindikatoren) sehr ökonomisch und einfach zu handhaben sind, steigt mit dem Abstraktionsgrad der Kodierungsgrundlage auch das Risiko subjektiver Beurteilungsfehler. Schwarzer & Schwarzer plädieren deshalb für eine Kodierung der Beobachtungsdaten auf mittlerem Abstraktionsniveau, d.h. für "konkretisierte Schätzungen" (vgl. Abbildung 11).

Die Methoden der Verhaltensbeurteilung können in Verfahren **ohne** vs. **mit Skalen** (Einschätzdimensionen) klassifiziert werden. Zur ersten Gruppe gehören sog. Check-list-Verfahren sowie verschiedene Formen von Beurteilungs- und Diagnosebögen, zur zweiten Gruppe alle skalenbasierten Schätzverfahren.

Analog zu den Beschreibungskategorien der Beobachtungsbögen werden in den Checklisten sowie Beurteilungsbögen einschließlich Diagnosebögen zur Schulleistungsbeurteilung mäßig- bis hochinferente Zuordnungskategorien verwendet. Die vom Beobachter bzw. Rater bei einer Person beobachteten Verhaltensweisen werden dabei vorgegebenen Einschätzkategorien zugeordnet, z.B. "technisch begabt", "intelligent", "kreativ" (hoch inferente Kategorien) oder "großer Wortschatz",

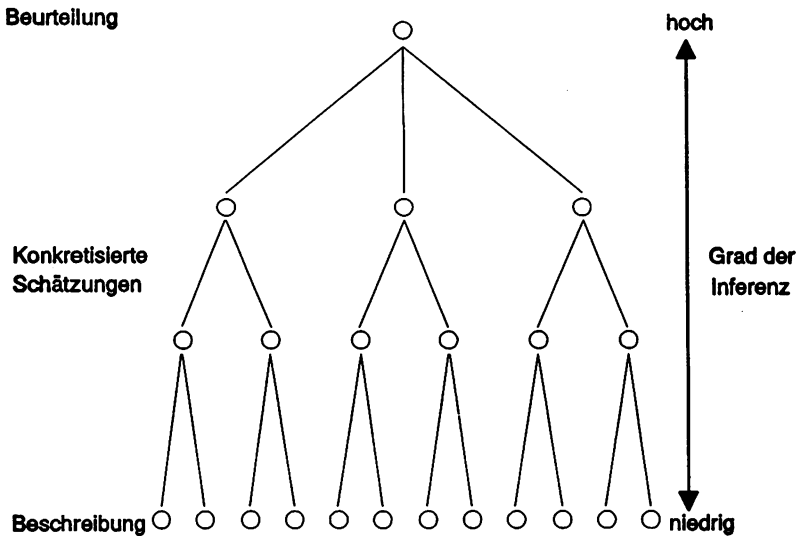


Abbildung 11: Beobachtungs- vs. Beurteilungsverfahren in Bezug auf den Abstraktionsgrad der Kodierungsgrundlage (n. Schwarzer & Schwarzer, 1977, S. 51)

"brilliert bei mathematischen Aufgaben", "denkt und arbeitet systematisch" (mäßig inferente Kategorien). Diese sind in einer **Eigenschaftsliste** (check-list) zusammengefaßt, deren Systematik ein explizites Konzept - z.B. von Hochbegabung - zugrundeliegen sollte, was in der Praxis jedoch häufig nicht kontrollierbar ist. Entsprechende Gefahren implizierter Persönlichkeitstheorien sind deshalb auch hier zu beachten. Ähnliche Vorbehalte gelten gegenüber **Beurteilungsbögen**, die sich nur durch Verwendung von Inferenzkategorien (statt verbaler oder adverbialer Beschreibungsmodi) von den Beobachtungsbögen unterscheiden. Die nach dem gleichen Prinzip erstellten **Diagnosebögen** werden vor allem zur Schulleistungsbeurteilung verwendet, wobei neuere Verfahren wegen der besseren Quantifizierbarkeit fast ausnahmslos mit **Skalen** arbeiten.

Vor der Konstruktion solcher Beurteilungsverfahren müssen nach Hasemann (1964, S. 828) folgende Fragen geklärt werden:

- Wer soll beurteilt werden (Probandenauswahl)?
- Welchem Zweck dient die Beurteilung (Problemstellung)?
- Wer soll beurteilen (Auswahl der Rater)?
- Welche Merkmale sollen beurteilt werden (Festlegung der Verhaltensdimensionen und Beobachtungssituation/en)?
- Wie sind die Merkmale bzw. ihre Ausprägungsgrade definiert (operationale Bestimmung)?

Bei der Auswahl der Merkmale und der Festlegung der Skalenstufen empfiehlt Hasemann (1983, S. 461), folgende Punkte zu beachten:

- Es sollten nur Merkmale beurteilt werden, die aus dem aktuellen oder vergangenen (beobachteten) Verhalten zu erschließen sind, d.h. Projektionen auf zukünftiges Verhalten dürfen nicht verlangt werden.
- Schließlich wird man in der Regel nur solche Verhaltensmerkmale via Rating erfassen, die nicht auf objektivere und zuverlässigere Weise mit anderen Instrumenten (z.B. Tests) meßbar sind.

A. Graphische Schätzskalen ohne Kategorien

Beispiel a)

Erfassungsdimension: Anstrengungsbereitschaft

+ _____

Beispiel b)

Erfassungsdimension: Leistungsmotivation

Der Schüler ist

intrinsisch motiviert extrinsisch motiviert

Beispiel c)

Mischform (Kombination von numerischer und graphischer Skala)

Der Schüler ist

-3 -2 -1 0 +1 +2 +3

intrinsisch motiviert extrinsisch motiviert

B. Graphische Schätzskalen mit verbalen Kategorien

Beispiel a)

Erfassungsdimension: Interesse an

recht un- uninter- etwas un- weder inter- etwas interessiert stark
interessiert essiert interessiert essiert noch interessiert
uninteressiert uninteressiert uninteressiert uninteressiert

Beispiel b)

Mischform

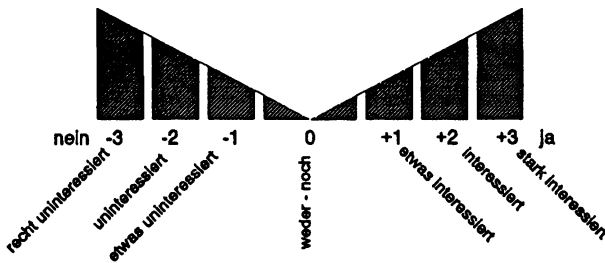


Abbildung 13: Graphische Ratingskalen (n. Osterland, 1976, S. 807f., bzw. Langer & Schulz v. Thun, 1974, S. 104)

Skalierte Ratingverfahren finden als numerische oder als graphische Schätzskalen - mit oder ohne verbale Kategorien - Verwendung. Desweiteren wäre

zwischen unipolaren und bipolaren Skalen zu unterscheiden, wobei letztere sich vor allem zur Einschätzung komplexer Merkmale eignen. Nach Langer & Schulz v. Thun (1974) ist bei der Entwicklung von Ratingskalen darauf zu achten, daß die Kriterien der Einfachheit, Übersichtlichkeit und "natürlichen" Abstufung eingehalten werden. Da die vom Rater markierte Skalenposition - Zahl oder Punkt auf der (kontinuierlichen) Skala - als Meßwert fungiert, ist die Auswertung solcher Ratings sehr einfach. Zugleich ist damit Objektivität gewährleistet. Ferner kann die Zuverlässigkeit über die Inter-Rater-Reliabilität bestimmt werden. Beispiele für numerische und graphische Schätzskalen sind in den Abbildungen 12 und 13 dargestellt.

Eine weitere interessante Form stellen sog. **Standard-Schätzskalen** (man-to-man) dar. Die Einstufung erfolgt hierbei durch direkten Vergleich von bekannten Personen, die als Maßstab dienen. Auf diese Weise kann die Zuverlässigkeit (Retestreliaibilität) der Beurteilung erhöht werden. Das in Abbildung 14 wiedergegebene Beispiel stammt aus dem später noch behandelten Angstfragebogen für Schüler (AFS) von Wieczerkowski et al. (1981; vgl. Abschnitt 3.4.4).

trifft gar nicht zu			durchschnittlich			trifft sehr stark zu
Schüler X			Schüler Y			Schüler Z

Abbildung 14: Standardskala "man-to-man" (n. Wieczerkowski et al., 1981)

Literaturhinweise

- Faßnacht, G. (1979). **Systematische Verhaltensbeobachtung**. München: Reinhardt.
- Hasemann, K. (1983). Verhaltensbeobachtung und Ratingverfahren. In K.J. Groffmann & L. Michel (Hrsg.), **Verhaltensdiagnostik (= Enzyklopädie der Psychologie, B II, Bd. 4)** (S. 434-488). Göttingen: Hogrefe.
- Langhorst, E. (1984). Beobachtung und Beurteilung des Schülerverhaltens im Unterricht. In K.A. Heller (Hrsg.), **Leistungsdiagnostik in der Schule** (S. 208-228). Bern: Huber.

3.2 Psychodiagnostisches Gespräch (Anamnese und Exploration)

Psychologische Gesprächsführung kann unterschiedliche Ziele verfolgen, wobei hier zwei Hauptfunktionen herausgestellt werden sollen: die **diagnostische Informa-**

tionsgewinnung (information getting interview) und das **Beratungsgespräch mit Interventionsabsicht**. In diesem Abschnitt wird nur das **diagnostische Interview** behandelt. Dabei bezeichnet der - aus der medizinischen Diagnostik stammende - Begriff der **Anamnese** das Bemühen, über wichtige Kontakt- oder Bezugspersonen des Probanden (z.B. Eltern) Informationen über den bisherigen Entwicklungsverlauf und eventuelle Besonderheiten zu erhalten. **Exploration** würde demgegenüber die Befragung des Probanden über sich selbst und/oder seine aktuelle Lebenssituation betreffen. Doch wird diese traditionelle Unterscheidung heute nur noch selten eingehalten. Im folgenden wollen wir deshalb die Begriffe "Exploration" und "psychodiagnostisches Gespräch" oder "psychodiagnostisches Interview" synonym verwenden. Der Anamnesebegriff sollte entsprechend seiner ursprünglichen Bedeutung vor allem zur Eruiierung ontogenetisch bedeutsamer Ereignisse bzw. Entwicklungsbedingungen oder auch zur Klärung des diagnostischen Untersuchungsanlasses in der Beratungssituation herangezogen werden.

In der Begabungsdiagnostik wird man neben Fähigkeitstests und anderen T-Datenquellen sowie Verhaltensbeobachtungen und Ratingverfahren (zur Erfassung von L-Daten) stets auch gesprächsdiagnostische Methoden (Q-Daten-Quellen) heranziehen. Hierunter wären nicht nur die in den letzten beiden Abschnitten dieses Kapitels behandelten Selbstbeurteilungsverfahren (Persönlichkeitsfragebögen) und standardisierten Fragebögen (Fremdratings) zur Erfassung familiärer und schulischer Lernumweltvariablen zu rechnen, sondern vor allem auch das hier erörterte diagnostische Interview. Während Fragebogentechniken gewöhnlich als Gruppenverfahren Verwendung finden, hat das psychodiagnostische Gespräch - mit seinen Varianten Exploration und Anamnese - in der **Einzelfalluntersuchung** seine Domäne. Es sollte praktisch in keiner Informationserhebung, also auch nicht bei begabungsdiagnostischen Untersuchungen, fehlen. Wegen seiner umfassenden Bedeutung wird es hier in einem eigenen Abschnitt behandelt.

Im Hinblick auf den Standardisierungsgrad lassen sich drei Varianten psychodiagnostischer Verfahren unterscheiden: die **freie Exploration**, die **halbstandardisierte** Befragung und die **vollstandardisierte** Fragebogentechnik (z.B. Persönlichkeitsfragebogen). Für alle drei Formen gelten prinzipiell die Forderungen der Objektivität, Reliabilität (Zuverlässigkeit) und Validität (Gültigkeit), d.h. die sog. Meß- oder Testgütekriterien (vgl. Kapitel 2 oben). Während diese für die vollstandardisierten Fragebögen und eingeschränkt auch für die halbstandardisierten Befragungen (mit vorformulierten Gesprächsleitfäden) kontrollierbar sind, ist dies bezüglich der **freien Exploration** - zumindest unter allgemeinen Praxisbedingungen - kaum möglich. Trotzdem sollte sich der Interviewer auch hier bemühen, soziale Vorurteile und andere einschlägige Fehler (etwa soziale Erwünschtheitsreaktionen beim Interviewten) zu vermeiden. Andererseits wird gerade in der **subjektiven Sichtweise** des Probanden, z.B. bezüglich des Begabungsselbstkonzepts, eigener Interessen und Motivationen, Kausalattributionen von Erfolg vs. Mißerfolg, von Zukunftserwartungen oder Befürchtungen, der Vorzug freier Explorationsen gesehen. Diese "subjektiven" Daten stellen somit eine wertvolle Ergänzung zu den "objek-

tiven" Dateninformationen (von Intelligenztests u.ä.) in begabungsdiagnostischen Untersuchungen dar.

Methodisch ergeben sich vor allem drei Probleme beim diagnostischen Einzelinterview: Probleme der Frageformulierung, der Gesprächsdurchführung und -protokollierung sowie der Verarbeitung gesprächsdiagnostischer Informationen. Die nachstehenden Ausführungen folgen einer früheren Darstellung des Erstautors (Heller, Nickel & Rosemann, 1978, S. 54ff.).

3.2.1 Zur Formulierung der Fragen

Allgemein gilt die Forderung, Inhalt und Form der Fragen auf den Untersuchungsanlaß und die Person des Befragten abzustimmen. Dabei sind komplizierte Satzkonstruktionen und unnötige Fachtermini ebenso zu vermeiden wie Doppelfragen, die mehrere Sachverhalte simultan zu erfassen versuchen. Ferner verbieten sich doppelte Verneinungen und Suggestivfragen. Konkrete Fragen sind gegenüber abstrakten vorzuziehen.

Im Hinblick auf die Funktion können folgende **Fragetypen** unterschieden werden: Einleitungs- und Übergangsfragen, Alternativ- und Filterfragen, geschlossene vs. offene sowie direkte vs. indirekte Fragen.

Geschlossene Fragen etwa zur Erfassung begabungsspezifischer Freizeitaktivitäten eines 16jährigen Gymnasiasten wären die folgenden: "Spielst Du ein Musikinstrument?" - "Haben Deine Eltern etwas dagegen, wenn du fast die gesamte Freizeit mit Klavierspielen zubringst und dadurch wichtige Schularbeiten vernachlässigst?"

Offene Fragen zum gleichen Themenbereich wären z.B.: "Ich habe gehört, daß Du Dich sehr stark mit Musik beschäftigst. Was macht Dir dabei am meisten Spaß?" - "Wie verbringst Du Deine Freizeit?" - "Wie verhalten sich Deine Eltern, wenn Du wegen Deines musikalischen Hobbys Schularbeiten zurückstellst?"

Wie an den Beispielen erkennbar, wird bei offenen Fragen vom Interviewer oder Gesprächsleiter nur das Rahmenthema (hier: Freizeithobby Musik) genannt. Der Befragte soll darauf mit seinen eigenen Worten frei antworten, wobei er für ihn bedeutsame Aspekte hervorheben kann. Dagegen engen geschlossene Fragen die Reaktionsmöglichkeit mehr oder weniger stark ein; im Extremfall würden nur bejahende oder verneinende Reaktionen möglich sein. Zugleich wird mit geschlossenen Fragen der Gesprächsfluß erheblich erschwert, wengleich bei sprachlich ungeübten oder gar behinderten Klienten diese Form oft die einzige Möglichkeit darstellt, ein Gespräch überhaupt zustande zu bringen. Andererseits dürfte für die meisten der Befragten die offene Frageform die Gesprächsbereitschaft steigern, da sich viele durch geschlossene Frageformulierungen in ihrer Reaktionsfreiheit eingeengt fühlen. Darüber hinaus läßt sich durch die Relation von offenen und gebundenen Fragen der Verlauf des diagnostischen Interviews steuern. Generell sollte man am Anfang eines Gesprächs eher offene Fragen stellen, wobei im wei-

teren Verlauf stärker zu gebundenen Fragen übergegangen werden kann - zumal wenn beide Seiten unter Zeitdruck stehen.

Unter dem Gesichtspunkt der Transparenz bzw. Befragungsintention wird die Unterscheidung von direkten und indirekten Fragen bedeutsam: Bei **indirekter** Befragung ist für den Befragten der Zweck einer Frage schwerer durchschaubar, was sozialen Erwünschtheitsreaktionen (als Fehlerquelle der Befragung) vorbeugen kann. Die Entscheidung für direkte vs. indirekte Fragen sollte jedoch auch unter inhaltlichen Gesichtspunkten getroffen werden. So eignen sich **direkte** Fragen besser zur Erfassung objektiver Tatbestände oder Fakten, während **indirekte** Fragen vor allem Informationen über subjektive Sichtweisen, Einstellungen, Werthaltungen oder auch Emotionen u.ä. vermitteln. Darüber hinaus kann der Gesprächsleiter in der indirekten Frageformulierung flexibler auf spontane Äußerungen des Probanden eingehen und neue Themen, die für den Untersuchungsanlaß bisher nicht in erkennbarem Zusammenhang standen, aber nunmehr bedeutsam erscheinen, aufgreifen. Auch hierzu einige Beispiele.

Direkte Fragen: "Wieviel Zeit verbringst Du täglich mit Klavierspielen?" - "Wann machst Du Deine Hausarbeiten für die Schule?" - "Hast Du noch andere Hobbies außer Musik?"

Indirekte Fragen: "Was hast Du gestern nach der Schule gemacht?" - "Wie hast Du das letzte Wochenende verbracht?" - "Welche öffentlichen Veranstaltungen hast Du in den letzten drei Monaten besucht?" Aus diesen und ähnlichen Fragen kann der Befragte nicht ohne weiteres schließen, daß der Interviewer etwas über die musikalische Betätigung und deren Inhalte erfahren möchte.

3.2.2 Planung und Durchführung psychodiagnostischer Einzelfallgespräche

Selbstverständlich ist auch bei der Exploration oder Anamneseerhebung darauf zu achten, daß ein angemessener äußerer Rahmen ungestörten Gesprächsablauf garantiert und für eine gute Atmosphäre sorgt. Zur Vorbereitung gehören auch Überlegungen über Inhalt und Form der Fragen sowie - bei Benutzung eines Gesprächsleitfadens - die kritische Reflexion des zugrundeliegenden (zumeist impliziten) theoretischen Konzepts. Je nach theoretischer Orientierung werden in solchen (in der Praxis sehr beliebten) Leitfäden andere Inhalte erfragt, z.B. auf lerntheoretischer vs. tiefenpsychologischer Basis. Über die inhaltlich-thematischen Fragestellungen hinweg empfiehlt Schraml (1969, S. 48) folgende - generellen - **Bereiche:**

- **Sozialisations-** bzw. **sozioökonomischer Hintergrund**, z.B. Schulbildung und Berufstätigkeit der Eltern, Wohnungssituation, Anzahl und Art der zur Verfügung stehenden Bücher und Zeitschriften, Freizeitaktivitäten in der Familie usw.
- **Individuelle Persönlichkeitsentwicklung** und **aktuelle Lebenssituation** des Befragten, z.B. Schilderung bestimmter Verhaltensweisen, Interessen und persönlicher Schwierigkeiten, Konflikte usw.

- **Körperliche Entwicklung**, z.B. Krankheiten, Besonderheiten des Eß- und Schlafverhaltens usw.

Diese Themenbereiche müssen jeweils im Hinblick auf die zentrale begabungsdiagnostische Fragestellung (Anlaßproblematik) gewichtet und erforderlichenfalls ergänzt werden. Dabei ist jedoch wiederum das implizite Bezugssystem des Diagnostikers zu reflektieren, um einseitige Frageinhalte zu vermeiden. Der Vergleich von früheren und gegenwärtigen Entwicklungsphasen bzw. -bedingungen gestattet gleichsam eine **Längsschnittbetrachtung**, die gerade im Kindes- und Jugendalter sehr oft bedeutsam wird.

Jedes Gespräch stellt eine **soziale Interaktion** dar, in dessen Verlauf beidseitig persönliche "Hypothesen" (oft implizit) aufgestellt und geprüft werden. Der Diagnostiker muß die erhaltenen Gesprächsinformationen fortlaufend sowohl im Hinblick auf neue Fragen als auch unter dem Gesichtspunkt der Hypothesenprüfung verarbeiten und eventuelle Zusatzfragen stellen. Auf der Klientenseite entstehen Überlegungen darüber, wie eine Frage wohl zu beantworten sei und welche Absicht dahinter steckt. Hierbei spielen normative Vorstellungen, persönliche Eigenarten des Interviewers und situative Kontextmerkmale eine Rolle. Die Wirksamkeit körperlicher Züchtigung wird heute kaum mehr von Eltern in der Befragung zugegeben (auch wenn die Erziehungspraxis anders aussehen mag), weil die Orientierung der Öffentlichkeit an sozial erwünschten Wertvorstellungen zu mächtig ist. In vielen Fällen ist es deshalb notwendig, die Aussagen der Befragten mit Verhaltensbeobachtungen in Real-life-Situationen zu vergleichen und/oder die Aussagen verschiedener Informanten einschließlich des Probanden zum gleichen Thema einander gegenüberzustellen. Schließlich können gesprächsbegleitende Beobachtungen der nonverbalen Kommunikation (Ausdrucksverhalten) wichtige Korrektivhinweise liefern, ohne deren Aussagekraft überbewerten zu wollen. Die Gefahr der Beeindruckung in der verbalen Interaktion besteht durchaus wechselseitig, worauf schon Kaminski (1970, S. 303) hingewiesen hat.

Obwohl Tonband- und Videoaufzeichnungen eine vollständige **Registrierung** der Explorationsdaten erlauben, werden diese Techniken überwiegend nur im Rahmen wissenschaftlicher Untersuchungen eingesetzt. In der diagnostischen Praxis begnügt man sich gewöhnlich mit Stichwortprotokollen, die nach Beendigung des Gesprächs - also bei noch frischer Erinnerung - aus dem Gedächtnis ergänzt werden sollten. Gegebenenfalls und mit dem Einverständnis des Klienten kann auch eine Kombination von Mitschrift und Tonbandaufzeichnungen erfolgen. Allerdings erfordert die Transkription von Tonbandprotokollen einen nicht unerheblichen Arbeitsaufwand.

3.2.3 Zur Verarbeitung gesprächsdiagnostischer Informationen

Es empfiehlt sich, die Angaben verschiedener Informanten getrennt zu registrieren und auszuwerten. Einem Hinweis Thomaes folgend fordern Tismer et al. (1976) die Sortierung **beschreibender** und **deutender** (bewertender) Verhaltensaussagen. Auf

diese Weise lassen sich die Explorationsdaten zuverlässiger interpretieren. Die Auswertung gerade auch heterogener Angaben der Befragten kann unter formalen und inhaltlichen Kriterien erfolgen, wobei natürlich die Prolemzentrierung nicht aus dem Auge verloren werden darf. Kaminski (1970, S. 97ff.) schlägt beispielsweise folgende Aussageebenen vor:

- Mitteilungen über **Fakten**, etwa Lebensalter, Schulabschluß, Berufsqualifikation, Zahl erlernter Fremdsprachen usw.
- **Aussagen über konkretes Verhalten** und dessen Kontextbedingungen. Folgende Beispiele aus dem Explorationsprotokoll einer elfjährigen Schülerin und der Anamnese einer Mutter zu ihrem elfjährigen Jungen (n. Tismer et al., 1976, S. 850) verdeutlichen das Gemeinte:

"Wenn ich eine Schulaufgabe verhaue, versuche ich, das möglichst von meiner Mutter fernzuhalten; die regt sich nämlich immer so auf. Ich tue dann, wenn ich aus der Schule komme, als wenn nichts wäre. Ich erzähle, was sonst in der Schule los war, nur nichts von der Aufgabe. Fragt sie nach der Schulaufgabe, sage ich einfach, wir haben sie noch nicht gekriegt."

"Peter kann beim Spielen nicht verlieren. Wenn er am Gewinnen ist, dann strahlt er. Aber wehe, der kleine Bruder hat mal einen Vorsprung, dann kann es passieren, daß er zu schreien anfängt und die Steine hinwirft. Erst gestern war es so. Wir haben zu viert 'Monopoli' gespielt. Nach dem zweiten Spiel sprang er auf einmal auf, daß alles umfiel, und lief aus dem Zimmer, nur weil er nicht gewonnen hatte."

- **Beurteilungen oder Einschätzungen**, z.B. "Unser Junge ist zu ängstlich; ich bin auch immer sehr schüchtern gewesen". Solche Sätze stellen keine Beschreibung konkreter Verhaltensweisen dar, sondern implizieren **Dispositionen**, die das betr. Verhalten - nach Meinung des Befragten - verursachen. Zugleich wird damit stillschweigend unterstellt, daß das angesprochene Problemverhalten auf eine persistente Persönlichkeitseigenschaft zurückzuführen sei (hier: Ängstlichkeit auf einen schüchternen Charakter). Extrapolationen situationsspezifischer Verhaltensweisen oder auch Erklärungsversuche mit Hilfe impliziter Persönlichkeitstheorien müssen vom Diagnostiker, der solche Aussagen kritisch zu bewerten hat, vorsichtiger behandelt werden als konkrete Verhaltensbeschreibungen oder gar Fakten.
- Schließlich führt Kaminski noch eine vierte Kategorie an, in die Berichte über **Wünsche, Befürchtungen, Zukunftsvorstellungen** sowie sonst nicht rubrizierbare Angaben einzuordnen wären.

Diagnostische Schlußfolgerungen aus Explorations- und Anamnesedaten folgen den üblichen Regeln diagnostischer Befunderstellung. Sofern auf relativ konsistente, überdauernde Verhaltensmerkmale oder Einstellungen geschlossen werden soll, müssen im Sinne des geforderten **Doppel- oder Mehrfachbeleges** mindestens zwei inhaltlich sich entprechende Aussagen verschiedener Informanten belegmäßig vorhanden sein. Ferner sind auch sich widersprechende Belege sorgfältig zu analysieren, z.B. im Hinblick auf Gültigkeit der Aussagen. Notfalls müssen unaufklärbare diskrepante Belege im Gutachten (Kapitel 6 in diesem Buch)

festgehalten werden, um sie vielleicht zu einem späteren Zeitpunkt plausibel interpretieren zu können (ausführlicher vgl. Tismer-Puschner et al., 1976).

Schließlich sei darauf hingewiesen, daß zur Vermeidung voreiliger Interpretationen die Protokollsätze - sofern nicht als wörtliche Rede gekennzeichnet - im **Konjunktiv** abgefaßt werden sollten. Auf diese Weise wird deutlicher erkennbar, daß es sich bei den betr. Aussagen um die Meinung der befragten Personen handelt, die vom Diagnostiker zusammen mit den übrigen Untersuchungsdaten in die Befunderstellung eingehen, nicht ohne vorher einer kritischen Prüfung bezüglich Zuverlässigkeit und Gültigkeit unterzogen worden zu sein. Dabei ist bei gesprächsdiagnostischen Belegen stets zu beachten, daß es sich um jeweils subjektive Sichtweisen handelt, die sich selbst oder auch im Vergleich zu anderen widersprechen können und nicht unbedingt objektive Tatbestände reflektieren. Als subjektive Informationsquellen können sie jedoch objektive Testdaten (nicht nur) um den Erlebensaspekt sinnvoll ergänzen.

Literaturhinweise

- Fisseni, H.J. (1990). **Lehrbuch der psychologischen Diagnostik (Kap 8)**. Göttingen: Hogrefe.
- Keßler, B.H. (1982). Biographische Diagnostik. In K.J. Groffmann & L. Michel (Hrsg.), **Persönlichkeitsdiagnostik (= Enzyklopädie der Psychologie, B II, Bd. 3)** (S. 1-55). Göttingen: Hogrefe.
- Tismer, K.-G., Fisseni, J. & Tismer-Puschner, I. (1976). Anamnese und Exploration in der schulpyschologischen Beratung. In K. Heller (Hrsg.), **Handbuch der Bildungsberatung, Bd. 3** (S. 837-856). Stuttgart: Klett.

3.3 Tests zur Erfassung kognitiver Fähigkeiten

Die Auswahl der nachstehend referierten Testverfahren erfolgte unter drei Kriterien: Erstens sollte damit einem möglichst umfassenden Konzept von Begabung (vgl. Kapitel 1) Rechnung getragen werden. Zweitens wurde auf die Zielgruppe dieses Lehrbuchs und entsprechende potentielle Untersuchungspopulationen Rücksicht genommen. Drittens sollten nur publizierte Verfahren behandelt werden. Daß mit diesen Auswahlkriterien, vor allem in bezug auf die ersten beiden, auch subjektive Bewertungen der Verfasser miteinfließen, ist trotz unserer Bemühungen um eine ausgewogene Darstellung wohl nicht völlig auszuschließen.

Zunächst seien einige technische und rechtliche Hinweise für Testbenutzer gegeben. Da sich das Testangebot und die Bezugspreise laufend ändern, verzichteten wir hier auf den Versuch einer vollständigen Testübersicht und auf Preisangaben. Diese können durch Anforderung der (kostenlosen) Jahreskataloge bei den nachstehend aufgeführten Testverlagen besorgt werden. Außerdem liegt mit dem von Brickenkamp (1975, 1983) herausgegebenen Handbuch psychologischer und pädagogischer Tests, eine relativ vollständige Testdokumentation für den deutschsprachigen Raum vor. Ferner sei auf den von Groffmann & Michel (1982) her-

ausgegebenen Enzyklopädie-Band 2 (Themenbereich B, Serie II) "Intelligenz- und Leistungsdiagnostik" verwiesen. Beide Werke sind im Verlag für Psychologie, Dr. C.J. Hogrefe, in Göttingen erschienen.

Die wichtigsten **Verlagsanschriften** für den (aus Gründen des Testschutzes nicht über den Buchhandel möglichen) Testbezug im deutschsprachigen Raum sind:

- **Beltz Test Gesellschaft**, Postfach 1120, D-6940 Weinheim.
- **Huber Verlag**, Länggäßstraße 76, CH-3000 Bern 9.
- **Verlag für Psychologie, Dr. Hogrefe**, Postfach 414, D-3400 Göttingen.
- **Testzentrale des Berufsverbandes Deutscher Psychologen**, Robert-Bosch-Breite 25, Postfach 3751, D-3400 Göttingen.

Der Jahreskatalog der **Testzentrale** enthält die in Deutschland, Österreich und der Schweiz am häufigsten eingesetzten psychologischen Testverfahren, und zwar aller Testverlage, sowie eine Auswahl wichtiger nichtdeutschsprachiger - vorab englischsprachiger - Tests.

Da aus Gründen des Testschutzes psychologische Tests nur direkt über den betreffenden Verlag oder die Testzentrale (TZ) ausgeliefert werden, seien hierzu die TZ-Richtlinien aus dem Jahreskatalog 1989/90 zitiert:

Bei Bestellungen von Tests (nicht Zeitschriften und Büchern), die von der Testzentrale gemäß ihren Vereinbarungen mit dem Berufsverband Deutscher Psychologen und mit Billigung des Börsenvereins des Deutschen Buchhandels nur an bestimmte zum Bezug berechnete Berufsgruppen ausgeliefert werden dürfen, wird der Interessent gebeten, Name und Beruf anzugeben, damit die Testzentrale anhand der vom Berufsverband Deutscher Psychologen ausgearbeiteten Listen der bezugsberechtigten Institutionen, Personengruppen und Personen zu prüfen vermag, ob eine Auslieferung erfolgen kann. Der Mißbrauch von Testverfahren zwingt leider zu einem solchen kontrollierten Vertrieb.

Für Schulpsychologen und Erziehungsberater empfiehlt sich deshalb bei **Testbestellungen**, Institutsbriefköpfe oder -stempel zu verwenden, gegebenenfalls durch eine Bestätigung der **Bezugsberechtigung** durch den diplomierten Beratungsstellenleiter. Analog sollten sich interessierte Studierende der Psychologie eine Bescheinigung der für die Ausbildung in psychologischer Diagnostik an der Universität zuständigen Institutsleitung besorgen.

Für Beratungslehrer und entsprechende Studierende, die nur eine eingeschränkte Bezugsberechtigung erlangen können, empfiehlt sich, bei Bestellungen von Leistungs- und Intelligenztests eine Bescheinigung der betreffenden Schulleitung oder Hochschulinstitution beizufügen. Der Erwerb von in diesem Lehrbuch dargestellten Testverfahren ist in der Regel auch für diese Zielgruppe möglich, sofern ein entsprechender Nachweis über die Bezugsberechtigung erbracht wird.

Interessenten, die sich allgemein über die **Rechtsgrundlagen und -bedingungen psychologischer Diagnostik**, insbesondere der Testdiagnostik, orientieren möchten, seien auf den Beitrag von Schumacher & Schumacher-Merz (1982) oder die entsprechenden Abschnitte in Jäger (1988) verwiesen. Speziell über juristische Aspekte der Testanwendung in der Schule informiert das Rechtsgutachten von Avenarius (1990).

Im folgenden werden nun vier Hauptkategorien von Testverfahren zur Erfassung psychologischer **Begabungsvariablen** behandelt, nämlich 1) Allgemeine Intelligenztests, 2) Differentielle Intelligenz- bzw. Kognitive Fähigkeitstests, 3) Spezielle Fähigkeitstests, 4) Kreativitätstests. Anschließend werden relevante nichtkognitive (Persönlichkeits-)Tests (Abschnitt 3.4) sowie Verfahren zur Erfassung sozialer Lernumweltvariablen (Abschnitt 3.5) dargestellt.

Die Darstellung zur ersten Testkategorie beginnt mit den Individualverfahren, die einzeln administriert werden müssen. Sofern die besprochenen Tests in einem anderen Verlag als den o.a. erschienen sind, wird als Bezugsquelle die Testzentrale (TZ) angegeben.

3.3.1 Tests zur Erfassung der Grund- oder allgemeinen Intelligenz

Diese Verfahren basieren auf der theoretischen Grundlage einer "Allgemeinen Intelligenz", etwa im Sinne des Generalfaktorenmodells von Spearman (vgl. Kapitel 1 in diesem Buch).

3.3.1.1 Individualverfahren

- (1) Die **Raven-Matrizen-Tests** von J.C. Raven, deutsche Bearbeitung von A. Schmidtke, S. Schaller & P. Becker (1980) und H. Kratzmeier & R. Horn (1979, 1987²; 1980) (Beltz)

Mit den Raven-Matrizen-Tests soll nach Angaben des Autors (Testmanual) "clear thinking" geprüft werden, wobei die Tests ursprünglich eher als Entwicklungstests denn als Intelligenztests konzipiert waren³. Da für die Tests sprachfreie Instruktionen vorliegen, kann mit den Raven-Matrizen-Tests die Fähigkeit zu logischem Denken auch bei gehörgeschädigten oder anderssprachigen Personen erfaßt werden.

Die erste Auflage des Progressive Matrices-Test (PMT) erschien 1938. Inzwischen sind zahlreiche Revisionen sowie Formen für Klein- und Grundschulkinder, überdurchschnittlich begabte Jugendliche und Erwachsene veröffentlicht worden: a) eine farbige Testform (**Coloured Progressive Matrices**, CPM; deutsche Bearbeitung von Schmidtke, Schaller & Becker, 1980) für 5- bis 11jährige Kinder mit drei Aufgabenreihen A, Ab und B; b) die Standardform (**Standard Progressive Matrices**; SPM, deutsche Bearbeitung von Kratzmeier & Horn, 1979, 1987²) mit 5 Aufgabensets A, B, C, D und E für Kinder, Jugendliche und Erwachsene ab etwa 10 Jahren; c) eine schwierigere Form für besser begabte Jugendliche und Erwachsene ab etwa 15 Jahren (**Advanced Progressive Matrices**, APM; deutsche Bearbeitung von Kratzmeier & Horn, 1980) mit zwei Aufgabenreihen (Set I und II).

³ Persönliche Information des Testautors.

Ursprünglich waren die Raven-Matrizen-Tests reine Power-Tests und als solche nur im Einzelverfahren durchführbar. In der deutschen Bearbeitung gilt dies für die CPM, die nach wie vor ausschließlich als reiner Power-Test, d.h. ohne zeitliche Limitierung bearbeitet werden. Für die SPM und die APM sehen die deutschen Bearbeiter auch eine Durchführung als Gruppentest vor, wobei für die Bearbeitung der SPM die Zeit nicht begrenzt wird, bei den APM jedoch eine Höchstzeit von 40 Minuten festgelegt wird. Wir haben die Raven-Matrizen-Tests dennoch zu den Individualverfahren gerechnet, weil dies zum einen u.E. der Intention der Tests gerechter wird, zum anderen sehen die deutschen Bearbeiter bei den APM die Validität des Tests durch die Begrenzung der Testzeit nicht eingeschränkt.

Testaufbau: Bei allen Raven-Matrizen-Tests werden jeweils Tafeln mit einem bestimmten Muster vorgegeben. Dieses ist unvollständig, d.h. ein Teil des Musters ist weggelassen. Der Proband soll nun aus den vorgegebenen 6 (CPM) bzw. 8 (SPM, APM) Teilstücken das jeweils passende herausfinden (vgl. dazu Abbildung 15). Die Lösungsnummer wird auf einem separaten Antwortbogen notiert. Bei jüngeren Kindern (CPM) notiert der Testleiter die Antworten.

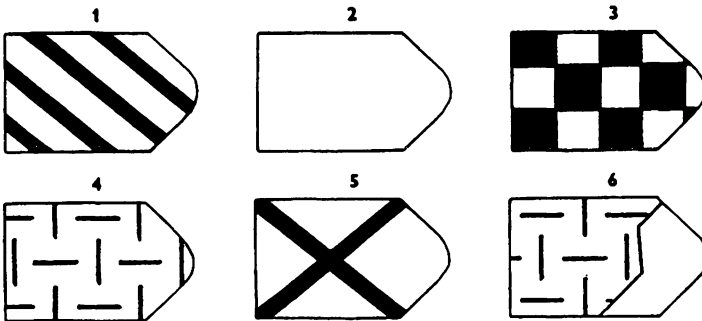
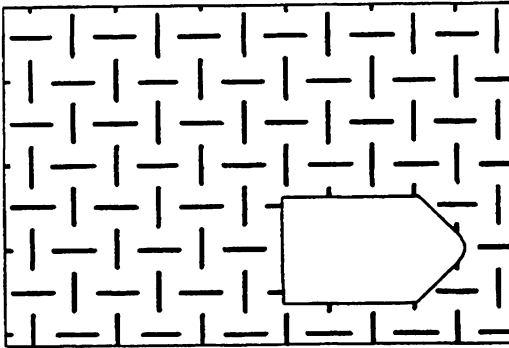


Abbildung 15: Übungsbeispiel zu den Standard Progressive Matrices

Die Aufgaben sind dem Schwierigkeitsgrad nach aufsteigend ("progressiv") angeordnet, sowohl innerhalb der einzelnen Sets als auch zwischen den Serien. Allerdings ließ sich diese Progression empirisch nur teilweise bestätigen.

Erfassungsdimensionen: Die Raven-Matrizen-Tests sollen klares, logisches Denken erfassen, wobei faktorenanalytische Untersuchungen den Nachweis erbrachten, daß die PMT sehr gut die **Allgemeine Intelligenz** (g-Faktor sensu Spearman et al.) erfassen. Darüber hinaus sind vor allem die Faktoren **Induktives Denken** und **Raumvorstellung** an der PMT-Leistung beteiligt. Für die Interpretation der PMT-Ergebnisse im Anwendungsfall bedeutet dies, daß - neben der Allgemeinbefähigung - folgende Intelligenzfunktionen in den PMT zum Ausdruck kommen: Unterscheidungsgenauigkeit (accuracy of discrimination) - vorab bei den einfacheren Aufgaben; Analogieschlußdenken, Regelerkennen bzw. Abstrahieren von Prinzipien und deren (richtige) Anwendung, induktives Denken sowie Raumvorstellungskomponenten - vorab bei den komplexen Aufgaben.

Praktische Handhabung: Bei allen drei Testreihen sind sprachliche und sprachfreie Instruktionen vorgegeben. Für die CPM sind etwa 25-40 Minuten, für die SPM bis zu 45 Minuten zu veranschlagen, bei den APM ist die Testzeit für Teil I auf 10 Minuten, für Teil II auf 40 Minuten begrenzt. Bei den SPM und APM sollte die Gruppengröße 15-20 Personen (je nach Zahl der Testleiter bzw. Helfer) nicht übersteigen. Die Auswertung wird dadurch erleichtert, daß die Antwortbögen als Durchschreibbögen gestaltet sind, so daß eine eigene Lösungsschablone überflüssig ist. Die Interpretation erfolgt mit Hilfe von Normtabellen. Im deutschen Handbuch der CPM finden sich als Interpretationshilfen Angaben zu Standardmeßfehler und kritischen Differenzen. Fallbeispiele sind keinem der Handbücher der deutschen Bearbeitungen zu entnehmen.

Zur **Reliabilität** der PMT finden sich unterschiedliche Angaben, wobei die berichteten Zuverlässigkeitskoeffizienten in etwa zwischen 0.9 und 0.7 schwanken, was gute bis noch befriedigende Reliabilität bedeutet. Der **Standardmeßfehler schwankt** zwischen 4.7 bzw. 8.2, d.h. bei einer fünf- bzw. zehnpromzentigen Irrtumswahrscheinlichkeit liegt der wahre Wert in einem Intervall von ± 9 bzw. ± 16 IQ-Punkten um den gemessenen Wert. Für die deutsche Ausgabe der CPM geben die Bearbeiter Split-Half- und Konsistenzreliabilitätsschätzungen von 0.49 bis 0.91 an, wobei sich die niedrigsten Werte bei den jüngsten Kindern ergaben. Verglichen mit anderen Verfahren erweisen sich die CPM-Scores der Kinder als erstaunlich stabil (zwischen 0.75 und 0.84 für einen Zeitraum von einem Jahr). Hingegen fand Knaack (1978) an einer Stichprobe von 5- und 6jährigen Kindern niedrigere Stabilitätskoeffizienten (0.54 für einen Zeitraum von 12 Monaten). Leider teilen die deutschen Bearbeiter der SPM und APM keine eigenen Reliabilitätsschätzungen mit. Bei deutschen Taubstummen fand Heller (1967, S. 227) Stabilitätskoeffizienten von 0.98 (Retestung nach 4 Wochen), 0.85 (Retestung nach 9 Monaten) und 0.72 (Retestung nach 2 Jahren). Bangen (1965, S. 41) berichtet bei 9-10jährigen Taubstummen eine Halbierungszuverlässigkeit von $r=0.90$. Somit ist die Zuverlässigkeit der PMT in der Anwendung bei gehörlosen (sowie resthörigen und schwerhörigen)

Kindern und Jugendlichen keinesfalls geringer als bei normalhörenden Probanden. Gleiches dürfte möglicherweise für anderssprachige Testpersonen gelten.

Für die deutsche Bearbeitung der CPM finden sich im Manual auch Prüfungen der Rasch-Homogenität der Aufgabenreihen. Für die meisten Modellgeltungstests haben die Autoren dabei leistungsstarke (bzw. ältere) und leistungsschwächere (bzw. jüngere) Stichproben gegenübergestellt. Wegen der Steigerung der Aufgabenschwierigkeit innerhalb der CPM-Aufgabenreihen ist bei diesem Vorgehen von vorneherein nicht mit vollständigem Nachweis der Rasch-Skalierbarkeit zu rechnen, da jeweils alle Items herausfallen, die in einer der einbezogenen Stichproben entweder von allen Probanden gelöst oder nicht gelöst wurden. Bei Aufteilung der Stichprobe nach Geschlecht erweisen sich jedoch praktisch alle Items der CPM als Rasch-homogen.

Die Validität der PMT ist in bezug auf die Erfassung der Allgemeinbegabung erstaunlich hoch. So korrelieren die PMT-Leistungen mit den wesentlich aufwendigeren Binet-Verfahren und dem WISC je nach Untersuchungsstichprobe zwischen 0.55 und 0.75, vereinzelt auch höher. Bezüglich der Subtests im WISC bzw. HAWIK ergaben sich die engsten Zusammenhänge mit dem Mosaiktest (MT). Schmidtke, Schaller & Becker (1980) fanden bei gemeinsamer Faktorenanalyse der HAWIK-Untertests und der CPM eine klare Zweifaktorenlösung, wobei die Tests des HAWIK-Verbalteils zusammen mit den CPM den ersten Faktor bildeten. Den Faktorenanalysen der CPM auf Itembasis, die dieselben Autoren mitteilen, sollte jedoch keine weitere Beachtung geschenkt werden. Wie bei relativ homogenen Items zu erwarten, fanden sich zwei bis drei Faktoren, auf denen jeweils Items ähnlicher Schwierigkeit laden (vgl. dazu auch Knaack, 1978; Rost & Gebert, 1980).

Für die SPM wird eine mittlere Korrelation von 0.5 zu den Untertests 3 und 4 des PSB mitgeteilt, während die Korrelationen mit der "Arbeitsprobe" im PSB (Untertests 9 und 10) mit 0.3 bzw. 0.2 am niedrigsten ausfällt. Die Zusammenhänge zu den verbalen PSB-Tests liegen in etwa dazwischen, sind aber insgesamt als eher gering einzustufen. Ähnlich geringe Zusammenhänge wie zum PSB 9 und 10 fanden sich auch zum d2. Allerdings weisen Kratzmeier & Horn (1987) darauf hin, daß in der deutschen Validierungsstudie die Bearbeitungszeit limitiert war.

In bezug auf Übereinstimmungs- oder prognostische Validität mit Schulleistungen werden unterschiedliche Validitätskoeffizienten berichtet. Für deutsche Stichproben finden sich in den Manualen lediglich zum CPM Korrelationskoeffizienten zur Übereinstimmung mit Schulleistungen. Demnach ergeben sich für die verschiedenen Altersgruppen niedrige Zusammenhänge (0.21 bis 0.38) mit der Deutsch- und mittlere (0.28 bis 0.57) mit der Mathematiknote. Knaack (1978) fand bei 5- und 6jährigen Vorschülern zwar eher keine oder niedrige, meist nichtsignifikante Zusammenhänge mit späteren Leistungsindikatoren ($r=0.33$ mit Mathematik, $r=0.29$ mit Deutsch nach 3 Jahren), angesichts der enormen Entwicklungseinflüsse, denen Kinder ab der Einschulung ausgesetzt sind, und der kleinen Stichprobe für die Berechnung der Indikatoren, würden wir diese Hinweise auf die Validität der CPM jedoch wesentlich günstiger bewerten als der Autor.

Im Bezug auf andere Intelligenztests fand Heller (1967, S. 227f.) bei zahlreichen PMT-Untersuchungen Gehörloser und Schwerhöriger Übereinstimmungskoeffizienten mit der Snijders Oomen Nicht-verbale Intelligenztestreihe (SON) zwischen 0.53 und 0.89 sowie mit dem HAWIK-Handlungsteil von 0.53 bis 0.80, was auf befriedigende bis gute, teilweise sogar sehr gute Validität der PMT hinweist. Auf der Grundlage spezifischer PMT-Normen für Hörgeschädigte (vgl. Heller, 1973) konnten bei einer größeren Schülerzahl mittlere Validitätskoeffizienten von 0.7 in der Relation PMT-SON sowie 0.8 zum HAWIK-Handlungsteil errechnet werden. Im Gegensatz dazu fand Evans (1966) in seiner Vergleichsuntersuchung mit dem PMT und WISC an 100 taubstummen und resthörigen (englischen) Schülern im Alter von 6 bis 15 Jahren nur sehr mäßige Zusammenhänge in der Höhe von 0.27 und 0.44, wobei möglicherweise Stichprobenartefakte eine Rolle gespielt haben. Überhaupt sollte man bei der Beurteilung dieser oder anderer Validitätskoeffizienten (stärker als es vielfach geschieht) stets die Größe vs. Kleinheit sowie die Repräsentanz vs. Irrepräsentativität der Untersuchungssamples berücksichtigen. Auch das Alter der jeweiligen Untersuchungen spielt natürlich eine entscheidende Rolle. Gerade im sonderpädagogischen und/oder klinischen Anwendungsbereich haften entsprechenden teststatistischen Aussagen aus begrifflichen Gründen nicht selten Stichprobenmängel an.

Die Original-Normen zu den Raven-Matrizen-Tests beziehen sich auf ca. 7000 englische Probanden und sind als Prozentränge für 6- bis 14jährige in Halbjahresschritten, für 20- bis 65jährige in Fünf-Jahresschritten angegeben. Diese Normen finden sich auch in den Handanweisungen der deutschen Bearbeitungen der PMT. Deutsche Eichwerte (IQ- und Prozentrangnormen) liegen für die CPM in Halbjahresschritten für Kinder von 4;9 bis 11;0 Jahren vor. Für die SPM wurden deutsche Prozentrangnormen für 10- und 11- sowie 12-, 13-, 14- und 15jährige berechnet, während im Manual der APM nur eine Prozentrangnormtabelle für Schüler der Klassen 9 bis 11 mitgeteilt wird. Dem Manual der SPM können auch die umfangreichen Alters- und schulartspezifischen Normen von Heller (1973) für hörgeschädigte bzw. rest- und schwerhörige sowie taube Kinder und Jugendliche entnommen werden.

Kritik: Die PMT stellen ökonomische und brauchbare Verfahren zur Erfassung der allgemeinen Intelligenz dar, umso mehr als die SPM und APM auch als Gruppentests durchführbar sind. Als nonverbale Tests sind die PMT nicht nur für den sonderpädagogischen und klinischen Anwendungsbereich (z.B. für die Intelligenzdiagnose Hör- und/oder Sprachgeschädigter) interessant, sie haben sich auch im Rahmen der Begabungs- und Schuleignungsermittlung in der Regelschule, etwa als nonverbales Intelligenzdiagnostikum bei sprachlich oder überhaupt sozio-kulturell benachteiligten Probanden als nützlich erwiesen (z.B. Aurin, 1966). Für Blinde liegt eine amerikanische Adaptation vor (Rich, 1963). Allerdings hat Taschinski (1985) die Brauchbarkeit der PMT für die Diagnose intellektueller Fähigkeiten bei Kindern türkischer Arbeitnehmer angezweifelt; vgl. hierzu auch die Kontroverse zwischen Taschinski (1987) und Friedrich & Müller (1987).

Auch wenn zu den Raven-Matrizen-Tests insgesamt eine Fülle von Untersuchungsergebnissen vorliegt, dürfen bei der Beurteilung der Testgütekriterien einige Schwachpunkte nicht übersehen werden: Bis auf das Manual zu den CPM enthalten die Handanweisungen zu den deutschen Bearbeitungen nur wenige Angaben zu Testgütekriterien bei deutschen Stichproben. Weder für die SPM noch für die APM werden Reliabilitätsprüfungen oder ausreichende Validitätsprüfungen an deutschen Stichproben mitgeteilt. Die Verweise auf Untersuchungen aus den 50er und 60er Jahren im anglo-amerikanischen Raum können hier eigene Untersuchungen nicht ersetzen. Zudem erschwert die Vielzahl der qualitativ unterschiedlichen Untersuchungen mit ihren verschiedenartigen Ergebnissen eine Gesamtbewertung der APM und SPM. Im Vergleich der drei deutschen Bearbeitungen zeigt sich die Adaptation der CPM am besten gelungen. Das Manual zu den deutschen CPM hebt sich deutlich von den Handanweisungen zu den APM und SPM ab, sowohl was die Sorgfalt der Testkonstruktion, als auch was die Darstellung von Schwachstellen des Verfahrens angeht (vgl. auch Jäger, 1981c). Dennoch sieht Knaack (1978) bezüglich der Verwendbarkeit des CPM bei Vorschülern enge Grenzen gesetzt.

- (2) **Der Hamburg-Wechsler-Intelligenztest für Erwachsene (HAWIE)** in der deutschen Bearbeitung von A. Hardesty & H. Lauber, herausgegeben von C. Bondy (1956b) (Huber)

David Wechsler, als klinischer Psychologe am Bellevue-Hospital in New York tätig, veröffentlichte 1939 die Wechsler Bellevue Intelligence Scale (WB), der 1944 die Wechsler Bellevue Adult Intelligence Scale (WAIS) folgte. 1981 schließlich wurde die Revisionsform WAIS-R vorgelegt, die im deutschen Sprachraum noch keine Entsprechung gefunden hat, trotz der Ankündigung bereits bei Brickenkamp (1975). Der Hamburg-Wechsler-Intelligenztest für Erwachsene (HAWIE) stellt folglich nach wie vor die deutsche Bearbeitung der WAIS durch A. Hardesty & H. Lauber dar und wurde 1956 von C. Bondy herausgegeben. Dem Test liegt derzeit als Manual ein unveränderter Nachdruck der Teile des alten Handbuchs von Wechsler (1961) bei, die sich auf Durchführung und Auswertung des HAWIE beziehen. Zusätzlich wird auf die Publikation von Matarazzo (1982, erste Auflage des englischen Originals 1972) verwiesen, der die meisten der unten aufgeführten Testkennwerte sowie Untersuchungsergebnisse zu den Testgütekriterien, insbesondere zur Validität, entnommen werden können. Im Handbuch findet sich der Hinweis, daß der Monographie von Matarazzo (1982) von den deutschen Übersetzern auch eine moderne Bibliographie angefügt worden sei. In dieser (Bechmann, 1982) sind zwar nur Publikationen bis zum 31.12.1974 erfaßt, allerdings werden die bis zu diesem Stichtag im deutschen Sprachraum erschienenen Untersuchungen zum HAWIE kurz und übersichtlich referiert.

Testaufbau: Der HAWIE besteht aus einem Verbalteil (V-Teil) und einem Handlungsteil (H-Teil). Die 5 bzw. mit dem zusätzlichen Subtest "Wortschatz" 6

Untertests des V-Teils und die 5 Untertests des H-Teils ergeben zusammen den Gesamttest, der demnach folgende 10 bzw. 11 Subtestreihen enthält:

Verbalteil:

- 1) **Allgemeines Wissen (AW):** Die Probanden müssen maximal 25 Fragen der Art "Wie groß ist die deutsche Frau im Durchschnitt?" beantworten (vorgesehene Antwort: 1.60 bis 1.68 m).
- 2) **Allgemeines Verständnis (AV):** Es sind 10 Fragen vorgesehen, die mit 0 bis 2 Punkten bewertet werden. Beispiel: "Warum soll man Steuern zahlen?".
- 3) **Zahlennachsprechen (ZN):** Es sind je zwei Zahlenreihen der Länge 3 bis 9 Zahlen vorwärts, je zwei der Länge 3 bis 8 Zahlen rückwärts verfügbar. Die zweite Zahlenreihe jeder Länge wird nur vorgegeben, wenn bei der ersten ein Fehler gemacht wurde. Als Score wird die Länge der jeweils längsten korrekt wiedergegebene Zahlenreihe herangezogen.
- 4) **Rechnerisches Denken (RD):** 10 Textrechenaufgaben mit Zeitbegrenzungen, z.B. "Sieben Briefumschläge kosten 25 Pfennige. Wieviele kann man für eine Mark bekommen?" Zusatzpunkte für schnelle Bearbeitung bei den letzten beiden schwierigsten Aufgaben.
- 5) **Gemeinsamkeitenfinden (GF):** 12 Items, bewertet mit jeweils 0 bis 2 Punkten; Beispiel: "Was ist das Gemeinsame bei einer Apfelsine und einer Banane?".
- 6) **Wortschatztest (WT):** 42 Items, bewertet mit jeweils 0 bis 2 Punkten; Beispiel: "Was ist eine Konfirmation?".

Handlungsteil:

- 1) **Zahlensymboltest (ZS):** Maximal 67 Symbole müssen Ziffern zugeordnet und im Protokollblatt eingezeichnet werden.
- 2) **Bilderordnen (BO):** 7 Bilderserien müssen so geordnet werden, daß sich eine sinnvolle Geschichte ergibt. Die Bewertung richtet sich nach der Qualität der gelegten Reihe und der benötigten Zeit.
- 3) **Bilderergänzen (BE):** Auf 15 Bildtafeln müssen fehlende Teile der dargestellten Gegenstände erkannt werden.
- 4) **Mosaiktest (MT):** Mit farbigen Würfeln müssen bis zu 7 Muster nachgelegt werden. Pro Muster gibt es 0 (Aufgabe nicht gelöst) oder 3 bis 6 Punkte je nach benötigter Zeit.
- 5) **Figurenlegen (FL):** 3 Figuren müssen aus Einzelteilen zusammengesetzt werden (wie bei einem Puzzle). Die Qualität der Lösung sowie die benötigte Zeit gehen in die Bewertung ein.

Erfassungsdimensionen: Da Wechsler zunächst ein differentielles Intelligenzkonzept für diagnostische Zwecke ablehnte (vgl. seine Intelligenzdefinition in Abschnitt 1.4.1 oben), sollte der HAWIE (WAIS) vorab die **allgemeine Intelligenz** sensu Spearman et al. erfassen. Gleichwohl werden den einzelnen Subtests - bis auf den heutigen Tag - teilweise verschiedene Fähigkeitsbereiche zugeordnet (siehe u.a. Bondy, 1956b, bzw. Wechsler, 1961; Priester, 1964; Matarazzo, 1982). Die Korrelationen zwischen den einzelnen Untertests im HAWIE und dem Gesamttestergebnis fallen unterschiedlich hoch aus. Die Interkorrelationen der Subtests sind

insgesamt relativ hoch - sowohl im Intertestvergleich als auch in bezug auf den Gesamttest abzüglich der Wertpunkte des jeweiligen Untertests (im Mittel um 0.68) - und sprechen somit für eine starke Sättigung auf dem g-Faktor; die Korrelation zwischen V-IQ und H-IQ beträgt sogar 0.82. Den einzelnen Subtests werden folgende Dimensionen zugesprochen:

- AW** Einblick in die Wissensbreite einer Person; Aufgeschlossenheit der Umwelt gegenüber; Genauigkeit und Ökonomie des Denkens; Interessenrichtungen. Die Bearbeitung von AW-Aufgaben setzt normale Fähigkeiten voraus, verbale Kenntnisse zu erwerben; in der AW-Leistung kann sich der persönliche und familiäre "Bildungsstandard" widerspiegeln, d.h. die Leistung in AW ist stärker vom sozio-kulturellen Milieu abhängig. Rigide Persönlichkeiten weisen in AW überdurchschnittlich hohe Leistungen auf.
Altersbeständigkeit⁴: positiv (d.h. die AW-Leistung ist gegenüber altersmäßigem Intelligenzabbau verhältnismäßig resistent).
Korrelation mit dem Gesamttest⁵: $r=0.79$ (hoch).
- AV** Test des "gesunden Menschenverstandes"; situatives Verständnis, d.h. Einsicht in Feldzusammenhänge; praktische Urteilsfähigkeit; Aufgeschlossenheit; Plandenken; Fähigkeit, frühere Erfahrungen auszuwerten und zu verbalisieren. Bei affektiven Störungen häufiger niedrige Leistungen in AV, die auch von sozialer Reife und emotionalen Einstellungen abhängig sind.
Altersbeständigkeit: positiv.
Korrelation mit dem Gesamttest: $r=0.70$ (mittel).
- ZN** Kurzzeitgedächtnis oder Merkfähigkeit; Aufmerksamkeit im Sinne mühelosen, mehr passiven Registrierens äußerer Reize (im Gegensatz zur Konzentration!); Einstellungs- und Umstellfähigkeit. Schwachsinnige erzielen oft hohe ZN-Leistungen, allerdings nur im Zahlennachsprechen vorwärts, während Neurotiker im intraindividuellen Vergleich öfters bessere Leistungen im Zahlennachsprechen rückwärts erzielen. - Auffällig niedrig fallen die ZN-Leistungen (vorwärts und rückwärts) bei hirnorganischen Störungen der Probanden aus.
Altersbeständigkeit: negativ (d.h. in hohem Maße vom altermäßigen Intelligenzabbau betroffen).
Korrelation mit dem Gesamttest: $r=0.55$ (niedrig).
- RD** Formallogisch-adäquates Denken; rechnerisches Denken (ohne anspruchsvolle rechnerische Kenntnisse); Abstraktionsvermögen auf Operationsbasis; geistige Wendigkeit; Konzentrationsvermögen (ZN und ZS mitbeachten!). Die RD-Leistung ist häufig stärker abhängig von affektiven Erfahrungen (z.B. in der Schule).
Altersbeständigkeit: eher negativ.

⁴ Die Untersuchungsergebnisse zur Altersbeständigkeit der HAWIE-Untertests stammen überwiegend aus den 50er Jahren und erscheinen methodisch nicht unproblematisch. Trotz dieser Vorbehalte haben wir sie der Vollständigkeit halber aufgenommen.

⁵ Hier und im folgenden ist jeweils die Korrelation zwischen Untertest und Gesamtskala minus Wertpunktzahl des betreffenden Subtests gemeint.

Korrelation mit dem Gesamttest: $r = 0,75$ (mittel).

- GF Abstraktlogisches Denken; Abstraktionsfähigkeit; Denkkonsequenz und Genauigkeit; schlußfolgerndes Denken; begriffliches Denken im Verbalen bzw. verbale Begriffsbildung. Nach Wechsler ist die GF-Leistung ein hypothetisches Maß der Intelligenzkapazität (im Gegensatz u.U. zur aktuellen Leistungsfähigkeit). Nach Rapoport (zit. bei Priester, 1964, S. 231) kann anhand der GF-Leistung noch der Entwicklungsstand des Abstraktionsvermögens des Probanden nach drei Stufen hin betrachtet werden: 1) die **konkrete** Stufe, auf der Probanden an mehr oberflächlichen Einzelheiten bzw. anschaulichen Gegebenheiten haften bleiben (z.B. "Apfelsine und Banane haben beide eine Schale"); 2) die **funktionale** Stufe, wo das Beziehungsdenken prävalent ist (z.B. "Apfelsine und Banane - beide kann man essen"); 3) die **abstrakte** Stufe (z.B. "Beide sind Früchte"), auf der begrifflich operiert wird, d.h. unabhängig von Anschauungsqualitäten und/oder Beziehungsdenken (zur eigenen Person) wesentliche Gemeinsamkeiten zweier Gegenstände oder Sachverhalte hervorgehoben werden.

Altersbeständigkeit: negativ.

Korrelation mit dem Gesamttest: $r=0.81$ (hoch).

- WT Sprachliche Ausdrucksfähigkeit; allgemeines Wissen und Vorstellungsbesitz (Begriffsschatz); denktypologische Artung geistiger Operationen; Anschaulichkeit des Denkens. Maß der Lernfähigkeit; Bestand an sprachlichen Kenntnissen (abhängig von der Lebenserfahrung des Probanden sowie von sozio-kulturellen Determinanten).

Nach Rapoport ist der WT einer der zuverlässigsten Untertests im HAWIE, er repräsentiere die "ursprüngliche" Intelligenz.

Altersbeständigkeit: positiv.

Korrelation mit dem Gesamttest: $r=0.89$ (sehr hoch).

- ZS Gedächtnisleistung; Speedfaktor im Psychomotorischen bzw. allgemeine psychomotorische Geschwindigkeit; Einbildungsfähigkeit für bildlich-sinnfreie Figurationen; Assoziationsfähigkeit (mit Symbolen); visuell-motorische Koordination; Lernfähigkeit und Konzentrationsvermögen (im Handlungsbereich). Neurotisch Unsichere weisen öfters niedrige Leistungen im ZS auf.

Altersbeständigkeit: negativ.

Korrelation mit dem Gesamttest: $r=0.63$ (niedrig).

- BO Visualität; Fähigkeit, soziale Gesamtsituationen zu erfassen bzw. zu verstehen und auszuwerten; Produktivität; Beobachtungsgabe; Kombinatorik; Planning ability and Anticipation ("Einstellung" sensu Rapoport). BO erfordert ähnliche Funktionen wie AV, wobei jedoch folgender Unterschied beachtet werden muß: Bei BO erfolgt die Lösung mehr emotional und intuitiv, bei AV mehr rational; ferner spielt bei BO der Auffassungstyp (analytisch versus ganzheitlich) eine Rolle. Niedrig Intelligente perseverieren bei der Lösung von BO gern; sofern jedoch BO gut gelöst wird, ist Intelligenzmangel bzw. Schwachsinn auszuschließen. Dies gilt auch im Fall sog. Pseudodebilität, wenn bei-

spielsweise verwehrloste Jugendliche bei insgesamt schlechten HAWIE-Leistungen in BO gut abschneiden (Contra-Indikator einer Lernbehinderung!).

Altersbeständigkeit: negativ.

Korrelation mit dem Gesamttest: $r=0.62$ (niedrig).

- BE** Visuelles Erkennen und Identifizieren bekannter Figuren und Gegenstände; Perzeption und Begriffsbildung; Beobachtungsgabe; optische Differenzierungsfähigkeit; Fähigkeit, Wichtiges von Unwichtigem zu unterscheiden. Test der willkürlichen, fixierenden Aufmerksamkeit (Konzentration); als Konzentrationstest u.a. geeignet zur Feststellung von Lernbehinderungen.

Altersbeständigkeit: positiv.

Korrelation mit dem Gesamttest: $r=0.70$ (mittel).

BE differenziert allerdings unzureichend zwischen hoher und sehr hoher Intelligenz resp. Konzentrationsfähigkeit ("Deckeneffekt").

- MT** Umstrukturierungsfähigkeit; räumliches Vorstellungsvermögen; Abstraktionsfähigkeit; Differenzierungsfähigkeit; Konstruktivität; Kombinationsfähigkeit. Synthetische und analytische Fähigkeiten; visuell-motorische Koordination. Allgemein wird die Leistung im MT als ein guter Maßstab für "general intelligence" (Allgemeinbegabung) angesehen. Sofern die MT-Leistung hoch ist und die Leistungen in den übrigen Subtests niedrig ausfallen, besteht Verdacht auf neurotische Störung(en). Die MT-Leistung ist im allgemeinen niedrig bei starken Intelligenzdefiziten, bei vorzeitigem Altersabbau bzw. vorliegenden Hirnerkrankungen.

Altersbeständigkeit: negativ.

Korrelation mit dem Gesamttest: $r=0.76$ (mittel/hoch).

- FL** Ekphorierfähigkeit; Klarheit und Schärfe bildlicher Vorstellungen; Art der Zielgerichtetheit; Antizipation; planende Phantasie; visuell-motorische Koordination. Aufmerksamkeit; Sorgfaltstreben; Arbeitshaltung.

Anmerkung: Die Charakteristik des Vorgehens bei der FL-Bearbeitung sollte durch eine (zusätzliche) Verhaltensbeobachtung festgehalten werden!

Altersbeständigkeit: positiv.

Korrelation mit dem Gesamttest: $r=0.54$ (niedrig).

Testgütekriterien: Da bisher keine Revisionsform des HAWIE vorgelegt wurde, muß nach wie vor auf ältere Untersuchungen aus den 50er und 60er Jahren zurückgegriffen werden, die z.B. auch der Monographie von Matarazzo (1982) im wesentlichen zugrunde liegen. Im Anschluß an diese Befunde werden Ergebnisse einiger neuerer Arbeiten aufgeführt.

Eine systematische Itemanalyse wurde bei der Testkonstruktion des HAWIE nicht durchgeführt. Lienert & Fickert (zit. nach Priester, 1964, S. 234) haben nach Erscheinen des HAWIE 1958 anhand einer - freilich nicht populationsrepräsentativen - Gymnasialstichprobe ($N = 217$) Schwierigkeit und Trennschärfe des verbalen Aufgabenteils kontrolliert. Dabei erwies sich der Subtest RD bei einem mittleren Schwierigkeitsindex von $P=78$ (bei Gymnasiasten) als relativ leicht, während der Subtest WT bei einem $P=47$ etwas zu schwer erscheint; unter altersspezifischem bzw. nicht schultypischem Normenbezug darf wohl eine noch größere WT-Schwierigkeit als angegeben angenommen werden.

Lienert & Fickert fanden ferner, daß die Hälfte der Aufgaben des HAWIE-Verbalteils hinlänglich trennscharf ist; nur 17% der Aufgaben zeigten unzureichende Trennschärfeindizes.

Bachmann & Rauchfleisch (1981) sehen die Itemkennwerte des HAWIE insgesamt als befriedigend an, schlagen aber eine Reihe von Überarbeitungen vor.

Die **Objektivität** der Testdurchführung kann als gesichert gelten. Bei der HAWIE-Auswertung sind hingegen subjektive Einflüsse des Testleiters, z.B. in AV, GF und WT, nicht immer auszuschließen. Auch wenn sich der Testleiter des HAWIE genau an die vorgeschriebene Instruktion und Auswertungskriterien (mit Beispielhilfen) hält, kann es zu Unregelmäßigkeiten kommen, so daß die Auswertungsobjektivität insgesamt nicht als gesichert gelten kann. Daher wären Untersuchungen zur Auswertungsreliabilität des HAWIE desiderabel bzw. eine Revision ähnlich wie beim HAWIK-R angezeigt.

Die **(Split-half-)Reliabilität** des HAWIE beträgt 0.94, berechnet an einer Stichprobe von 370 Probanden im Alter von 20 bis 34 Jahren. Die Retest-Reliabilität liegt in bezug auf den Verbalteil zwischen 0.92 und 0.95, in bezug auf den Handlungsteil zwischen 0.80 und 0.86, in bezug auf die Gesamt-Punkte zwischen 0.90 und 0.94. Ähnlich hohe, d.h. befriedigende Stabilitätskoeffizienten wurden nur noch bei den Subtests AW (0.83-0.91), RD (0.73-0.84) und GF (0.85-0.90) gefunden, während ZN und BO die niedrigsten (Retest-)Reliabilitätskoeffizienten überhaupt aufwiesen. Kangas & Bradway (1971; zit. nach Matarazzo, 1982) fanden bei 48 Personen, die mit 29 und 42 Jahren mit den WAIS getestet wurden, für den Verbalteil eine Korrelation von 0.57, für den Handlungsteil von 0.70 sowie für den Gesamttest eine Korrelation von 0.73 zwischen erster und zweiter Testung.

Wenngleich die **Validität** des HAWIE bislang noch nicht systematisch untersucht worden ist, findet sich in der Literatur eine Fülle von Einzelergebnissen (unterschiedlichster Provenienz). Die bei Matarazzo (1982) zusammengefaßten Angaben weisen Korrelationen zu einzelnen Binetskalen zwischen 0.7 und 0.8 auf. Zum Schulerfolg werden - je nach Schultyp bzw. Erfolgskriterium und Probandengruppe - Zusammenhänge zwischen 0.4 und 0.6 mitgeteilt. Mit dem Intelligenz-Struktur-Test (IST) von Amthauer korreliert der HAWIE mit $r=0.74$.

Faktorenanalytische Befunde zum HAWIE haben für deutsche Stichproben z.B. Lienert & Fickert (1958) sowie Riegel (1960) vorgelegt. Demnach können vier Faktoren(gruppen) in der HAWIE-Leistung unterschieden werden: 1) ein **genereller (g-)Faktor**, der etwa 50% der Gesamtvarianz einnimmt; 2) ein Faktor **Sprachbeherrschung** (verbales Verständnis): "Sprachliches Verständnis ist die Fähigkeit, den Bedeutungsinhalt von einzelnen oder kombinierten Worten zu verstehen. Dieser Faktor wird am besten im WT, AW, AV und GF repräsentiert. Nach aphasischen Störungen zu beobachtende Defekte scheinen darauf hinzuweisen, daß dieser Faktor eine kortikale Lokalisierung aufweist" (Priester, 1964, S. 239); 3) ein **Handlungsfaktor** (nichtverbale Organisation): "Er wird besonders deutlich beim FL und beim MT. Das BO und das BE zeigen ebenfalls geringe Ladungen ... Hauptbestandteil scheint die Fähigkeit zur Organisation einzeln räumlich wahrgenommener, Einheiten zu größeren Ganzen oder Konfigurationen zu sein" (a.a.O.); 4) ein **indifferenzierter Gedächtnisfaktor**, von manchen Autoren auch als **Rechenfertigkeit** interpretiert: "Die höchsten Ladungen mit ihm zeigen ZN, ZS und bei einigen Altersgruppen RD und AW. Dieser Faktor repräsentiert nach Wechsler eine Art allgemeines Behaltenkönnen (general retentiveness), das operational wohl am besten als assoziatives Gedächtnis identifiziert werden kann. Interessanterweise steigt die Faktorenladung an diesem Faktor mit zunehmendem Alter deutlich an. Der ältere Mensch ist mehr und mehr abhängig vom früheren Erfahrungsgut (aufgespeicherten Wissen) und weniger von primären Fähigkeiten. Sobald ein gutes Gedächtnis ihn verläßt, ist nicht mehr viel vorhanden" (a.a.O.). Schließlich fanden Lienert & Riegel noch eine Reihe zusätzlicher (spezifischer) Faktoren, die

jedoch weniger als 2% der Gesamtvarianz ausmachen, somit die **Faktorenstruktur** des HAWIE nur unwesentlich beeinflussen.

In der Testliteratur finden sich immer wieder Hinweise für eine syndromatische Ordnung der HAWIE-Befunde auf der Basis der - zweifelhaften - Profilanalyse. Häufig werden beispielsweise folgende genannt (vgl. Matarazzo, 1982; Bechmann, 1982):

Organikersyndrom: niedrige HAWIE-Leistungen in ZS, FL und MT. Allerdings haben Scheller & Sittauer (1974; zit. nach Bechmann, 1982) eine bessere diskriminanzanalytische Trennung von Hirnorganikergruppen untereinander erreicht, wenn der MT nicht in die Analyse einbezogen wurde.

Schwachsinigensyndrom: sehr niedrige bzw. (intraindividuell) niedrige Leistungen in RD, ZS und AW (Faktor bzw. Faktorengruppe der "Lernfähigkeit").

Aphasiesyndrom: niedrige HAWIE-Leistungen in WT, AW, AV und GF (siehe oben).

Oberschulsyndrom (Gymnasialeignung): hohe Leistungen in AW, RD, GF und WT.

Bei der Verwendung dieser Indikationen sollte man stets die Reliabilität der betreffenden Subtests berücksichtigen.

Die drei wichtigsten differentialdiagnostischen **Profilauswertungen** zu den Wechsler-Tests seien - zusammen mit den vorgebrachten Kritiken - in aller Kürze erörtert.

a) Ein **Vergleich zwischen Verbal-IQ und Handlungs-IQ** ist die einfachste und wohl unproblematischste Form der HAWIE- bzw. HAWIK-Profilauswertung. So wird etwa ein Testleistungsdefizit im Wechsler-Verbalteil (V-IQ deutlich, d.h. mindestens 10 bis 15 Punkte kleiner als H-IQ) fallweise interpretiert als schlechte oder inadäquate Schulbildung, als Underachievement, als extrem praktische Befähigung, als soziale Fehlanpassung, als kortikale Schädigung (der dominanten Hirnhemisphäre) und dergleichen mehr. In der Regel wird sich der Stellenwert des Befundes (V-IQ kleiner als H-IQ) also erst aus dem übergeordneten Kontext ermitteln lassen. Entsprechend werden Testleistungsdefizite im Handlungsteil (H-IQ deutlich kleiner als V-IQ) fallweise als stärker reduziertes Allgemeininteresse (z.B. bei Depressiven), als einseitige (Schul-)Bildungseffekte versus hypertrophierte Anpassungstendenzen und dergleichen mehr gedeutet.

Im Hinblick auf die **Schuleignungsermittlung** sind vielleicht noch folgende Hinweise interessant: Gymnasiasten weisen öfters - aber keineswegs immer! - einen höheren Verbal-IQ im intraindividuellen Vergleich zum Handlungs-IQ auf, wohingegen **Realschülern** häufiger ein höherer Handlungs-IQ (als Verbal-IQ) nachgesagt wird. Nach einer Untersuchung von Pawlik zeigte eine Gruppe von Oberschulversagern ebenfalls bessere HAWIE-Leistungen im Handlungsteil im Vergleich zum Verbalteil. Solche oder ähnliche - mehr oder minder statistisch kontrollierte - Erfahrungshinweise bedeuten gelegentlich eine Hilfe im psychodiagnostischen Entscheidungsprozeß. Speziell bei eignungsdiagnostischen Fragestellungen wird man nach verlässlicheren Maßstabskriterien Ausschau halten müssen; insonderheit gilt dies für die Treffsicherheit von Diagnosen versus Prognosen im **Einzelfall**.

b) Vorab in der klinischen Diagnostik spielt der von Wechsler vorgeschlagene **Abbauindex** (Deterioration Index) eine Rolle. Dieser Index (DI) basiert auf der Annahme, daß die einzelnen psychischen Leistungsfunktionen vom (normalen) Altersabbau unterschiedlich stark betroffen seien. "Im allgemeinen wird ein Intelligenzabbau am besten durch eine Messung der Beantwortungsschwierigkeit, der Lernfähigkeit und der Fähigkeit, neue, besonders räumliche Konfigurationen zu erfassen, aufgedeckt" (Wechsler, 1961, S. 81). Nach Wechsler werden vom altersmäßigen Intelligenzabbau besonders stark betroffen die HAWIE-Subtests ZN, RD, ZS, MT und GF (BO ist hier fraglich), während die Subtests

AW, AV, BE und WT relativ stabil, d.h. gegen Abbau resistent sein sollen. Der Verfall kann als Prozentdifferenz

$$DI = \frac{\text{Summe WP stab. Tests} - \text{Summe WP instab. Tests}}{\text{Summe WP stab. Tests}} \cdot 100$$

oder als Abbauquotient

$$DI = \frac{\text{Summe WP instab. Tests}}{\text{Summe WP stab. Tests}} \cdot 100$$

ausgedrückt werden. "Um ein Maß des Abbaus zu gewinnen, vergleicht man die Summe der Wertpunkte der 'beständigen' Tests mit denen der 'nicht-beständigen' Tests. Dabei muß man Unterschiede in der Anzahl der Tests jeder Gruppe berücksichtigen. Für einen genaueren Vergleich stellt man die Summe der ersten vier 'beständigen' Tests derjenigen der ersten vier 'nicht-beständigen' Tests gegenüber. Das Ergebnis kann entweder als Quotient oder als Differenz beider Summen ausgedrückt werden. Sofern das Ergebnis als Differenz ausgedrückt wird, muß es in eine Prozentdifferenz umgerechnet werden, um die absolute Größe der verglichenen Summen zu berücksichtigen. Wenn also die Summe der 'beständigen' Untertests einer Person 50 ist und die Summe der 'nicht-beständigen' Untertests 40, beträgt der Verlust durch Abbau 20% und der Abbauquotient 80. Hoher Verlustprozentsatz oder niedriger Abbauquotient (erheblich unter 100) zeigen Abbauerscheinungen an" (S. 77f.). Diagnostisch ist natürlich der vorzeitige Abbau, etwa als Symptom hirnorganischer Krankheitsprozesse, interessant. Die Verwendung des Abbauindex ist verschiedentlich kritisiert worden, wobei der Einwand, daß Wechsler's Befunde zum Abbauindex auf Querschnittstudien (statt auf hier relevanteren Längsschnitten) basieren, wohl am schwersten wiegt. Auch konnte in verschiedenen Kontrolluntersuchungen die Reliabilität und Validität der indexmäßigen Subtestanalyse nicht eindeutig gestützt werden, so daß für entsprechende Interpretationsversuche mit Hilfe des Abbauindex größte Vorsicht am Platz ist, d.h. weitere Indikatoren berücksichtigt werden müssen.

c) Häufig wird auch die Wertpunkt-Variabilität der Subtests selbst zum Anlaß genommen, eine auf die oben beschriebenen Subtestdimensionen des HAWIE bezogene Profilanalyse durchzuführen, obwohl ja die Wechsler'skalen ursprünglich nicht multifaktoriell konzipiert waren. Die Verwendung des HAWIE als differentielles Intelligenzdiagnostikum ist jedoch nach zahlreichen empirischen Untersuchungen (verschiedener Autoren) keineswegs unproblematisch.

An eine profilanalytische Testauswertung werden allgemein folgende Forderungen gestellt: "Man erwartet, daß jeder Einzeltest einigermaßen reliabel ist und daß er mit jedem anderen möglichst niedrig korreliert. Desgleichen erwartet man, daß jeder Test unabhängig vom anderen eine ausreichende logische oder empirische Validität nachweisen kann" (Lienert, 1969, S. 366f.). Diese Voraussetzungen sind beim HAWIE nur bedingt erfüllt. So sind einmal die Testinterkorrelationen vergleichsweise hoch, zum andern ist die Subtestzuverlässigkeit (Split-half-Reliabilität) teilweise unzureichend für eine derartige profilanalytische Interpretation, wie etwa die Untersuchung von Priester & Kerekjarto (1960) zeigt. Die Profiltzuverlässigkeit der Wechsler-Tests, die sowohl die Subtestreliabilitäten (die hoch sein sollen) als auch die Intertestkorrelationen (die möglichst niedrig sein sollen) berücksichtigt, erreicht die für eine Profilauswertung allgemein geforderte Mindestgröße von $r_{\text{prof.}} = 0.5$ gerade noch, was aber vielfach als nicht ausreichend erachtet wird (z.B. Zimmermann, Kornmann & Lorenz, 1971, S. 49). Berücksichtigt man schließlich noch die faktorenanalytischen Befunde zum HAWIE und HAWIK (alt), dann können wir abschließend festhalten: Die Wechsler-Tests messen recht gut und zuverlässig die "allgemeine" Intelligenz sensu Spearman et al. Sie gestatten darüber hinaus einigermaßen gesichert Aussagen über die verbalen Fähigkeiten versus die Handlungintelligenz (mehr

praktische Fähigkeiten) einer Person. Spezielle Fähigkeiten wie Gedächtnis, Rechenfähigkeit und dergleichen mehr, sind jedoch nur bedingt mit Hilfe des HAWIE diagnostizierbar.

In neuerer Zeit haben z.B. Hartje & Orgass (1972, 1974), Bachmann & Rauchfleisch (1981), Baud & Rauchfleisch (1982), Sturm & Büssing (1982) sowie Scroko (1983) Arbeiten zur Diagnostik psychisch Kranker und neurologischer Patienten sowie hirnorganischer Störungen mit dem HAWIE bzw. der HAWIE-Kurzform WIP vorgelegt.

Die Normen des Wechsler-Tests sind für die einzelnen Subtests in WP-Einheiten (Wertpunkten) dargestellt, die durch einfache lineare Transformation aus den Rohpunkten gewonnen wurden ($M=10$; $s=3$). Sie basieren auf den HAWIE-Leistungen von 370 Probanden im Alter von 20 bis 34 Jahren der Standardisierungsstichprobe. Die WP-Summen des Verbal-, Handlungs- und Gesamttests ergeben - durch IQ-Transformation - den Verbal-IQ, Handlungs-IQ und Gesamt-IQ. "Zur Aufstellung der drei IQ wurden bei allen Altersstufen (Alter 10-59) die Wertpunktverteilungen ermittelt und deren statistische Konstanten berechnet. Die Ergebnisse zeigten, daß die Mittelwerte der Wertpunktsummen mit dem Alter fortschreitend bis zu etwa 28 Jahren ansteigen, um dann deutlich wieder abzufallen. Dieser Anstieg der Mittelwerte bis zum Alter 28 steht im Widerspruch zu den Normen anderer Tests. Bei der neuen Standardisierung des amerikanischen Originaltests (Wechsler, 1961) wurden ähnliche Ergebnisse gefunden, wobei der Anstieg der mittleren Werte von manchen Autoren (Anastasi, 1976) als ein Ergebnis einer breiteren modernen Erziehung angesehen wird. Nach unserer Meinung ist dieser Anstieg ebenfalls als Funktion der neuartigen Aufgabenzusammenstellung und der adäquateren Stichprobenerhebung anzusehen. Die Variationskoeffizienten der Wertpunkte sinken bei den Jugendlichen und steigen bei Erwachsenen mit dem Alter. Diese Resultate decken sich mit anderen Untersuchungen, sind aber bisher nicht eindeutig interpretiert worden. Die IQ beim HAWIE werden als Abweichungs-Intelligenzquotienten bezeichnet mit dem Mittelwert 100 und der Standardabweichung 15. Bei der Benutzung des HAWIE-IQ sollte immer beachtet werden, daß die Bewertung eines Probanden beständig mit der mittleren Leistung seiner eigenen Altersgruppe verglichen wird und nicht - wie etwa bei den Binet-Skalen - mit einem hypothetischen durchschnittlichen Erwachsenen-Intelligenzalter. Bei Vergleichen zwischen HAWIE- und anderen Intelligenzeinstufungen ist darauf zu achten, daß bei gleicher, konstanter oder objektiver Leistung eines Probanden der HAWIE diesen mit steigendem Alter höher einstufen wird, da seine Testbewertung mit ständig sinkenden Altersnormen verglichen wird. Wenn z.B. ein 20jähriger Proband den IQ 60 erhält, wird er bei gleicher Leistung im Alter von 50 den IQ 70 erhalten" (Priester, 1964, S. 235 f.). Sämtliche Normdaten, d.h. WP-Tabelle und IQ-Tabellen (Altersnormen von 10;0 bis 59 J.) finden sich im Textband zum HAWIE (vgl. Bondy, 1956b bzw. Wechsler, 1961; Matarazzo, 1982) bzw. im Separatdruck Wechsler (1982), der heute dem Test beiliegt.

Anwendungsbereiche: Der HAWIE gehört im klinischen Bereich nach wie vor zu den meistverwendeten (Individual-)Tests zur Messung der Intelligenz im späten Kindes-, Jugend- und Erwachsenenalter, also vom 10. Lebensjahr aufwärts bis zum 60. Lebensjahr und darüber. Obwohl schon von Brickenkamp (1975) angekündigt, ist bis heute keine deutsche Revisionsform vorgelegt worden. Die Normen und übrigen Testkennwerte, aber auch die Testkonstruktion (verwendetes Wertpunkt-konzept) müssen als veraltet gelten, weshalb sich die Verwendung des HAWIE in der Schulberatung nicht mehr empfiehlt und stattdessen auf die Revisionsform des HAWIK, den HAWIK-R (siehe unten) verwiesen sei.

(3) Der Hamburg-Wechsler-Intelligenztest für Kinder (HAWIK) in der deutschen Bearbeitung von F.P. Hardesty & H.J. Priester (1956, 1966³) (Huber)

Der HAWIK (Hardesty & Priester, 1966, bzw. Bondy, 1956a) stellt die deutschsprachige Adaptation der Wechsler Intelligence Scale for Children (WISC) dar, die Wechsler 1949 quasi als Erweiterung und Ergänzung der WAIS für jüngere Jahrgänge herausbrachte. Inzwischen ist jedoch als Revision des WISC die WISC-R (Wechsler, 1974) erschienen, und der HAWIK-R (Tewes, 1985²) hat im deutschsprachigen Raum den HAWIK weitgehend abgelöst. Deshalb soll nur eine knappe Darstellung des HAWIK gegeben werden. Für eine umfassende Kritik und neuere Forschungsergebnisse (aus den 70er Jahren) sei auf die ausgezeichnete Monographie von Kubinger (1983b) verwiesen.

Testaufbau: Die WISC bzw. der HAWIK sind im Aufbau mit der WAIS bzw. dem HAWIE nahezu identisch. Wie der HAWIE besteht der HAWIK aus einem Verbal- und einem Handlungsteil, die zusammen den Gesamttest ergeben. Statt des im HAWIE obligatorischen ZN und fakultativen WT wird im HAWIK der Subtest WT obligatorisch und ZN fakultativ (als Zusatztest) angeboten. Ansonsten gelten praktisch die oben angeführten Erfassungsdimensionen der 10 bzw. 11 Subtests auch hier.

Testgütekriterien: Für den alten HAWIK können in etwa die beim HAWIE ermittelten Daten zur Standardisierung, Objektivität, Reliabilität und Validität des Verfahrens - sinngemäß - übertragen werden. Retest-Reliabilität sowie interne Konsistenz des HAWIK in bezug auf die Gesamtskala sind durch Koeffizienten um 0.9 (0.91, 0.92 und 0.93 für die 7-, 10- und 13jährigen) ausgewiesen. Für die WISC wurde noch nach vier Jahren eine Retestreliabilität (Zeitstabilität) von $r=0.77$ ermittelt. Weitere Standardisierungsinformationen finden sich bei Priester (1958) sowie bei Wewetzer (1964). Nach Wewetzer (1964, S. 210) liegen die Interkorrelationen beim HAWIK für die Gruppe der 7-, 10- und 13jährigen Probanden bei 0.33, 0.42 und 0.39 (beim WISC zum Vergleich: 0.34, 0.43 und 0.39). Obgleich die Faktorenstruktur des HAWIE cum grano salis auch für den HAWIK relevant ist, scheint hier doch stärker der g-Faktor zu dominieren, was ganz im Sinne der Wechslerschen Intelligenzdefinition liegen dürfte. WISC und HAWIK wurden mit zahlreichen anderen Intelligenztests als Außenkriterien korreliert. Im allgemeinen zeigte sich eine gute bis sehr gute Übereinstimmung zu den verschiedenen Binetskalen, zum California Test of Mental Maturity von W.W. Clark et al. u.ä. Tests, wobei die Validitätskoeffizienten nicht selten Werte um 0.8 und höher erreichen. Allerdings streuen die mitgeteilten Übereinstimmungskoeffizienten teilweise beträchtlich, doch lassen sich mindestens zwei Trends beobachten: 1) die Abnahme der durchschnittlichen Übereinstimmungsvalidität bei ansteigendem Bildungsniveau (z.B. 0.8 bei Sonderschülern versus 0.6 bei Gymnasiasten) nach einer Untersuchung von Priester (1959) mit dem HAWIK bzw. Stanford-Intelligenztest sensu Lückert. In allen drei untersuchten Schülerstichproben (Lernbehinderte, Volksschüler, Gymnasiasten) lagen bei der Mehrzahl der Fälle, besonders deutlich bei den Oberschülern, die HAWIE-IQ über den Stanford-IQ, und zwar im Mittel bei 66% der Probanden. Im Gegensatz dazu zeigte sich 2) fast bei sämtlichen (übrigen) Binet-Test- und HAWIK- bzw. WISC-Vergleichen, daß der Wechsler-Test "schwerer" ist, wobei sich die größten Differenzen bei hohem Intelligenzniveau manifestierten (z.B. 130 Wechsler-IQ ~ 146 Stanford-Binet-IQ). Nach einer Untersuchung von Scheller (1973) fallen die Ergebnisse im HAWIE und HAWIK bei Oberschülern sehr ähnlich aus, lediglich im Verbalteil schneiden sie im HAWIK um ca. 3 IQ-Punkte besser ab. Bei Sonderschülern dagegen ergeben sich sowohl im Verbal- und Handlungsteil als auch in der Gesamtleistung jeweils um etwa 6 IQ-Punkte höhere Durchschnittswerte. Je nach Schultyp ist der HAWIE bzw. HAWIK mit dem Schulerfolg zwischen 0.4 und 0.6 korreliert.

(4) **Der Hamburg-Wechsler-Intelligenztest für das Vorschulalter (HAWIVA)**
in der deutschen Bearbeitung von D. Eggert (Hrsg.) (1975, 1978²) (Huber)

Der HAWIVA stellt eine deutsche Bearbeitung der Wechsler Preschool and Primary Scale of Intelligence (WPPSI) dar. Bisher ist jedoch nur eine sog. "Experimentalform" erschienen, die zur Frühdiagnose der Intelligenz im Alter von 4 bis 6;6 Jahren und zur frühen Erfassung geistiger Entwicklungsstörungen dienen soll.

Testaufbau: Die Autoren haben im Rahmen umfangreicher und sorgfältiger Arbeiten der Testentwicklung verschiedene Vorformen erprobt und letztlich eine Reduzierung auf acht Subtests vorgenommen, von denen die Untertests **Allgemeines Wissen (AW)**, **Wortschatz (WS)** und **Allgemeines Verständnis (AV)** dem **Verbalteil**, **Labyrinth (LA)**, **Figurenzeichnen (FZ)** und **Mosaiktest (MT)** dem **Handlungsteil** zugeordnet werden. Die Subtests **Rechnerisches Denken (RD)** und **Tierhäuser (TH)** werden keinem der beiden Testteile zugeordnet. Die Untertests AW, WS, AV, MT und RD entsprechen denen der übrigen Wechsler-Tests (HAWIE, HAWIK bzw. HAWIK-R).

Beim Untertest Labyrinth müssen mit dem Farbstift Labyrinth durchfahren werden, ohne die Begrenzungslinien zu berühren. Der Untertest Tierhäuser entspricht dem Zahlensymboltest im HAWIE bzw. HAWIK-R, nur müssen hier vier verschiedenen Tieren Farben zugeordnet werden.

Erfassungsdimensionen: Im Gegensatz zu allen anderen Wechsler-Tests und im Widerspruch zur Intelligenztheorie Wechslers wird beim HAWIVA auf die Ermittlung eines Gesamt-IQ verzichtet, da dieser wissenschaftlich im Lichte neuerer Ergebnisse der Intelligenzforschung unbefriedigend sei und gerade bei Selektionsentscheidungen im Rahmen sonderpädagogischer Diagnostik mißbraucht werden könne. Die Bearbeiter versuchten, diesbezüglicher Kritik am HAWIK Rechnung zu tragen, und meinen, daß eine "differenzierte Angabe von Leistungsdispositionen ... zu einer differenzierten Betrachtung des Leistungsspektrums des Schülers" anrege (Eggert, 1978, S. 7). Erfasst werden demnach die Leistungen im Verbal- und Handlungsteil sowie die Ergänzungstests Rechnerisches Denken und Tierhäuser als eigene Indikatoren für Leistungsdispositionen.

Der **Verbalteil** mißt nach Angaben im Manual den Grad des Sprach- und Wissenserwerbs auf dem Hintergrund sozialer Bezüge sowie schlußfolgerndes Denken auf der sprachlichen Ebene, der **Handlungsteil** die visuelle und visuomotorische Koordination, das Erfassen von Größen- und Lageverhältnissen sowie die Fähigkeit zur Durchgliederung komplexer Strukturen. Zu den Untertests Rechnerisches Denken und Tierhäuser vgl. den HAWIK-R (Untertests RD und ZS).

Handhabung: Der HAWIVA wird als Einzeltest durchgeführt und dauert 30 bis 45 Minuten, die auf mehrere Testsitzungen aufgeteilt werden können. Der Testleiter muß mit Durchführung und Handhabung sehr vertraut sein und für einen reibungslosen Ablauf der Untersuchung sorgen; die Aufgabenbearbeitung wird auf einem Antwortbogen mitprotokolliert.

Gütekriterien: Die Bearbeiter haben versucht, die **Durchführungs- und Auswertungsobjektivität** gegenüber dem amerikanischen Original zu steigern, jedoch

hängen beide davon ab, inwieweit sich der Testleiter exakt an die Vorgaben des Handbuchs hält. Die durchschnittlichen Schwierigkeiten und Trennschärfen der Items sowie die Reliabilität der einzelnen Skalen können befriedigen, die innere Konsistenz ist in allen Altersgruppen befriedigend bis hoch (um 0.9 für den Verbalteil, etwas darunter für den Handlungsteil). Bei verschiedenen Behinderten- gruppen ergaben sich für eine Vorform nach zwei bis sechs Wochen Stabilitäts- koeffizienten zwischen 0.83 (mittel) und 0.97 (sehr hoch).

Zwischen dem HAWIVA und dem HAWIK (alt) ergab sich eine Übereinstim- mungsalidität von $r=0.84$ für den Verbal- und von $r=0.71$ für den Handlungs- teil, für die meisten Subtests liegen die Werte etwas niedriger ($r=0.42$ bis $r=0.80$). Mit anderen Intelligenztestverfahren ergaben sich mittlere Zusammen- hänge (z.B. um 0.6 für die CMM). Prüfungen der faktoriellen Validität erbrachten eine Bestätigung der Testdimensionen von Verbal- und Handlungsteil.

Entsprechend der "Experimentalform" des Tests stehen nur nicht-repräsentative Normen (Prozentränge) für 4 bis 6;6jährige im halbjährlichen Abstand zur Ver- fügung.

(5) **Der Hamburg-Wechsler-Intelligenztest für Kinder - Revision (HAWIK-R)** in der deutschen Bearbeitung von U. Tewes (1983, 1985³) (Huber)

1974 legte Wechsler mit den WISC-R eine revidierte Form des WISC vor, wobei die Aufgaben des Handlungsteils überarbeitet und der Test neu normiert wurde. Auch an der deutschen Ausgabe wurde Kritik laut, vor allem galten die deutschen Normen als veraltet. Für eine Literaturübersicht und eine fundierte Kritik des HAWIK aus unterschiedlicher Perspektive sei auf Kubinger (1983b) verwiesen. Tewes (1985) hat die deutsche Version des Tests noch stärker verändert als Wechsler die WISC-R, vor allem wurden viele ideologisch vorbelastete Items überarbeitet. Dennoch sind Aufbau und Aufgabenstellung der Revisionsform HAWIK-R mit dem alten HAWIK (Hardesty & Priester, 1966) weitgehend identisch geblieben.

Testaufbau: Wie der HAWIE gliedert sich der HAWIK-R in einen Verbal- und einen Handlungsteil. In der Revisionsform enthält der HAWIK folgende Subtests (die Nummern beziehen sich auf die Reihenfolge bei der Durchführung).

Verbalteil:

2. **Allgemeines Wissen (AW):** 33 Fragen zum Allgemeinwissen (z.B. "Wieviele Beine hat ein Hund"). Bis auf 7 Items aus WISC-R bzw. HAWIK wurden die Aufgaben neu entwickelt.
7. **Allgemeines Verständnis (AV):** 20 Fragen wie "Warum sollen kranke Kinder zuhause bleiben?", die mit 0 oder 1 Punkt bewertet werden. Außer 2 Items wurden alle neu konstruiert.
6. **Rechnerisches Denken (RD):** Maximal 29 Zähl- und Textrechenaufgaben, bis auf 4 Items des WISC-R alle neu entwickelt.

10. **Gemeinsamkeitenfinden (GF):** 25 Items der Art "Sage mir jetzt bitte, was eine Säge und eine Kneifzange gemeinsam haben. Nenne ein gemeinsames Wort für beide!" 2 Punkte werden jeweils für einen Oberbegriff vergeben, 1 Punkt für ein gemeinsames Merkmal. Bis auf 2 Aufgaben wurde der Test neu entwickelt.
4. **Wortschatz-Test (WT):** Maximal 44 Fragen, davon 5 aus dem HAWIK, der Form "Was ist eine Fremdsprache?" werden den Kindern zur Beantwortung vorgelesen. Bewertung mit 0 oder 1 Punkt.
- (3.) **Zahlennachsprechen (ZN):** Es werden je zwei Zahlenreihen der Längen 3 bis 9 (vorwärts) bzw. 2 bis 8 (rückwärts) vorgelesen, die von den Kindern vorwärts bzw. rückwärts wiederholt werden sollen (maximal 28 Aufgaben). Für jede korrekte Wiedergabe wird ein Punkt vergeben. Die Zahlenreihen, nicht jedoch Darbietungsmodus und Auswertung sind identisch mit dem alten HAWIK.

Handlungsteil:

11. **Zahlen-Symbol-Test (ZS):** Bei Kindern bis 8 Jahren müssen maximal 93 Symbole einfachen, auf dem Protokollbogen aufgezeichneten Gegenständen zugeordnet und auf dem Bogen eingetragen werden. Ältere Kindern erhalten statt Gegenständen Zahlen vorgelegt. Bewertet werden die Zahl der richtigen Lösungen und die Bearbeitungsgeschwindigkeit. Der Test wurde unverändert aus dem WISC-R bzw. HAWIK übernommen.
1. **Bilderergänzen (BE):** Auf 33 Bildvorlagen müssen fehlende Details erkannt werden. Die Bilder des alten HAWIK wurden durch neue ersetzt.
9. **Bilderordnen (BO):** 12 Bilderserien sollen in eine sinnvolle Reihenfolge gebracht werden. Für schnelle Aufgabenlösungen gibt es außer bei den 3 leichtesten Items 3 bis 5 Zusatzpunkte. Bei den 3 schwersten Items wird auch je eine alternative Lösung anerkannt.
5. **Mosaik-Test (MT):** Mit 4 bzw. 9 Klötzen mit farbigen Mustern sollen maximal 17 Vorlagen (alle aus HAWIK, HAWIE bzw. WISC-R) nachgelegt werden. Bewertet wird auch die Schnelligkeit der Lösung (Zusatzpunkte).
8. **Figurenlegen (FL):** Maximal 10 neuentwickelte Vorlagen sollen aus Einzelteilen nachgelegt werden. Zusatzpunkte für schnelle Aufgabenlösung(en).

Erfassungsdimensionen: Die Gesamtleistung wird im Sinne Wechslers als allgemeine und übergeordnete Einheit gesehen, die vielfältige Erscheinungsformen intellektueller Leistungen annehmen kann und durch vielfältige Einflüsse in ihrer Entwicklung beeinflusst wird. Insofern ist auch der HAWIK-R einem Intelligenzkonzept verpflichtet, das allgemeine Intelligenz und geistige Fähigkeit gleichsetzt und spezifischen Fähigkeiten eine eher geringe Bedeutung zuspricht. Dennoch werden den einzelnen Untertests folgende Erfassungsdimensionen bzw. erfaßte Fähigkeiten zugeordnet (genauere Überlegungen zu den einzelnen Untertests finden sich bei Titze & Tewes, 1987):

AW Allgemeines Wissen: Breite des Allgemeinwissens (Faktenwissen), Aufgeschlossenheit gegenüber der Umwelt. Gute Leistungen sind von der Lernfä-

higkeit und der familiären und schulischen Sozialisation, aber auch vom Sprachvermögen abhängig.

- AV Allgemeines Verständnis:** Praktische Urteilsfähigkeit, Lernen aus Erfahrung(en), Denken in Ursache-Wirkungszusammenhängen, Erfassung von Situationen des Lebensalltags. Sprachvermögen, kulturelle Einflüsse und soziale Intelligenz beeinflussen das AV-Ergebnis.
- RD Rechnerisches Denken:** Durchführen numerischer Operationen im Kopf. Zahlengedächtnis, Konzentrationsvermögen, Sprachverständnis und Belastbarkeit bestimmen die Leistung mit.
- GF Gemeinsamkeitenfinden:** Logisches und abstraktes Denken in Kategorien. Die Leistung hängt auch vom sprachlichen Denk- und Ausdrucksvermögen sowie vom sprachlichen Abstraktionsvermögen ab; wesentliche und unwesentliche Merkmale müssen unterschieden werden können.
- WT Wortschatz-Test:** Allgemeine Sprachentwicklung, Bestand sprachlicher Kenntnisse. Das Lern- und Erinnerungsvermögen sowie die Entwicklung der Begriffsbildung spielen eine wichtige Rolle, auch kulturelle Einflüsse, Interessen sowie schulische und familiäre Sozialisation beeinflussen die WT-Leistung.
- ZN Zahlennachsprechen:** Die Leistung hängt von der akustischen Merkfähigkeit, der Aufmerksamkeit und Selbstkontrolle sowie der Übung im Umgang mit Zahlen ab. Ängstlichkeit und mangelne Streßbewältigung in Prüfungssituationen wirken sich negativ aus.
- ZS Zahlen-Symbol-Test:** Psychomotorische Geschwindigkeit, visuell-motorische Koordination, Konzentrationsvermögen bei Routineaufgaben. Die Belastbarkeit bei Zeitdruck, die geistige Flexibilität und das visuelle Kurzzeitgedächtnis beeinflussen die ZS-Leistung.
- BE Bilderergänzen:** Erkennen bekannter Formen, Gegenstände oder Figuren und Unterscheidung wesentlicher und unwesentlicher Details. Organisation der Wahrnehmungsprozesse, Flexibilität der Wahrnehmung, schnelles Erfassen von Figur-Hintergrund-Relationen spielen eine Rolle für die BE-Leistung.
- BO Bilderordnen:** Erfassen sozialer Handlungsabläufe auf visuellem Weg, Herstellen von Ordnungen und Sequenzen unter Unterscheidung wesentlicher und unwesentlicher Details. Die Leistung ist auch von der Organisation der visuellen Wahrnehmung, vom logischen Denkvermögen, von der Lernstrategie und der sozialen Intelligenz sowie von der geistigen Flexibilität, kreativen Fähigkeiten und familiären und kulturellen Einflüssen abhängig.
- MT Mosaik-Test:** Räumliches Vorstellungsvermögen, psychomotorische Koordination und Kombinationsfähigkeit. Durch Beobachtung wird erkennbar, ob nach einer "Versuch-Irrtum"-Strategie vorgegangen wird, das Kind aus Erfahrungen lernt, bei schwierigeren Aufgaben nach neuen Strategien sucht (Flexibilität) oder erfolgreiche Strategien aus leichteren Aufgaben beibehält.
- FL Figurenlegen:** Wahrnehmung und Reproduktion konkreter Figuren. Die visuell-motorische Koordination, das räumliche Vorstellungsvermögen, die Flexibilität des Denkens bei der Suche nach Lösungsstrategien sowie Kon-

zentrationenvermögen und Belastbarkeit unter Zeitdruck spielen eine Rolle bei der Bearbeitung der FL-Aufgaben.

Im Falle starker Profilschwankungen sollte der Testauswerter überlegen, ob die Leistungsausfälle auf eine gemeinsame Ursache zurückgeführt werden können. Im Ergänzungsband zum HAWIK-R (Titze & Tewes, 1987²) werden Vorschläge für eine solche "Profilauswertung" vorgestellt. Dazu werden Differenzen zwischen Subtestkombinationen ermittelt und auf Signifikanz geprüft. Entsprechend hohe (positive) Differenzen sollten im Hinblick auf Förderung oder Therapie beachtet werden. Diese **Profilauswertungen** können inzwischen auch auf den normalen Protokollbögen des HAWIK-R durchgeführt werden:

Verbalteil vs. Handlungsteil: Titze & Tewes (1987) warnen aufgrund der bisherigen Ergebnisse zum HAWIK-R vor der üblichen Interpretation, eine Leistungsminderung im Verbalteil als Anzeichen für mangelnde Förderung oder soziokulturelle Benachteiligung zu interpretieren. Eine solche Interpretation sei am ehesten angebracht, wenn Leistungsminderungen (nur) in RD und AW festgestellt werden.

Durchhaltevermögen: Erste vs. zweite Testhälfte. Leistungsminderungen bei zunehmender Testdauer können als mangelndes Durchhaltevermögen oder geminderte Belastbarkeit interpretiert werden. Es ist darauf zu achten, ob ein kontinuierlicher Abfall oder einzelne Leistungseinbrüche vorliegen.

Leistungsvermögen unter Zeitdruck: (AW+AV+RD+GF+WT[+ZN]+BE) vs. (ZS+BO+MT+FL). Zeitdruck kann sich motivierend oder als Störvariable bzw. Anzeichen für eine geminderte Belastbarkeit auswirken.

Lern- und milieuabhängige Leistungen: (AV+GF[+ZN]+ZS+BE+BO+MT+FL) vs. (AW+RD+WT). Titze & Tewes (1987) weisen explizit darauf hin, daß diese Interpretation auf Plausibilitätsüberlegungen und nicht auf empirischen Befunden beruht.

Konzentrationsvermögen und Belastbarkeit: (AW+AV+GF+WT+BE+BO+MT+FL) vs. (RD+ZN+ZS).

Visuell-motorische Koordination: (AW+AV+RD+GF+WT[+ZN]+BE+BO) vs. (ZS+MT+FL). Eine entsprechende Interpretation sollte nur im Zusammenhang mit genauer Verhaltensbeobachtung erfolgen. Damit soll vermieden werden, daß z.B. Zeitdruck fälschlicherweise als Störung der visuell-motorischen Koordination interpretiert wird.

Unterscheidung von Wesentlichem und Unwesentlichem: (AW+AV+RD+WT[+ZN]+ZS+MT+FL) vs. (GF+BE+BO). Auch hier bemerken Titze & Tewes (1987), daß diese Interpretation auf Plausibilitätsüberlegungen und nicht auf empirischen Befunden beruht.

Integrierende Hirnfunktion: (AW+AV+RD+GF+WT[+ZN]+BE+FL) vs. (ZS+BO+MT). Nach Titze & Tewes (1987) haben Untersuchungen - an allerdings kleinen Stichproben - erste empirische Belege für diese Interpretationsmöglichkeit erbracht.

Titze & Tewes (1987) weisen ausdrücklich darauf hin, daß diese Untertestkombinationen keine "Kochbuchanweisungen" (S. 106) darstellen und durch entspre-

chende weitere Befunde abgesichert werden müssen. Die Unterschiede zwischen den genannten Subtestkombinationen sind zwar durchaus plausibel, doch gilt es bei deren Interpretation zu beachten, daß jeweils auch ganz andere Syndrome zugrunde liegen können. Dies wird bereits darin sichtbar, daß sich einzelne der oben aufgeführten Untertestkombinationen nur geringfügig unterscheiden. Dennoch kommt u.E. Überlegungen darüber, wie bestimmte Leistungsausfälle zustande gekommen sein könnten, durchaus ein sinnvolle **hypothesebildende** Funktion zu.

Handhabung: Die Aufgaben des HAWIK-R werden in der Regel einzeln vorgelesen bzw. dargeboten, was eine hohe Vertrautheit mit dem Material erfordert. Bei den meisten Subtests werden nicht alle Aufgaben vorgelegt, sondern nur so viele, wie der Proband lösen kann. Nach Fehlleistungen von drei bis fünf Aufgaben hintereinander, wird der Versuch abgebrochen. Im Gegensatz zum HAWIE gibt es beim HAWIK-R für jede Altersgruppe eine eigene Wertpunkttabelle. Durchführung und Auswertung des HAWIK-R kann in Einzelfällen zwei bis drei Stunden in Anspruch nehmen. Die praktische Ringbindung des Manuals erleichtert die Durchführung und Auswertung des Tests. Besonders bei jüngeren, leistungsstarken Kindern kann die Testdurchführung sehr lange dauern, da bei den einzelnen Subtests mit leichten Aufgaben begonnen und das Abbruchkriterium erst spät erreicht wird. Für eine korrekte Auswertung und sinnvolle Interpretation muß die Publikation von Titze & Tewes (1987) herangezogen werden.

Testgütekriterien: In bezug auf die **Durchführungs- und Auswertungsobjektivität** wurde der HAWIK-R gegenüber älteren Versionen der Wechsler-Tests entscheidend verbessert. Bis auf den Untertest Gemeinsamkeitenfinden, für den wie bei den übrigen Subtests genaue Auswertungsrichtlinien für die Vergabe von 0, 1 oder 2 Punkten pro Aufgabe angegeben sind, werden die Aufgaben mit 0 oder 1 Punkt bewertet, dazu kommen gegebenenfalls noch Zusatzpunkte für die Schnelligkeit der Aufgabenlösung. Bei geübten Testanwendern dürfte die Auswertungsobjektivität somit gewährleistet sein. Dies wird auch durch entsprechende Untersuchungen belegt (Titze & Tewes, 1987, S. 47ff.).

Die **inneren Konsistenzen** der Subtests des HAWIK-R, geschätzt nach der Formel 20 von Kuder-Richardson, liegen nach Titze & Tewes (1987) mit Ausnahme von Bilderordnen (0.67) zwischen 0.71 und 0.89 (gemittelte Koeffizienten). Für den Verbalteil wurde eine Konsistenz-Reliabilität von 0.96, für den Handlungsteil von 0.91 und für die Gesamtleistung von 0.97 ermittelt. Da die Koeffizienten für jede Altergruppe getrennt berechnet wurden, was eine Einschränkung der Varianz bedeutete, können diese Werte als gut bis sehr gut beurteilt werden. Die Standardmeßfehler liegen entsprechend zwischen 1 und 1.8 Wertpunkten für die Untertests und zwischen 2.6 und 4.5 IQ-Punkten für die Testteile und den Gesamttest. Allerdings wird weder im Testmanual noch im Ergänzungsband von Titze & Tewes (1987) auf die Voraussetzungen zur Anwendung der Formel 20 von Kuder-Richardson eingegangen, obwohl diese (z.B. gleiche Iteminterkorrelationen bzw. Trennschärfen) bei den vorgenommenen Analysen eher nicht gegeben gewesen sein dürften. Überhaupt fehlen Angaben zu Itemkennwerten wie Schwierigkeit, Trennschärfe und Iteminterkorrelationen.

Die Stabilitätskoeffizienten der Untertests nach 7 bis 9 Monaten liegen erheblich niedriger: sie reichen von 0.03 für Bilderergänzen bei 14jährigen bis zu 0.74 für den Mosaik-Test bei derselben Stichprobe. Insgesamt sind von 33 Stabilitätskoeffizienten 14 nicht größer als 0.40, 11 liegen zwischen 0.40 und 0.60, 8 zwischen 0.60 und 0.80. Dieses Ergebnis stellt Selektionsentscheidungen mit dem HAWIK-R grundsätzlich in Frage. Die Korrelationen der Untertests untereinander liegen durchweg im niedrigen und mittleren Bereich.

Zur Validität des HAWIK-R werden im Ergänzungsband von Titze & Tewes (1987) nur wenige empirische Befunde mitgeteilt. Die HAWIK-R-Leistungen lernbehinderter Sonderschüler liegen bis zu 2 Standardabweichungen (Allgemeines Verständnis, Gemeinsamkeitenfinden) unter dem Mittelwert der jeweiligen Altersgruppe. Unterschiede zwischen den Testteilen konnten im Gegensatz zum alten HAWIK und zu einer Stichprobe von verhaltensgestörten Schülern (etwas niedrigere Leistungen im Verbal-IQ) jedoch nicht gefunden werden. Die Übereinstimmungsvalidität mit Schulnoten wurde für eine kleinere Stichprobe untersucht (Stichprobengröße für einzelne Korrelationen zwischen 56 und 92): Demnach korrelieren vor allem die Untertests des Verbalteils in mittlerer Höhe mit den Zensuren, die des Handlungsteils hängen eher niedrig mit den Schulnoten zusammen. Die multiplen Korrelationen der Noten mit den Untertests liegen außer für Sport und Musik zwischen 0.60 und 0.70. Schüler verschiedener Schularten unterscheiden sich erwartungsgemäß in der Leistungshöhe, es ergeben sich jedoch keine schularttypischen Profile. Die Autoren berichten ferner zur diskriminanten Validität des HAWIK-R, daß Korrelationen mit den konstruktfernen Skalen des Persönlichkeitsfragebogens für Kinder PFK (Seitz & Rausche, 1991) erwartungsgemäß durchweg niedrig (unter 0.30) waren.

Unterschiede zwischen HAWIK und HAWIK-R: Titze & Tewes (1987) referieren eine Untersuchung, nach der Kinder und Jugendliche im HAWIK-R einen um etwa 12.4 IQ-Punkte niedrigeren Gesamtwert (Verbal-IQ 7.1, Handlungs-IQ 15.4 Punkte) erzielen als im alten HAWIK, wobei bei der zitierten Untersuchung neue HAWIK-R-Protokolle mit älteren HAWIK-Ergebnissen verglichen wurden. Eggert, Liman & Schirmacher (1984) fanden bei 25 sprachgestörten Kindern, die im Abstand von 2 Wochen zunächst mit dem HAWIK und dann mit dem HAWIK-R getestet worden waren, noch größere Differenzen: Im Gesamt-IQ betrug die Differenz 15 Punkte, für den Verbalteil wurde ein Unterschied von 9.2, für den Handlungsteil von 18.4 IQ-Punkten gefunden. Die Korrelation zwischen den beiden Versionen des Tests betrug $r=0.77$ und $r=0.62$ für den Verbal- und Handlungsteil sowie 0.75 für die Gesamtleistung.

Schallberger (1987a u. b) hat in seiner Vergleichsstudie an einer Schweizer Stichprobe von Grundschulern versucht, mögliche Lerneffekte bei der Durchführung des HAWIK und HAWIK-R innerhalb von 3 bis 4 Wochen zu kontrollieren. Demnach zeigen sich sehr große Differenzen zwischen erster und zweiter Messung: Für den Verbal-IQ betragen diese 11.4, für den Handlungsteil 27.3 und für die Gesamtleistung 21.5 IQ-Punkte. Schallberger vermutet angesichts dieser erheblichen Differenzen, daß die 1956er Eichstichprobe des HAWIK nicht reprä-

sentativ, sondern eher leistungsschwach gewesen sein muß. Die Korrelationen zwischen den beiden Testversionen liegen für die Testteile und den Gesamttest zwischen $r=0.65$ und $r=0.70$, für die Untertests z.T erheblich niedriger (am niedrigsten für AV, RD, BE und FL, jeweils kleiner $r=0.40$). In Faktorenanalysen wurde eine zufriedenstellende Übereinstimmung in den Faktorenladungsmustern von HAWIK und HAWIK-R gefunden, was für eine gute inhaltliche Vergleichbarkeit der beiden Tests spricht.

Die Normierung erfolgte an einer Stichprobe von 1898 Schülern im Alter von 6 bis 15 Jahren, wobei sich die Testautoren im Rahmen der datenschutzrechtlichen Bestimmungen um Repräsentativität der Stichprobe bemüht haben. Tabellen für die Wertpunkte werden für Altersgruppen von jeweils 4 Monaten angeführt, so daß die einzelnen Normierungsstichproben in der Regel weniger als 70 Schüler enthalten haben müßten, was für einen Test der Bedeutung und Funktion des HAWIK-R zu wenig ist. Das Manual enthält schließlich Tabellen zur Umrechnung der Wertpunktsummen in die IQ-Werte für Verbal- und Handlungsteil sowie die Gesamtleistung. Im Manual finden sich ferner eine Umwertungstabelle für die Untertests des Verbalteils für den Fall der Reduzierung auf 5 Untertests, eine Umrechnungstabelle von Wert- in IQ-Punkte und Prozentränge sowie Testalteräquivalente für die Wertpunkte.

Kritik zum HAWIK-R: Zunächst mußten die Testautoren - auch aus Gründen des Copyrights - die testtheoretisch problematische Konzeption mit den Zusatzpunkten für schnelle Aufgabenlösungen, die Verrechnung über Wertpunkte etc. vom alten HAWIK übernehmen. Etwas unverständlich ist, warum weder in das Testmanual noch in den Ergänzungsband von Titze & Tewes (1987) Angaben zu den Itemkennwerten aufgenommen wurden. Überhaupt müßten alle für die Durchführung, Auswertung und Interpretation notwendigen Fakten im Testmanual enthalten sein; die Informationen des Ergänzungsbandes von Titze & Tewes (1987) sollten deshalb in das Handbuch integriert werden oder der Band jedem verkauften Exemplar des Tests beiliegen. So kennen viele Sonderpädagogen beispielsweise diesen Ergänzungsband immer noch nicht, obwohl er bereits 1984 erschienen ist. Faktorenanalytische Untersuchungen der Untertests wurden nicht vorgenommen. Die Validität des HAWIK-R ist insgesamt zu wenig empirisch belegt.

Von seiten der sonderpädagogischen Diagnostik wurde der HAWIK-R stark kritisiert. Ahrbeck, Lommatzsch & Schuck (1984) warnen insbesondere vor der Interpretation von Differenzen im unteren Extrembereich (unter $IQ=80$) und befürchten, daß aufgrund des HAWIK-R mehr Schüler als vor Erscheinen des Tests eine Intelligenzminderung bescheinigt bekommen könnten⁶ (vgl. auch Eggert, Liman & Schirmacher, 1984). Der Test scheint in der Tat für jüngere und leistungsschwächere Kinder zu wenig Aufgaben angemessener Schwierigkeit zu enthalten (Munz, 1984), was auch an den starken Sprüngen in den Wertpunktstabellen ersichtlich wird. Dieselben Autoren bemängeln ferner die Wertmaßstäbe der ver-

⁶ Die Testautoren haben dem insofern Rechnung getragen, als sie alle Kultusministerien bzw. Schulbehörden von den Normverschiebungen zum alten HAWIK unterrichteten.

balen Untertests, die nach wie vor an der Mittelschicht orientiert seien und Kinder unterer Sozialschichten benachteiligten. Sühlig & Sühlig (1984) zeigen dies im Detail für den Untertest Bilderordnen auf und können mit den meisten Bildergeschichten belegen, daß die Inhalte der Geschichten aufgrund der enthaltenen Rollenklischees sexistischen Charakter aufweisen. Vgl. hierzu auch die Stellungnahme von Kormann (1987c), der diese Darstellungen als "peinlich" (S. 60) bezeichnet. Kleber (1984) sieht im HAWIK-R insgesamt eine größere Gefahr für die Umschulungsdiagnostik als im alten HAWIK und bezeichnet die HAWIK-Revision als "unfairen Test" (S. 721). Für den sonderpädagogischen Bereich wäre deshalb angesichts der Schwächen der einschlägigen Testverfahren nicht nur eine gründlichere psychodiagnostische Ausbildung künftiger Sonderschullehrer, sondern auch eine kontinuierliche Fortbildung der in der sonderpädagogischen Diagnostik tätigen Lehrer zu fordern.

Zusammenfassende Kritik zu den Wechsler-Tests (HAWIE, HAWIK, HAWIVA und HAWIK-R): Die Wechsler-Skalen gehören zu den meistverwendeten (Individual-)Intelligenztests. Jedoch müssen der HAWIE und wohl auch die Kurzform WIP (Dahl, 1968, 1972) als veraltet gelten und sollten in der Schulberatung nicht mehr verwendet werden. Es wäre zu wünschen, daß bald eine Revisionsform des HAWIE vorgelegt wird, nachdem bereits 1981 die WAIS-R als Revision der WAIS erschienen ist. Die deutsche Revision HAWIK-R verfügt über eine **hinreichende Objektivität und Reliabilität** - sowohl des Gesamttests als auch seiner Teile. Auch die **Standardisierung und Testeichung** des HAWIK-R kann als ausreichend gelten. Damit stehen dem Diagnostiker mit dem HAWIK-R **vielseitige Anwendungsmöglichkeiten** offen. Der HAWIK bzw. HAWIK-R ist beispielsweise nach wie vor eines des Standardverfahren in der sonderpädagogischen Diagnostik, ohne daß freilich den genannten Kritikpunkten immer hinreichend Rechnung getragen wird.

So müssen wie bei früheren Versionen der Wechsler-Tests mindestens drei nicht unerhebliche **Nachteile** genannt werden (für eine umfassende Auseinandersetzung mit dem HAWIK vgl. Kubinger, 1983b): 1) die mehr oder weniger **willkürliche Auswahl** der (genauer: einzelner) **Intelligenzfaktoren** und die immer noch sehr **lückenhafte Testvalidierung**; 2) die teilweise zu **hohen Testinterkorrelationen**, die 3) eine - von Wechsler selbst befürwortete - **Profilauswertung** fragwürdig machen. Vielfach wurde auch der Einwand erhoben, daß die HAWIK-Leistungen zu sehr **milieuabhängig** ausfielen, was sich etwa in der verbalen Leistungsüberlegenheit (V-Teil gegenüber H-Teil) der Mittelschichtkinder gegenüber Unterschichtkindern manifestiere.

Ad 1: Die Wechsler-Skalen nehmen hinsichtlich ihrer **faktoriellen Erfassungsstruktur** gewissermaßen eine Zwischenstellung zwischen Binetarium als Form der traditionellen (allgemeinen) Intelligenzmessung und in Anlehnung an Faktorenmodelle der Intelligenz konzipierten differentiellen Fähigkeitstests ein. Diese Zwitterstellung kommt bereits in dem doppelten Anspruch Wechslers zum Ausdruck, einerseits die **Allgemeinbefähigung** (Intelligenzdefinition auf der Basis der g-Fak-

torhypothese) und andererseits wie auch immer umschriebene Strukturmomente der Intelligenz (Profilanalyse) zu erfassen. Unter den bisher faktorenanalytisch nachgewiesenen Faktoren beansprucht jedoch der g-Faktor mit Abstand den Löwenanteil (quer durch die einzelnen Subtests), weshalb wir HAWIE und HAWIK(-R) - entsprechend der Wechsler'schen Intelligenzdefinition - den Verfahren zur Messung der Allgemeinbegabung zuordneten.

Ad 2: Die teilweise hohe **Redundanz der Untertests**, die sich u.a. schon in den Korrelationskoeffizienten der Subtests in bezug auf den Gesamt-IQ andeutet und eo ipso auf eine hohe interne Konsistenz des (Gesamt)-Tests hinweist (vgl. auch die Beschreibung der Erfassungsdimensionen zum HAWIE oben und deren inhaltliche Überschneidung in den Subtests), legte die Erstellung von Kurzformen der Wechsler-Tests nahe. Nach Pawlik (1970, S. 139; vgl. auch Dahl, 1968, 1972) erwies sich dabei für den HAWIE folgende Subtestkombination als die **beste altersunabhängige Kurzfassung**: AW, AV, BO und FL. Die Ergebnisse dieser vier Unterteststufen korrelieren mit dem Gesamt-IQ (der Normal- oder Langform) mit 0.93, mit dem Verbal-IQ mit 0.91 und mit dem Handlungs-IQ mit 0.88. Allerdings können sich für unterschiedliche Altersstufen unterschiedliche Subtestzusammenstellungen als optimal erweisen. Aufgrund eigener empirischer Untersuchungen schlug Pawlik (a.a.O.) für die Altersgruppe der **Siebenjährigen** beispielsweise folgende Subtestkombinationen vor:

- a) AW, RD, ZS und FL (maximale Korrelation, d.h. optimaler Zusammenhang mit dem Gesamt-IQ = 0.95);
- b) AW und WT (maximale Korrelation mit dem Verbal-IQ = 0.85);
- c) BO und FL (maximale Korrelation mit dem Handlungs-IQ = 0.89).

Ad 3: Zweifellos die schwerwiegendsten Einwände wurden gegen die Wechsler-Tests als **Profilverfahren** vorgebracht. Eine Profilauswertung für differentialdiagnostische Zwecke erscheint von der Anordnung der Subtests bzw. Testteile her intuitiv sehr verlockend. Dementsprechend sind profilanalytische Auswertungsvorschläge auch zum HAWIK bis auf den heutigen Tag im Gebrauch (vgl. Titze & Tewes, 1987), gleichwohl sind sie aber auch umstritten (vgl. z.B. die Kontroverse zwischen Wallasch & Dony, 1980, und Kubinger, 1983a).

(6) Das **Adaptive Intelligenz Diagnostikum (AID)** von K.D. Kubinger und E. Wurst (1985, 1988²) (Beltz)

Das oder "der" AID wurde in Anlehnung an Wechsler's Intelligenzkonzept entwickelt. Ein wesentliches Ziel war dabei die Entwicklung eines Verfahrens, das zum einen der Vorliebe der Praktiker für die Wechsler-Tests bezüglich der Auswahl der Untertests Rechnung trägt, das aber gleichzeitig den Anforderungen moderner Testtheorien und Testkonstruktion genügt. Die Konstruktion der meisten Skalen nach dem Rasch-Modell ermöglichte als wesentliche Neuerung die adaptive Darbietung der Aufgaben, wobei die Testpersonen nicht alle Aufgaben eines Untertests bearbeiten müssen, sondern jeweils nur Items vorgelegt bekommen, die vom

Schwierigkeitsgrad her am besten zu ihren Fähigkeiten passen. Zudem ist bei einzelnen Untertests die Durchführung einer Kurzform als Screening-Verfahren (d.h. zur vorläufigen groben Identifizierung des Intelligenzniveaus) bzw. einer Über-Langform zum genaueren Erfassen einzelner intellektueller Fähigkeiten möglich.

Testaufbau: Ähnlich wie die oben beschriebenen Wechsler-Tests werden die Untertests des AID zwei Fähigkeitsbereichen zugeordnet, nämlich den **verbal-akustischen Fähigkeiten** mit 6 Skalen und den **manuell-visuellen Fähigkeiten** mit 5 Skalen. Die Zuordnung der Aufgabenreihen zu den beiden Testteilen entspricht ziemlich genau dem Aufbau des Verbal- bzw. Handlungsteils beim HAWIE oder HAWIK-R (die Nummern entsprechen der Abfolge der Subtests bei der Testdurchführung):

Verbal-akustische Fähigkeiten	Manuell-visuelle Fähigkeiten
1. Alltagswissen	2. Realitätssicherheit
3. Angewandtes Rechnen	4. Soziale und Sachliche Folgerichtigkeit
5. Unmittelbares Reproduzieren	7. Kodieren und Assoziieren
6. Synonyme-Finden	8. Antizipieren und Kombinieren
9. Funktionen-Abstrahieren	10. Analysieren und Synthetisieren
11. Soziales Erfassen und Sachliches Reflektieren	

Zusätzlich ist die Beurteilung der **Arbeitshaltung** des Probanden via Fremdrating durch den Testleiter möglich.

Erfassungsdimensionen: Die Autoren nehmen für den nach der probabilistischen Testtheorie (hier: Rasch-Modell) konstruierten AID in Anspruch, daß der Test zunächst ähnliche Fähigkeiten bzw. Fähigkeitsbündel wie die Wechsler-Tests erfaßt: Fertigkeiten von Kindern im Umgang mit verbalem und symbolischem Material sowie bei der Handhabung von Gegenständen und bei der Erfassung visueller Gestalten. Dennoch liege dem AID der "pragmatische Standpunkt" zugrunde, ein möglichst breites Spektrum von "Fähigkeiten, die (a-priori) für 'intelligentes' Verhalten verantwortlich scheinen, zu erfassen" (Handbuch, S. 15). Gleichzeitig wird die therapie- bzw. förderbezogene differential-diagnostische Interpretation der einzelnen AID-Untertests betont; auch die intellektuelle Lernfähigkeit soll mit Hilfe des **Lernquotienten** bestimmt werden können. Dieser letztgenannte Indikator wird als Differenz der Leistung zwischen der ersten und zweiten Testhälfte im Untertest "Kodieren und Assoziieren" bestimmt.

Als Indikator für die allgemeine Leistungsfähigkeit schlagen Kubinger & Wurst (1988) für den verbal-akustischen Teil, den manuell-visuellen Teil und den Gesamttest die **Intelligenzquantität** vor. Damit ist gemeint, daß anstelle der üblichen Berechnung der Testgesamtleistung als Durchschnitt bzw. Summe aller Untertests der niedrigste T-Wert herangezogen wird. Ähnlich wie eine Kette nur so stark sei wie ihr schwächstes Glied, hängt nach Ansicht der Autoren des AID die intellektuelle Leistungsfähigkeit, die sich z.B. in Schulleistungen ausdrückt, von vorhandenen Leistungstiefen ab. Die Schwächen könnten nicht ohne weiteres durch Stärken bei anderen Fähig- oder Fertigkeiten kompensiert werden, so daß nach Ansicht der Autoren des AID die übliche Bestimmung eines Gesamtwertes in der

Regel inhaltlich wenig sinnvoll ist. Dennoch kommen die Autoren des AID den Bedürfnissen der Praktiker mit einem über die Summe aller T-Werte bestimmten **Gesamt-IQ** entgegen.

Schließlich spiegelt der **Range der Intelligenz**, definiert als die Differenz zwischen dem kleinsten und größten T-Wert, die Spannbreite der intellektuellen Leistungsfähigkeit wider. Für die drei "Intelligenzquantitäten" wie für den "Range" der Intelligenz finden sich eigene Prozentrangnormen. Die 11 Untertests des AID werden folgenden Dimensionen zugeordnet.

Untertest Alltagswissen: Hiermit soll Aufschluß darüber gegeben werden, inwieweit sich ein Kind Sachkenntnisse über in unserer Gesellschaft alltägliche Inhalte erworben hat und erwerben kann. Diese Skala entspricht damit dem Untertest Allgemeines Wissen des HAWIK-R. Aus einem Aufgabenpool von 60 Items werden dem Kind 15 Fragen (Kurzform: 10 Fragen) gestellt (Beispiel: "Womit kannst Du riechen?"), die mündlich zu beantworten sind.

Untertest Realitätssicherheit: Analog zum HAWIK-R-Untertest Bilderergänzen prüft diese Skala, inwieweit der Proband die "Wirklichkeit um Dinge des Alltags" versteht und kontrollieren kann. Dem Kind werden dabei Bildkarten mit Gegenständen vorgelegt, auf denen fehlende Details entdeckt werden müssen. Von den insgesamt 20 Aufgaben werden 10 zur Bearbeitung vorgegeben (Über-Langform: 15 Items).

Untertest Angewandtes Rechnen: Hier müssen die Kinder wie beim Subtest Rechnerisches Denken des HAWIK-R 15 (von insgesamt 60; Kurzform: 10) Textaufgaben lösen, mit denen unabhängig von schulischen Rechenfertigkeiten die Problemlösefähigkeit der Kinder bei alltäglichen Aufgabenstellungen (Schlußfolgerungen und Anwendung entsprechender Rechenoperationen) geprüft werden soll. Beispiel: "Auf einem Leuchter brennen 4 Kerzen. Ich blase 3 aus. Wie viele brennen noch?"

Untertest Soziale und Sachliche Folgerichtigkeit: Die Fähigkeit, die "Abfolge sozialen Geschehens bzw. alltäglicher Sachgegebenheiten zu verstehen und zu kontrollieren" (S. 17), wird mit dieser Skala erfaßt. Wie beim Untertest Figurenlegen des HAWIK-R müssen die Kinder dabei ungeordnete Bildfolgen von Geschichten in die richtige Reihenfolge bringen. Von insgesamt 20 Aufgaben werden 7 (Über-Langform: 9) Items bearbeitet.

Untertest Unmittelbares Reproduzieren: Der verbal-akustische Aspekt der Konzentrationsfähigkeit eines Kindes soll hier ermittelt werden. Analog zum Zahlennachsprechen beim HAWIK-R müssen Zahlenreihen aus zwei bis neun Zahlen (beim HAWIK-R maximal nur acht Zahlen) vorwärts und rückwärts nachgesprochen werden. Der Untertest, der nicht Rasch-skaliert werden konnte, wird abgebrochen, wenn das Kind alle drei Reihen einer Länge nicht reproduzieren kann. Sowohl für "vorwärts" wie für "rückwärts" wird ein eigener T-Wert berechnet.

Untertest Synonyme Finden: Mit dieser Skala, die dem Wortschatz-Test des HAWIK-R entspricht, soll der passive Wortschatz der Testpersonen geprüft werden. Dazu muß das Kind zu einem vorgegebenen Wort ein anderes mit derselben

Bedeutung nennen (Beispiel: "Sag' mir ein anderes Wort für kräftig!"). 15 der 60 Items sind zu bearbeiten (Kurzform: 10 Aufgaben).

Untertest Kodieren und Assoziieren: Damit wird die "latente Lernfähigkeit" (Lernquotient) des Probanden ermittelt. Diese sei die "Fähigkeit, in bestimmten Problemsituationen selbständig Lösungsstrategien oder Fertigkeiten (Assoziationen) zu entwickeln" (S. 18). Daneben wird die Schnelligkeit der Informationsverarbeitung bei symbolischem Material im manuell-visuellen Bereich erfaßt. Dazu muß das Kind zunächst zwei Minuten lang nach einer Vorlage einfache geometrische Symbole 12 anschaulichen Objekten (Zitrone, Apfel, Schirm usw.) zuordnen und einzeichnen, wobei der Versuchsleiter nach einer Minute die bis dahin bearbeitete Kodiermenge notiert. Anschließend wird die Vorlage weggenommen und überprüft, wieviele der 12 Symbole den Objekten noch zugeordnet werden können, d.h. wieviele der Zuordnungen (Assoziationen) gelernt wurden. Die Kodiermenge und die Anzahl der ausgebildeten Assoziationen werden gesondert ausgewertet. Der Lernquotient wird als Verhältnis richtig bearbeiteter Symbole zwischen der ersten und zweiten Halbzeit beim Kodieren ermittelt. Die Aufgabenstellung, nicht jedoch die Auswertung und die inhaltliche Einordnung, entspricht dem Zahlen-Symbol-Test des HAWIK-R. Der Untertests ist nicht Rasch-skaliert.

Untertest Antizipieren und Kombinieren: Hiermit soll das schlußfolgernde Denken geprüft werden, insbesondere die Fähigkeit, "Teile eines (konkreten) Ganzen zu erkennen und dieses Ganze zu gestalten" (S. 18), d.h. das Ganze aus den Teilen zusammensetzen. Analog zum Figurenlegen des HAWIK-R muß eine Figur wie bei einem Puzzle aus Einzelteilen zusammengesetzt werden, allerdings wird beim AID keine Vorlage gezeigt und der Gegenstand nicht benannt. Dafür erhalten die Kinder bei jeder Aufgabe ein "Anker"-Teil mit der Instruktion, daß alle anderen Teile daran angefügt werden können. Jedes Kind bearbeitet 7 der insgesamt 11 Aufgaben, wobei die Geschwindigkeit der Lösung jeweils mitbewertet wird.

Untertest Funktionen Abstrahieren: Wie der HAWIK-R-Untertest Gemeinsamkeitenfinden zielt diese AID-Aufgabenreihe auf die Erfassung der Begriffsbildungsfähigkeit, d.h. durch die Abstraktion der wesentlichen Funktionen zweier Objekte soll deren gemeinsame Eigenschaft gefunden werden (Beispiel: "Was ist das Gemeinsame an Pinsel und Kreide?"). Jeder Proband bearbeitet 15 von 60 Aufgaben (Kurzform: 10 Aufgaben).

Untertest Analysieren und Synthetisieren: Es werden geometrische Muster vorgegeben, die - wie beim Mosaiktest des HAWIK-R - mit entsprechenden Würfeln nachgelegt werden sollen. Erfaßt wird die Fähigkeit, "komplexe (abstrakte) Gestalten durch eine geeignete Strukturierung zu reproduzieren" (S. 19). Die Kinder bearbeiten nach zwei Aufwärmaufgaben 7 von 20 Aufgaben (Über-Langform: 9 Aufgaben).

Untertest Soziales Erfassen und Sachliches Reflektieren: Hiermit soll wie mit dem HAWIK-R-Subtest allgemeines Verständnis geprüft werden, d.h. inwieweit die Kinder in der Lage sind, Sachzusammenhänge der gesellschaftlichen Umwelt zu begreifen und inwieweit ihnen sozial angepaßte Verhaltensweisen sowie gesell-

schaftliche Bedingungen bekannt sind. Wieder werden 15 Aufgaben aus dem 60 Items umfassenden Itempool vorgegeben (Kurzform: 10 Aufgaben; Beispiel: "Warum sollte man Obst vor dem Essen waschen?").

Praktische Handhabung: Der AID erfordert vom Testleiter aufgrund der adaptiven Vorgabe ein hohes Maß an Vertrautheit mit dem Testmaterial. Fast alle Untertests bestehen aus mehreren Gruppen von jeweils annähernd schwierigkeits-homogenen Aufgaben. Je nach Alter der Schüler beginnt man mit einer einfacheren oder schwierigeren Aufgabengruppe. Je nachdem, wieviele Aufgaben pro Gruppe korrekt gelöst wurden, wird zu einer Gruppe mit leichteren, gleichschweren oder schwierigeren Aufgaben verzweigt. In der Regel arbeiten die Kinder zweier oder drei Aufgabengruppen durch. Abbildung 16 zeigt die Anordnung und Verzweigungsmöglichkeiten der Aufgabengruppen für die Untertests mit 60 Items. Die Angabe im Manual, wonach die Durchführung der Normalform 60 Minuten in Anspruch nehme, scheint aufgrund praktischer Erfahrungen etwas zu niedrig angesetzt zu sein.

Die Bewertung der Aufgaben als richtig bzw. falsch sowie ein Teil der Summenbildungen erfolgen wegen der adaptiven Darbietung bereits während der Testdurchführung. Zu jeder für Normal-, Kurz-, Über-Lang- oder Parallelform vorgesehenen Kombination von Aufgabengruppen findet sich eine Tabelle zur Um-

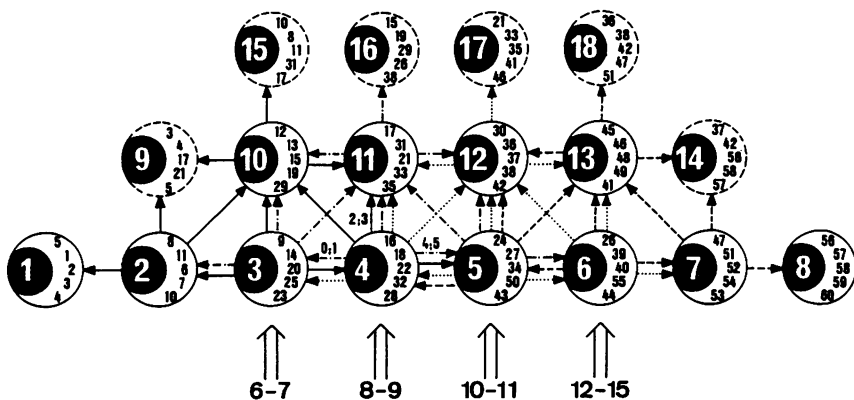


Abbildung 16: Testvorgabe beim AID nach dem Prinzip des "branched testing"

rechnung des Rohwertes in den Fähigkeitsparameter. Mit Hilfe der Fähigkeitsparameter kann man innerhalb probabilistischer Testmodelle die Leistungen von Schülern vergleichen, auch wenn sie unterschiedliche Aufgaben bearbeitet haben. Den Fähigkeitsparametern schließlich können altersspezifische T-Werte zugeordnet werden, aus denen die Testkennwerte zur Intelligenzquantität etc. bestimmt werden können. Für diese wiederum finden sich Tabellen mit Prozenträngen. Die Interpretation wird durch ein Fallbeispiel unterstützt.

Testgütekriterien: Durchführungs- und Auswertungsobjektivität scheinen durch genaue Instruktionen und Auswertungsrichtlinien im Vergleich zu anderen Individualverfahren recht gut gewährleistet. Dies zeigen auch die Prüfungen der Testleiter- und Übungseffekte nach dem multiplikativen Poissonmodell von Fischer (1977). Mit Beeinträchtigungen der Objektivität ist demnach vor allem bei den Untertests Synonyme-Finden und Funktionen-Abstrahieren zu rechnen.

In der **testtheoretischen Analyse** der einzelnen AID-Skalen wird mittels Maximum-Likelihood-Modellgeltungs-Tests sowie graphischen Modelltests überzeugend nachgewiesen, daß die meisten Untertests als Rasch-skaliert angesehen werden können, womit die Voraussetzung für die adaptive Aufgabenvorgabe nach dem Prinzip des "branched testing" gegeben ist.

Obwohl für Tests, die nach probabilistischen Testmodellen wie dem Rasch-Modell konstruiert sind, die Bestimmung der **Reliabilität** eigentlich nicht vorgesehen ist, finden sich im Manual Angaben zur Split-Half- und Retest-Reliabilität. Bis auf die Untertests Realitätssicherheit ($r=0.70$) und Antizipieren und Kombinieren ($r=0.81$) liegen alle Split-Half-Koeffizienten zwischen 0.91 und 0.95, womit eine sehr hohe Zuverlässigkeit deutlich wird, was sich in entsprechend niedrigen Standardmeßfehlern ausdrückt. Dieses Ergebnis kommt aber wohl auch dadurch zustande, daß die Eichstichprobe ($N=1460$) mit ihrer entsprechend großen Varianz für diese Berechnungen herangezogen wurde. Auch die einzelnen Stabilitätskoeffizienten (aus Retestung bzw. kombinierter Re- und Paralleltestung) liegen bis auf die Realitätssicherheit durchweg über 0.8 oder 0.9. Allerdings wird nicht angegeben, wie sich die Stichprobe zur Ermittlung der Retest-Reliabilität zusammensetzte.

Die **inhaltliche Gültigkeit** der AID-Items wurde durch Expertenurteil (Fachbeirat, Klinische Psychologen, Schulpsychologen) überprüft. Die **Übereinstimmungsvalidität** im Rahmen der Konstruktvalidität wurde durch Korrelation der AID-Scores von 153 Kindern mit den PSB-Untertests überprüft. Während die Intelligenzquotienten mit 0.62 korrelieren, liegen die Koeffizienten für die Zusammenhänge der Untertests wesentlich niedriger. Mittlere Koeffizienten über 0.5 finden sich nur zwischen dem PSB Untertest 1+2 (Allgemeinbildung) und den AID Skalen Alltagswissen und Synonyme-Finden. Im Sinne einer **Extremgruppenvalidierung** konnte gezeigt werden, daß der AID zwischen Schülern aus Lernbehindertenschulen und Kindern hochbegabter Eltern (MENSA-Mitglieder) zu differenzieren imstande ist.

Im Zuge der **Konstruktvalidierung** wurden die AID-Untertests faktorenanalysiert. Die vier extrahierten Faktoren werden mit "Informationsverarbeitung in der gesellschaftlichen Umwelt" (Untertests 1, 2, 3, 6, 9 und 11), "Informationsverarbeitung neuer Inhalte" (Kodieren und Assoziieren), "Konzentrationsfähigkeit" (Unmittelbares Reproduzieren) und "(Re-)Produktionsfähigkeit durch Strukturierung" (Untertests 8 und 10) benannt, wobei kritisch anzumerken ist, daß wir angesichts des Eigenwertverlaufes für eine Ein-Faktor-Lösung plädiert hätten. Bei einer schrittweisen Regression des globalen Intelligenzkoeffizienten - gegen dessen Brauchbarkeit die Autoren erhebliche Einwände vorbringen - auf die AID-Unter-

tests werden durch Angewandtes Rechnen bereits 57 Prozent der Varianz aufgeklärt. Bezieht man die nächsten vier Subtests (Funktionen-Abstrahieren, Kodiermenge, Antizipieren und Kombinieren sowie Alltagswissen) ein, so steigt der Anteil gemeinsamer Varianz bereits auf 88 Prozent.

In die Normierung des AID gingen insgesamt 2144 Kinder (davon über 1000 Testungen aus der Analysestichprobe) aus der Bundesrepublik Deutschland und Österreich ein. Es finden sich Jahresnormen für 6- bis 15jährige Schüler im jährlichen Abstand, so daß mit Ausnahme der 15jährigen (N=70) alle Normierungs-Teilstichproben aus mehr als 200 Schülern bestehen.

Kritik: Mit dem AID haben Kubinger & Wurst (1988) einen Fähigkeitstest vorgelegt, der aufgrund der sorgfältigen Konstruktion nach der probabilistischen Testtheorie und der adaptiven Vorgabe nach dem Prinzip des "branched testing" gravierende Nachteile anderer Individualverfahren weitgehend vermeidet. Von daher ist dem AID eine weite Verbreitung in der Schul- und Erziehungsberatung zu wünschen. Dieses positive Urteil wird jedoch durch eine Reihe von Schwachpunkten getrübt.

Insgesamt halten die Testautoren trotz erheblicher Bedenken am Testkonzept Wechslers fest. Jeder Subtest entspricht weitgehend einem Untertest des HAWIK-R. Im Hinblick auf die erfaßten Indikatoren fällt immer wieder der Widerspruch zwischen verschiedenen Forderungen auf: Zum einen stellen die Autoren zu recht hohe Ansprüche an die Konstruktion und die aus Untertests abgeleiteten Indikatoren, zum anderen wird im Hinblick auf die Praxis doch immer wieder auf explizit abgelehnte Konzepte zurückgegriffen.

Schließlich ist fraglich, ob die "Intelligenzquantität" tatsächlich jene Rolle spielt, die die Autoren ihr zuweisen. So kann die unspezifische Intelligenzquantität genauso wie ein mittlerer T-Wert im Einzelfall die genauen Problembereiche verschleiern. Je nach Leistungsbereich können wohl unterschiedliche Fähigkeiten eine entscheidende Rolle spielen. Zudem wird das Konzept der Intelligenzquantität dadurch fragwürdig, daß einige Untertests gerade im unteren Leistungsbereich nur schlecht differenzieren, womit die Einsatzmöglichkeiten in der Lernbehinderten-diagnostik eingeschränkt werden. Auch die Interpretation des Lernquotienten als Maß für die intellektuelle Lernfähigkeit scheint uns etwas gewagt. Die von der Forschungsgruppe um Guthke entwickelten Testverfahren zum Lerntestkonzept scheinen auf einer besser abgesicherten Grundlage zu stehen.

Kormann (1987c) hat darauf verwiesen, daß die materielle Qualität des Testmaterials zum Teil den Ansprüchen der Praxis nicht genügt. Der für den Anwender gravierendste Nachteil dürfte jedoch darin zu sehen sein, daß im Hinblick auf die differential-diagnostische Verwendung des AID so gut wie keine empirischen Ergebnisse vorliegen. Stattdessen finden sich hierzu lediglich Plausibilitätsüberlegungen, die von den operationalen Definitionen der Untertests ausgehen, zum Teil wird auch auf die Validität entsprechender Subtests des HAWIK verwiesen. Viele praktisch tätigen Psychologen stören sich auch an der für sie umständlichen Auswertung, was allerdings zum überwiegenden Teil auf mangelnde Vertrautheit mit der probabilistischen Testtheorie zurückzuführen sein dürfte. Viele Anwender

scheinen die Möglichkeiten der flexiblen Arbeit mit dem AID noch gar nicht entdeckt zu haben. Ähnlich wie auch beim (differentiellen) Kognitiven Fähigkeitstest (KFT 4-13+) gilt es hier, die Mühe der Einarbeitung in das adaptive Verfahren auf sich zu nehmen, um dessen Vorzüge für eine differenzierte Erfassung intellektueller Fähigkeiten voll zu nutzen. Von einer einfach zu handhabenden Pocketkamera wird man schließlich auch nicht gleiche Bildqualitäten erwarten wie von der einen gewissen Bedienungsaufwand erfordernden Hochleistungskamera. Wenn man trotzdem die erste Alternative bevorzugt, sollte man sich aber der Konsequenzen bewußt und nicht vom Ergebnis enttäuscht sein!

3.3.1.2 Gruppenverfahren

- (1) Die **Culture Fair Intelligence Tests (CFT)** von R.B. Cattell in den deutschen Bearbeitungen von R.H. Weiß & J. Osterland sowie R.H. Weiß (TZ)

Von R.B. Cattell stammen mehrere ursprünglich als kulturunabhängig (culture free) bezeichnete Intelligenztests, so der "IPAT Culture Free Intelligence Test, Scale 1" für Kinder im Alter von 4 bis 8 Jahren, der "IPAT Culture Free Intelligence Test, Scale 2" für Kinder im Alter von 8 bis 13 Jahren sowie der "IPAT Culture Free Test, Scale 3" für Jugendliche und Erwachsene. Hierzu sind inzwischen drei, für deutsche Verhältnisse adaptierte Testversionen erschienen (CFT 1, CFT 20, CFT 3). Während der CFT 20 und der CFT 3 bis auf den Schwierigkeitsgrad identisch sind, unterscheidet sich der CFT 1 von den beiden anderen Versionen deutlich.

- a) Grundintelligenztest Skala 1 (CFT 1) von R.H. Weiß & J. Osterland (1977, 1980⁴) für Vorschulkinder (ab 5 Jahren) und die ersten drei Grundschulklassen

Obwohl der CFT 1 grundsätzlich eine Adaptation des "Culture Fair Intelligence Test" von Cattell (1960) darstellt, mußten im Hinblick auf die deutsche Zielgruppe praktisch alle Testaufgaben bei der Bearbeitung neu konstruiert werden, so daß nur das Grundprinzip der Subtests von der amerikanischen Testversion übernommen werden konnte.

Testaufbau: Der CFT 1 gliedert sich in fünf Untertests mit folgender Aufgabenstellung: Beim Subtest 1 **Substitutionen** müssen Symbole unter zugehörige figürliche Darstellungen gesetzt werden (reproduktiver Aspekt der Wahrnehmung). Anschließend sollen beim Subtest 2 **Labyrinth** vier Labyrinthreihen richtig durchfahren werden (produktiver Aspekt der Wahrnehmung sowie visuelle Orientierung und Aufmerksamkeit). Beim Subtest 3 **Klassifikationen** ist eine Figur von 4 merkmalsähnlichen abzugrenzen (Klassifizieren, beziehungsstiftendes Denken bei figuralem Material), während bei Subtest 4 **Ähnlichkeiten** eine vorgegebene Zeichnung genau erfaßt und aus fünf Auswahlzeichnungen, von denen vier im Detail verändert sind, wieder herausgefunden werden muß. Schließlich müssen

beim Subtest 5 **Matrizen** gegebene Muster vervollständigt werden (Erkennen von Regeln und Zusammenhängen bei figuralen Problemstellungen).

Erfassungsdimensionen: Die Gesamtleistung (sog. Summe 1 bzw. $\Sigma 1$) erfährt die Grundintelligenz sensu Cattell, d.h. die Fähigkeit, "in neuartigen Situationen und anhand von sprachfreiem, figuralem Material, Denkprobleme zu erfassen, Beziehungen herauszustellen, Regeln zu erkennen, Merkmale zu erkennen und rasch wahrzunehmen" (Weiß & Osterland, 1980, S. 4). Die Summe der ersten beiden Subtests (Summe 2 bzw. $\Sigma 2$) mißt die Ausprägung des Wahrnehmungsumfangs und das Wahrnehmungstempo. Hier dürfte aufgrund des Speedcharakters der soziokulturelle Einfluß der Motivation zum tragen kommen. Bei der Summe der letzten drei Subtests (Summe 3 bzw. $\Sigma 3$) soll eine geringere kulturelle Beeinflussung vorliegen, so daß hiermit die grundlegende intellektuelle Leistungsfähigkeit erfährt werden kann.

Handhabung: Der Test kann als Einzel- oder Gruppentest durchgeführt werden. Die Kinder kreuzen die richtigen Lösungen direkt im Testheft an. Bei Vorschulkindern bzw. bei Schülern aus Schulen für Lernbehinderte soll die Gruppengröße 6 bis 8 Kinder nicht übersteigen. Bei Grundschulern soll die Gruppenstärke unter 15 (1. Klasse) bzw. 20 Kindern (2. und 3. Klasse) bleiben. Der Zeitbedarf liegt bei 50 bis 60 Minuten, wobei für Vorschulkinder und Lernbehinderte bei einzelnen Subtests mehr Bearbeitungszeit vorgesehen ist. Zum Test liegt eine Pseudoparallellform vor, die sich nur durch die Anordnung der Items unterscheidet, die aber das Abschreiben erschweren soll. Im Handbuch finden sich ein Interpretationsbeispiel und Hinweise für die Anwendung des CFT 1 bei speziellen Fragestellungen (Legasthenie, Förderdiagnostik, Sonderschulbedürftigkeit, Ausländerkinder).

Gütekriterien: Genaue Anweisungen zur Durchführung und Auswertung lassen die **Objektivität** als gesichert erscheinen. Die **Itemschwierigkeiten** liegen für die Untertests 3 bis 5 in etwa zwischen 0.2 und 0.8, für ältere Schüler scheint der Test allerdings etwas zu leicht zu sein; die **Trennschärfekoeffizienten** dieser Untertests sind hoch bis sehr hoch. Das Manual enthält darüber hinaus die Standardmeßfehler und Anleitungen zur Bestimmung des Vertrauensintervalls sowie kritische Differenzen für interindividuelle Vergleiche. Es liegen nur für die $\Sigma 3$ bzw. die daran beteiligten Subtests **Reliabilitätskoeffizienten** (innere Konsistenz) vor, da für die Speed-Tests (Substitutionen und Labyrinth) korrekterweise auf die Schätzung der Zuverlässigkeit mit Konsistenzmethoden verzichtet wurde. Für die $\Sigma 3$ ergeben sich in verschiedenen Stichproben Reliabilitätskoeffizienten nach Spearman-Brown von 0.90 bis 0.96, was als sehr gut gelten kann.

Im Hinblick auf die **Validität** wurden die 5 Untertests des CFT 1 faktorisiert, wobei bei der Zweifaktorenlösung die Subtests 1 und 2 einerseits sowie die Untertests 3 bis 5 andererseits die beiden Faktoren bilden. Allerdings wäre auch eine Einfaktorenlösung angemessen, so daß insgesamt die Bildung der drei Summenwerte damit gerechtfertigt wird. Bei gemeinsamen Faktorenanalysen mit den HAWIK-Untertests bilden die CFT-Untertests zusammen mit dem Mosaiktest den varianzstärksten Faktor. Die Interkorrelationen der Subtests sind niedrig (Regelschüler) bis mittel (Schüler von Lernbehindertenschulen).

Zur **Übereinstimmungsvalidität** mit anderen Intelligenztests und Schulleistungsindikatoren kann zunächst festgehalten werden, daß der CFT höher mit dem HAWIK-Handlungsteil und der Gesamtleistung (jeweils $r=0.66$) als mit dem Verbalteil ($r=0.48$) zusammenhängt. Mit dem ebenfalls sprachfreien BT liegen die Korrelationen allerdings niedriger ($r=0.50$ bis 0.55), was die Autoren mit einer hohen Beteiligung des BT auf einem Verbalfaktor erklären. Mit Einschulungstests werden niedrige (RTS, KST) bis mittlere/hohe Zusammenhänge (DVET, WTA), mit dem DRT 2 Korrelationen etwa zwischen 0.2 und 0.3 berichtet. Insgesamt korreliert der CFT mit konstrukt näheren Verfahren enger als mit konstruktferneren. Bezüglich Schulleistungen ergaben sich etwas höhere Zusammenhänge mit der Mathematik- ($r=0.44$) als mit der Deutsch- ($r=0.35$) und Sachkundenote ($r=0.39$).

Gruppenunterschiede: Die CFT-Leistungen von Legasthenikern liegen knapp, jedoch nicht signifikant unter dem Durchschnitt. Dasselbe gilt für Kinder ausländischer Arbeitnehmer im Vergleich zu deutschen Kindern. Bildet man Gruppen nach der sozialen Herkunft, so ergeben sich z.T. signifikante Unterschiede zugunsten höherer Schichten. Hingegen unterscheiden sich gleichaltrige Vorschul- und Grundschul Kinder nicht in ihren CFT-Leistungen.

Die **Normierung** erfolgte an insgesamt 6078 Kindern aus verschiedenen westdeutschen Bundesländern. Da sich keine Unterschiede zwischen Jungen und Mädchen sowie zwischen den Testformen A und B ergaben, wurde auf eine getrennte Normierung verzichtet. Die Normierung erfolgte für Altersgruppen im halbjährlichen Abstand sowie für die ersten drei Klassenstufen der Grundschule und die ersten vier Klassen der Lernbehindertenschule. Die meisten Stichproben, insbesondere die zur Ermittlung der Alters- und Klassennormen im Grundschulbereich, sind genügend groß (363 bis 1710 Schüler).

- b) Grundintelligenztest Skala 2 (CFT 20) von R.H. Weiß (1978, 1987a³) für Schüler des dritten bis zehnten Schuljahres; als Ergänzungstests hat Weiß (1987b) einen Wortschatztest (WS) und einen Zahlenfolgentest (ZF) herausgebracht.
- c) Grundintelligenztest Skala 3 (CFT 3) von R.H. Weiß (1971) für Schüler und Erwachsene ab 14 Jahren

Testaufbau: Die CFT-Skalen CFT 20 und CFT 3 bestehen jeweils aus zwei Testteilen (Teil 1 und 2), die zwar häufig als Parallelförmigkeiten benutzt werden, ursprünglich jedoch von Cattell nicht als solche, sondern als ergänzende - sukzessiv darzubietende - Skalenhälften vorgesehen waren. Jede Form besteht aus 4 Subtests mit einmal 14 (Classifications), zweimal 12 (Series, Matrices) und einmal 8 Aufgaben (Topologies). Die Lösung der Aufgaben erfolgt sprachfrei oder nonverbal, nicht jedoch die Instruktion der Aufgaben. Das Aufgabenmaterial vom CFT 20 und CFT 3 besteht ausschließlich aus figuralem Material, wobei folgende Aufgabenstellungen zu bearbeiten sind: Reihenfortsetzen (Untertest 1, Series), figurale

Klassifikationen (Untertest 2, Classifikations), Matrizenaufgaben (Untertest 3, Matrices) und topologische Schlußfolgerungen (Untertest 4, Topologies). Beim Untertest 4 muß das Kästchen ausgewählt werden, in dem der Punkt genauso zwischen die geometrischen Elemente gesetzt werden kann wie in der Vorlage. Zur Veranschaulichung seien einige **Übungsbeispiele** abgebildet (vgl. Abbildung 17).

Beim als Ergänzungstest vorgesehenen **Wortschatztest** muß zu 30 vorgegebenen Wörtern aus jeweils fünf Antwortalternativen dasjenige ausgewählt werden, das die gleiche Bedeutung wie das vorgegebene Schlüsselwort aufweist (Beispiel: Acker - a) Pferd b) Traktor c) Landwirt d) Feld e) Kartoffel). Beim **Zahlenfolgentest** müssen die Probanden zu einer Folge von 6 Zahlen die 7. aus einer Reihe von 5 Alternativen wählen.

Erfassungsdimensionen: Der CFT wird im allgemeinen als guter Indikator der **allgemeinen Intelligenz** (g-Faktor sensu Spearman et al.) angesehen. Er wird insbesondere dann eingesetzt, wenn sozio-kulturelle Einflüsse der Intelligenz-entwicklung isoliert werden sollen. Beispielsweise ist in einer Reihe von Bundesländern beim Sonderschulüberweisungsverfahren von Kindern ausländischer Arbeit-

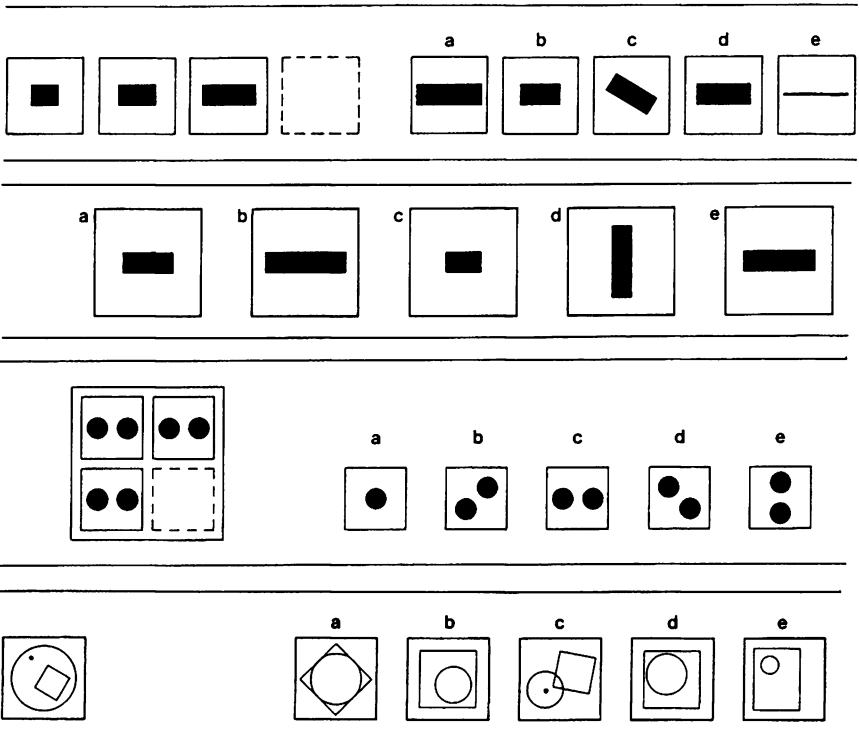


Abbildung 17: Beispielaufgaben zum CFT 20

nehmer ein sprachfreier Test wie der CFT vorgeschrieben. Trotzdem: Der CFT ist **kein** kultur-unabhängiger Test.

Aufgrund eigener faktorenanalytischer Untersuchungen zum CFT 3 kommt Weiß auf drei die CFT-Leistung wesentlich konstituierende Dimensionen, die er zusammenfassend - sensu Cattell - "Grundintelligenz" nennt: 1) Reasoning, 2) Beziehungsstiftendes Denken 3) Interferenzneigung/funktionelle Störbarkeit. Diese drei Faktoren umfassen zusammen rund 2/3 der Gesamtvarianz des CFT (Weiß, 1971).

In der Neuauflage des CFT 20 bezieht sich Weiß (1987a) stärker auf das Berliner Intelligenzmodell von Jäger. Die Aufgaben des CFT 20 werden nunmehr der figuralen Verarbeitungskapazität zugeordnet. Der Ergänzungstest Wortschatz wird als Indikator nicht nur für Wortschatz, sondern auch für Allgemeinbildung und die verbale Verarbeitungskapazität (nach Jäger) reklamiert. Der Zahlenfolgentest soll die numerische Verarbeitungskapazität nach Jäger (Erkennen von Regeln und Gesetzmäßigkeiten bei einfachen und komplexen numerischen Aufgabenstellungen) erfassen. Nach Ansicht des Autors dürfte auch eine gewisse Beziehung zu der Operation "Verarbeitungsgeschwindigkeit" im Jägerschen Intelligenzmodell bestehen.

Handhabung: Die Lösung der einzelnen Aufgaben wird auf einem separaten Antwortbogen, bei den Ergänzungstests zum CFT 20 direkt auf dem Aufgabenblatt eingetragen: der Proband streicht oder kreuzt die unter den angebotenen Antworten jeweils (einzig) richtige Lösung an (Mehrfachwahl-Antwortform). Für die Durchführung des CFT 20 müssen etwa 60 Minuten, für den CFT 3 etwas weniger veranschlagt werden (bei Durchführung nur eines Testteils jeweils etwa die Hälfte), für den Ergänzungstest Wortschatz zum CFT 20 müssen nochmals 15 Minuten, für den Zahlenfolgentest 20 bis 25 Minuten eingeplant werden. Auswertung und Interpretation erfolgen mittels Lösungsschablonen und Normtabellen im Manual. Der CFT 3 kann alternativ als Speed- und als Power-Test durchgeführt werden. Zwar wurde auf Fallbeispiele verzichtet, aber im Handbuch des CFT 20 finden sich wichtige Interpretationshinweise für die Schullaufbahnberatung, Legastheniediagnose und die Überprüfung der Sonderschulbedürftigkeit.

Testgütekriterien: Objektivität und Ökonomie der beiden CFT-Versionen können als gesichert gelten. Für den CFT 20 und auch für den CFT 3 werden Ergebnisse von Itemanalysen mitgeteilt: Sowohl Schwierigkeiten als auch Trennschärfen sind für den CFT 20 durchweg günstig bzw. hoch, wenngleich bei den älteren Schülern die Aufgaben als etwas zu leicht erscheinen. Im CFT 3 scheinen einige zu leichte Items enthalten zu sein, Informationen zu den Trennschärfeindizes fehlen hier. Die entsprechenden Kennwerte, insbesondere die Trennschärfen, fallen für die Ergänzungstests (Wortschatz, Zahlenfolgen) ungünstiger aus.

Im Manual werden für die amerikanische Vorform des CFT 20 **Reliabilitätskoeffizienten** zwischen 0.70 und 0.92 erwähnt, eigene Untersuchungen der Testautoren erbrachten Split-Half-Koeffizienten von über 0.90 und eine Retestreliabilität (Testwiederholung mit Pseudoparallelförmigkeit nach 2 Wochen) von $r=0.77$ bei 13- bis 15jährigen. Der Standardmeßfehler beträgt demnach für den Gesamttest 3.35 IQ-Punkte, die kritische Differenz zwischen erstem und zweitem Testteil 13 IQ-Punkte.

Für interindividuelle Vergleiche liegt die kritische Differenz für den Gesamttest bei 9 IQ-Punkten. Der Übungsgewinn bei zweimaliger Durchführung wird mit 5.4 Rohpunkten angegeben.

Für den CFT 3 werden im Manual lediglich Reliabilitätsschätzungen der amerikanischen Vorlage mitgeteilt (Split-Half-Methode), die zwischen 0.70 und 0.95 liegen. Der Standardmeßfehler für den Gesamttest wird mit 4.5 IQ-Punkten angegeben. Die kritische Differenz zwischen erster und zweiter Testhälfte beträgt dementsprechend 12.5 IQ-Punkte für die Absicherung interindividueller Unterschiede auf dem 5-Prozent-Niveau, für den intraindividuellen Vergleich (Gesamtleistung) 15 IQ-Punkte.

Für den Ergänzungstest Wortschatz werden Split-Half-Koeffizienten zwischen 0.8 und 0.9 mitgeteilt, die entsprechenden Werte für die Zahlenfolgen liegen etwas höher. Auch alle angegebenen Retest-Reliabilitätskoeffizienten liegen über 0.80, so daß die Zuverlässigkeit der Ergänzungstests insgesamt als noch gut bezeichnet werden kann. Im Manual finden sich auch genaue Anleitungen zur Berechnung von Vertrauensintervallen und kritischen Differenzen. Wortschatz und Zahlenfolgen korrelieren untereinander niedrig ($r=0.35$), mit dem CFT-Gesamtwert in mittlerer Höhe ($r=0.48$ bzw. 0.57).

Zur Validität von CFT 20 und CFT 3 wurde zunächst die Faktorenstruktur überprüft. Für die beiden CFT-Tests konnte dabei - auch unter Einbezug der Untersuchungen Cattells - gezeigt werden, daß die Untertests - wie postuliert - hohe Ladungen auf einem g-Faktor der Intelligenz (fluid intelligence) und entsprechend niedrige auf sprachgebundenen Faktoren aufweisen (verbale und numerische Intelligenzskalen, Rechtschreiben, Schulleistungen in Deutsch und Englisch etc.). Die Faktorenstruktur erwies sich als konstant über verschiedene Stichproben hinweg. Versucht man den Hauptfaktor "flüssige Intelligenz" weiter aufzugliedern, so kann man drei Faktoren unterscheiden: 1) Erlernen von Regelmäßigkeiten und Gesetzmäßigkeiten (Reasoning, Subtests 1 und 3), 2) Beziehungsstiftendes Denken, Erfassen von und Operieren mit Relationen (Subtest 4) sowie 3) Interferenz- oder Störbarkeit (Subtest 2).

Analog gilt für Untersuchungen der Übereinstimmungsvalidität an verschiedenen Stichproben (von Grundschulern über Haupt-, Real- und Berufsschüler sowie Gymnasiasten bis hin zu Bundeswehrsoldaten), daß die CFT-Tests mit konstrukt-näheren Merkmalen bzw. Variablen (z.B. CFT 20 mit CPM $r=0.58$) deutlich höher korrelieren als mit konstruktfernen (z.B. CFT 20 mit RST $r=0.09$).

Mit dem IST (Gesamtstandardwert) korreliert der CFT 3 zwischen 0.6 und 0.7 mit dem PSB (Gesamtleistung) mit etwa 0.40 bis 0.65 (Weiß, 1971, S. 38). Die Zusammenhänge mit konstruktferneren Tests zur Erfassung von Konzentration oder Persönlichkeitsmerkmalen sind eher niedrig. Unter den Schulleistungsfächern ist Mathematik am höchsten mit der CFT-3-Gesamtleistung korreliert, nämlich mit 0.53 (Gymnasialzensuren) und 0.58 (Realschulzensuren). Die Zusammenhänge mit anderen Noten sowie mit der Durchschnittsnote sind niedriger (etwa 0.2 bis 0.4).

Der CFT 20 korreliert mit dem PSB-Gesamtwert mit $r>0.60$, mit dem LPS mit $r=0.53$ und mit dem WIP mit $r=0.65$. Insgesamt liegen die Korrelationen mit

traditionellen mehr- oder eindimensionalen Begabungstests in mittlerer Höhe ($r=0.5$ bis 0.6), einige sind als hoch zu bezeichnen ($r \geq 0.6$). Dagegen sind die Zusammenhänge mit konstruktfernen Verfahren zur Erfassung von Aspekten wie Schulleistungen, Konzentration oder auch verbalen Intelligenzskalen deutlich niedriger (zwischen $r=0.1$ und 0.4). Wie für den CFT 3 gilt auch für den CFT 20, daß die Schulleistungen in Mathematik enger mit der CFT-Leistung zusammenhängen (um oder höher als $r=0.4$) als mit der Deutschnote (die meisten Koeffizienten kleiner als $r=0.4$). Die Zusammenhänge nehmen dabei mit ansteigender Klassenstufe ab.

Im Hinblick auf die **prognostische Validität** mit Schulleistungen führt Weiß (1987a) zunächst einige Ergebnisse an, die die Gültigkeit von Prognosen mit dem CFT 20 allerdings nur über relativ kleine Zeiträume (1 bis 8 Monate) belegen. In einem Schulversuch über mehr als 10 Jahre konnte nachgewiesen werden, daß sich später erfolgreiche und weniger erfolgreiche Schüler zu Beginn der Untersuchung im CFT 20 unterschieden hatten. Allerdings kommt Langfeldt-Nagel (1982c) mit demselben Datenmaterial zu dem Schluß, daß dem CFT nur eine geringe prognostische Gültigkeit für spätere "kristallisierte" Leistungen (operationalisiert nicht über Schulleistungen, sondern über PSB-Untertests) zukommt. Auch die Korrelation des CFT 20 (erhoben in der 4./5. Klasse) von $r=0.20$ mit der Mathematiknote in der 9. Klasse, die Heller, Rosemann & Steffens (1978) für dieselbe Stichprobe mitteilen, ist eher niedrig.

Lederle-Schenk & Marschner (1980) prognostizierten bei Viert- und Fünftklässlern die mit dem KS 4/5 erfaßten Schulleistungen über einen Zeitraum von einem Schuljahr und fanden mittlere bis hohe Korrelationen (0.73 in Grundschulklasse 4, 0.6 für Hauptschüler und 0.5 für die Realschüler und Gymnasiasten in Klasse 5).

Für den CFT 20 liegt eine Reihe von Studien vor, die **Unterschiede** in den CFT-Leistungen für verschiedene **Gruppen** untersuchen: Zwar treten auch beim CFT 20 Unterschiede zwischen Kindern aus unterschiedlichen Sozialschichten zutage, doch nimmt der Testautor in Anspruch, daß diese geringer seien als bei vergleichbaren Intelligenztests. Ähnliches gelte für Kinder aus städtischen und ländlichen Regionen (letztere schneiden schlechter ab). Dagegen scheinen nach Ansicht des Testautors weder Legastheniker, noch Schüler mit Konzentrationschwächen, sehschwache, linkshändige und Schüler mit sensumotorischen Störungen im CFT in dem Sinne benachteiligt zu werden, daß sie niedrigere Leistungen erzielen. Dagegen erzielen Kinder mit Rechenschwächen niedrigere Durchschnittsleistungen, was aber natürlich nicht für jeden Einzelfall gilt. 28 Kinder ausländischer Arbeitnehmer der Eichstichprobe schnitten zwar im Durchschnitt um 5 Rohpunkte schlechter ab, die Unterschiede waren aber nicht signifikant. Hier und in einer größeren Studie ($N=153$) wurde deutlich, daß solche Kinder im zweiten Testteil besser abschneiden, so daß der Testautor empfiehlt, bei Ausländerkindern den zweiten Testteil zur Beurteilung der Grundintelligenz heranzuziehen. Allerdings erreichen sie auch dann im Schnitt nur einen IQ von 93 bis 94, was nach Ansicht des Testautors mit den Werten deutscher Stichproben entsprechender Berufsgruppen übereinstimmt. Dennoch zieht Gözlü (1986) aus den Ergebnissen ihrer Studie mit türkischen Grundschulern den Schluß, daß auch der "kulturfaire" CFT türkische

Kinder benachteiligt, wobei die Autorin allerdings auf das u.E. eher fragwürdige Identitätskonzept der Testfairneß zurückgreift (vgl. Abschnitt 2.5 in diesem Buch).

Die Ergänzungstests Wortschatz (WS) bzw. Zahlenfolgen (ZF) korrelieren erwartungsgemäß mittel bis hoch mit konstruktnahen Kriterien ($r=0.5$ bis 0.7 , z.B. WS mit der Deutschnote durchschnittlich $r=0.50$, ZF mit der Mathematikzensur $r=0.60$) und niedrig mit konstruktferneren (in der Regel unter $r=0.4$). Mit Persönlichkeitsvariablen (PFK, AFS, FSL) ergaben sich praktisch keine Zusammenhänge. Im Hinblick auf Gruppenunterschiede schnitten Kinder niedriger sozialer Schichten und Ausländerkinder schlechter ab, letztere erzielten im Wortschatztest die schlechtesten Leistungen. Schulartunterschiede zeigten sich in der erwarteten Richtung. In einer Faktorenanalyse bildete der Wortschatztest zusammen mit dem verbalen und nonverbalen Teil der KFT-Kurzform und dem WST einen Faktor, auf einem zweiten Faktor laden neben den Zahlenfolgen der quantitative Teil der KFT-Kurzform und der d2.

An der Normierung des CFT 20 nahmen im Jahr 1977 4400 Schüler teil. Es stehen populationsrepräsentative Normen für 8- bis 18jährige für beide Testteile und den Gesamttest zur Verfügung. Für Erwachsene wurde nur Teil 1 normiert. Altersnormen (IQ- und T-Werte, Prozentränge) liegen im halbjährlichen Abstand (für die jüngsten Kinder) bis 10jährigem Abstand (für Erwachsene) vor. Schul- und Klassennormen (IQ-Werte, Prozentränge) finden sich für die Grundschule, die Sekundarstufe I und die ersten beiden Berufsschulklassen.

Zum CFT 3 liegen nur wenige Normtabellen vor: IQ- und T-Werte für die Durchführung als Speed-Test können für 16- bis 19jährige, für Berufsschüler, für Schüler mit mittlerer Reife sowie für Schüler der 12. Gymnasialklasse nachgeschlagen werden, für Studierende unterschiedlicher Fachrichtungen finden sich Mittelwerte und Standardabweichungen. Power-test-Normen liegen für 15- bis 16jährige sowie für Schüler der 12. Gymnasialklasse vor. Die Eichstichproben sind genügend groß; die Zusammensetzung gerade der Berufsschulstichproben sollte vor jeder Testinterpretation genau inspiziert werden, um Fehlinterpretationen zu vermeiden.

Für die Ergänzungstests Wortschatz und Zahlenfolgen finden sich Altersnormen (T-Werte, Prozentränge) für 8- bis 15jährige im halbjährlichen (Grundschule) bzw. jährlichen (Sekundarstufe) Abstand. Schulstandardwerte wurden für die dritte und vierte Klasse der Grundschule, sowie für Haupt- bzw. Realschulen und Gymnasien berechnet. Speziell für bayerische Verhältnisse, wo die Realschule erst ab der 7. Klasse einsetzt, wurden Normen für Volksschüler der Klassenstufen 5 und 6 bestimmt.

Anwendungsbereiche für die CFT-Reihen: Die CFT-Skalen empfehlen sich überall dort, wo man nonverbale Intelligenzfaktoren, vorab Reasoning und "Beziehungsstiftendes Denken", oder Begabungspotenzen (z.B. zur Aufhellung sog. Underachievement-Probleme) diagnostizieren will. Durch relativ geringfügige Modifikationen an der Instruktion dürfte die Anwendung des CFT auch bei Hör- und/oder Sprachgeschädigten möglich sein und als Indikator der "allgemeinen intellektuellen Leistungsfähigkeit" interessant werden. Im Rahmen der Schul- und

Studieneignungsermittlung spielt der CFT etwa bei der Erfassung vorhandener, sozio-kulturell jedoch mehr oder weniger vernachlässigter Begabungspotenzen eine Rolle. In manchen Bundesländern ist - wie bereits erwähnt - beim Sonderschulüberweisungsverfahren für Kinder mit nichtdeutscher Muttersprache ein sprachfreies Intelligenztest vorgeschrieben.

Kritik: Die CFT-Skalen sind keinesfalls - wie der ursprüngliche Name beansprucht - kulturunabhängige Intelligenzmeßverfahren. Kulturunabhängige Intelligenztests wird es im strengen Sinne auch nie geben. Aber auch die Kulturfairneß der CFT-Skalen wurde z.B. von Gözli (1986) angezweifelt. Andererseits zeigte sich in dieser Untersuchung, daß die Sprache der Testinstruktion im Gegensatz zu den Deutschkenntnissen der Kinder keinen Einfluß auf die CFT-Ergebnisse hatte.

In bezug auf sonderpädagogische Diagnostik verdient eine Untersuchung von Lauth & Wiedl (1985) Beachtung, nach der bei Sonderschülern ausschließlich durch Intensivierung der Testinstruktion eine Steigerung des IQ zwischen zwei Testungen erreicht werden konnte, die bei Regelschülern auch ohne diese Bedingung auftrat. Die Autoren interpretieren dies dahingehend, daß das Instruktionsverständnis auch beim "kulturfairen" CFT eine entscheidende Rolle spielt. Auch Mönikes & Mönikes (1975) fanden eine Abhängigkeit der CFT 2-Ergebnisse vom Instruktionsverständnis gerade jüngerer Schüler mit Lernbehinderungen, wengleich der CFT sich in wesentlich geringerem Ausmaß als kultur- und sprachabhängig erwies als der HAWIK.

Langfeldt-Nagel (1982a, b u. c) hat gezeigt, daß auch die CFT-Leistungen Umwelteinflüssen unterliegen, sogar im höheren Maß als eine Reihe von PSB-Untertests, die im Sinne der Theorie Cattells eher kristallisierte Intelligenz erfassen. Dieselbe Autorin meldet in den genannten Artikeln weiter Zweifel an der Validität des CFT bzw. der Cattellschen Intelligenztheorie an.

Trotz aller Vorbehalte können u.E. vor allem der CFT 1 und der CFT 20 als ökonomische und objektive Verfahren zur Erfassung der allgemeinen flüssigen Intelligenz im Sinne der Zweifaktorentheorie von Cattell angesehen werden. Besonders für den CFT 3 würde der Testanwender sich jedoch mehr Untersuchungsergebnisse zu Reliabilität und Validität wünschen, die an deutschen Stichproben gewonnen wurden.

Die Ergänzungstests Wortschatz und Zahlenfolgen können zwar zusätzliche Informationen liefern, die für die Schulberatung von großer Wichtigkeit sind; ob sie aber die Einbeziehung bewährter verbaler und quantitativer Tests überflüssig machen, scheint uns mehr als fraglich. Die Zuordnung beider Tests zur Operation Verarbeitungskapazität sensu Jäger muß zumindest für den Wortschatztest angezweifelt werden, da keine Leistungen wie Analogien oder Klassifikationen verlangt werden, die hohe Verarbeitungskapazität voraussetzen. Für die Schuleignungsprognose im Rahmen der Bildungsberatung kann den CFT-Skalen somit nur partielle Bedeutung zuerkannt werden.

(2) **Der Zahlen-Verbindungs-Test (ZVT)** von W.D. Oswald & E. Roth (1978, 1987²)

Testaufbau: Der ZVT wurde in Anlehnung an den "Trail Making Test" (Reitan, 1956) entwickelt. Der Test setzt sich aus vier parallelisierten Zahlenmatrizen zusammen, bei denen möglichst viele der Zahlen von 1 bis 90 innerhalb von 30 Sekunden verbunden werden müssen. Der erreichte Wert für jede Matrize wird mittels Tabelle in den Rohwert (Anzahl verarbeiteter Bits pro Sekunde) umgerechnet. Anschließend wird der ZVT-Gesamtwert als Durchschnitt über diese vier Scores berechnet. Der ZVT ist auch als Einzeltest durchführbar. In diesem Falle werden die Zeiten notiert, die für die Bearbeitung jeweils der gesamten Matrizen benötigt werden.

Erfassungsdimensionen: Nach Oswald & Roth (1987) dient der ZVT der sprachfreien Erfassung der basalen, allen Intelligenzleistungen zugrundeliegenden, weitgehend milieunabhängigen und genetisch bedingten kognitiven Leistungs- und **Informationsverarbeitungsgeschwindigkeit**. Diese basale Informationsverarbeitungsgeschwindigkeit wird durch die Bestimmung der Anzahl binärer Entscheidungen pro Sekunde operationalisiert.

Gütekriterien: Die Durchführungsobjektivität ist durch Vorlesen einer standardisierten Instruktion gesichert. Einschränkungen der Objektivität können sich nach unseren Erfahrungen ergeben, wenn im Gruppenversuch Schüler nach Ablauf der 30 Sekunden weiterarbeiten bzw. der Versuchsleiter die Zeitgrenzen nicht exakt einhält.

Die **Stabilität** des ZVT wird mit 0.81 (Retesting nach 6 Wochen) angegeben, die errechneten Maße für die **Konsistenzreliabilität** liegen in etwa zwischen 0.8 und 0.9, was insgesamt zufriedenstellend ist (Gruppenversuch). Die Werte für den Einzelversuch liegen noch deutlich darüber. Perleth & Sierwald (1991) bzw. Perleth (1991) ermittelten bei Gruppenadministration in eigenen Untersuchungen für Sekundarstufenschüler unterschiedlicher Jahrgänge Konsistenzschätzungen knapp unter 0.9 und Stabilitäten zwischen 0.69 und 0.85 für den Zeitraum eines Jahres sowie von 0.65 bis 0.73 für zweijährigen Abstand zwischen den Messungen.

Im Hinblick auf die **Übereinstimmungsvalidität** haben die Testautoren verschiedene Untersuchungen mit z.T. kleineren Stichproben durchgeführt, die im allgemeinen mittlere, bisweilen hohe Zusammenhänge (z.B. $r=0.83$ mit dem PSB für allerdings nur 24 Realschüler) ergaben. Mit Konzentrationstests korreliert der ZVT niedrig bis mittel. Die berichteten Zusammenhänge mit Schulleistungen sind eher niedrig (<0.4). Zusammenhänge mit physiologischen Meßwerten wurden zwar identifiziert, blieben aber gering. Bei Faktorenanalysen mit dem PSB zeigten sich Affinitäten des ZVT mit der PSB-Arbeitsprobe (Untertests 9 und 10). Auch andere Faktorenanalysen bestätigten nach Ansicht der Testautoren, daß der ZVT kognitive Leistungsgeschwindigkeit erfaßt. Weitere im Manual angeführten Untersuchungsergebnisse belegen eine stärkere genetische Verankerung der ZVT-Leistungen sowie die mögliche Verwendung des ZVT als Diagnostikum für Hirnleistungsstörungen.

Nach den Untersuchungsergebnissen von Perleth & Sierwald (1991) bzw. Perleth (1991) stellt der ZVT keinen guten Prädiktor für Schulleistungen dar; die meisten Korrelationen - bis auf jene mit der Mathematiknote - waren nicht signifikant. Gleiches galt für außerschulische Aktivitäten und Leistungen. Daneben fanden wir, daß die ZVT-Leistung mit anderen Indikatoren für kognitive Leistungsfähigkeit kovariert, die höchsten Korrelationen ergeben sich erwartungsgemäß mit dem sog. "Straßenplan", einem Speed-Test zur Erfassung schneller räumlicher Orientierung (Scanning). Die im Manual des ZVT mitgeteilten Korrelationen zu anderen Tests zur Erfassung intellektueller Leistungen konnten jedoch in der dort berichteten Höhe nicht bestätigt werden. Im Gegenteil zeigte sich, daß die ZVT-Leistung mit psychomotorischen Tests in etwa ebenso hoch korreliert ist (zwischen $r=0.21$ bis 0.38 mit Handsicherheit) wie mit der KFT-Gesamtleistung (um $r=0.30$). Somit scheinen psychomotorisch geschickte Schüler beim ZVT entgegen den Annahmen der Testautoren doch einen Vorteil zu besitzen. Diese Ergebnisse ließen sich auch faktorenanalytisch in etwa bestätigen.

Normen zum ZVT gibt es für 8- bis 11- und 12- bis 16jährige (Gruppenversuch), sowie für 8- bis 11-, 12- bis 16- und 16- bis 60jährige (Einzelversuch).

Kritik: Der ZVT zeichnet sich bei äußerst kurzer Durchführungszeit und geringem Materialverbrauch durch hohe Ökonomie aus und scheint hinreichend reliabel und zuverlässig Aspekte basaler kognitiver Verarbeitungsgeschwindigkeit zu erfassen. Allerdings sollten mögliche Anwender beachten, daß die Autoren des ZVT selbst den Untertitel des Tests von "Ein sprachfreier Intelligenz-Schnell-Test" (Oswald & Roth, 1978) bei der Neuauflage des Tests 1987 in "Ein sprachfreier Intelligenz-Test zur Messung der 'kognitiven Leistungsgeschwindigkeit'" umformuliert haben. Dies haben wohl auch die Untersuchungen von Gaußmann, Hochhausen & Schmidt-Rogge (1978) nahegelegt. Auch wir müssen angesichts unserer Erfahrungswerte davor warnen, den ZVT im schulischen Bereich als ökonomischen Ersatz für einen der bewährten Intelligenztests heranzuziehen. Eine solche Verwendung ist von den Testautoren auch nicht intendiert.

Tabelle 7: Übersicht über Tests zur Erfassung der Grund- oder Allgemeinen Intelligenz

Autor und Erscheinungsjahr	Testbezeichnung	Erfassungsdimensionen	Alters- und Zielgruppe	Zeitbedarf
A. Hardesty & H. Lauber (1956)	Hamburg-Wechsler-Intelligenztest für Erwachsene (HAWIE)	Allgemeine Intelligenz mit 11 Untertests	Erwachsene ab 20 Jahren	keine Angaben
U. Tewes (1983, 1985 ³)	Hamburg-Wechsler-Intelligenztest für Kinder (HAWIK-R)	Allgemeine Intelligenz mit 11 Untertests	6- bis 15jährige Kinder und Jugendliche	mindestens 90 Minuten, bis 3 Stunden in Einzelfällen

D. Eggert (1975, 1978 ²)	Hamburg-Wechsler-Intelligenztest für das Vorschulalter (HAWI-VA) - Experimentalform	Allgemeine Intelligenz mit 8 Subtests, getrennt für Verbal- und Handlungsteil (kein Gesamtwert)	Vorschulkinder von 4 bis 6;6 Jahren, auch ältere Lern- und Geistigbehinderte	in der Regel 30-40 Minuten, Aufteilung auf mehrere Sitzungen empfohlen
A. Schmidtke, S. Schaller & P. Becker (1980)	Coloured Progressive Matrices (CPM)	Klares Denken ("clear thinking")	Vor- (ab 4;9 Jahren) und Grundschüler (bis 11 Jahre)	25-40 Minuten, keine Zeitbegrenzung
H. Kratzmeier & R. Horn (1979, 1987 ²)	Standard Progressive Matrices (SPM)	Klares Denken ("clear thinking")	10-15jährige, englische Normen auch für Erwachsene	45 Minuten, in Einzelfällen länger
H. Kratzmeier & R. Horn (1980)	Advanced Progressive Matrices (APM)	Klares Denken ("clear thinking")	Klassen 9-11, gut begabte Jugendliche und Erwachsene	60 Minuten
K. Kubinger & E. Wurst (1983, 1988 ²)	Adaptives Intelligenz-Diagnostikum (AID)	Allgemeine Intelligenz nach Wechsler, 11 Untertests zur Differentialdiagnose	6- bis 15jährige Kinder und Jugendliche	mindestens 60 Minuten, in Einzelfällen länger
R. Weiß & J. Osterland (1977, 1980 ⁴)	Grundintelligenztest Skala 1 (CFT 1)	Flüssige Intelligenz sensu Cattell	Vorschule ab 5;6 Jahren, Grundschule bis Klasse 3	50-60 Minuten je nach Altersgruppe
R. Weiß (1978, 1987a ³ , 1987b)	Grundintelligenztest Skala 2 (CFT 20); Ergänzungstests: Wortschatz (WS) und Zahlenfolgen (ZF)	Flüssige Intelligenz sensu Cattell; Ergänzungstests: verbale und numerische Verarbeitungskapazität	8-18jährige, Grundschule Klasse 4, Sekundarstufe I, Berufsschule	CFT 20: 60 Minuten; Ergänzungstests: 35-40 Minuten
R. Weiß (1971)	Grundintelligenztest Skala 3 (CFT 3)	Flüssige Intelligenz sensu Cattell	16-19jährige Jugendliche, Gymnasiasten, Berufsschüler	50 Minuten
W.D. Oswald & E. Roth (1987 ²)	Zahlen-Verbindungs-Test (ZVT)	Kognitive Leistungsgeschwindigkeit	8-60jährige	5-10 Minuten

R. Kühn & R. Heck-Möhling (1976) bzw. K. Ingenkamp (Hrsg.) (1976)	Bildertest 1-2 (BT 1-2) und Bildertest 2-3 (BT 2-3)	Allgemeinbegabung	1. bis 3. Klasse (Gruppentest)	110 Minuten, verteilt auf 2 Tage (BT 1-2) bzw. 90 Minuten (BT 2-3)
---	---	-------------------	--------------------------------	--

3.3.2 Tests zur Erfassung differentieller kognitiver Fähigkeiten

(1) Das Leistungsprüfsystem (LPS) von W. Horn (1962, 1983²) (Hogrefe)

Das LPS ist in Anlehnung an die Primary Mental Abilities (PMA) Testreihe von Thurstone 1962 von W. Horn entwickelt worden. Eine 2., mit technischen Angaben und zusätzlichen Auswertungshinweisen erweiterte Auflage erschien 1983.

Testaufbau: Die Testbatterie besteht aus 14 Subtests (ohne Arbeitskurve), zu denen in Abbildung 18 jeweils die Übungsaufgaben abgebildet sind. Jeder der 14 Subtests im LPS enthält 40 Aufgaben, der gesamte Test umfaßt somit 560 Items (ohne Arbeitskurve). Beim LPS-Test 1 und 2 werden Hauptwörter im Singular angeboten, wobei jedes Wort (Item) einen Druckfehler enthält, der vom Probanden durchzustreichen ist. Die Tests 3 und 4 repräsentieren Klassifikations- und Analogietestitems bzw. Aufgaben zum Reihenfortsetzen; das nicht passende Glied einer jeden (Figuren- bzw. Zahlen- oder Buchstaben-)Reihe ist wieder zu markieren. LPS-Test 5 ist ein Anagrammtest; die richtige Reihenfolge der durcheinandergewürfelten Buchstaben ist zu erraten und der Anfangsbuchstabe des richtigen (herausgefundenen) Lösungswortes zu markieren. Im LPS-Subtest 6 sollen in bestimmten Zeiteinheiten möglichst viele Wörter zu vorgegebenen Anfangsbuchstaben gefunden werden. In Subtest 7 sind spiegelbildlich gedrehte Ziffern und Buchstaben aus plan verschobenen herauszufinden. Subtest 8 enthält Aufgaben zur Oberflächenabwicklung; der Schüler muß vorgegebene Oberflächennetze in der Vorstellung falten und die Ziffern der Netze in Buchstaben dreidimensional abgebildeten Körpern zuordnen. Bei Subtest 9 muß die Flächenzahl vorgegebener Körper angegeben werden.

In Subtest 10 sind bestimmte, in ablenkende Gesamtfiguren eingebettete Figuren-Items zu identifizieren. Subtest 11 ist als Gestaltergänzungstest aufgebaut; der Anfangsbuchstabe der betreffenden Gegenstandsbezeichnung ist wieder durchzustreichen. Ähnlich sind im Subtest 12 unvollständige Wortgestalten zu ergänzen bzw. die Rechtschreibfehler zu markieren. Die Subtests 13 und 14 verlangen die Durcharbeitung diverser Buchstaben- und Ziffernreihen, wobei bestimmte Regeln beachtet werden müssen. Subtest 15 repräsentiert eine Arbeitskurve oder Arbeitsprobe; analog zum Pauli-Test sind hier Additionen von 10 einstelligen Zahlen in bestimmten Zeiteinheiten auszuführen.

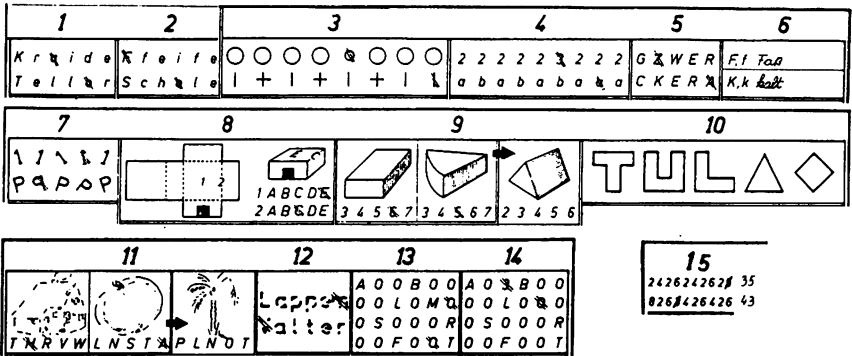


Abbildung 18: Beispielaufgaben der LPS-Untertests

Erfassungsdimensionen: W. Horn (1983, Testmanual, S. 20, 57f.) ordnet den einzelnen Subtests im LPS folgende Primärfähigkeiten sensu Thurstone zu:

- Subtests 1+2 **Verbal Factor (V):** Allgemeinbegabung oder "Bildungsstandard" (kristallisierte Intelligenz sensu Cattell), Rechtschreibkenntnisse; Wort- und Begriffsschatz (vermutlich ist auch der Faktor W beteiligt).
- Subtests 3+4 **Reasoning Factor:** allgemeine, relativ sprach- und bildungsunabhängige Denkfähigkeit (flüssige Intelligenz sensu Cattell); logisches und schlußfolgerndes Denken (sehr komplexer Faktor).
- Subtests 5+6 **Wordfluency (W):** Worteinfall und Wortflüssigkeit (mehr assoziative, relativ inhalts- bzw. sinnunabhängige sprachliche Reproduktionsleistung).
- Subtests 7-10 **Space 1 (LPS 7) und 2 (LPS 8, 9); Closure 2 (LPS 10).** Technische Begabung; Raumvorstellung, Visualisation sensu French; technisch-konstruktives Denken, mehr theoretisch-technische Findigkeit.
- Subtests 11+12 **Closure, Closure 1:** Ratefähigkeit; visuelle bzw. optische Gestaltauffassung; Flexibilität der Gestaltwahrnehmung sensu French; bei Subtest 12 ist der Faktor V beteiligt.
- Subtests 13+14 **Perceptual Speed: Accuracy;** Wahrnehmungstempo; Fehlerbemerken bzw. Richtigkeit des Reagierens.
- Subtests 1-14 **Gesamtleistung (GL):** Begabungsniveau.
- Arbeitskurve 1+2 **Number Factor:** Rechenfertigkeit (**nicht: Rechenfähigkeit!**); Hinweise auf Leistungsmotivation, Ausdauer, Konzentration, Belastbarkeit, Güte der Arbeitsleistung (Arbeitsqualität).
- Arbeitskurve 1-10 **Verlauf der Kurve:** länger fortlaufende Arbeitsqualität; Sorgfaltsstreben, Leistungsmotivation; Konzentration, Ausdauer, Belastbarkeit.

In der Anwendung des LPS bei 13jährigen Volksschülern konnte Tent (1969, S. 92 ff.) folgende **psychologische Dimensionen** aufgrund eigener Faktorenanalysen bestimmen:

- 1) **Sprachliche Leistungsfähigkeit** mit den höchsten - ihrem Gewicht zufolge bereits geordneten - Ladungen auf den LPS-Subtests 1, 2, 12, 5 und 6;
- 2) **Schlußfolgerndes Denken** mit den höchsten Ladungen auf den LPS-Subtests 3, 4, 13, 8 und 15;
- 3) **Räumliches Vorstellen** mit den höchsten Ladungen auf den LPS-Subtests 8, 11, 10 und 9;
- 4) **Leistungsmotiviertheit** mit den höchsten Faktorladungen in den LPS-Subtests 14 und 15;
- 5) **Wiedererkennen von Symbolen** mit den höchsten Ladungsgewichten auf den LPS-Subtests 7, 12 und 5;
- 6) **Schulisches Lernen** mit den höchsten Ladungsgewichten in den LPS-Subtests 15, 9 und 6.

"Zusammenfassend läßt sich das LPS für die untersuchte Altersstufe als ein psychologisch gut gegliederter und reichhaltiger Indikator des Leistungspotentials kennzeichnen. Der Umstand, daß die von Horn postulierte Faktorenstruktur des LPS innerhalb der Batterie nur teilweise reproduziert werden konnte, ist demgegenüber unerheblich" (Tent, 1969, S. 100f.). Auf die Funktion des LPS als **Schuleignungsdiagnostikum** werden wir weiter unten zu sprechen kommen. Nach Ansicht Tents wären für **Legastheniker** Ausfallserscheinungen in den LPS-Subtests 5, 7 und 12 zu erwarten.

Handhabung: Die Auswertung des LPS, d.h. die Auszählung der Rohscores (Plus- und Fehlerpunkte), erfolgt mittels Schablonen. Die Rohpunktsummen werden getrennt für jeden Subtest in ein Diagramm (Plus-Spalte) eingetragen; abschließend wird ein Gesamt-Summenscore gebildet und in die GL-Spalte (Gesamtleistung) übertragen, was zwar üblich, aber streng genommen mit den Konstruktionsprinzipien eines differentiellen Fähigkeitstests unvereinbar ist. Nach der Rohwert-Transformation in Centil-Werte ist eine **Profilzeichnung** möglich. Die gesamte LPS-Durchführung dauert inklusive der ersten beiden Spalten der Arbeitsprobe zwei Unterrichtsstunden (90 Minuten), bei Durchführung der gesamten Arbeitsprobe muß eine Schulstunde mehr eingeplant werden; der Test kann als Gruppen- oder Individualverfahren administriert werden. Für die formale LPS-Auswertung benötigt eine eingearbeitete Kraft 10 bis 15 Minuten pro Test-Protokoll (Form A 1 und A 2 versus Parallellform B 1 und B 2). Häufig werden von der Arbeitsprobe nur die ersten beiden - von insgesamt 10 - Aufgabenreihen durchgeführt. Im Handbuch finden sich keine Fallbeispiele, wohl aber Interpretationshinweise.

Testgütekriterien: Die Objektivität der LPS-Durchführung und -Testauswertung ist bei korrekter Einhaltung der Instruktions- und Auswertungsbestimmungen sehr gut. Itemkennwerte werden für das LPS auch in der Auflage von 1983 nicht mitgeteilt.

Die **Reliabilität** der einzelnen Subtests liegt zwischen 0.90 und 0.99 (Halbierungszuverlässigkeit), die des Gesamttests bei 0.99 (korrigierte Split-Half-Reliabili-

tät). In der LPS-Anwendung bei 10jährigen Probanden (4. Grundschulklasse) fand Tent (1969, S. 87 ff.) Halbierungskoeffizienten, die nur geringfügig unter den oben genannten, von Horn bei Erwachsenen gefundenen Werten liegen, nämlich Koeffizienten zwischen 0.88 und 0.97, bezogen auf LPS-GL sogar 0.98, die andeuten, "daß das LPS auch für die Kindergruppe eine beachtlich hohe innere Konsistenz seiner Meßfunktion behält". Allerdings ist kritisch anzumerken, daß bei Tests mit hoher Speed-Komponente wie sie die LPS-Untertests darstellen, die Anwendung des Split-Half-Koeffizienten zur Reliabilitätschätzung in der Regel zu einer Überschätzung der Zuverlässigkeit führt (vgl. Abschnitt 2.2).

Die Retest-Reliabilitätskoeffizienten, die Tent bei einem Sample von 13jährigen Schülern nach 32monatigem Testintervall ermittelte, lagen im Mittel der 15 Subtests immerhin noch bei 0.6, bezogen auf LPS-GL bei 0.83; ebenso ergab sich in den Untersuchungen Tents eine genügend hohe Paralleltestreliabilität. Die Subtests korrelieren durchschnittlich um 0.45 miteinander, was durch teilweise identische Faktorenladungen in dieser Höhe bedingt sein mag. Tent fand allerdings niedrigere Interkorrelationen der LPS-Subtests im 5. und 6. Schuljahr. Mit der Gesamtleistung korrelieren die LPS-Subtests im Mittel um 0.7. Der Meßfehler des LPS beträgt demnach etwa 0.3 C-Werte und bezieht sich auf die LPS-Anwendung im 4. Grundschuljahr sowie die LPS-Durchführung in der Sekundarstufe I (Hauptschule, Gymnasium). Horn selbst teilt in der Neuauflage des Tests etwas höhere Retest-Reliabilitätskoeffizienten mit, die an einer nicht näher spezifizierten Stichprobe von 91 erwachsenen Männern ermittelt wurden. Auffallend sind jedoch die hohen Standardabweichungen der einzelnen Untertests in dieser Stichprobe, was die Koeffizienten nach oben gedrückt haben mag.

Eine unerläßliche Voraussetzung für die Profilinterpretation eines differentiellen Intelligenztests ist seine **Profilreliabilität**, die gewöhnlich niedriger ausfällt als die Reliabilität des Gesamttests versus der Untertests. Die Profilreliabilität des LPS liegt nach unseren Berechnungen (vgl. Heller, 1973, S. 198) - je nach Untersuchungspopulation und/oder Reliabilitätskriterium (z.B. Halbierungs- oder Retest-Reliabilität) - zwischen 0.91 und 0.47; bei Grundschulern der 4. Klasse beträgt die Profilreliabilität des LPS 0.87, bei Sekundarschülern (Hauptschülern, Realschülern, Gymnasiasten) 0.91 bzw. 0.66 (bei nicht minderungskorrigierten Reliabilitätskoeffizienten bei der Berechnung der Profilreliabilität).

Zur **Validität** des LPS liegen verhältnismäßig wenige Untersuchungsergebnisse vor. Die von Horn (1962, S. 24) mitgeteilten Übereinstimmungs-Validitätskoeffizienten variieren zwischen 0.61 (in der Relation LPS 1-6 und Lehrerurteil über Intelligenz) und 0.17 (in der Relation LPS 7-10 und Lehrerschätzurteil); in bezug auf LPS-GL und Lehrerurteil wurden Validitätskoeffizienten um 0.5 ermittelt. Diese Ergebnisse beziehen sich auf Stichprobenerhebungen Horns im 4. Grundschuljahr (N=175) sowie Untersuchungssamples der Haupt- und Realschule (N=354). Je nach Untersuchungszweck und Probanden-Gruppe liegen also die Validitätskoeffizienten des LPS - wie bei allen vergleichbaren Intelligenztests - höher oder niedriger, was aber nicht unbedingt dem LPS-Diagnostikum angelastet werden muß, sondern auch an der Wahl des Außenkriteriums liegen kann. Bei-

spielsweise läßt sich der niedrige Zusammenhang vom LPS 7-10 und Lehrerurteil dahingehend interpretieren, daß in den betreffenden LPS-Subtests andere (Intelligenz-)Faktoren erfaßt werden, als sie für die Schulleistung relevant bzw. für die allgemeine Begabung charakteristisch erscheinen. Tent (1969, S. 110 ff.), der anhand seines Untersuchungsmaterials die LPS-Gesamtleistung zur Schulleistung in der 5. Klassenstufe des Gymnasiums, der Real- und Hauptschule in Beziehung setzte, errechnete entsprechende Validitätskoeffizienten von maximal etwa 0.5 und 0.6, wobei allerdings die LPS-GL mit den Einzelfachzensuren nur zwischen 0.31 (Naturkunde bzw. Geschichte) und 0.55 (Raumlehre) korreliert war. Der Zusammenhang der Gesamtleistung mit der Durchschnittsnote betrug $r=0.56$. Diese Werte wurden allerdings bereits vor 25 Jahren gewonnen, ohne daß vom Testautor neue Erhebungen inzwischen vorgenommen wurden.

Die **Prognosegültigkeit** hinsichtlich der Schulbewährung auf den weiterführenden Bildungseinrichtungen kann nach Tent (1969) als zufriedenstellend bezeichnet werden. "Faßt man die Einteilung in Volksschulen, Realschule und Gymnasium als Trichotomisierung einer kontinuierlichen, hinreichend normal verteilten Variablen auf, die man 'quantitativ abgestufte schulische Leistungsforderung' nennen könnte, läßt sich der Zusammenhang zwischen LPS-Gesamtpunktwert und dem Schulbesuch im 5. Schuljahr durch einen triserialen Korrelationskoeffizienten (Jaspens, 1946) von $r_{\text{tris}}=0.59$ wiedergeben" (a. a. O.). Ähnliche Erfahrungen mit dem LPS wurden von Aurin (1966) und Heller (1970, 1973) berichtet. Ob damit die Validität des LPS im Hinblick auf die Begabungsermittlung bzw. Schuleignungsprognose auch heute noch gesichert ist, muß bezweifelt werden, da die Grundstruktur des LPS praktisch nicht verändert wurde und keine neuen Untersuchungen vom Testautor vorgelegt wurden. Eher dagegen sprechen beispielsweise die Ergebnisse von Tent, Fingerhut & Langfeldt (1976) sowie Kühn (1984), in deren multiplen Regressionsanalysen zusammen mit anderen Verfahren der LPS-Gesamtwert nur wenig (zusätzliche) Varianz an den Schulleistungen aufklärte, wobei sich unter den einbezogenen Verfahren auch Schulleistungstests oder schulnahe Intelligenztests (AzN) befanden.

Für das LPS liegen Altersnormen in C-Werten ($M=5$; $s=2$) ab 9 Jahren (für GL) bzw. 10 Jahren (für die Subtests) vor, die zunächst in Jahresabständen, vom 21. bis zum 50. Lebensjahr in 10-Jahresschritten tabelliert sind. Daneben finden sich Mittelwerte und Streuungen für verschiedene Berufsgruppen, allerdings ohne Angabe, wann diese gewonnen wurden.

Anwendungsbereiche: Das LPS stellt ein objektives, zuverlässiges und hinreichend gültiges Intelligenzleistungsdiagnostikum dar, das sowohl die allgemeine Begabung als auch differentielle (teilweise relativ bildungs- und sprachunabhängige) Fähigkeiten erfaßt. Es empfiehlt sich von daher grundsätzlich für differentielle Intelligenzdiagnosen. Für Schuleignungsprognosen liegen inzwischen jedoch tauglichere Verfahren vor.

Kritik: Die Vorteile des LPS liegen einmal in seiner Ökonomie (der Testdurchführung und -auswertung) sowie in den relativ guten Testreliabilitäten, zum anderen repräsentiert das LPS ein brauchbares Konzept zur Testung wichtiger Primärfähigkeiten nach Thurstone. Daneben eignet sich der Test zur **Profilanalyse**. Freilich

liegen hierin auch noch gewisse Schwächen, die sich etwa in den vergleichsweise (im Hinblick auf die Konstruktionsprinzipien eines Profilverfahrens immer noch zu) hohen Testinterkorrelationen sowie der beachtlichen Korrelationen der Subtests mit der Gesamtleistung ausdrücken. Insbesondere ist zu kritisieren, daß vom Testautor auch in der Neuauflage von 1983 keine Schulstandards mitgeteilt werden, obwohl entsprechende Eignungskennbereiche von Heller (1973) publiziert wurden. Diese werden deshalb - unter dem Vorbehalt möglicher Überalterung - auszugsweise in Abschnitt 4.1.5 unten wiedergegeben.

Inwieweit das LPS noch heutigen Ansprüchen der Schuleignungsprognose gerecht wird, kann angesichts der schlechten "Testpflege" nicht beurteilt werden. So nimmt der Testautor in der zweiten Auflage von 1983 neuere Forschungsergebnisse zum LPS nicht in das Handbuch auf und meint zum Problem der Überalterung der Normen, daß man "die ursprünglichen Eichgruppen als feste Bezugspunkte ansehen" solle (S. 46). Damit wird der Diagnostiker in den 90er Jahren ebenso wenig zufrieden sein wie mit dem Fehlen von Schuleignungskennwerten oder gar Schultypnormen. Ebenso sind die Interpretationshinweise und viele der Berufsempfehlungen, die Horn möglichen LPS-Ergebnissen zuordnet, dadurch völlig entwertet, daß der Autor sich auf eigene Untersuchungen beruft, von denen nicht gesagt wird, wann, wo und mit welchen Stichproben sie gewonnen wurden. Überhaupt entspricht das Testhandbuch der Neuauflage so wenig den heute üblichen Standards, daß man sich wundert, daß der Verlag diese Neuauflage mitgetragen hat.

In der Praxis der Schullaufbahnberatung hat sich das im Anschluß beschriebene PSB, das praktisch eine Kurzform des LPS darstellt, wesentlich stärker als das LPS durchgesetzt. Weitere Kritikpunkte werden deshalb im Anschluß an das PSB genannt.

(2) Das Prüfungssystem für Schul- und Bildungsberatung (PSB) von W. Horn (1969) (Hogrefe)

Das PSB (Horn, 1969) ist aus dem LPS entstanden und stellt praktisch eine Kurzform desselben dar. Aufgrund der Befunde umfangreicher Diskriminanzanalysen beim LPS im Rahmen der baden-württembergischen Bildungsberatung (vgl. Aurin, 1966, S. 147ff.) wurden hier diejenigen LPS-Subtests zum sog. PSB zusammengefaßt bzw. ausgewählt, denen im Hinblick auf die Begabungs- und Schuleignungsdifferenzierung die beste diakritische Funktion zukommt.

Testaufbau: Die PSB-Subtests 1+2 sowie 3 und 4 sind praktisch, d.h. von kleineren Veränderungen abgesehen, mit den LPS-Subtests 1+2 sowie 3 und 4 identisch. PSB-Subtest 5 entspricht LPS-Subtest 6. PSB-Subtest 6 ist wiederum mit dem LPS-Subtest 12, die PSB-Subtests 7 und 8 sind mit den LPS-Subtests 8 bzw. 10 identisch. Lediglich die Subtests 9 und 10 im PSB wurden gegenüber der Arbeitsprobe bzw. dem Subtest 14 im LPS stärker verändert. Die Übungsbeispiele zu den PSB-Subtests 5 bis 10 sind in Abbildung 19 wiedergegeben; bezüglich der Übungsbeispiele zu den PSB-Subtests 1 bis 4 vgl. obige Ausführungen zum LPS.

5		6		7		8			9		10	
F.f-Tab	Lappeta	[Diagram: A cross-shaped arrangement of five squares, with the central square containing the number '1' and the two squares to its right containing the numbers '2' and '3' respectively.]		[Diagram: A 3D cube with the letters A, B, C, D, E, F on its faces.]		T U L A \triangle \diamond			73468293	73488293	73488293	73488293
K.k-halt	Walter	[Diagram: A cross-shaped arrangement of five squares, with the central square containing the number '1' and the two squares to its right containing the numbers '2' and '3' respectively.]		1 ABCDE 2 ABDE					6372859	6437859	6437859	6437859

Abbildung 19: Beispielaufgaben des PSB (Auswahl)

Erfassungsdimensionen: Analog zu den entsprechenden LPS-Subtestdimensionen formuliert Horn (1969, S. 18) für den PSB folgende Erfassungsdimensionen oder Primärfähigkeiten sensu Thurstone (in Klammern die PSB-Subtests, auf denen die betreffenden Faktoren laden):

- Subtests 1+2, 5, 6: Verbalität
- Subtests 3, 4: Denkfähigkeit
- Subtests 7, 8: Technische Begabung
- Subtest 8: Closure
- Subtests 9, 10: Wahrnehmungstempo

Eisebitt (1986) konnte diese Faktorenstruktur in etwa bestätigen, allerdings fand er keinen eigenen Closure-Faktor; und Subtests 9 und 10 bildeten zusammen einen Faktor.

Testgütekriterien: Im großen und ganzen beansprucht der Testautor dieselben Gütekriterien wie beim LPS (siehe oben), obwohl einige Untertests doch verändert wurden. Im Hinblick auf die **Durchführungsobjektivität** stimmt bedenklich, daß der Testautor den Versuchsleitern je nach Probandengruppe Veränderungen der Instruktion anheimstellt.

Eigene Untersuchungen zur **Reliabilität** des PSB werden nicht mitgeteilt, stattdessen wird auf die Kennwerte entsprechender LPS-Subtests verwiesen, ohne diese jedoch im Manual aufzuführen. Die durchschnittliche **Halbierungszuverlässigkeit** aller Subtests im PSB, gegen deren Berechnung dieselben Vorbehalte wie oben beim LPS angemeldet werden müssen, liegt nach Aurin (1966) bei 0.92 (Viertkläßkinder) und nach Horn (1969) bei 0.96 (Erwachsene). Das arithmetische Mittel aller Subtestkorrelationen im PSB beträgt nach unseren früheren Berechnungen (Heller, 1973), bezogen auf die von Horn adaptierten LPS-Werte für das PSB 0.51 bzw. auf die von Aurin bei 907 Viertkläßkindern ermittelten PSB-Interkorrelationen 0.41, was die Homologie der Strukturen beider Horn-Tests bestätigt. Dementsprechend ermittelten wir eine **PSB-Profilreliabilität** für die Anwendung im 4. Schuljahr von 0.86 (nach den Hornschen Interkorrelationen) und 0.84 (nach den Aurinschen Interkorrelationen); vgl. Heller (1973, S. 139).

Kastner (1978) berichtet hingegen eine geringere mittlere Interkorrelation von 0.26. Langfeldt-Nagel (1982a) fand mittlere Test-Retest-Korrelationen für die PSB-Untertests von 0.29 bis 0.36, wobei zwischen erster und zweiter Testung (im Jahre 1975) allerdings acht Jahre lagen. In der Untersuchung von Haenisch & Klaghofer (1979) lagen die Retestkoeffizienten für ein Intervall von zwei Jahren bis auf PSB 1+2 (0.71) alle unter 0.6.

In neuerer Zeit hat Marschner (1982) für männliche Facharbeiter **Retest-Reliabilitätskoeffizienten** berechnet (2- und 3jährige Intervalle zwischen den Testungen). Demnach liegt die Wiederholungszuverlässigkeit der PSB-Dimensionen zwischen 0.61 (PSB 7+8) und 0.85 (PSB 1+2), für die Gesamtleistung wurde eine Reliabilität von 0.81 ermittelt. Auch eine spätere Untersuchung Marschners (1985) belegt die Zeitstabilität der PSB-Skalen.

Abel (1988) hat die Reliabilitätskoeffizienten des PSB für eine Stichprobe aus 9 Hauptschul-, Realschul- sowie Gymnasialklassen über einen Zeitraum von 4 Wochen berechnet. Für die Subtests ergab sich eine durchschnittliche Stabilität von 0.73, für die Gesamtleistung eine von 0.93 bei Vorgabe derselben Testform. Für die Schüler, denen bei der zweiten Messung die Parallelförmigkeit vorgelegt wurde, lauten die entsprechenden Koeffizienten 0.64 und 0.85. Auch dies spricht für eine zufriedenstellende Zeitstabilität des PSB. Dagegen scheinen die Untertests 3 und 7 der beiden Testformen nach den Ergebnissen von Abel (1988) nicht ausreichend parallel zu sein.

In bezug auf die **Kriteriumsvalidität** wird im Handbuch auf entsprechende Korrelationen der z.T nicht identischen LPS-Untertests verwiesen, wengleich die Untertests des PSB aufgrund der Untersuchung von Aurin (1966; z.B. für den Zusammenhang der Untertestkombination 3+4 mit der Mathematiknote $r=0.6$, minderungskorrigiert $r=0.8$) zusammengestellt wurden. Der Testautor faßte dabei diejenigen Testteile des LPS zusammen, die für schulische Zwecke besonders geeignet schienen bzw. die sich in Diskriminanzanalysen als besonders trennscharf erwiesen. Noch unbrauchbarer sind Hinweise auf Untersuchungen aus den 50er Jahren mit Testverfahren, die nach Ansicht des Autors einzelnen PSB-Subtests ähnlich seien. Auch die "Vermutungen" des Autors über die mögliche Höhe von Validitätskoeffizienten in bestimmten Klassenstufen tragen wenig zur Erhellung der Validität des Verfahrens bei.

In jüngerer Zeit errechneten Haenisch & Klaghofer (1979) niedrige bis mittlere **Übereinstimmungsvaliditäten** der PSB-Untertests mit Schulnoten. Dabei zeigte sich, daß die Zusammenhänge der Mathematiknote mit PSB 3 und PSB 4 mit 0.33 und 0.36 (4. Schuljahr) bzw. 0.42 und 0.47 (6. Klasse) entgegen der Vermutung Horns (1969) nicht viel niedriger ausfallen als mit dem Wortschatztest PSB 1+2 (0.33 und 0.39). Eisebitt (1986) fand für eine Beratungsstichprobe einen Zusammenhang von $r=0.36$ mit der Deutsch- und von $r=0.24$ mit der Mathematiknote, diese Zusammenhänge sind gleichzeitig höher als alle Einzelkorrelationen der Untertests. So korrelierten hier PSB 3 und 4 mit der Mathematiknote nur zu $r=0.13$ und $r=0.18$, d.h. niedriger als PSB 5 oder 9 (jeweils $r=0.20$) bzw. PSB 7 ($r=0.22$). In Diskriminanzanalysen zur Vorhersage von Schulleistungsgruppen erwiesen sich PSB 1+2 und PSB 7 als trennstärkste Variablen, wobei mit der Gesamtleistung keine besonders gute Gruppentrennung gelang. Kastner (1978) teilt für die PSB-Gesamtleistung einen Zusammenhang mit Mathematik von 0.62 und mit Deutsch von 0.53 mit. Allerdings hält Kastner nicht alle Subtests des PSB für schulische Zwecke für relevant, und Abel (1988) hält die Zuordnung des PSB-Untertests 7 zum Primärfaktor Raumvorstellung für falsch. Heller, Rosemann &

Steffens (1978) berichten für die PSB-Gesamtleistung (erhoben in der 4. Klasse) einen Zusammenhang von $r=0.14$ mit der Deutschnote und $r=0.21$ mit der Mathematiknote in der 9. Klasse, also fünf Jahre später. Bei der Bewertung dieser Ergebnisse ist allerdings nicht nur die große Zeitspanne zwischen der Erhebung der Prädiktoren und der Erhebung der Kriteriumsvariablen zu berücksichtigen, sondern auch die Notenmaßstabdifferenz der Sekundarschultypen (Hauptschule, Realschule, Gymnasium).

Zum PSB liegen wiederum **Altersnormen** für das 9. bis 20. Lebensjahr vor, für die Altersgruppen 9 bis 11 jeweils in Vierteljahresintervalle unterteilt. Die Centil-Normierung des LPS wurde auch hier beibehalten, d.h. die Normen sind in C-Werten angegeben. Abel (1988) hält eine Neunormierung des PSB für geboten.

Anwendungsbereiche: Das PSB empfiehlt sich besonders in der Grundschule (Klasse 4) sowie auf der Sekundarstufe I überall dort, wo intelligenzdiagnostische und begabungs- oder schuleignungsprognostische Fragen vorliegen. Seine Durchführung beansprucht die Hälfte der LPS-Bearbeitungszeit (1 Stunde). Die PSB-Zuverlässigkeit liegt zumindest auf der Grund- und Hauptschule nur geringfügig unter der LPS-Reliabilität. Bei Gymnasiasten und Realschülern der Oberstufe wurde im Rahmen von Begabungsuntersuchungen mitunter eine gegenüber dem LPS verminderte Reliabilität festgestellt.

Kritik: Die bei der Erörterung des LPS vorgetragenen Kautelen versus Verbesserungsvorschläge gelten sinngemäß auch hier. Darüber hinaus muß im Hinblick auf einen Einsatz der Horn-Tests in der Schulberatung festgehalten werden, daß sämtliche Untertests gerade bei jüngeren Kindern eine starke Speed-Komponente aufweisen, wodurch reflexive oder ängstliche Testpersonen möglicherweise benachteiligt werden.

Wir haben oben darauf hingewiesen, daß die von Horn verwendete Split-Half-Reliabilitätsschätzung bei Geschwindigkeitstests in der Regel eine Überschätzung der Zuverlässigkeit bedingt. Insofern überrascht den testtheoriekundigen Anwender auch nicht, daß die Untertests mit der stärksten Speed-Komponente (Subtests 9 und 10) die relativ höchste, die mit größerer Power-Komponente (Subtests 3 und 4) die relativ niedrigsten Konsistenzkoeffizienten aufweisen. In ähnlicher Weise können viele der im Manual mitgeteilten Ergebnisse unter statistisch-methodischem Blickwinkel nicht befriedigen.

Die Normen des PSB, insbesondere zu den Subtests 9 und 10, müssen als veraltet gelten, wie auch die meisten Untersuchungen zum PSB aus den 60er Jahren datieren. Gerade die Beliebtheit des PSB bei Übertrittsentscheidungen stimmt von daher bedenklich. Vorbehalte müssen auch gegenüber dem Fehlen quantitativer Fähigkeitsdimensionen erhoben werden. Dieses Manko kann auch nicht durch den Reasoning-Faktor kompensiert werden, ebensowenig durch Subtest 9 oder 10. Angesichts der weiten Verbreitung und Beliebtheit des Verfahrens bei Praktikern verwundert es, daß das Testmanual, das bei weitem nicht den Ansprüchen eines modernen, sorgfältig konstruierten Tests entspricht, seit 1969 unverändert geblieben ist. Der Käufer des Tests findet nicht einmal eine aktualisierte Literaturliste in den

Testunterlagen vor. Eine Überarbeitung oder Revision des Verfahrens scheint gerade wegen seiner Verbreitung dringend erforderlich.

(3) Die Kognitive Fähigkeitstestbatterie (KFT) von Heller et al. (Beltz)

Die KFT-Batterie ist eine deutschsprachige Adaptation der in den Vereinigten Staaten weitverbreiteten Cognitive Abilities Tests (CAT) von Thorndike & Hagen (1971). Der KFT umfaßt drei Testversionen, wobei die ersten beiden in der Aufgabenstruktur und im Aufbau untereinander vergleichbar sind, sich gegenüber der dritten "großen" Version jedoch deutlich unterscheiden.

- a) Kognitiver Fähigkeitstest für das Kindergartenalter (KFT-K) von K.A. Heller & H.-J. Geisler (1983b) (Beltz)
- b) Kognitiver Fähigkeitstest für die Grundschule (KFT 1-3) von K.A. Heller & H.-J. Geisler (1983a) (Beltz)

Testaufbau: Der KFT-K besteht wie der KFT 1-3 aus vier Subtests mit je 15 Items. Beim Untertest **Sprachverständnis** müssen die Kinder aus jeweils fünf Abbildungen des Testheftes Handlungen, Gegenstände oder deren Verwendung herausuchen und ankreuzen. Die Aufgaben des Untertests **Beziehungserkennen** erfassen die Fähigkeit, zeitliche, räumliche, größen- oder mengenmäßige Relationen auf bildlichen Vorlagen zu erkennen. Beim **Schlußfolgerndem Denken** sollen die Kinder aus jeweils fünf Bildern das herausuchen, das nicht zu den anderen paßt. Schließlich prüft der Untertest **Rechnerisches Denken**, inwieweit die Kinder mit Zahlbegriffen und mengenmäßigen Vorstellungen umgehen können. Beispiele für die Aufgabenstellungen können Abbildung 20 entnommen werden.

Erfassungsdimensionen: Während mit dem KFT-K eindimensional die **allgemeine Intelligenz** bzw. das **kognitive Fähigkeitsniveau** von Kindergartenkindern und Vorschülern erfaßt wird, geben die Untertests 1, 2 und 4 des KFT 1-3 nach Ansicht der Autoren über die **schulische Lernfähigkeit** Auskunft, während Untertest 3 eher die **allgemeine Denkfähigkeit** mißt. Die Gesamtleistung im KFT 1-3 wird als Indikator für das **kognitive Fähigkeitsniveau** angesehen.

Handhabung: Beide Tests sind reine Papier- und Bleistifttests und prinzipiell als Gruppentests durchführbar, wobei die mögliche Gruppengröße vom Alter und Leistungsstand der Kinder abhängt. Jüngere Kindergartenkinder sollten jedoch immer im Einzelversuch getestet werden, wobei die Autoren empfehlen, die vier Testteile an aufeinanderfolgenden Tagen vorzugeben. Bei Grundschulern sollte die Dauer der Pausen (5 bis 15 Minuten) dem Alter und der Leistungsfähigkeit der Kinder angepaßt werden. Beide Tests sind reine Power-Tests, der Testleiter liest die Aufgabenstellung laut vor, und die Kinder kreuzen jeweils im Testheft die richtige Lösung an. Der Zeitbedarf liegt bei Grundschulern, je nach Leistungsstärke und Gruppengröße, zwischen 45 und 60 Minuten oder auch darüber, jedenfalls

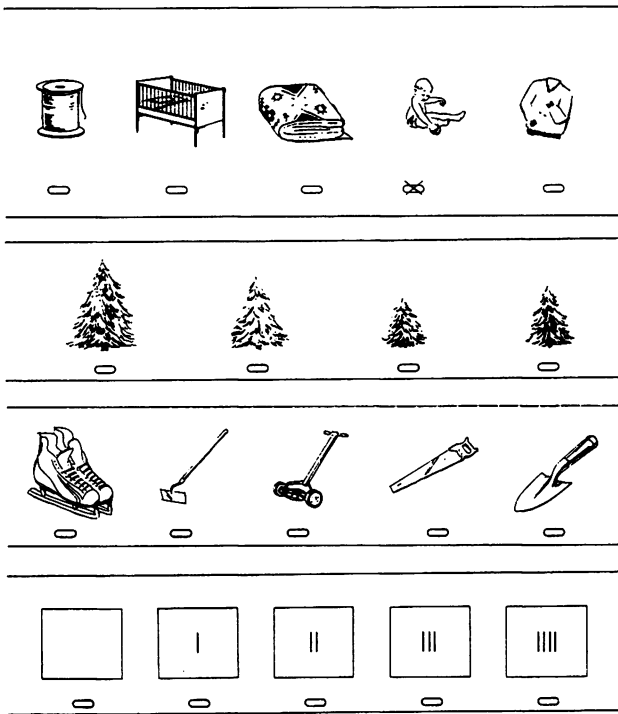


Abbildung 20: Beispiele für Testaufgaben des KFT-K und KFT 1-3

sollen die Kinder nicht unter Zeitdruck gesetzt werden. Die Auswertung erfolgt mit Hilfe eines Lösungsschlüssels, pro Untertest wird die Summe der richtig gelösten Aufgaben und die Gesamtleistung auf dem Antwortbogen notiert. Den Normtabellen können klassen- und altersspezifische Normen (T-Werte) entnommen werden. In den Manualen finden sich Tabellen mit Meßfehlern und Anleitungen zur Erstellung von Vertrauensintervallen, beim KFT 1-3 auch kritische Differenzen.

Gütekriterien: Durchführungs-, Auswertungs- und Interpretationsobjektivität dürften aufgrund der detaillierten Instruktion und Auswertungsanweisungen gesichert sein. Die Interpretationsobjektivität wird durch Angaben zu Vertrauensintervallen und kritischen Differenzen noch erhöht.

Konsistenzschätzungen zur Reliabilität des KFT-K bzw. KFT 1-3 ergaben für die Untertests Koeffizienten zwischen 0.53 bzw. 0.59 und 0.88 bzw. 0.77, für die Gesamtleistung liegen die Werte um 0.90 bzw. 0.84. Die Wiederholungsreliabilität des KFT-K (Testwiederholung nach 5 Wochen) liegt bei 5- bis 6jährigen Kindern

um 0.8 für die Untertests und bei 0.93 für die Gesamtleistung, was für sehr hohe Stabilität spricht. Bei Grundschulern wurden Stabilitätskoeffizienten zwischen 0.58 und 0.80 (Gesamtleistung) ermittelt (Retestung ebenfalls nach 5 Wochen). Bei drei- bis vierjährigen Kindergartenkindern liegen die Retestrelabilitätskoeffizienten deutlich niedriger (0.53 für die Gesamtleistung), auch die Konsistenzkoeffizienten können für diese Altersgruppe nicht befriedigen. Zwei negative Konsistenzreliabilitätsschätzungen weisen darauf hin, daß die Voraussetzungen zur Anwendung der Formel 20 nach Kuder-Richardson nicht gegeben war. Die Autoren raten deshalb vom Einsatz des KFT-K bei 3- bis 4jährigen Kindern ab.

Die innere und äußere **Übereinstimmungsvalidität** wurde für den KFT-K durch Korrelation und gemeinsame Faktorenanalysen mit dem CFT 1 und dem Duisburger Vor- und Einschulungstest DVET untersucht. Für die Validierung des KFT 1-3 wurden ebenfalls der CFT 1, der Bildertest BT 2-3 sowie der AST-2 und Schulzensuren herangezogen. Beide Tests (Gesamtleistung) korrelieren hoch mit den CFT-Summenwerten 1 und 3, was für eine gute Erfassung der sog. Grundintelligenz spricht (KFT-K: $r=0.64$ mit Summe 1, $r=0.66$ mit Summe 3; KFT 1-3: $r=0.59$ mit Summe 1, $r=0.61$ mit Summe 3). Dagegen fällt die Korrelation beider KFT-Versionen mit der Wahrnehmungsgeschwindigkeit (CFT-Summe 2) deutlich niedriger aus (KFT-K: $r=0.42$; KFT 1-3: $r=0.29$). Für den Zusammenhang des KFT 1-3 mit dem BT 2-3 wurde ein $r=0.67$ ermittelt. Auf Untertestebene fällt beim KFT 1-3 auf, daß sich für Wortschatz die niedrigsten Korrelationen ergeben (alle $r < 0.25$). Eine kanonische Korrelation zwischen dem BT, den CFT-Summen 2 und 3 auf der einen sowie den Untertests des KFT 1-3 auf der anderen Seite ergab ein $r=0.73$.

Im Bezug auf Übereinstimmungsvalidität mit **Schulleistungen** korreliert die KFT-K-Gesamtleistung mit dem eher schulleistungsnahen DVET mit $r=0.62$; die Untertests liegen darunter, am niedrigsten korreliert der DVET mit dem Schlußfolgernden Denken ($r=0.42$). Beim KFT 1-3 zeigen die Untertests Beziehungserkennen und Rechnerisches Denken die engsten Zusammenhänge (in etwa $0.35 < r < 0.55$) mit den Schulleistungen (Lehrerurteil, Zensuren, AST 2-3). Eine fächerbezogene Prognose mit dem KFT 1-3 scheint allerdings nicht möglich zu sein. Für Schüler der dritten Klassen ergab eine kanonische Analyse zwischen Zensuren und den KFT-Untertests eine kanonische Korrelation von $r=0.62$, was einer gemeinsamen Varianz von 38.6% entspricht. Zur Ermittlung der prognostischen Validität des KFT 1-3 wurden Korrelationen zwischen Testergebnissen in der dritten mit den Leistungen der vierten Klasse berechnet. Die Zusammenhänge unterschieden sich dabei nur wenig von denen zur Übereinstimmungsvalidität.

Ergebnisse von **Faktorenanalysen** belegen für den KFT-K Eindimensionalität, für den KFT 1-3 laden bei Faktorisierung aller bei der Validierung verwendeten Variablen die Untertests Sprachverständnis und Beziehungserkennen auf einem Schulleistungsfaktor, die Subtests Schlußfolgerndes Denken und Rechnerisches Denken auf einem Intelligenzfaktor.

Im Rahmen der **Konstruktvalidierung** konnte für den KFT-K nachgewiesen werden, daß Kinder mit höherem Alter, längerem Kindergartenbesuch (Testtermin im

Herbst/Winter oder Frühjahr/Sommer) und steigender Sozialschicht höhere Werte in den Untertests und damit auch in der Gesamtleistung erzielen. Analoge Unterschiede fanden sich beim KFT 1-3, zusätzlich wurden Geschlechtsunterschiede deutlich. Die Unterschiede im Hinblick auf die Sozialschicht sind für das Schlußfolgernde Denken bei beiden Verfahren am niedrigsten (ausführlicher vgl. Geisler, 1981).

Den Normen zum KFT-K liegen die Testergebnisse von 509 Kindern überwiegend aus dem Einzugsbereich größerer Städte in Nordrhein-Westfalen im Alter von 5;1 bis 7;0 zugrunde, so daß die Repräsentativität der Eichstichprobe eingeschränkt ist. Zudem sind die Stichproben zum Teil sehr klein. Für Altersgruppen im halbjährlichen Abstand (jeweils für Testtermine im Herbst/Winter und im Frühjahr/Sommer) wurden flächentransformierte T-Werte berechnet. Für die Normierung des KFT 1-3 bekamen Ende 1978/Anfang 1979 insgesamt 4592 Kinder aus der gesamten Bundesrepublik (einschließlich Westberlin) den Test vorgelegt, so daß die Eichstichprobe des KFT 1-3 als sehr groß gelten kann. Für Jungen und Mädchen getrennt liegen Klassen- und Altersnormen in halbjährlichem Abstand vor (flächentransformierte T-Werte).

Kritik: Otto (1984) regt an, daß in einer späteren Neuauflage des KFT-K die Kennwerte und Subtestinterkorrelationen der zur Validitätskontrolle herangezogenen Tests in das Manual mit aufgenommen werden sollen. Diese sind in zwei (allerdings nicht veröffentlichten) Examensarbeiten niedergelegt, die im KFT-Beiheft (vgl. S. 8ff.) genannt werden (Lambert, 1982; Puls, 1982). Hany & Rennen-Allhoff (1987) bemerken, daß angesichts festgestellter Abhängigkeit der KFT-K-Ergebnisse von Schichtvariablen schichtspezifische Normen hätten berechnet werden sollen.

- c) **Kognitiver Fähigkeitstest für 4. bis 13. Klassen sowie Berufsschüler und Studierende (KFT 4-13+)** von K. Heller, A.-K. Gaedike & H. Weinläder (1976, 1985²) (Beltz)

Testaufbau: In Anlehnung an die amerikanische Vorlage gliedert sich der KFT 4-13+ in drei große Testteile.

Der verbale Teil (V-Teil) enthält folgende vier Subtests:

- V1 Beim Untertest **Wortschatz** ist jeweils zu einem vorgegebenen Wort aus einer Reihe von fünf Auswahlwörtern ein Synonym oder Oberbegriff herauszusuchen. Beispiel: "Rose: a) Musik b) Blume c) Speise d) Aussicht e) Lasso".
- V2 **Satzergänzungen:** Hier müssen die Probanden dasjenige von fünf Auswahlwörtern heraussuchen, das einen Satz am sinnvollsten vervollständigt. Beispiel: "Das Feuer ist ____: a) feucht b) grün c) heiß d) flüssig e) kräftig".
- V3 Beim Untertest **Wortklassifikationen** müssen die Schüler aus fünf Auswahlwörtern dasjenige auswählen, das in die gleiche Wortklasse bzw. unter den gleichen Oberbegriff fällt wie drei vorgegebene Wörter. Beispiel: "Maus Wolf Bär: a) Rose b) Löwe c) laufen d) hungrig e) braun".

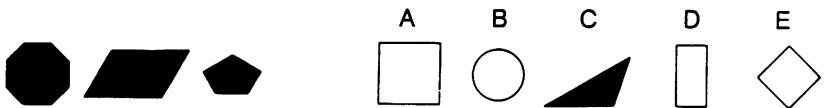
V4 Bei den **Wortanalogien** sind jeweils zwei Wörter gegeben, die in einer bestimmten Relation zueinander stehen. Zu einem dritten Wort muß nun aus fünf Auswahlwörtern dasjenige herausgesucht werden, das mit diesem im gleichen Verhältnis steht wie die Wörter des vorgegebenen Paares. Beispiel: "groß zu riesig wie klein zu: a) Junge b) erwachsen c) winzig d) wenig e) mehr".

Im **quantitativen Teil (Q-Teil)** finden sich folgende Untertests:

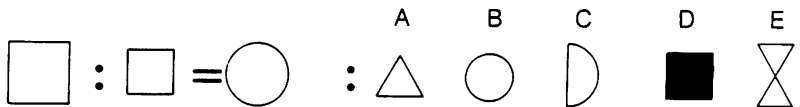
- Q1 **Textrechenaufgaben:** Neben vier Auswahlantworten ist als fünfte Auswahlmöglichkeit jeweils die Alternative "keines von diesen" (k.v.d.) vorgegeben. Beispiel: "Frau Müller kauft 1 Pfund Kartoffeln für 20 Pf und 1 Pfund Spinat für 35 Pf. Wieviel hat sie ausgegeben? a) 20 Pf b) 45 Pf c) 50 Pf d) 60 Pf e) k.v.d".
- Q2 Beim Subtest **Mengenvergleiche** müssen die Schüler für jeweils zwei Mengen (Terme, Flächen, Winkel, Strecken etc.) entscheiden, welches die größere ist oder ob beide gleich groß sind.
- Q3 Um die Aufgaben des Untertests **Zahlenreihen** zu lösen, muß jeweils zu einer Reihe von fünf oder sechs Zahlen erkannt werden, welcher Regel der Aufbau der Reihe folgt. Aus fünf Antwortalternativen muß sodann die richtige Fortsetzung der Reihe herausgefunden werden. Beispiel: "10 12 14 16 18 20: a) 21 b) 22 c) 23 d) 24 e) 25".
- Q4 Beim Untertest **Gleichungenbildern** müssen vorgegebene Zahlen und Operationszeichen so kombiniert werden, daß sich als Ergebnis der Rechnung eine von fünf Auswahlantworten ergibt. Beispiel: "2 2 3 + · : a) 6 b) 8 c) 9 d) 10 e) 11".

Der **nonverbale Teil (N-Teil)** enthält schließlich folgende Subtests:

N1 Ähnlich wie beim Untertest V3 Wortklassifikationen muß bei **Figurenklassifikation** diejenige von fünf Auswahlfiguren ausgewählt werden, die zur Klasse der vorgegeben Figuren paßt. Beispiel:



N2 Analog zu V4 müssen beim Subtest **Figurenanalogien** Analogien von Figurenpaaren ergänzt werden. Beispiel:



N3 **Figurensynthese:** Die Schüler sollen untersuchen, ob die angegebenen Flächen aus vorgegebenen Formen zusammengesetzt werden können.

Gegebene Stücke:



Zusammengesetzte Formen:

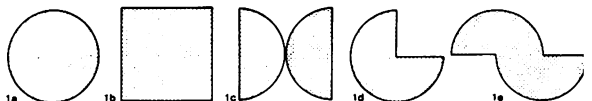


Tabelle 8: Untertests und Erfassungsdimensionen des KFT 4-13+

Erfassungsdimensionen	Aufgabenbezeichnungen	Vergleichbare Skalen	Mögliche Faktoren nach			
			T	G	J	F
Sprachverständnis	V1 Wortbedeutung	PSB 1+2; MIT-WB	V	CMU	F6	V
	V2 Satzergänzen	PSB 1+2; MIT-SW	V	CMU	F6	V
Sprachgebundenes Denken	V3 Wortklassifikationen	PSB 1+2, 4, 6	V	CMC	F6	V, I3
	V4 Wortanalogien	MIT-WV; PSB 1+2, 4, 6	V, R	CMR	F4, F6	V, I3
Arithmetisches Denken	Q1 Textrechenaufgaben	PSB 4	R, V, N	CMS	F5	R1, N
	Q2 Mengenvergleiche		N	ESR	F4, F5	R1, N
Rechenfähigkeit	Q3 Zahlenreihen	PSB 3,4; MIT-Z	R, N	CSS	F4, F5	I3, R
	Q4 Gleichungen bilden		N	NSR	F4, F5	R1, N
Anschauungsgebundenes Denken	N1 Figurenklassifikation	MIT-FR; CFT 1; PMT; PSB 3+4	R, S	CFC	F1, F4 (F2)	I3
	N2 Figurenanalogien	PSB 3, 4, 7, 8	R	CFR	F1, F4	I3
Konstruktive Fähigkeiten	N3 Figurensynthese	PSB 7, 8; MIT-MO; CFT 4	S2, S1	CFU	F1, F4	S1, Vz

Legende: Faktoren nach

Thurstone:

N: Number

R: Reasoning

S: Space

S1: Räumliche Beziehungen

S2: Veranschaulichung

V: Verbal Comprehension

Guilford:

Denkoperationen:

C: Cognition

N: Convergent

Production

E: Evaluation

Denkinhalte:

F: Figural

S: Symbolic

M: Semantic

Denkprodukte:

U: Units

C: Classes

R: Relations

S: Systems

Jäger:

F1: Anschauungsgebundenes Denken

F2: Einfallsreichtum und Produktivität

F4: Verarbeitungskapazität, Formallogisches Denken

F5: Zahlengebundenes Denken

F6: Sprachgebundenes Denken

French:

I3: Induction

Figure Classification

N: Numerical Facility

R: Reasoning

R1: Mathematical Aptitude

S: Spatial Orientation

V: Verbal Comprehension

Vz: Visualization

Der KFT 4-13 ist als Staffeltest gestaltet: Von der Gesamtzahl der pro Untertest zur Verfügung stehenden Items bearbeiten die Schüler der einzelnen Klassenstufen jeweils nur einen Teil, der ihrem Leistungsstand entsprechen soll. In jeder höheren Klassenstufe fallen die drei bis fünf leichtesten Items weg, und es kommt die gleiche Zahl schwierigerer Aufgaben hinzu, wobei sich die einzelnen Aufgabensets teilweise überlappen. Damit soll erreicht werden, daß jüngere Schüler nicht von zu schweren Aufgaben überfordert und ältere Schüler nicht mit zu leichten Items gelangweilt werden. Im einzelnen bearbeitet jede Alterstufe bei den Untertests V1, V2, V3, Q2, N1 und N2 25 von insgesamt 70 Aufgaben, bei den Subtests V4 und Q3 20 von insgesamt 56 Items, bei Q1 12 von 39, bei Q4 15 von 42 und bei N3 28 von 64 Aufgaben.

Trotz der zeitlichen Begrenzung für jeden Subtest wird der KFT 4-13+ von den Autoren eher als Power- denn als Speed-Test angesehen.

Erfassungsdimensionen: Tabelle 8 (vgl. Manual, S. 6) gibt einen Überblick über Untertests, Anzahl der Items und Erfassungsdimensionen des KFT 4-13+.

Handhabung: Der KFT 4-13+ kann als Einzel- oder Gruppentest durchgeführt werden. Für die Gruppenadministration oder Wiederholungsuntersuchungen stehen zwei Testformen A und B zur Verfügung. Für jeden Subtest sind Zeitobergrenzen festgelegt, die so gewählt sind, daß etwa 80 Prozent der Schüler alle Aufgaben in dieser Zeit bearbeiten können. Der KFT sei daher "als kombinierter Power-Speed-Test (eher Power- als Speed-Test) zu betrachten" (Manual, S. 6). Die Durchführung beansprucht laut Handbuch 3 Schulstunden (reine Testzeit 100 Minuten), wobei einschließlich Pausen unseren Erfahrungen zufolge mindestens 4 Schulstunden angesetzt werden müssen. Die Durchführung einer Kurzform (ca. 2 Schulstunden) ist möglich. In diesem Fall werden nur die Subtests V1, V4, Q2, Q3, N1 und N2 bearbeitet. Die Auswertung erfolgt wie üblich mittels Lösungsschablonen⁷, aus den Untertestrohwerten werden die Dimensionsrohwerte errechnet. Eigentlich im Widerspruch zur faktorenanalytischen Konzeption des differentiellen Tests ist auch die Berechnung eines Gesamtleistungswertes vorgesehen (siehe auch unten).

Für die Dimensionen und die Gesamtleistung können Alters- und Klassennormen (T-Werte) nachgeschlagen werden. Leider ist hier die Bezeichnung etwas unglücklich gewählt: Mit Hilfe der Altersnormen können die Schüler mit der Gesamtpopulation des jeweiligen Schuljahrgangs (T_{AN}) verglichen werden, die Klassennormen (T_{KN}) sind schulartspezifische Normen für jede Klassenstufe. Das Ergebnis wird graphisch dargestellt. Anleitung zur Berechnung von Vertrauensintervallen und kritischen Differenzen finden sich im Handbuch ebenso wie Hilfen zur schulartspezifischen Profilverinterpretation (sog. AUKL-Normen) und Hinweise zur Diagnose bei lese-rechtschreibschwachen Schülern. Als Interpretationshilfen sind acht, z.T. ausführliche Fallbeispiele zu unterschiedlichen diagnostischen Fragestellungen aufgeführt.

Gütekriterien: Durch die festgelegten Instruktionen und die Auswertungsschablonen sind **Durchführungs- und Auswertungsobjektivität** gegeben. Über die **Itemanalysen** im Rahmen der Testkonstruktion wird im Überblick berichtet. Mitgeteilt werden pro Klassenstufe Schwierigkeits- und Trennschärfeindizes für den Untertest Q1, die jedoch z.T. sehr ungünstig ausfallen. Nach Berechnungen von Perleth (1985) sowie Perleth et al. (1989) ist Q1 gerade der Untertest mit den schlechtesten Itemkennwerten und der geringsten Reliabilität.

Die Angaben zur **Reliabilität** (Konsistenz) der Skalen liegen um $Rel_X=0.8$, wobei die Werte von 0.46 bis 0.95 schwanken (ohne Klasse 13, wo die Stichprobe zu klein war). Da die Berechnungen nach Klassenstufen und Schularten getrennt vorgenommen wurden, was mit erheblicher Varianzeinschränkung verbunden ist, können sie als durchaus befriedigend bezeichnet werden. Eine kombinierte Bestimmung der Re- und Paralleltestreliabilität (Administration der Parallelförmung nach

⁷ Es sei darauf hingewiesen, daß sich bei älteren Exemplaren des Tests einzelne Fehler in die Lösungsschablonen sowie Druckfehler in die Normtabellen eingeschlichen haben, die aber inzwischen beseitigt sind.

3 Wochen) ergab einen mittleren Koeffizienten (Median) von 0.78 für die Normal- und 0.81 für die Kurzform, die Werte für die Subtests liegen z.T. deutlich darunter. Aber auch diese Werte sind angesichts der strengen Prüfung nicht unbefriedigend (Varianzeinschränkung durch homogene Stichproben, Maximierung der Fehlervarianz durch Vorgabe der Parallelförmigkeit beim Retest). Die Profiltestreliabilität ist nach Lienert (1969) mit Werten von in der Regel über 0.7 voll ausreichend.

Untersuchungen in Rahmen der Münchner Längsschnittstudie zur Hochbegabung (Heller, 1990; Perleth, 1991) mit einer Spezialform des KFT 4-13 ergaben insgesamt ein ähnliches Bild der Reliabilität des Verfahrens. Der Großteil der berechneten Konsistenzschätzungen nach Formel 20 von Kuder-Richardson liegt etwa zwischen 0.7 und 0.9, während die Stabilitätskoeffizienten für die Gesamtleistung (Re-Testung nach 1 bzw. 2 Jahren) um 0.7 liegen.

Perleth et al. (1989) fanden für eine Stichprobe bayerischer Gesamtschüler hohe Stabilitätskoeffizienten für die KFT-Gesamtleistung (0.82 bis 0.87 über 1 und 2 Jahre), während die Werte für die Dimensionen z.T. erheblich geringer ausfielen (0.4 bis 0.8), wobei sich vor allem für das Arithmetische Denken (Q1+Q) und die Konstruktiven Fähigkeiten (N3) geringe Stabilitätskoeffizienten ergaben. Auch mittels eines linearen Strukturgleichungsmodells (LISREL) zur kombinierten Abschätzung von Reliabilität und Stabilität der Dimensionen des KFT 4-13 konnten die Autoren zeigen, daß der erfaßte Anteil allgemeiner Intelligenz über einen Zeitraum von 2 Jahren (5. bis 7. Klasse) hochstabil ist (zur Methode vgl. Jöreskog, 1979). Da ein Großteil der Varianz in dem geprüften Modell der allgemeinen Intelligenz zugeordnet wurde, fielen die Meßfehler der Untertests z.T. recht groß aus.

Die Untersuchungen zur **Validität** wurden klassenstufen- bzw. schulartspezifisch bei insgesamt 1180 Schülern durchgeführt (vgl. Gaedike, 1976). Zur Abschätzung der Übereinstimmungsvalidität mit inneren Kriterien wurden die KFT-Untertests mit denen des PSB und des CFT in Beziehung gesetzt. Als äußere Kriterien wurden Schulleistungen und Einschätzungen der Intelligenz durch Lehrer verwendet. Die Ergebnisse sind im Handbuch sowie bei Gaedike (1976) ausführlich dokumentiert und brauchen hier nicht im einzelnen dargestellt zu werden. Alles in allem ergeben sich in bezug auf die Intelligenztests mittlere bis enge Zusammenhänge mit konstruktnahen und eher niedrige mit konstruktferneren Verfahren. So variieren z.B. die meisten Übereinstimmungskoeffizienten mit den inhaltlich vergleichbaren PSB-Subtests zwischen $r=0.40$ und $r=0.70$, wobei zur PSB-Gesamtleistung sogar einzelne Zusammenhangswerte bis zu $r=0.80$ gefunden wurden. Inhaltlich konnten die in der Tabelle 8 oben dargestellten Erfassungsdimensionen bestätigt werden. In bezug auf Schulleistungen sehen die Autoren die Dimensionen Sprachverständnis (V1+V2) und Arithmetisches Denken (Q1+Q2) in engerem Zusammenhang mit Schulleistungen als die Dimensionen Sprachgebundenes Denken (V3+V4) und Rechenfähigkeiten (Q3+Q4).

Ähnliche Befunde liegen zur amerikanischen Originaltestversion CAT (Cognitive Abilities Test) vor. So berichten Kail & Pellegrino (1988, S. 48ff.) Validitätskoeffizienten des CAT bezüglich anderer Intelligenztests zwischen $r=0.60$ und

$r=0.74$ sowie bezüglich Schulnoten als Leistungsprognosekriterium zwischen $r=0.50$ und $r=0.60$.

In der oben bei der Besprechung des PSB erwähnten Studie von Eisebitt (1986) konnten mit dem KFT diskriminanzanalytisch 67 Prozent der Schulleistungsvarianz (dichotomisierte Durchschnittsnote) aufgeklärt werden; zusammen mit dem PSB stieg dieser Wert sogar auf 77 Prozent. Als mittlere Korrelation der KFT-Untertests mit Schulnoten gibt Eisebitt (1986) $r=0.37$ an, wobei jedoch - erwartungsgemäß - die Subtest-Kriteriumsbeziehungen fachspezifisch zwischen $r=0.04$ (z.B. V3+V4 zur Mathematiknote) und $r=0.54$ variieren. Die entsprechenden PSB-Vergleichswerte lauten $r=0.20$ (mittlere Korrelation) bzw. $r=-0.03$ (PSB 8 zur Erdkundenote) und $r=0.37$ für die betr. Subtest-Fachnotenbeziehungen (Eisebitt, 1986, S. 40).

Perleth (1991) fand für die Schüler der Münchner Hochbegabungsstudie niedrige bis mittlere Zusammenhänge der KFT-Leistungen mit Schulnoten, wobei wie erwartet verbale Intelligenz am stärksten mit der Deutschensur und quantitative Intelligenz mit der Mathematiknote korrespondierten. Mit konstrukt näheren Verfahren (ZVT, MTP) ergaben sich niedrige bis mittlere, für konstruktferne Verfahren (Kreativität) keine Zusammenhänge mit der KFT-Gesamtleistung.

Die faktorielle Validität des KFT 4-13 kann als gut gelten. In einer Reihe von Faktorenanalysen wurde ein stabiles Drei-Faktoren-Muster über die verschiedenen Altersstufen hinweg gefunden: 1) Sprachgebundenes Denken (Untertests des V-Teils), 2) Formallogisches Denken (N-Teil) und 3) Zahlengebundenes Denken (Subtests des Q-Teils). Auch bezüglich des CAT wurde diese Drei-Faktorenstruktur vielfach bestätigt (vgl. Kail & Pellegrino, 1988, S. 49). In neueren Untersuchungen zum KFT (z.B. Steffens, Hospelt & Heller, 1983; Perleth & Geisler, 1986) mußte jedoch der Untertest Q3 vor allem bei leistungsstärkeren Schülern dem Reasoning-Faktor zugeordnet werden. Da die zur Lösung der Aufgaben von Q3 notwendigen arithmetischen Operationen eher einfacher Natur sind, ist dieses Ergebnis durchaus plausibel. In den Faktorenanalysen zum KFT - auch in späteren, z.T. unveröffentlichten Untersuchungen - zeigte sich regelmäßig eine starke erste Hauptkomponente, was die Bestimmung einer Gesamtleistung trotz prinzipiell mehrfaktoriellem Ansatz zulässig erscheinen läßt (vgl. auch die CAT-Befunde sowie zur theoretischen Bezugsbasis das Jägersche Intelligenzmodell, z.B. Jäger, 1982, 1984).

Die Normierung des KFT 4-13 erfolgte im Herbst 1974 in allen Bundesländern. Normierungen wurden für die beiden Testformen, die Schuljahrgangsstufen 4 bis 13 (sog. Altersnormen) und innerhalb jeder Klassenstufe für die einzelnen Schularten (sog. Klassennormen) vorgenommen. Eigene Normtabellen liegen für Gesamtschüler, Berufsschüler, Höhere Handelsschüler und Studenten vor. Speziell für Berufsschüler kann auch auf die von Perleth & Geisler (1986) publizierten Normen zurückgegriffen werden, die auf einer großen Stichprobe bayerischer Berufsschüler aller Sparten beruhen. Die Schüler dieser Stichprobe haben - besonders wegen der Anforderungen im Q-Teil in höheren Klassenstufen - die Aufgaben für Klasse 9 bearbeitet, während jene Schüler, denen die Normen im Manual zugrunde liegen (Einstellungstest bei einem großen rheinischen Industrieunternehmen), überwiegend Aufgabensets für Klasse 10 vorgelegt bekamen; gerade im Q-Teil seien dabei große

Motivationsprobleme zu verzeichnen gewesen⁸. Für Jungen und Mädchen wurden keine getrennten Normtabellen berechnet, da die Geschlechtsunterschiede entweder nicht signifikant waren oder von den Autoren als nicht gravierend beurteilt wurden (Differenzen in etwa 1 Rohpunkt).

Für Schüler der 6. und 8. Klasse wurden Schuleignungskennbereiche mit dem AUKL-Programm (Gruppierungsanalyse) ermittelt. Anhand dieser Kennbereiche können die Testprofile von Schülern mit den Ergebnissen der per AUKL ermittelten "idealen" Hauptschüler, Realschüler und Gymnasiasten verglichen werden.

Im Hinblick auf mögliche Normverschiebungen berichten Praktiker, daß der KFT die Leistungen in der Dimension Arithmetisches Denken eher unterschätzt, die im nonverbalen Teil eher überschätzt. Entsprechende Hinweise haben auch Perleth et al. (1989) für ihre - nicht repräsentative - Gesamtschulstichprobe gefunden, wobei andererseits deutlich wurde, daß zumindest für diese Stichprobe in den 80er Jahren kaum Normverschiebungen stattfanden.

Kritik: Zwar leidet die Ökonomie des KFT 4-13 durch die lange, für schwächere Schüler teilweise belastende Durchführungszeit, andererseits können mit dem KFT wichtige und nützliche Informationen für die Schulberatung gewonnen werden. Vom PSB hebt sich der KFT 4-13 z.B. dadurch ab, daß er eher als Power- denn als Speed-Test konzipiert ist und einen anspruchsvollen Q-Teil sowie einen differenzierten V-Teil enthält. Durch die Staffelung der Aufgaben kann erreicht werden, daß Schüler Testaufgaben bearbeiten, die eher zu ihrem Leistungsstand passen. Andererseits handelt man sich damit natürlich den Nachteil ein, daß die Leistungen von Schülern unterschiedlicher Klassenstufen nicht auf Rohwertbasis miteinander verglichen werden können. Dies stellt aber eher eine Einschränkung für den Forscher (z.B. bei Längsschnittstudien, die auf die Untersuchung einer Entwicklungsfunktion abzielen) als für den Praktiker dar.

Für die Ökonomie des KFT spricht andererseits die Möglichkeit, eine Parallel- und Kurzform durchzuführen. Auch bietet das Manual eine Reihe von Hilfen zur Testinterpretation, die deutlich über vergleichbare Verfahren hinausgehen. Dazu gehören auch so nützliche Details wie die Tabelle auf S. 6 des Manuals, die die mit dem KFT erfaßten Fähigkeiten in Relation mit einschlägigen Intelligenztheorien setzt. Überhaupt bietet das Manual eine Fülle von (auch statistischen) Informationen, die dem Benutzer einerseits eine eigene Beurteilung des Verfahrens ermöglicht, andererseits aber Anwender ohne intensive testtheoretische Ausbildung wie Beratungslehrer eher überfordern dürfte. Dennoch muß hervorgehoben werden, daß damit im Manual Stärken und Schwächen des Tests nicht verschleiert, sondern transparent gemacht werden.

Die Kritik Kubingers (1986) an der Neuauflage des KFT 4-13+ kann dahingehend zusammengefaßt werden, daß der KFT nach der Klassischen Testtheorie und nicht nach dem Rasch-Modell konstruiert wurde. Ansonsten weist der Autor darauf hin, daß seines Erachtens die Untertests zu lange und manche Aufgaben des V- und

⁸ Mündliche Mitteilung von Frau Hospelt-Renette (Juni 1984), die die Normierung für Berufsschüler durchgeführt hat.

Q-Teils zu schwer sind. Auch hält er die Zusammenfassung von je zwei Subtests zu den Testdimensionen nicht für ausreichend begründet. Dennoch sei der KFT 4-13+ ein wichtiges Hilfsmittel bei der Schullaufbahn- und Bildungsberatung.

In letzter Zeit haben wir im Rahmen der Münchner Längsschnittstudie zur Hochbegabung (Heller, 1990; Perleth, 1991) eine **Spezialform zur Anwendung bei besonders begabten Schülern** zusammengestellt und überprüft (Untertests V1, V2, Q2, Q4, N1 und N2). Diese Zielgruppe bearbeitet im V- und N-Teil Aufgaben, die normalerweise erst von Schülern der übernächsten Klassenstufe bearbeitet werden. Diese Form hat sich in unseren Studien bewährt und kommt derzeit nach erfolgreicher Übersetzung bzw. Übertragung ins Russische in einer (Teil-)Replikation der Münchner Begabungsstudie im Großraum Moskau zum Einsatz. Dabei überraschte, daß die Kennwerte des Verfahrens in der russischen Analysestichprobe mit denen der deutschen Schüler nahezu völlig übereinstimmen (Perleth, Averina & Scheblanowa, 1991).

Tabelle 9: Übersicht über differentielle Fähigkeitstests

Autor und Erscheinungsjahr	Testbezeichnung	Erfassungsdimensionen	Alters- und Zielgruppe	Zeitbedarf
W. Horn (1962, 1983 ²)	Leistungsprüfungssystem (LPS)	Primärfähigkeiten nach Thurstone	9-17jährige	2 Schulstunden, mit kompletter Arbeitskurve 3 Schulstunden
W. Horn (1969)	Prüfsystem für Schul- und Bildungsberatung (PSB)	Primärfähigkeiten nach Thurstone	9-20jährige	1 Stunde
R. Amthauer (1970, 1973 ⁴)	Intelligenz-Struktur-Test (IST 70)	Sprachliche, rechnerische und räumliche Intelligenzaspekte, Gedächtnis	12-60jährige, v.a. Berufsberatung	90 Minuten
W. Conrad, G. Eberle, L. Hornke, B. Kierdorf & B. Nagel (1976)	Mannheimer Intelligenztest für Kinder und Jugendliche (MIT-KJ)	Befähigung in wichtigen Bereichen intellektuellen Verhaltens	9-15jährige	1 Stunde
H. Kratzmeier (1982)	Heidelberger Intelligenztest (HIT 3-4)	Allgemeine kognitive und kreative Leistungsfähigkeit	1. und 2. Klassen	90 Minuten

E. Hylla & B. Kraak (1968), Neueichung von R. Kühn (1976)	Aufgaben zum Nachdenken 4+ (AzN 4+)	Rechnerische Begabung und schulische Leistungsvor- aussetzungen	Klassen 4-5	110 Minuten
K. Heller & H.-J. Geisler (1983)	Kognitiver Fähigkeits- Test - Kinder- gartenform (KFT-K)	Wortschatz, Relationen, Schlußfol- gerndes und Rechnerisches Denken, Allgemeinbe- gabung	Vorschüler im Alter von 4;7 bis 7;0 Jahren	ca. 60 Minu- ten, verteilt auf mehrere Test- termine
K. Heller & H.-J. Geisler (1983)	Kognitiver Fähigkeits- Test - Grund- schulform (KFT 1-3)	Wortschatz, Relationen, Schlußfol- gerndes und Rechnerisches Denken, Allgemeinbe- gabung	Schüler der 1. bis 3. Klasse	je nach Pbn- Alter/Klas- senstufe 45-60 Minuten
K. Heller, A.- K. Gaedike & H. Weinländer (1976, 1985 ²)	Kognitiver Fähigkeits-Test für 4.-13. Klas- sen (KFT 4-13+)	Verbale, quantitative und nonver- bale (Reason- ing-) Fähig- keiten, Allge- meinbegabung	Schüler der 4. bis 13. Klasse aller (Regel-) Schularten, Berufsschüler, Studenten	je nach Pbn- Alter/Klas- senstufe 3-4 Schulstunden, Kurzform ca. 2 Schulstunden

3.3.3 Spezielle Fähigkeitstests

- (1) Der Mechanisch-Technische Verständnistest (MTVT) von G.A. Lienert (1964, 1970²) (Hogrefe)

Testaufbau: Der MTVT besteht aus 32 praktisch-technischen Problemaufgaben, z.B. Hebelaufgaben, Räderwerkitems u.ä. (vgl. Abbildung 21). Trotz der - aus ökonomischen Gründen - vorgesehenen zeitlichen Limitierung auf 45 Minuten ist der MTVT eher ein Power- oder Niveautest.

Erfassungsdimensionen: Der MTVT soll das mechanisch-technische Verständnis oder Denken prüfen, das nach Lienert einen wichtigen Aspekt der "praktischen Intelligenz" repräsentiert. Die Faktorenanalyse mit einer Testvorform erbrachte folgendes Ergebnis: "Der MTVT ist durch einen Faktor C, interpretiert als die Fähigkeit des zweckmäßigen Kombinierens und sinnvollen Ergänzens, zu 67 Prozent ($r=0.82$) bestimmt. Ein weiterer Faktor A, interpretiert als die Fähigkeit des schlußfolgernden Denkens, ist mit einem Anteil von 33% ($r=0.57$) an der MTVT-Leistung beteiligt. Ein dritter Faktor B, interpretiert analog dem kin-

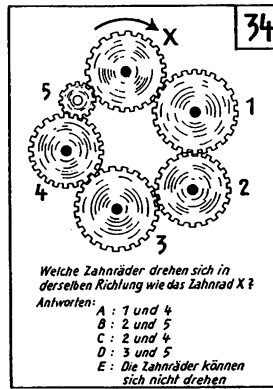
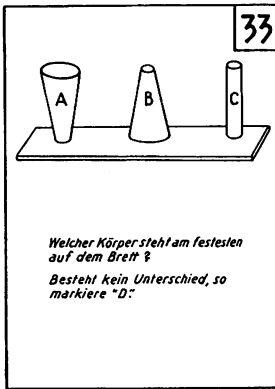


Abbildung 21: Beispiele für Aufgaben des MTVT

aesthetischen Faktor von Thurstone, ist unter der MTVT-Leistung nicht wirksam" (Lienert, 1964, S. 7).

Testgütekriterien: Der MTVT verfügt über gute **Objektivität**. Die **Reliabilität** dagegen ist nur befriedigend. So wurden vom Testautor Retest-Zuverlässigkeitskoeffizienten von $Rel_{\chi} = 0.76$ bzw. 0.82 ermittelt, wobei der Übungsgewinn (der Wiederholungstestung nach 7 Wochen) bei durchschnittlich 6 Z-Werten (Mittelwert: 100, Standardabweichung: 10) lag. Die Halbierungszuverlässigkeit des MTVT beträgt $r = 0.81$ bzw. 0.86 , die innere Konsistenz $r = 0.84$. Der **Standardmeßfehler** liegt dementsprechend bei 5 Z-Werten.

Abgesehen von der faktoriellen Gültigkeit des MTVT (siehe oben) liegen Informationen zur inneren und äußeren **Validität** vor. Die Korrelationen des MTVT zu verschiedenen Subtests der Mechanical Aptitude Battery von Thurstone variieren zwischen 0.24 und 0.43 , zu verschiedenen Subtests des LPS zwischen 0.32 und 0.40 (Lienert 1964, S. 7f.). Wir selbst ermittelten in der Relation MTVT und LPS 7-10 folgende Zusammenhänge: $r = 0.28$ (bei 881 Gymnasiasten und 523 Realschülern) sowie $r = 0.30$ (bei 2664 Hauptschülern); die von Lienert bei 209 Oberschülern gefundenen Zusammenhangswerte von MTVT und LPS konnten somit nicht ganz bestätigt werden (vgl. Heller 1970, S. 103). Mit dem FLT (Form-Lege-Test von G.A. Lienert) ist die MTVT-Leistung mit $r = 0.44$ korreliert. Insofern kann man der Meinung des Testautors beipflichten, wonach der MTVT ein "objektiver, im Verhältnis zu den meisten Tests dieser Art hoch zuverlässiger Test" ist, dessen Gültigkeit "innerhalb und faktoriell begründet" wird.

Die **Normen** werden in PR- und Standardwert- bzw. Z-Einheiten angegeben. Es liegen Schultyp-, Berufsgruppen- und Altersgruppen-Normen vor. Da die Normen jedoch aus den frühen 60er Jahren stammen, scheint eine Neunormierung oder zumindest Überprüfung der Normen erforderlichlich.

Anwendungsbereiche: Der MTVT dient der Erfassung mehr praktisch-technischer Fähigkeiten und ist ab dem 13. Lebensjahr auch im Sekundarstufenbereich einsetzbar, wo er eine gute Ergänzung zu üblichen Intelligenztests, besonders bei Haupt- und Realschülern darstellt. Für die interschulische Begabungsdifferenzierung im gegliederten Sekundarschulsystem (besonders zwischen Realschülern und Gymnasiasten) ist er hingegen weniger geeignet (Heller, 1973).

Testkritik: Der MTVT kann zwar hinsichtlich seiner Treffsicherheit nicht allen Ansprüchen moderner Testkritik genügen, auf der Hauptschulabschlußstufe sowie in der Eignungsermittlung von Berufsschülern (Lehrlingen und Anlernlingen) ist der Test jedoch insgesamt ein recht brauchbares Hilfsmittel zur Information über die praktisch-technische Begabung von Schülern und Jugendlichen.

(2) **Der Mannheimer Test zur Erfassung des physikalisch-technischen Problemlösens (MTP)** von W. Conrad, E. Baumann & V. Mohr (1980) (Hogrefe)

Testaufbau: Der Test besteht aus 26 Testaufgaben, die nach Lösungsschwierigkeiten geordnet sind. Pro Item müssen die Probanden die richtige Antwort zu einem physikalisch-technischen Problem aus 5 Antwortalternativen herausuchen (vgl. Abbildung 22). Die beiden zur Verfügung stehenden Testformen unterscheiden sich in Details der Aufgabenstellung und in der Anordnung der Lösungsalternativen.

Welche beiden Zahnräder drehen sich in der gleichen Richtung wie das Zahnrad x?

- A: G und K
- B: G und J
- C: H und K
- D: H und J
- E: keine von diesen



Abbildung 22: Beispielaufgabe zum MTP

Erfassungsdimension: Mit dem MTP soll die allgemeine Fähigkeit zur Lösung physikalisch-technischer Problemstellungen im Sinne eines "latent-trait" nach dem Rasch-Modell erfaßt werden.

Handhabung: Die Durchführung des Verfahrens beansprucht inklusive Instruktion 30 Minuten. Mit Hilfe von Lösungsschablonen erfolgt die Bestimmung von Rohwerten, die anhand von Tabellen in Fähigkeitsparameter transformiert werden (der MTP wurde Rasch-skaliert). Daneben stehen populationspezifische Vergleichsdaten zur Verfügung.

Gütekriterien: Durchführungs- und Auswertungsobjektivität sind aufgrund entsprechender Standardisierung gegeben. Der MTP wurde nach dem Rasch-Modell konstruiert, wobei von ursprünglich 66 Items nach unterschiedlichen Selektions- und Prüfschritten letztendlich 24 übrigblieben. Bei der Selektion der Items wurden (etwas im Widerspruch zu den Zielen eines Rasch-skalierten Tests) auch Itemkennwerte der Klassischen Testtheorie herangezogen. Nach Methoden der Klassischen

Testtheorie wurde auch die **Reliabilität** bestimmt: Die Schätzungen für die innere Konsistenz liegen zwischen 0.8 und 0.9, die Stabilitätskoeffizienten in derselben Größenordnung (für die Gesamtgruppe wegen der größeren Heterogenität der Stichprobe bei 0.91).

Auch die **Validität** des Verfahrens ist gut belegt: Der MTP korreliert hoch bis mäßig mit konstrukt-nahen Tests ($r=0.77$ mit dem PTV, $r=0.69$ mit dem MTVT, $r=0.43$ mit dem LPS 8+9) und niedriger mit konstruktferneren Variablen ($r=0.18$ mit dem LPS 1+2, $r=0.03$ mit dem d2, $r=0.36$ mit dem Figure Reasoning Test, geringe bzw. verschwindende Zusammenhänge mit Persönlichkeitsvariablen). Bei Faktorenanalysen mit vorwiegend Intelligenzskalen bildeten der MTP und der MTVT einen Faktor "Physikalisch-technisches Problemlösen". Schließlich fanden sich die erwarteten Unterschiede zwischen Stichproben unterschiedlicher Ausbildungsniveaus und Fachrichtungen.

Normen stehen für Berufs-, Fach- und Fachhochschüler/Studenten sowie unterschiedliche Studienrichtungen zur Verfügung. Die Größe der Eichstichproben variiert für die einzelnen Gruppen zwischen 56 (Fachschüler der Datentechnik) und 1359 (Studenten der Elektrotechnik u.ä.) beträchtlich, wobei Stichproben unter 100 kaum mehr aussagekräftig sind.

Tabelle 10: Übersicht über Tests zur Erfassung spezieller Fähigkeiten

Autor und Erscheinungsjahr	Testbezeichnung	Erfassungsdimensionen	Alters- und Zielgruppe	Zeitbedarf
E. Fürntratt (1969)	Differentieller Wissenstest (DWT)	Wissen aus 11 Bereichen	ab 14 bis 17 Jahren	60-90 Minuten
H. Steingrüber (1971)	Hand-Dominanz-Test (HDT)	Händigkeit	6-10 Jahre	10 (Einzeltest) bis 15 Minuten (Gruppentest)
G. Bäumlner (1974)	Lern- und Gedächtnistest (LGT 3)	Gedächtnis (figural, verbal, numerisch)	ab 16 Jahren	40 Minuten
W. Conrad, E. Baumann & V. Mohr (1980)	Mannheimer Test zur Erfassung des physikalisch-technischen Problemlösens (M-T-P)	Fähigkeit zur Lösung physikalisch-technischer Problemstellungen	Berufs-, Fach- und Fachhochschüler, Studenten	30 Minuten
G.A. Lienert (1970)	Mechanisch-technischer Verständnistest (M-T-V-T)	Praktische Intelligenz	Sekundarstufenschüler ab dem 13. Lebensjahr	45 Minuten

R. Amthauer (1972)	Ein Test zur Untersuchung des praktisch-technischen Verständnisses (PTV)	Praktisch-technisches Verständnis	13-50jährige	25 Minuten ohne Instruktion
H. Stumpf & E. Fay (1983)	Schlauchfiguren. Ein Test zur Beurteilung des räumlichen Vorstellungsvermögens	Räumliches Vorstellungsvermögen	Haupt- und Realschulabsolventen, Lehrstellenbewerber, Abiturienten	12 Minuten ohne bzw. 20 Minuten mit Instruktion
E.J. Kiphard & F. Schilling (1974)	Körperkoordinationstest (KTK)	Gesamtkörperbeherrschung/Grobmotorik	5-14jährige, ältere behinderte Kinder	keine Angaben
Ch. Butsch & H. Fischer (1966)	Seashore-Test für Musikalische Begabung	Grundfähigkeiten musikalischer Begabung	10-19jährige	ca. 1 Stunde

3.3.4 Kreativitätstests

(1) Der Verbale Kreativitäts-Test (V-K-T) von K.-J. Schoppe (1975) (Hogrefe)

Testaufbau: Der VKT besteht aus 9 Untertests, mit denen verbale Produktivität erfaßt werden soll: Beim Untertest **Wortanfänge (WA)** müssen die Probanden möglichst viele Wörter aufschreiben, die mit einer gegebenen Buchstabenkombination beginnen. Beim Untertest **Wortenden (WE)** müssen analog Wörter gefunden werden, die mit einer bestimmten Buchstabensequenz enden. Als nächstes (Subtest 3) müssen die Probanden zu vier vorgegebenen Buchstaben (z.B. E-I-O-V) so viele verschiedene **Vier-Wort-Sätze (VS)** wie möglich aufschreiben, wobei jeder Buchstabe genau einmal als Anfangsbuchstabe eines der vier Wörter vorkommen muß (z.B. "Oma ißt viele Eier" oder "Ist Onkel ein Verbrecher"?). Bei Untertest 4 **Namen-Erfinden (NE)** bekommen die Probanden "Abkürzungen" vorgelegt, für die sie Namen erfinden müssen. Untertest 5 **Gleiche Eigenschaften (GE)** verlangt die Produktion möglichst vieler Begriffe zu einer vorgegebenen Eigenschaft (z.B. "Teller", "Uhr", "Knopf" zu "rund"), während für den Subtest 6 **Ähnlichkeiten (ÄK)** möglichst viele Synonyme zu einem vorgegebenen Begriff gefunden werden müssen. Einfälle zu möglichst unüblichen Verwendungen alltäglicher Gegenstände (z.B. "leere Konservendose") werden beim Untertest 7 **Ungewöhnliche Verwendungsarten (UV)** verlangt, während beim Subtest 8 **Utopische Situationen (US)** die Folgen aufzuschreiben sind, die einem zu zwei ungewöhnlichen Situationen (z.B. "einen Monat ohne Strom leben") einfallen. Abschließend sollen sich die Probanden beim Subtest 9 **Spitznamen-Erfinden (SE)** originelle Spitznamen für

alltägliche Gegenstände (z.B. "Geld" = "Flöhe", "Mäuse", "Piepen", "Eier",...) ausdenken.

Erfassungsdimensionen: Der VKT dient nach Ansicht des Autors zur Erfassung der **Verbalen Produktivität**. Diese wird als ein Hauptfaktor des Intelligenzspektrums in Anlehnung an Jägers (1967) Faktor "Einfallsreichtum und Produktivität" oder Guilfords "Divergent-Production"-Kategorie verstanden. Insofern liegt dem VKT nach Schoppe (1975) kein eigentliches Kreativitätsmodell zugrunde, vielmehr wird ein solches Modell auch nicht als unverzichtbare Grundlage eines Kreativitätstests angesehen (vgl. Schoppe, 1975, S. 35).

Schoppe (1975) betont ausdrücklich, daß mit der VKT-Testserie lediglich Produktivität, nicht aber Originalität erfaßt werden solle, da seiner Meinung zufolge "nur Produktivität meßbar ist, während die Beurteilung der Originalität als qualitative Wertzumessung nicht die diagnostische Intention eines Psychologen sein kann" (Schoppe, 1975, S. 31). Auch die Definition von Originalität über die statistische Häufigkeit bzw. Seltenheit der jeweiligen Produktion lehnt Schoppe für verbales Material ab, nicht zuletzt weil ein derartig operationalisiertes Merkmal hoch mit der Produktion korreliert. Ähnlich - so der Autor des VKT - lassen sich auch Scores kognitiver Flexibilität nicht unabhängig von der Produktivität erfassen, zudem seien die Operationalisierung und Handhabung entsprechender Kategoriensysteme im Rahmen psychometrischer Diagnostik nicht vernünftig zu realisieren.

Handhabung: Der VKT liegt in zwei Parallelförmigkeiten vor und kann als Einzel- oder Gruppentest angewendet werden, wobei ausdrücklich keine Zielgruppen genannt werden. Diese ergäben sich aus Schwierigkeit und Konstruktionsart der Items. Für die einzelnen Items des VKT sind zwischen 90 und 240 Sekunden vorgesehen, die einzelnen Untertests benötigen jeweils 2 bis 8 Minuten Bearbeitungszeit. Die gesamte Durchführung dauert nach Angaben des Autors eine Stunde, wobei der Test auch in zwei Etappen (Unterbrechung nach Untertest 6) administriert werden kann. Die Auswertung erfolgt nach rein quantitativen Gesichtspunkten: Pro Subtest wird die Anzahl gültiger Produktionen ermittelt, allerdings muß der Auswerter falsche Produktionen streichen (Richtlinien für die Anerkennung richtiger Produktionen finden sich im Anhang des Manuals). Nach kurzer Einarbeitung beträgt die Auswertungszeit pro Testheft nach Angaben im Manual etwa 3 Minuten. Weitere 2 Minuten werden für die Ermittlung und Übertragung der Rohwerte sowie das Nachschlagen der Standardwerte veranschlagt. Die Subtests 1 und 2, 3 und 4 sowie 5 und 6 werden zusammengefaßt, so daß sich insgesamt 6 Standardwerte ergeben, aus denen ein sog. Kreativitäts-Index (KI) als Mittelwert berechnet wird. Auf eine eigene Normierung dieses Mittels wird aus praktisch-ökonomischen Überlegungen verzichtet.

Gütekriterien: Die **Auswertungsobjektivität** kann für ein derartiges Verfahren als gut bezeichnet werden. Schoppe teilt Inter-Auswerter-Korrelationen von 0.93 bis 0.97 mit. Zur Abschätzung der **Zuverlässigkeit** des VKT wurden Korrelationen zwischen den Parallelförmigkeiten berechnet. Die **Paralleltestreliabilität** liegt demnach für die sechs Skalen zwischen 0.83 (ÄK) und 0.91 (WE), für die Gesamtleistung KI beträgt der Wert 0.91. Eine Schätzung der Reliabilität über eine Retestung mit

der Parallellform ergab Werte zwischen 0.52 (US) und 0.84 (UV) sowie von 0.81 für die Gesamtleistung KI über einen Zeitraum von sechs Wochen. Nach einem Jahr oder mehr ergaben sich niedrigere Koeffizienten (für KI aber immer noch 0.85). Bei Retestung mit derselben Testform zur Bestimmung der **Stabilität des VKT** ergeben sich für die meisten Scores niedrigere Werte, einige liegen sogar unter 0.4. Dabei ist allerdings zu berücksichtigen, daß die Probanden bei der Retestung ihre Leistungen teilweise um 100 Prozent steigern konnten, was sicherlich damit zusammenhängt, daß bei den verwendeten Aufgabenstellungen mit erheblichen Erinnerungseffekten zu rechnen ist.

Im Hinblick auf **Gruppenunterschiede** teilt Schoppe (1975) mit, daß Alter, Geschlecht und - mit Einschränkungen - Intelligenzniveau keinen Zusammenhang mit den VKT-Leistungen zeigten. Wir selbst (Perleth, 1991) fanden für eine etwas modifizierte Version des Untertests Vierwortsätze für die Teilnehmer der Münchner Längsschnittstudie zur Hochbegabung (Heller, 1990) Geschlechtsunterschiede zugunsten der Mädchen, die mit steigender Klassenstufe deutlicher hervortraten.

Im Hinblick auf die **faktorielle Validität** des Verfahrens konstatiert Schoppe (1975), daß sich bei allen Faktorenanalysen zusammen mit anderen Begabungs- und Leistungstests ein Faktor "Verbale Produktivität", gebildet durch die VKT-Subtests sowie LPS-6 (Wortflüssigkeit), ergab. Leider folgt die Darstellung des Ergebnisses der Faktorenanalysen im Manual nicht dem üblichen Muster, was die Beurteilung dieses Befundes erheblich erschwert.

Zur Beurteilung der **Übereinstimmungsvalidität** wurden die VKT-Untertests mit Intelligenzskalen und Fähigkeitstests (LPS, IST, PMT, MTVT u.a.), Interessenskalen (DIT), Persönlichkeitsskalen und Schulnoten korreliert. Die VKT-Skalen zeigten in den Analysen für Abiturienten praktisch keine bedeutsamen Zusammenhänge mit den genannten Kriterien, niedrige Korrelationen ergaben sich mit einzelnen LPS-Skalen und dem LPS-Gesamtwert, wobei sich mit den Untertests "Ähnlichkeiten" und "Gleiche Eigenschaften" etwas höhere Zusammenhänge als mit den anderen VKT-Untertests ergaben. Die höchste Korrelation findet sich zwischen den beiden Gesamtleistungen von IST und VKT ($r=0.44$). Bei einer Stichprobe von Arbeitern zeigten sich hingegen deutliche Zusammenhänge zwischen VKT und Intelligenzskalen, beispielsweise korreliert LPS 8+9+10 (technische Begabung, Raumvorstellung etc.) mit der VKT-Gesamtleistung KI mit $r=0.74$. Die Zusammenhänge scheinen mit dem Alter zuzunehmen. Mit Lehrereinschätzungen der Kreativität ergaben sich niedrige bis mittlere Zusammenhänge.

Normen liegen für Arbeiter (18 bis 25, 26 bis 35 und 36 bis 55 Jahre) Jugendliche mit mittlerer Reife, Abiturienten und Hochschulüler vor.

Kritik: Der VKT scheint zur Erfassung der verbalen Produktivität geeignet zu sein. Inwieweit dieses Konstrukt jedoch Prognosen auf kreatives Verhalten zuläßt, müßte noch weiter untersucht werden. Die zur Verfügung stehenden Normen sind für den Einsatz in der Schulberatung nicht ausreichend, insbesondere fehlen Normen für jüngere Schüler.

(2) Der Test zum divergenten Denken (Kreativität) für 4. bis 6. Klassen (TDK 4-6) von U. Mainberger (1977) (Beltz)

Testaufbau: Der TDK 4-6 enthält 11 Aufgaben, die zu 5 Aufgabengruppen mit gleicher Aufgabenstellung zusammengefaßt werden. Bei der ersten Aufgabe müssen die Kinder in Gedanken Figuren aus drei länglichen Bausteinen legen und diese dann auf das karierte Papier des Antwortheftes aufzeichnen. Diese besonders leichte Aufgabe soll auch zur Motivation gerade schwächerer Kinder dienen. Es folgen zwei Aufgaben zu **Verrückten Geschichten**: Die Kinder müssen sich ausdenken, was geschehen könnte, wenn man frei in der Luft schwebte (Aufgabe 2) oder sich von drei Tabletten am Tag ernährte (Aufgabe 4). Bei den folgenden drei Aufgaben werden jeweils 16 einfache geometrische Figuren (z.B. Ovale, Trapeze) vorgegeben, die in **Zeichnungen** irgendwelcher Dinge integriert werden sollen. Die beiden Aufgaben 7 und 8 **Lustige Vergleiche** erfordern, daß die Kinder sich möglichst viele Gemeinsamkeiten zu einem Paar von Gegenständen (Kerze, Ofen) oder Tieren (Katze, Maus) ausdenken. Bei der letzten Aufgabengruppe **Bilderraten** müssen die Kinder Ideen dazu produzieren, was drei angefangene Bilder darstellen könnten.

Erfassungsdimensionen: Der TDK 4-6 soll bei Kindern der 4. bis 6. Klassen die Fähigkeit zum divergenten Denken als wesentliche kognitive Voraussetzung für kreatives Verhalten und kreative Leistungen erfassen. Divergentes Denken wird dabei als Problemlösungsverhalten betrachtet, das viele Lösungen bzw. Lösungsversuche hervorbringt. Auch wenn der TDK 4-6 in Anlehnung an Guilford konzipiert wurde, war es nicht Absicht der Autorin, alle Faktoren des Guilfordschen Modells, die sich auf divergentes Denken beziehen, zu erfassen. Vielmehr wird, nicht zuletzt wegen der relativ geringen Anzahl der Aufgaben, ein Gesamtwert gebildet und als Fähigkeit zum divergenten Denken interpretiert.

Handhabung: Das Handbuch gibt ausführlich und übersichtlich Anleitungen zur Vorbereitung, Durchführung und Auswertung des TDK 4-6. Die Durchführung dauert 45 bis 50 Minuten und sollte nach Ansicht der Autorin in die ersten beiden Schulstunden gelegt werden. Ausgewertet wird zunächst die Anzahl akzeptabler Antworten, wobei für die einzelnen Aufgaben anschaulich und anhand instruktiver Beispiele beschrieben wird, welche Antworten oder Produktionen gewertet werden. Da in der Regel "Punktobergrenzen" z.B. für Produktionen derselben Kategorie festgelegt werden, gehen implizit auch Aspekte der Flexibilität und Originalität in den Gesamtwert ein. Die Handhabung der Normentabellen wird ebenfalls anschaulich, anhand dreier Beispiele erläutert.

Gütekriterien: Die einzelnen Schritte der Testkonstruktion über mehrere Vorformen sind gut nachvollziehbar dargestellt, allerdings beziehen sich nicht alle Angaben zu den Gütekriterien auf die endgültige Form des TDK 4-6, sondern überwiegend auf die dritte und vierte Versuchsform. In der Phase der Testkonstruktion wurde neben dem Produktivitätswert auch ein Originalitätsscore bestimmt, der jedoch hoch mit dem Produktionsmaß übereinstimmte und überdies geringere

Validitätskoeffizienten aufwies, so daß aufgrund der komplizierten Bestimmung für die Endform darauf verzichtet wurde.

Bei Tests, die vom Auswerter eine Beurteilung der Antwortqualität erfordern, ist mit Einschränkungen der Auswertungsobjektivität zu rechnen. Für den TDK 4-6 ergab sich aber eine zufriedenstellende Inter-Rater-Reliabilität von 0.93. Zur Zuverlässigkeit wird für die dritte Versuchsform eine **Split-Half-Reliabilität**, bestimmt nach der odd-even-Methode, von 0.90 berichtet, was als gut zu bezeichnen ist⁹.

Im Hinblick auf die **Konstruktvalidität** zeigte sich, daß die Aufgaben des TDK 4-6 mit Intelligenztests und Schuleistungen niedriger korrelieren als untereinander bzw. mit anderen verbalen Kreativitätstests. Dies konnte auch faktorenanalytisch bestätigt werden. Die Korrelationen des TDK-Gesamtwertes mit dem GI-Wert von PSB bzw. BBT liegen um 0.4, die mit dem verbalen Kreativitätstest von Wallach & Kogan bei 0.65. **Gruppenunterschiede** im TDK 4-6 ergaben sich zugunsten von älteren Mädchen und Gymnasiasten, dagegen wurde keine Schichtabhängigkeit der Ergebnisse gefunden.

Die **Normierung** des TDK 4-6 erfolgte zum Jahreswechsel 1976/1977 an insgesamt 2262 Kindern aus Grund- und Hauptschulen sowie Gymnasien aller westdeutschen Bundesländer ohne Stadtstaaten und Schleswig-Holstein. Normen wurden für Klassenstufen und Schularten, in Klasse 6 auch für die Geschlechter getrennt berechnet. Die Stichprobengrößen für die einzelnen Normierungsgruppen liegen zwischen 142 (Mädchen 6. Klasse) und 550 (Grundschule 4. Klasse). Durch die linksgipflige Verteilung differenziert der TDK 4-6 im oberen Bereich besser. Neben T-Werten können in den Normentabellen Prozentrang- und T-Wert-Bänder (Vertrauensintervalle) abgelesen werden.

Kritik: Mit dem TDK 4-6 liegt ein Kreativitätstest zur Erfassung des divergenten Denkens (Produktivität) vor, der einigermaßen sorgfältig konstruiert und leicht zu handhaben ist. Für einen Einsatz in der Schul- und Erziehungsberatung wären jedoch noch weitere Informationen zur Validität, insbesondere zur prognostischen Kriteriumsvalidität, erforderlich.

Tabelle 11: Übersicht über Tests zur Erfassung kreativer Fähigkeiten

Autor und Erscheinungsjahr	Testbezeichnung	Erfassungsdimensionen	Alters- und Zielgruppe	Zeitbedarf
K.-J. Schoppe (1975)	Verbaler Kreativitäts-Test (VKT)	Verbale Produktivität	Jugendliche und Erwachsene	ca. 1 Stunde

⁹ Im Manual findet sich auch eine Angabe zur Split-Half-Reliabilität (Rel=0.95), jedoch erscheint uns zweifelhaft, ob durch eine einfache Testhalbierung beim TDK 4-6 zwei parallele Testhälften gebildet werden können.

Testgütekriterien: Die Objektivität der Testdurchführung und -auswertung ist beim KLT gewährleistet. Die **Reliabilität** der KLT-Mengenleistung ist gut; die Stabilitätskoeffizienten liegen bei sofortiger Retestung nach 1 Stunde bei 0.92, bei Retestung nach 8 Wochen zwischen 0.86 und 0.88. Die entsprechenden Retestreliabilitäten in bezug auf den Fehlerprozentsatz im KLT sind dagegen weniger gut; die Koeffizienten liegen hier zwischen 0.74 und 0.05 nach acht Wochen bzw. bei 0.47 nach sofortiger (1stündiger) Retestung. Allerdings ist der Übungsgewinn bei sofortiger Wiederholung des KLT beträchtlich, er beträgt zunächst (Retestung nach 1 Stunde) rund 25 Prozent und sinkt dann allmählich, bis er nach 2 Jahren praktisch völlig verschwunden ist (Angaben der Testautoren). Der **Standardmeßfehler** wird mit $s_E=3.3$ Z-Werten für die KLT-Mengenleistung angegeben.

Über die **Validität** des KLT ist - wie bei den meisten Tests dieser Art - wenig bekannt. Im allgemeinen korreliert die KLT-Leistung mit den Schulzensuren mäßig bis niedrig, am höchsten (0.4) mit der Mathematikzensur (besonders bei den jüngeren Probanden des 4. bis 6. Schuljahrs). Auch zum IST von Amthauer wurden keine engeren Beziehungen der KLT-Leistung ermittelt. Die auf Lehrerurteil basierenden Gültigkeitskoeffizienten liegen maximal bei $r=0.50$, durchschnittlich jedoch nur bei 0.25; bezüglich des Fehlerwertes F% wurden überwiegend negative Korrelationskoeffizienten gefunden. Wir selbst ermittelten bei 220 Gymnasiasten und Realschülern der 10. und 12. Klasse in der Relation von KLT-Leistung und LPS-Arbeitskurve 1+ (4 Min) relativ hohe Korrelationskoeffizienten zwischen 0.79 und 0.49 oder im Mittel $r=0.69$; in der Relation von KLT-Leistung und LPS-Arbeitskurve 1+2+ (8 Min) errechneten wir sogar Korrelationen zwischen 0.81 und 0.51 oder im Mittel $r=0.74$ (Heller 1970, S. 102f.; vgl. auch Heller, 1973).

Zum KLT werden diverse PR- und Standardwert- oder Z-Normen ($M=100$, $s=10$) angeboten, und zwar Altersnormen für männliche vs. weibliche Berufs- und Berufsfachschüler, Klassennormen für männliche vs. weibliche Volks- und Oberschüler sowie diverse Berufsgruppen-Normen für männliche vs. weibliche Lehrlinge und Anlernlinge. Für Mittel- bzw. Realschüler stehen keine Normen im Testmanual. Anhand einer Realschulstichprobe der 10. Klasse ($N=52$) errechneten wir deshalb folgende Statistiken: $M=110$ RP und $s=40$ RP für die KLT-Mengenleistung sowie $M=11$ Fehler für die KLT-Fehlervariable. Die beobachteten Differenzen bezüglich der KLT-Leistung und KLT-Fehlerprozent zwischen Jungen und Mädchen waren statistisch nicht zu sichern (a.a.O.).

Anwendungsbereiche: Der KLT ist ab der 4./5. Klasse zur Erfassung der "allgemeinen psychischen Leistungsfähigkeit", also der psychischen Aktivität oder "Willensstärke", der Anstrengungsbereitschaft, der Konzentration, Ausdauer u.ä. Eigenschaften der Sekundärintelligenz sensu Mierke (1963), geeignet. Sein Einsatz dürfte ohne größere Mühe auch in der Sonderschule, etwa bei Hör- und Sprachgeschädigten, möglich und informativ sein, wenngleich hierzu keine Normen vorliegen.

Testkritik: Während die Reliabilität des Faktors "Leistungsmenge" im KLT als gut bezeichnet werden kann, schwankt die Zuverlässigkeit der Fehler-Variablen teilweise beträchtlich und genügt somit nicht immer den teststatistischen Anforde-

rungen. Als Indikator der Konzentrationsleistung scheint der KLT brauchbar, wohingegen diagnostische Aussagen über die Arbeitsqualität nur unter Vorbehalt möglich sind. Hierzu bedarf es zusätzlicher Informationsquellen, etwa Verhaltensbeobachtung oder Exploration. Siehe auch Bartenwerfer (1983, S. 495f.) oder Kleber (1982).

- (2) **Der Pauli-Test.** Anweisung zur sachgemäßen Durchführung, Auswertung und Anwendung des Kraepelinschen Arbeitsversuches nach R. Pauli, herausgegeben von W. Arnold (1961³, 1975⁵) (TZ)

Testaufbau: Der Pauli-Test - eine 1936 von R. Pauli erstmals in standardisierter Form vorgestellte Variante des Kraepelinschen Arbeitsversuches aus den neunziger Jahren des vorigen Jahrhunderts - besteht aus einem Bogen voller einstelliger Zahlen, die kolonnenförmig angeordnet sind. Jeweils zwei übereinanderstehende Zahlen sind fortlaufend zu addieren und die letzte Stelle des Summenergebnisses zwischen die beiden Zahlen zu schreiben. Der Pauli-Test ist ein reiner Speedtest und dauert bei voller Durchführung 60 Minuten; alle 3 Minuten wird vom Probanden auf ein Glockenzeichen des Testleiters hin die zuletzt bearbeitete Addition markiert. Für die Auswertung und Interpretation interessieren somit u.a. die Gesamtzahl der Additionen, die Fehlerzahl und der Arbeitsverlauf (3 Minuten-Leistungseinheiten). Informationen über systematische Untersuchungen zur Standardisierung des Pauli-Tests finden sich in der Monographie von Arnold (a.a.O.).

Erfassungsdimensionen: Mit Hilfe des Pauli-Tests werden praktisch dieselben Merkmale wie beim KLT und ähnlichen Verfahren erfaßt, also Konzentration, Belastbarkeit, Ausdauer, Arbeitsorgfalt etc. Nach einer faktorenanalytischen Untersuchung zum Pauli-Test (unter dem Aspekt der "Leistungsmenge") von Bäumler (1964) wird vor allem die "Grundfähigkeit des rechnerischen Umgangs mit Zahlen" im Pauli-Test geprüft. Diese Interpretation dürfte jedoch nur für jüngere bzw. rechnerisch wenig geübte Probanden zutreffen.

Testgütekriterien: Die **Objektivität** der Testdurchführung und Testauswertung ist gesichert. Die Angaben zur **Reliabilität** des Pauli-Tests schwanken, zumal auch hier wieder ein mehr oder minder großer Übungs- und Lerngewinn veranschlagt werden muß. So fand Plössl (1941, zit. nach Pawlik, 1970, S. 196) "für einen Index aus Menge und Güte eine Retest-Korrelation von 0.96 (Abstand ein Monat). Für andere Indizes ist die Zuverlässigkeit geringer. Z.B. haben Bartenwerfer und Plössl für die Fehlerzahl Koeffizienten von 0.7 bis 0.8 gefunden". Noch dürftiger sind die Angaben zur **Validität** des Pauli-Tests (vgl. Bartenwerfer, 1983, S. 485ff.).

Günstiger als die teststatistischen Angaben zu den Gütekriterien der Reliabilität stellt sich hier das **Normen-Problem** dar. Die Monographie von Arnold enthält beispielsweise rund zwei Dutzend Pauli-Testnormen, teilweise auf unterschiedliche Untersuchungsbedingungen bezogen. In der diagnostischen Praxis wird man trotzdem sehr oft auf die Erstellung populationspezifischer Normen angewiesen sein,

worauf auch Bartenwerfer (1983, S. 488) hinweist. Für Untersuchungen bei Hörschädigten werden Pauli-Testnormen von Heller (1973, S. 242ff.) mitgeteilt.

Testkritik: Gegenüber vergleichbaren anderen Verfahren bietet der Pauli-Test u.a. den Vorteil, detaillierte Verlaufsanalysen wichtiger quantitativer und qualitativer Eigenschaften des Arbeits-Leistungsverhaltens bereits vom Grundschulalter an zu ermöglichen. Relativ ausgedehnte Verhaltensstichproben (Bearbeitungszeit 1 Stunde) garantieren im allgemeinen hinreichend zuverlässige Aussagen über die Arbeitsfähigkeit und Leistungsqualität eines Probanden, wengleich die qualitative Analyse auch hier - wie bei allen vergleichbaren Verfahren - die meisten Probleme in sich birgt.

(3) Der Aufmerksamkeits-Belastungs-Test (Test d2) von R. Brickenkamp (1962, 1981⁷) (Hogrefe)

Der d2 stellt eine Weiterentwicklung des Durchstreichtests dar, den Bourdon um die Jahrhundertwende entwickelte und der in der Folge Vorbild für eine Reihe von Konzentrationstests wurde. Ursprünglich zur Überprüfung der Fahrtauglichkeit entwickelt, dürfte der d2 im Rahmen der Schulberatung heute einer der verbreitetsten Konzentrationstests sein.

Testaufbau: Der d2 besteht aus 14 Zeilen mit jeweils 47 Zeichen (d oder p). Jedes dieser Zeichen ist mit 0 bis 3 Strichen versehen, die sich über und unter dem Zeichen befinden. Die Probanden müssen alle "d" anstreichen, die mit 2 Strichen versehen sind, egal ob beide Striche oberhalb, unterhalb oder je einer ober- und unterhalb gedruckt ist.

Erfassungsdimensionen: Der d2 mißt nach Brickenkamp (1975) die Schnelligkeit und Genauigkeit bei der Unterscheidung ähnlicher visueller Reize, auch wenn der Autor im Manual hervorhebt, daß für die praktische Verwendung eines Tests die Kenntnis der Gütekriterien wichtiger sei als die theoretische Fundierung. Die erfaßten Aspekte der Konzentrationsleistung sind die Bearbeitungs- und Lösungsmenge (korrekte Bearbeitungen), die Fehlermenge und die Schwankungsbreite der konzentrativen Leistung.

Handhabung: Der d2 kann als Einzel- oder Gruppentest administriert werden, eine Parallelform liegt nicht vor. Für jede Zeile ist 20 Sekunden Zeit, so daß die gesamte Testdurchführung inklusive Instruktion etwa 8 Minuten beansprucht (reine Testzeit: 4 Minuten und 40 Sekunden). Die Auszählung der Fehler (Auslassungen und falsche Anstreichungen) erfolgt mittels Schablonen. Die Gesamtzahl der bearbeiteten Zeichen wird mit GZ (Anzahl der Zeichen bis zum letzten angestrichenen), die Anzahl der Fehler mit F, der Prozentanteil der Fehler an der Gesamtleistung GZ mit F% sowie die Gesamtzahl der korrekt bearbeiteten Zeichen mit GZ-F bezeichnet. Die Schwankungsbreite SB bestimmt sich als Differenz der Anzahl bearbeiteter Zeichen der Zeilen mit der höchsten und geringsten Mengenleistung. In das Manual wurden auch spezielle Auswertungshinweise zu Übungs-

effekten bei Testwiederholung, Erfahrungen bei der Diagnose von Schizophrenen sowie Auswirkungen von Umwelteinflüssen auf die d2-Leistung aufgenommen.

Gütekriterien: Aufgrund der standardisierten Durchführung und Auswertung sowie der Normtabellen kann die **Objektivität** des d2 grundsätzlich als gesichert gelten. Allerdings haben kürzlich Oehlschlägel & Moosbrugger (1991) sowie Berg & Erlwein (1991) darauf hingewiesen, daß es zu Fehldiagnosen kommen kann, falls Probanden Zeichen ohne durchzusehen überspringen oder willkürlich anstreichen bzw. am Ende der Bearbeitungszeit jeder Zeile weit rechts noch ein Zeichen anstreichen. Dadurch kann der GZ-Wert viel zu hoch ausfallen. Eine hohe Zahl von Auslassungsfehlern bei niedriger Zahl von Verwechslungsfehlern stellt daher nach Berg & Erlwein (1991) ein Indiz für flüchtiges Arbeiten dar, während viele Auslassungs- und Verwechslungsfehler für einen eher zufälligen Anstreichmodus sprechen. Oehlschlägel & Moosbrugger (1991) schlagen aus ähnlichen Überlegungen heraus eine Korrektur des Wertes GZ-F vor. Da der korrigierte Wert jedoch hoch mit dem einfachen GZ-F-Wert korreliert, kann man nach Brickenkamp (1991a) darauf verzichten. Brickenkamp (1991b) betont, daß die Auswertung des d2 jedenfalls nicht mechanisch erfolgen dürfe, sondern einen kompetenten Beurteiler erfordere.

Die **Konsistenzreliabilität**, berechnet nach der Zeitfraktionierungstechnik, beträgt nach Angaben im Manual für die Scores GZ und GZ-F um bzw. über 0.9 (hoch), auch die meisten **Stabilitätskoeffizienten** (Retestung nach 5 Stunden bis 1 Jahr) liegen zwischen 0.7 und 0.9. Niedrigere, teilweise sogar unbrauchbare Zuverlässigkeitsschätzungen ergaben sich hingegen für F% und besonders für SB (die meisten mitgeteilten Koeffizienten von SB < 0.3). Dingel (1971) und Eser (1987) konnten diese Angaben bei Kraftfahrereignungsuntersuchungen bzw. für Kinder und Jugendliche in etwa bestätigen. Rauchfleisch (1983) konnte zeigen, daß der d2 auch in der Anwendung bei Psychiatriepatienten hochreliabel ist. Es sei angemerkt, daß die Fehlerzahl bzw. F% im d2 mit dem GZ- oder GZ-F-Wert kaum korreliert, so daß langsames Arbeiten nicht einfach mit geringer Fehlerzahl einhergeht und umgekehrt.

Korrelationsstudien zur **Übereinstimmungsvalidität** belegen, daß die im d2 erfaßten Leistungen kaum (HAWIE, IST) oder nur gering (BTS) mit Intelligenztestleistungen zusammenhängen. Dagegen scheinen solche Zusammenhänge bei psychiatrischen Patienten höher zu sein (z.B. Rauchfleisch, 1983). Zwischen d2 und dem K-L-T, dem Pauli-Test, dem KVT und anderen Konzentrationsverfahren ergaben sich in der Regel mittlere Zusammenhänge. Während der GZ-Wert mit dem Lehrerurteil über Aktivität in Leistungssituationen korrespondiert, konnten für F% Zusammenhänge mit dem Lehrerurteil über Willenskraft und Ausdauer gefunden werden.

Bewährungskontrollen zeigten, daß signifikante Zusammenhänge zwischen den d2-Leistungen und der Kraftfahrtauglichkeit sowie der Leistung von Berufsgruppen wie Fernsehprüferinnen oder Datentypistinnen bestehen. Rauchfleisch (1983) wies mit Diskriminanzanalysen den Wert des d2 zur Differentialdiagnose von Störungen hirnorganischer und neurotischer Art nach.

Die Normen basieren auf einer Stichprobe von 3132 Schülern und 3132 Erwachsenen. Eigene Normtabellen finden sich für 9- bis 15jährige Volks-, 15- bis 18jährige Berufs- und 11- bis 18/19jährige Oberschüler(innen) im 2jährigen (für Berufsschulen im 3jährigen) sowie für Erwachsene im 10jährigen Abstand. Wegen der geringeren bzw. niedrigen Zuverlässigkeit der F%- und SB-Werte ermittelte der Testautor hierfür nur Prozenträge bzw. noch gröbere Perzentilnormen.

Kritik: Der d2 erfaßt u.E. zuverlässig einen bestimmten Aspekt konzentrativer Leistung, der jedoch nicht mit dem Begriff Konzentration, wie er alltagssprachlich - auch unter Lehrern - verwendet wird, verwechselt werden darf. Im Hinblick auf den Einsatz in der Schulberatung wären weitere Validierungsstudien, insbesondere zur prognostischen Validität, wünschenswert.

Ein generelles Problem von Konzentrationstests stellt deren **ökologische Validität** dar. Dies gilt insbesondere für den Einsatz in der Schulberatung, vorab bei älteren Schülern. So kann die Konzentrationsleistung ein und desselben Schülers je nach Unterrichtsfach und auch in Abhängigkeit von der Intelligenz bzw. vom intellektuellen Anforderungsniveau des Lerninhalts stärker variieren, weshalb auf Verhaltensbeobachtungen im Unterricht (als Informationsquelle für die Beurteilung des Aufmerksamkeits- oder Konzentrationsverhaltens) in den seltensten Fällen verzichtet werden kann.

- (4) Die vom Deutschen Institut für Internationale Pädagogische Forschung in Frankfurt/M. herausgegebenen Konzentrationstests für das Vor- und Grundschulalter
 - a) Der Frankfurter Tests für Fünfjährige - Konzentration - (FTF-K) von U. Raatz & R. Möhling (1971) (Beltz)
 - b) Der Konzentrationstest für das erste Schuljahr (KT 1) von R. Möhling und U. Raatz (1974) (Beltz)
 - c) Der Konzentrationstest für 3. und 4. Klassen (KT 3-4) von R. Heck-Möhling (1986) (Beltz)

Testaufbau: Bei allen drei Verfahren handelt es sich um Durchstreichtests, wobei beim FTF-K und KT 1 die Kinder aus 14 Reihen mit je 12 Zeichnungen von Birnen und Äpfeln die Birnen herausuchen und anstreichen müssen. Beim KT 3-4 sind Würfel auf Übereinstimmung mit vier vorgegebenen Musterwürfeln zu überprüfen und identische wieder anzustreichen.

Erfassungsdimensionen: Erfasst wird Konzentration als ein "aktiver, zielgerichteter, gewollter und bewußter, über eine bestimmte Zeit hin anhaltender psychischer Vorgang" (Heck-Möhling, 1986, S. 4).

Handhabung: Alle drei Verfahren sind auch für die Hand des Lehrers bestimmt. Durchführung und Auswertung sind entsprechend gut verständlich beschrieben. Die

Durchführung des FTF-K beansprucht 10 Minuten, die des KT 5 Minuten, während für den KT 3-4 laut Manual eine Schulstunde zu veranschlagen ist (reine Testzeit 20 Minuten). Wie beim d2 werden ein GZ-Wert, die Anzahl der richtigen Lösungen sowie Fehlerwerte (Auslassungen und falsche Anstreichungen) bestimmt. In den Manualen finden sich detaillierte Interpretationshinweise.

Gütekriterien: Die Objektivität der Verfahren scheint durch Standardisierung von Durchführung und Auswertung gesichert. Die Zuverlässigkeit der Tests kann als befriedigend bezeichnet werden, auch wenn beim KT 3-4 die Berechnung der Split-Half-Reliabilität mit Bildung der Testhälften durch gerade und ungerade Items einen schweren methodischen Fehler darstellt. Die so ermittelten Werte von $Rel_x=0.99$ sind deshalb unbrauchbar.

Neben der inhaltlichen Gültigkeit wurde für den KT 3-4 die **Konstruktvalidität** durch Korrelation mit anderen Konzentrationstests sowie den AzN 4+ bestimmt. Es ergaben sich durchweg niedrige bis mittlere Zusammenhänge mit den Scores von d2 und KLT, keine bzw. niedrige Korrelationen mit den AzN 4+.

Normen liegen für das erste und zweite Halbjahr des ersten Schuljahres sowie das erste und zweite Halbjahr des dritten und vierten Schuljahrs vor (hier auch geschlechtsspezifisch). Für die Fünfjährigen wurde nur eine Grobklassifikation der Leistungen vorgenommen.

(5) Der Konzentrations-Verlaufstest (KVT) von D. Abels (1961², 1974³) (Hogrefe)

Testaufbau: Beim KVT müssen die Probanden 60 auf der Rückseite nummerierte Kärtchen, die mit jeweils 36 zweistelligen Zahlen bedruckt sind, auf vier Stapel danach sortieren, ob die Zahl 43, die Zahl 63, sowohl 43 als auch 63 oder keine der beiden Zahlen unter den 36 aufgedruckten enthalten sind.

Erfassungsdimensionen: Der KVT stellt eine Modifikation des Zahlensortierversuchs von Münsterberg dar, der in den 20er Jahren am Leipziger Psychologischen Institut verwendet wurde. Neben Rechen- (KLT, Pauli-Test) und Durchstreichaufgaben (d2) stellt die Sortieraufgabe des KVT demnach ein drittes Paradigma zur Erfassung von Konzentration dar.

Gütekriterien: Die Meßgenauigkeit des KVT ist nach Brickenkamp (1975) als gering zu werten. Zwar teilt der Autor keine empirischen Gültigkeitskontrollen mit, doch liegt eine Reihe von Untersuchungen vor, die die Validität des KVT belegen (vgl. zusammenfassend Brickenkamp, 1975, S. 228ff., sowie Bartenwerfer, 1983, S. 492ff.). Die Normierungsstichprobe ist zwar groß, jedoch fehlen Angaben zu ihrer Zusammensetzung, so daß die Normen eigentlich unbrauchbar sind. Für hörgeschädigte Schüler hat Heller (1973, S. 239ff.) KVT-Normen publiziert, ebenso zum Bourdon-Figuren-Durchstreichtest (1973, S. 238).

Tabelle 12: Übersicht über Tests zur Erfassung von Konzentration

Autor und Erscheinungsjahr	Testbezeichnung	Erfassungsdimensionen	Alters- und Zielgruppe	Zeitbedarf
B. Bourdon (1955)	Bourdon-Test	Aufmerksamkeit und Konzentration	keine Angabe	keine standardisierte Testvorgabe
E.W. Kleber & G. Kleber (1974)	Differentieller Leistungstest - KE (DL-KE)	Konzentrationsleistung und Belastbarkeit	5;7 bis 6;6 Jahre	25 Minuten
E.W. Kleber & G. Kleber (1975)	Differentieller Leistungstest - KG (DL-KG)	Konzentrationsleistung und Belastbarkeit	6;6 bis 10;5 Jahre bzw. Klassen 1 bis 4	21 Minuten ohne Instruktion
U. Raatz & R. Möhling (1971)	Frankfurter Tests für Fünfjährige - Konzentration - (FTF-K)	Konzentration	5;0 bis 6;0jährige Vorschüler	10 Minuten
R. Möhling & U. Raatz (1974)	Konzentrations-test für das erste Schuljahr (KT 1)	Konzentration	Erstkläßler	5 Minuten
R. Heck-Möhling (1986)	Konzentrations-test für 3. und 4. Klassen (KT 3-4)	Konzentration	Schüler der 3. und 4. Klasse	maximal 1 Schulstunde (reine Testzeit: 20 Minuten)
H. Düker & G. Lienert (1965 ²)	Konzentrations-Leistungs-Test (KLT)	Koordination, Konzentration	ab 4. Schuljahr	knapp 40 Minuten
D. Abels (1961 ² , 1974 ³)	Konzentrations-Verlaufs-Test (KVT)	Konzentration	keine Angabe	keine Angabe, etwa 20 Minuten
W. Arnold (1961 ³ , 1975 ⁵)	Pauli-Test	Konzentration, Leistungsverlauf	ab 7. Lebensjahr (Addieren muß beherrscht werden)	60 Minuten
G. Marschner (1972, 1980)	Revisions-Test	anhaltende Konzentration bei Tempoarbeit	10-75 Jahre	15 Minuten
R. Brickenkamp (1962, 1981 ⁷)	Test d2. Aufmerksamkeits-Belastungs-Test	Schnelligkeit und Genauigkeit visueller Reiz-Diskrimination	9 bis 60 Jahre	8 Minuten

3.4.2 Motivationstests

(1) Der Anstrengungsvermeidungstest (AVT) von B. Rollett & M. Bartram (1977, 1981²) (TZ)

Testaufbau: Der AVT enthält 41 Feststellungen, die von den Schülern mit "stimmt" oder "stimmt nicht" beantwortet werden sollen. 20 Items bilden eine sog. Anstrengungsvermeidungsskala, 10 Items eine Pflichteiferskala (beide Skalen sind Rasch-homogen). Die 11 weiteren Items, die sich nicht Rasch-skalisieren ließen, sollen Hinweise für therapeutische Maßnahmen liefern.

Erfassungsdimensionen: Der AVT soll die Anstrengungsvermeidungstendenz erfassen. Diese Verhaltensdisposition äußert sich nach Rollett & Bartram (1981) bei Schülern vor allem darin, sich Leistungsanforderungen generell zu entziehen, aber auch darin, Aufgaben und Probleme ökonomisch zu lösen. Daneben wird Pflichteifer erfaßt.

Handhabung: Der AVT kann als Einzel- oder Gruppentest durchgeführt werden. Die Auswertung erfolgt mittels Schablone und Normtabellen im Manual. Im Manual finden sich auch detaillierte Hinweise zur Indikation und Durchführung therapeutischer Maßnahmen anhand von Fallbeispielen.

Gütekriterien: Die Objektivität des Verfahrens ist durch die standardisierten Durchführungs- und Auswertungsrichtlinien gesichert. Trotz des Nachweises der Rasch-Homogenität der AVT-Skalen wurde eine Schätzung der Reliabilität nach der Konsistenzmethode vorgenommen: Es ergaben sich eher niedrige Koeffizienten von 0.8 für die Anstrengungsvermeidungsskala und 0.69 für die Pflichteiferskala.

Untersuchungen zur Validität ergaben niedrige bis mittlere Zusammenhänge mit Lern- und Behaltenstests, eine negative Korrelation von $r = -0.33$ mit der Leistungsmotivation sowie niedrige Zusammenhänge mit Intelligenzvariablen (< 0.3). Ähnliche Ergebnisse fanden sich in bezug auf Schulnoten, während elterliche Strenge und Anstrengungsvermeidung zwischen 0.22 und 0.38 korrelieren.

Die Normierung erfolgte an einer repräsentativen Stichprobe von 3093 Schülern der 5. bis 9. Klasse, wobei die Klassen 5 und 6 sowie 7 bis 9 zusammengefaßt wurden. Für die Anstrengungsvermeidungsskala wurden auch geschlechts- und schulartspezifische Normen berechnet.

Kritik: Der AVT kann wichtige Informationen im Rahmen einer Einzelfallberatung liefern, jedoch sollten neben der eher niedrigen Reliabilität und der ungünstigen Verteilungseigenschaften der Skalen bei der Interpretation auch Schwächen des Verfahrens in bezug auf die theoretische Fundierung berücksichtigt werden. Kühn (1982) und insbesondere Jopt (1982) kritisieren die mangelnde Einordnung des Konstrukts "Anstrengungsvermeidung" in eine Theorie der Lernmotivation (vgl. auch die Kontroverse zwischen Bartram & Rollett, 1984, und Jopt, 1984). Darüber hinaus erscheint es fragwürdig, ab Prozentrang 75 eine Therapie als hilfreich, ab Prozentrang 95 als notwendig zu bezeichnen. Die Verknüpfung von Diagnose und Intervention im Manual hebt das Verfahren jedoch von üblichen Motivationstests ab.

(2) Der Attribuierungsfragebogen für Erfolg und Mißerfolg in der Schule für 5. bis 7. Klassen (AEM 5-7) von H. Widdel (1977a) (Beltz)

Testaufbau: Beim AEM 5-7 müssen die Schüler für 14 Erfolgs- (+) und 14 Mißerfolgssituationen (-) auf einer 5stufigen Skala ("trifft ganz zu" bis "trifft gar nicht zu") einschätzen, inwiefern die Faktoren Begabung (B), Anstrengung (A), Aufgabenschwierigkeit (S) und Zufall (Z) für Erfolg bzw. Mißerfolg als Ursachen in Frage kommen. Es ergeben sich somit acht Skalen: B+ = Attribuierung von Erfolg auf Begabung, B- = Attribuierung von Mißerfolg auf mangelnde Begabung, A+ = Attribuierung von Erfolg auf Anstrengung etc.

Erfassungsdimensionen: Mit dem AEM 5-7 sollen Attribuierungstendenzen im Sinne von Weiner et al. (1971) erfaßt werden (vgl. Tabelle 13).

Tabelle 13: Attributionsmodell nach Weiner et al. (1971)

Stabilität	Personabhängigkeit	
	internal	external
stabil	Begabung	Aufgabenschwierigkeit
variabel	Anstrengung	Zufall

Handhabung: Die Durchführung als Gruppentest wird empfohlen, wobei etwa eine Schulstunde (30-45 Minuten) veranschlagt werden müssen. Die Auswertung erfolgt mittels Schablonen und Normtabellen im Manual, wobei alle Schritte vom Testautor eingehend erläutert werden. Auch zur Interpretation finden sich ausführliche Hinweise im Manual, jedoch kein Fallbeispiel.

Gütekriterien: Die **Objektivität** des AEM 5-7 kann als gegeben gelten. Die **Reliabilitäten** (Konsistenzschätzungen) liegen bis auf Z- zwischen 0.7 und 0.8 (noch ausreichend). Aspekte der **Konstruktvalidität** werden durch Korrelationsstudien der AEM-Skalen mit Variablen des schulischen Leistungsmotivs (niedrige bis mittlere Zusammenhänge), Fremdbeurteilungen des Leistungsverhaltens (eher niedrige Zusammenhänge), Schulangst (niedrige Zusammenhänge), Selbstkonzept und Einstellung zur Schule (niedrige bis mittlere Zusammenhänge), Intelligenz (niedrige bis mittlere Zusammenhänge) und Schulleistungen (niedrige bis mittlere Zusammenhänge) untersucht, wobei die Korrelationen für Hauptschüler jeweils etwas höher ausfallen.

Die **Normierung** des AEM 5-7 erfolgte um die Jahreswende 1976/1977 an 2346 Schülern aus 7 westdeutschen Bundesländern. Neben T-Werten finden sich T-Wertbänder und Prozentrangbänder, so daß bei der Auswertung nicht eigene Vertrauensintervalle berechnet werden müssen.

Kritik: Mit dem AEM 5-7 liegt ein Verfahren vor, durch das wichtige Informationen im Rahmen der Schulberatung, insbesondere der Individualberatung gewonnen werden können. In diesem Kontext kann der AEM auch zur Entscheidung über bzw. Planung von Fördermaßnahmen herangezogen werden.

Tabelle 14: Übersicht über Tests zur Erfassung von Motivation

Autor und Erscheinungsjahr	Testbezeichnung	Erfassungsdimensionen	Alters- und Zielgruppe	Zeitbedarf
B. Rollett & M. Bartram (1977, 1981 ²)	Anstrengungsvermeidungstest (AVT)	Aspekte der Leistungsbereitschaft: Anstrengungsvermeidung, Pflichteifer	5.-9. Klassen	ca. 20 Minuten
J.W.L. Wagner (1977a)	Fragebogen Einstellung zur Schule für 4. bis 6. Klassen (FES 4-6)	Verschiedene Einstellungsdimensionen gegenüber Schule, Lehrer und Lernen	4.-6. Klassen	ca. 30 Minuten
H. Widdel (1977a)	Fragebogen zur Attribuierung von Erfolg und Mißerfolg in der Schule für 5. bis 7. Klassen (AEM 5-7)	Attribuierungsmuster (internal vs. external, stabil vs. variabel)	5.-7. Klassen	1 Schulstunde
H. Widdel (1977b)	Fragebogen zur Erfassung des schulischen Leistungsmotivs für 5. bis 7. Klassen (FSL 5-7)	Leistungsmotivation	5.-7. Klassen	25 Minuten
G. Krampen (1981)	IPC-Fragebogen zu Kontrollüberzeugungen ("Locus of Control")	Locus of Control: internale vs. externale Kontrollüberzeugung	Erwachsene ab 18 Jahren	10-15 Minuten
H. Schmalt (1976)	Das LM-Gitter. Ein objektives Verfahren zur Messung des Leistungsmotivs bei Kindern	Leistungsmotivation (v.a. Hoffnung auf Erfolg, Furcht vor Mißerfolg)	3.-5. Klassen	ca. 45 Minuten
H.J.M. Hermans (1976) (Deutsche Fassung von U. Undeutsch)	Leistungsmotivationstest für Jugendliche (LMT-J)	Komponenten der Leistungsmotivation	12-16 Jahre	20-50 Minuten

3.4.3 Interessentests

(1) Der Berufs-Interessen-Test II (BIT II) von M. Irle & W. Allehoff (1984) (Hogrefe)

Der erstmals 1955 veröffentlichte BIT dient der Erkundung beruflicher Interessenrichtungen bei älteren Schülern (etwa ab dem 14./15. Lebensjahr) sowie Jugendlichen und Erwachsenen. Ursprünglich mehr für die Berufsberatung (z.B. im Rahmen der Berufseignungsfindung) konzipiert, hat der BIT im Laufe der Jahre zahlreiche Erprobungen auch im schulischen Bereich - mit unterschiedlichem Erfolg (vgl. Heller 1970, 1973) - erfahren. Inzwischen ist 1984 eine überarbeitete Version des Verfahrens als BIT II erschienen, die sich in einigen Punkten vom alten BIT unterscheidet, ohne die wesentlichen Prinzipien aufzugeben. Da für den BIT II eine Reihe von Items neu formuliert wurde, sollte die alte Form nicht mehr angewendet werden.

Testaufbau: Der BIT II ist in der Form eines vollstandardisierten Fragebogens angelegt und liefert ipsative Werte (siehe unten) über verschiedene (Berufs-)Interessenrichtungen. Der Fragebogen besteht aus einem (früher zwei) Bogen mit je 100 Kästchen oder Testitems, die in 10er Reihen bzw. Kolonnen angeordnet sind, z.B. "Druckfehler in Zeitungsartikeln berichtigen", "Pflastersteine für den Straßenbelag behauen", "In Zügen Fahrkarten der Reisenden prüfen", "Eine Segeljacht steuern" (vgl. auch Abbildung 23). Der Proband soll nun an jedem der 81 fortlaufend nummerierten (im abgebildeten Beispiel blanc gelassenen) Schnittpunkte ("Straßenkreuzung" eines Viererblocks) die von den vier angebotenen Tätigkeiten jeweils **bevorzugte** mit einem "+" und die **am wenigsten zusagende** mit einem "-" markieren. Der Proband wird also vor eine **relative Entscheidung** gestellt und muß sich immer sowohl positiv, als auch negativ für **eine** der vier Wahlmöglichkeiten entscheiden (Wahlzwang bzw. Forced-Choice).

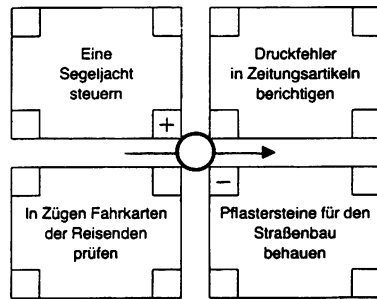


Abbildung 23: Beispielitems zum BIT II

"Der Wahlzwang bei den einzelnen Items zieht einige Folgen nach sich, die wir betrachten müssen, um die Resultate ... sinnvoll interpretieren zu können. Weil der Proband seine Vorlieben für Tätigkeiten, die verschiedenen Testkategorien angehören, irgendwie in eine Rangordnung bringen muß, kennzeichnen seine Gesamtwerte in erster Linie natürlich nur die relativen Positionen in bezug auf die verschiedenen Interessensvariablen.

Solche relativen Angaben von Ausprägungsgraden nennt man **ipsativ** (Cattell, 1944). Die Meßwerte gewöhnlicher Art, die nicht aus Items mit Wahlzwang hergeleitet werden, zeigen solche Ausprägungsgrade an, die sich um den Mittelwert der allgemeinen Population verteilen. Sie heißen daher **normative Werte** ... Weiter folgt aus dem Wahlzwang, daß die Durchschnittsergebnisse bei jeder Person etwa gleich sind. Das trifft natürlich nicht den wirklichen Tatbestand der Dinge... Ein Nachteil des ganzen Verfahrens

ist der, daß das Profil nicht solche Wesenszüge angemessen wiedergibt, die ausschließlich stark oder ausschließlich schwach sind. Denn zweifellos gibt es Menschen, deren Interessen überwiegend stark ausgeprägt sind, und andere, die durchgehend nur mäßig beruflich interessiert sind. Informationen hierüber sind aber oft sehr wichtig" (Guilford, 1965, S. 203ff.).

Der BIT II kann neben dieser "Forced-Choice"-Form (zwei Wahlzwang-Parallelformen AA und AB) in einer "Free-Choice"-Form (Parallelformen BA und BB) durchgeführt werden, bei der die Probanden zu jedem der 81 Items auf einer 5stufigen Skala angeben können, wie gerne sie die Tätigkeit ausüben würden.

Erfassungsdimensionen: Im BIT II werden individuelle Interessenausprägungen in bezug auf folgende Tätigkeitsfelder erfaßt: **TH** = Technisches Handwerk, **GH** = Gestaltendes Handwerk, **TN** = Technische und naturwissenschaftliche Berufe, **EH** = Ernährungshandwerk, **LF** = Land- und forstwirtschaftliche Berufe, **KB** = Kaufmännische Berufe, **VB** = Verwaltende Berufe, **LG** = Literarische und geisteswissenschaftliche Berufe, **SE** = Sozialpflege und Erziehung. Diese 9 Berufsinteressenrichtungen repräsentieren je 9 berufliche Einzelstätigkeiten.

Handhabung: Für die Gruppen- oder Wiederholungstestung stehen sowohl für die ipsative als auch die normative Fragebogen-Version zwei Testformen zur Verfügung. Die BIT-Auswertung geschieht mit Hilfe von Auswertungsschablonen und liefert entweder ipsative Scores (Forced-Choice-Versionen AA und AB) oder normative Werte (Free-Choice-Versionen BA und BB) zu 9 verschiedenen Berufsinteressenrichtungen (siehe Erfassungsdimensionen). Für den BIT II sind keine Zeitgrenzen festgelegt, die durchschnittliche Bearbeitungszeit wird mit 20-80 Minuten für die Forced-Choice-Version und mit 10-40 Minuten für die Free-Choice-Form angegeben.

Testgütekriterien: Der BIT II verfügt sowohl hinsichtlich seiner Durchführung als auch bezüglich der Auswertung über eine gute **Objektivität**. Angaben zur **Reliabilität** fehlen im Manual. Der Zweitautor nennt in seiner Monographie (Allehoff, 1985) Korrelationen zwischen den Testformen zwischen 0.7 und 0.9 für die Forced-Choice-Form und 0.6 bis 0.75 für die Free-Choice-Form.

Die Interkorrelationen der 9 Interessenrichtungen tendieren für die Formen AA und AB leicht zu negativen Werten, was Irle & Allehoff (1984) mit der Abhängigkeit der Skalen interpretieren (Forced-Choice-Verfahren). Weitergehende Angaben zur **Validität** des BIT II enthält die Handanweisung nicht. Bei Allehoff (1985) finden sich Ergebnisse von Faktorenanalysen (auch geschlechtsspezifisch). Für beide Testversionen ergab sich eine Drei-Faktorenlösung, bei Aufteilung nach Geschlecht mußte auf Zwei-Faktorenlösungen zurückgegriffen werden. Für die Free-Choice Version ergab sich für alle Probanden ein Faktor kaufmännisch/verwaltungsberuflicher Interessen, ein Faktor technisch/naturwissenschaftlich/land- und forstwissenschaftlicher Interessen und ein Faktor mit den übrigen Interessengebieten. Für die Jungen ergab sich ein Faktor mit technisch/naturwissenschaftlichen Interessen und ein zweiter mit den übrigen Gebieten, für die Mädchen ein Faktor mit technisch/naturwissenschaftlichen und kaufmännisch/verwaltungsberuflichen Interessen und ein zweiter mit den restlichen Interessengebieten.

Die große Bedeutung der Interessenrichtungen für die Berufswahl belegt Allehoff (1985) mit einer Reihe von Strukturgleichungsmodellen. Dabei ergab sich, daß auch fehlende Interessen für die Berufswahl von Bedeutung zu sein scheinen: Z.B. scheint die Bereitschaft zum Ergreifen eines Berufs des sozialen bzw. Erziehungsbereichs von der Ablehnung kaufmännischer Interessen abzuhängen.

Als Normen zum BIT II liegen für alle Formen geschlechtsspezifische Prozentränge für jede der 9 Interessenkategorien vor, so daß für jeden einzelnen Probanden ein Profil der Berufsinteressen erstellt werden kann. Die Normen beruhen auf den BIT II-Ergebnissen von jeweils mehr als 2000 weiblichen und männlichen Jugendlichen aus allen Schularten (9. bis 13. Klasse). Scores unter PR = 15 bzw. 5 werden als niedrige bzw. sehr niedrige Interessen, Scores über PR = 85 bzw. 95 als hohe bzw. sehr hohe Interessen interpretiert, während mittlere Scores zwischen dem 15. und 85. Prozentrang durchschnittliche Interessenausprägungen indizieren. Die Autoren weisen ausdrücklich darauf hin, daß aufgrund der BIT II-Ergebnisse allein keine Klassifikations- und schon gar keine Selektionsentscheidung getroffen werden könne, sondern die Testresultate im Einzelfall als ein Ergebnis im Rahmen einer umfangreicheren diagnostischen Untersuchung eine Rolle spielen könnten. Interessenprofile für einzelne Berufsgruppen haben die Autoren in den BIT II nicht mehr aufgenommen.

Testkritik: Als "direkte" Methode vermittelt der BIT II nur dann zuverlässige und gültige Informationen, wenn sichergestellt ist, daß der Proband gemäß der Instruktion sorgfältig und ehrlich seine Wahlen trifft und in etwa schon eine Vorstellung über die gefragten Tätigkeiten besitzt. Unter diesen Voraussetzungen eignet sich der BIT II zur Erfassung berufsrelevanter Interessenrichtungen bei Haupt- und Berufsschülern, eventuell auch bei Realschülern, wohingegen er hinsichtlich seiner Iteminhalte weniger gut die Bildungs- und Berufsinteressen bzw. Berufsmöglichkeiten von Gymnasiasten und Studenten trifft. Diese Kritik schränkt die Anwendung des BIT II z.B. bei der Schul- und Studieneignungsermittlung ein.

Neue interessante Anwendungsaspekte ergeben sich gegenüber dem alten BIT durch die Formen BA und BB (Free-Choice-Version), mit denen nun zu denselben Interessenbereichen wie bisher auch normative Kennwerte für Schüler ermittelt werden können. Insofern kann mit dem BIT II nicht nur die ipsative Struktur der Interessen, sondern auch die Höhe der Interessen im Vergleich zur Normpopulation abgeschätzt werden.

Leider wird der Wert des BIT II dadurch eingeschränkt, daß im Manual so gut wie keine Ergebnisse zur Zuverlässigkeit und Validität des Verfahrens mitgeteilt werden.

(2) Der Differentielle Interessentest (DIT) von E. Todt (1967, 1972²) (Huber)

Testaufbau: Der DIT besteht aus 390 Items, die auf vier Materialbereiche (Tätigkeiten, Berufe, Bücher, Zeitschriften) und 11 Interessenrichtungen (s.u.) verteilt

sind. Die Probanden geben zu jedem Item anhand einer 5stufigen Skala (sehr gern bis sehr ungern) ein Urteil über ihre Zu- bzw. Abneigung ab.

Erfassungsdimensionen: Todd (1967) versteht Interessen "als Verhaltens- oder Handlungstendenzen, die variieren auf den Dimensionen der Stabilität, der Generalität, der Intensität (bipolar: Zuneigung, Abneigung) und die qualitativ gekennzeichnet sind durch ihre Richtung" (S. 9). Im DIT werden die Interessenrichtungen Sozialpflege und Erziehung (SE), Politik und Wirtschaft (PW), Verwaltung und Wirtschaft (VW), Unterhaltung (UN), Technik und exakte Naturwissenschaft (TN), Biologie (BI), Mathematik (MA), Musik (MU), Kunst (KU), Literatur und Sprache (LS) sowie Sport (SP) berücksichtigt. Neben berufsbezogenen sollten mit dem DIT auch Freizeitinteressen erfaßt werden. Zudem sollten mehrere "Materialarten" herangezogen werden, an denen sich Interessen manifestieren können" (S. 9).

Handhabung: Zielgruppe des DIT sind Realschüler, Gymnasiasten und Wirtschaftsgymnasiasten von 15-20 Jahren. Der Test kann als Einzel- oder Gruppenverfahren administriert werden. Eine Kurzform berücksichtigt nur den Materialbereich Tätigkeiten. Die Durchführungsdauer beträgt etwa 30-45 Minuten (Kurzform: 8-15 Minuten), die Auswertungszeit 8-10 Minuten (Kurzform 2-3 Minuten). Hierzu werden die Rohpunkte mittels Schablonen ermittelt und auf dem Auswertungsbogen für die jeweilige Bezugsgruppe eingetragen. Diesem Auswertungsbogen können direkt Standardwerte ($M=100$, $s=10$) und Prozentränge sowie das Profil des Probanden entnommen werden.

Gütekriterien: Die Objektivität der Durchführung und Auswertung kann durch die Standardisierung als gegeben gelten. Die Reliabilitätskoeffizienten (Split-Half) liegen für die einzelnen Subskalen um 0.95 (sehr hoch), die Stabilitätskoeffizienten können für die Skalen der Langform ebenfalls als ausreichend bis hoch bewertet werden (ab 0.8).

Zur Prüfung der Validität des DIT wurden zunächst Faktorenanalysen durchgeführt, die zeigen, daß die Interessenrichtungen des DIT weitgehend unabhängig voneinander sind. Mit dem BIT ergeben sich für vergleichbare Interessenrichtungen mittlere, ansonsten niedrige Korrelationen. Die Zusammenhänge mit Skalen von Wissenstests erwiesen sich als niedrig, die höchsten Korrelationen ergaben sich für vergleichbare Gebiete. Auch die Korrelationen mit Skalen des Wilde-Intelligenz-Tests (WIT) fielen eher niedrig aus. Die Zusammenhänge zwischen DIT und Schulnoten erwiesen sich ebenfalls als eher gering. Weiter finden sich im Manual Hinweise zu den Interessenprofilen von Achievern und Underachievern sowie Schülern verschiedener Schularten und Klassenstufen.

Normen wurden für männliche und weibliche Schüler sowie für männliche Probanden nach Schularten getrennt berechnet. Eine Normierung nach Altersgruppen erwies sich nicht als notwendig.

Kritik: Den Angaben im Manual zufolge können mit dem DIT zuverlässig Interessen bei älteren Realschülern und Gymnasiasten ermittelt werden. Die mit dem DIT erfaßten Interessen erwiesen sich in den Validitätsstudien jedoch als so unabhängig von Leistungskriterien, daß Ergebnisse nicht überinterpretiert werden sollten. Zudem stammen die im Manual mitgeteilten Untersuchungen alle aus den

60er Jahren. Der potentielle Anwender sollte zudem den Hinweis Todts (1967) beachten, daß die Validität des DIT in Wettbewerbssituationen nicht mehr gegeben sein muß.

(3) **Das Testverfahren für Interessen: Beruf - Schule (TIBS 1) von K. Weible & H. Bethäuser (1986) (Beltz)**

Mit Hilfe des TIBS 1 sollen Schulpsychologen, Bildungsberater und Beratungslehrer, aber auch Hauptschullehrer im Rahmen allgemeiner Beratungsaufgaben bei Hauptschulabgängern (Ende Klasse 8, Anfang Klasse 9) Hinweise über deren Motivation zu beruflicher versus schulischer Aus- und Weiterbildung gewinnen können. Gleichzeitig will der Test Informationen über Interessenschwerpunkte im Bezug auf die Ausbildungsrichtung erfassen. Die Ergebnisse im TIBS sollen damit zum einen als Entscheidungshilfe bei der Wahl der Art der Ausbildung dienen (betriebliche Ausbildung mit begleitendem Berufsschulbesuch versus Aus- und Weiterbildung in Vollzeitschulen wie Berufsfachschulen oder im Berufsgrundbildungsjahr). Zum anderen sollen die Berater bzw. die Schüler Informationen an die Hand bekommen, ob bestimmte Berufsgruppen den Interessen und Neigungen der Schulabgänger entsprechen.

Testaufbau: Im Testbogen finden sich insgesamt 120 Items, die sich laut Manual auf zwei Teile und sieben Skalen folgendermaßen aufteilen:

- Motivation zu beruflicher Ausbildung (Skala B, 15 Items) und Motivation zu schulischer Weiterbildung (Skala S, 16 Items);
- Skalen zu Interessen an Metalltechnik (Skala M, 15 Items), Elektrotechnik (Skala E, 20 Items), Wirtschaft (Skala W, 20 Items), Hauswirtschaft/Erziehung (Skala H, 17 Items) und Landwirtschaft (Skala L, 15 Items).

Laut Auswertungsschablone wird ein weiteres Item (Nummer 62) zur L-Skala herangezogen, Item 63 ist keiner Skala zugeordnet. Da bereits die Skalierung (einfache Summenbildung) geschlechtsspezifisch erfolgt, reduzieren sich die Skalenlängen um bis zu sieben Items: Die Skala E für Jungen umfaßt nach Angaben im Manual nur 13 Items (laut Lösungsschablone allerdings 14 Items). Zum TIBS existiert keine Parallelform.

Erfassungsdimensionen: Das TIBS rekurriert nicht auf eine bestimmte Motivations- oder Interessentheorie, sondern ist aus der Praxis der Bildungsberatung heraus entstanden. Leitend war für die Autoren dabei der Gedanke, pädagogisch-psychologische Diagnostik im Dienste der optimalen Förderung aller Möglichkeiten des einzelnen Schülers innerhalb des gegebenen Schulsystems einzusetzen.

Zur Konstruktion der Items des Interessenteils zogen die Testautoren neben Schülerbefragungen nach besonderen Kennzeichen und bedeutsamen Inhalten der jeweiligen Fachrichtung eine Analyse von Stoffverteilungsplänen der Berufsfachschulen in Baden-Württemberg heran. Die Brauchbarkeit der Items wurde in Zusammenarbeit mit Fachlehrern an Berufsfachschulen überprüft. Die endgültige Auswahl erfolgte unter Berücksichtigung von Itemschwierigkeiten und Trennschärfen, aber auch unter Einbeziehung der Rückmeldungen der beteiligten Lehrer und Schüler. Die Autoren weisen darauf hin, daß die fünf Interes-

senskalen des TIBS ohne große Einschränkungen in den Bundesländern Bayern, Rheinland-Pfalz, Bremen und Hamburg verwendet werden können, in den übrigen Bundesländern könnten Informationen über die Schülerinteressen zu den Bereichen Wirtschaft und Hauswirtschaft gewonnen werden.

Handhabung: Das TIBS kann sowohl als Einzel-, als auch als Gruppentest durchgeführt werden. Die Schüler können die Instruktionen den Testheften entnehmen, die Autoren empfehlen aber, die Instruktion zum Motivationsteil laut vorzulesen. Die gesamte Durchführungszeit dürfte eine Schulstunde nicht übersteigen, auch wenn kein Zeitlimit vorgesehen ist.

Die Auswertung erfolgt für Jungen und Mädchen getrennt mit Hilfe von Schablonen. Die Rohwerte werden in Prozentränge und T-Werte transformiert, T-Wert-Bänder (Vertrauensbereiche zur Irrtumswahrscheinlichkeit von 68 Prozent) können ebenfalls direkt aus den Normtabellen abgelesen werden. Für die Schüler ist ein eigenes Blatt mit graphischer Darstellung ihrer Ergebnisse und Interpretationshinweisen vorgesehen. Für die gesamte Auswertung (einschließlich Ergebnisblatt für den Schüler) veranschlagen die Autoren zehn Minuten.

Das Manual enthält generelle Hinweise zu den Prozenträngen und T-Werten bzw. T-Bändern sowie zur Interpretation der einzelnen Skalen. Anhand von Fallbeispielen wird die Umsetzung der Ergebnisse für Beratungszwecke demonstriert. Auf die Grenzen des Verfahrens wird von den Autoren hingewiesen.

Gütekriterien: Durchführungs- und Auswertungsobjektivität können als gegeben angesehen werden (schriftliche Instruktion, Auswertung mittels Schablonen). Zur Beurteilung der **Reliabilität** teilen die Autoren - getrennt für Jungen und Mädchen - Konsistenz- und Stabilitätskoeffizienten (Retesting nach zwei Monaten) mit, wobei über die konkret verwendete Berechnungsart nichts mitgeteilt wird. Die Konsistenzkoeffizienten liegen zwischen 0.79 (Skala S, Mädchen) und 0.94 (Skala E, Jungen) und können somit als zufriedenstellend bis gut bezeichnet werden (Mädchen: N=178; Jungen: N=218). Die Stabilitätskoeffizienten liegen durchweg niedriger zwischen 0.52 (Skala B, Mädchen, sowie Skala S, Jungen) und 0.88 (Skala S, Mädchen) (Mädchen: N=27; Jungen: N=28).

Die **prognostische Validität** der Skalen B und S wurde drei Monate nach Schulabschluß überprüft, indem eine Gruppe Auszubildender (N=86) einer Gruppe Vollzeitschüler (N=77) gegenübergestellt wurde. Neben der Prüfung der Mittelwertsunterschiede, die in der erwarteten Richtung ausfielen (Auszubildende erzielten höhere Werte in der B-Skala, Schüler höhere in der S-Skala), wurden biseriale Korrelationen zwischen den Skalen und der gewählten Ausbildungsrichtung berechnet. Zusätzlich wird im Manual angegeben, wieviel Prozent gemeinsame Varianz Skala und Kriterium aufweisen (Ω -Prozent). Im einzelnen ergeben sich für die Skala B $r_{\text{bis}}=0.69$ und $\Omega\text{-}\%=30$, für Skala S $r_{\text{bis}}=0.47$ und $\Omega\text{-}\%=14$. Mit derselben Methode wurden innerhalb der Gruppe der Schüler die Zusammenhänge zwischen den Interessenskalen und den gewählten Berufsrichtungen untersucht. Auch hier zeigten sich die erwarteten Zusammenhänge, lediglich das $r_{\text{bis}}=0.26$ bei Skala H ist nicht signifikant. Zu Skala L wird allerdings nichts mitgeteilt.

Im Bezug auf die Dauer des Schulbesuchs zeigte sich, daß Jugendliche, die das einjährige Berufsgrundschuljahr absolvieren, eine höhere Motivation zu beruflicher

Tätigkeit aufweisen als die Berufsfachschüler (zweijährige Schulzeit), während letztere auf der Skala S höhere Werte erzielten. Ein Vergleich mit einer Gruppe von Schülern, die lieber eine betriebliche Ausbildung gemacht hätten (N=43), ergab, daß deren schulische und berufliche Motivation eher den Auszubildenden als den Vollzeitschülern entsprach.

Zur Bestimmung der **faktoriellen Validität** des Interessenteils wurden die Items der Skalen M, E, W, H und L (hier mit Item 63) einer Faktorenanalyse mit anschließender Varimaxrotation unterzogen. Insgesamt fanden sich in den berechneten sechs Faktoren die Interessensskalen in etwa wieder, wenngleich die Autoren feststellen, daß Items zu einfachen manuellen Tätigkeiten stärker gewichtet in die Faktoren eingehen als Items zu eher theoretischen Aspekten der Fachrichtungen. Leider gebrauchen die Autoren den Begriff Gesamtvarianz etwas mißverständlich: Faktor 1 repräsentiert 39.8 Prozent der durch die Analyse aufgeklärten Varianz und nicht der gesamten in die Analyse eingehenden Varianz. Auf letztere bezogen erklärt er 15.8 Prozent. Gleiches gilt entsprechend für die anderen Angaben (so beträgt der Anteil der ersten drei Faktoren an der erklärten Varianz 81 Prozent und 39.6 Prozent an der gesamten Varianz). Zudem legt der Eigenwertverlauf eine Lösung mit weniger Faktoren nahe (bereits der vierte der sechs berechneten Faktoren repräsentiert nur noch einen Anteil von weniger als 3 Prozent der gesamten Varianz).

Die **Normierung** des TIBS erfolgte im Herbst 1982 an einer Stichprobe von 396 Schülern (178 Mädchen und 218 Jungen) aus baden-württembergischen Hauptschulen. Nicht zuletzt um Verfälschungen entgegenzuwirken, wurden nur solche Schüler in die Eichstichprobe einbezogen, die über mögliche Berufsfelder bereits informiert waren und Interesse an einer Beratung hatten. Mit den Schülern wurden in einem Beratungsgespräch die Testergebnisse durchgesprochen.

Kritik: Das TIBS zeichnet sich durch eine schnelle und einfache Durchführung sowie Auswertung aus. Der Test eignet sich wegen der einfach gehaltenen Hinweise zu Auswertung und Interpretation auch für die praktische Anwendung durch Lehrer.

Das Bemühen der Autoren um jeden einzelnen an der Testentwicklung beteiligten Schüler muß zunächst positiv hervorgehoben werden. Wenn allerdings mit Schülern auf der Basis der Ergebnisse im TIBS ein Beratungsgespräch zu Fragen der Wahl der weiteren Ausbildung geführt wurde, ist die Bestimmung der prognostischen Validität nach der oben angeführten Methode wenig aussagekräftig. Zudem ist die Berechnung der biserialen Korrelation im vorliegenden Fall nicht angezeigt, da Ausbildungsrichtungen und Schularten nicht als Dichotomisierungen normalverteilter metrischer Variabler aufgefaßt werden können.

Der Umfang der Normierungsstichprobe erscheint zu klein, Angaben über deren genauere Zusammensetzung fehlen. Positiv fällt dagegen die sorgfältige Auswahl und vorbildliche Ergebnisrückmeldung der Eichstichprobe auf. Den Umfang der Rückmeldung sollten die Autoren allerdings nicht als Begründung für den geringen Umfang der Eichstichprobe heranziehen (S. 11).

An einigen Stellen läßt das Manual die zur Beurteilung der statistischen Kennwerte sowie der angewendeten Methoden notwendige Exaktheit vermissen. So bleibt

die Art der Berechnung der Reliabilitätskoeffizienten oder das verwendete Faktorenanalyseverfahren im dunkeln, und aus Text und Tabellen ist nicht schlüssig ersichtlich, ob die mitgeteilten Trennschärfen korrigiert oder unkorrigiert berechnet wurden. Selbst bei den T-Werten fehlt der Hinweis, ob Flächentransformationen vorgenommen wurden (letzteres muß nach Inspektion der Normtabellen angenommen werden). Angaben zur Beurteilung der Formen der Rohwertverteilungen fehlen.

Auch wenn dem Verfahren keine Theorie zugrunde liegt, scheint das TIBS im Rahmen seiner Möglichkeiten für die Praxis gut verwendbar zu sein (abgesehen von den zum Teil zu geringen Stabilitätskoeffizienten). Die Items und Skalen sind konsequent an den Inhalten relevanter Berufsschulrichtungen orientiert, auch wenn so wichtige Bereiche wie vor allem die Bauberufe, die Bereiche der freien Berufe (Arzthelferinnen etc.) oder das Hotel- und Gaststättengewerbe fehlen. Diese Ausrichtung an der beruflichen und berufsschulischen Praxis läßt das TIBS bei seiner spezifischen Zielgruppe den üblichen Interessentests als überlegen erscheinen.

Tabelle 15: Übersicht über Tests zur Erfassung von Interessen

Autor und Erscheinungsjahr	Testbezeichnung	Erfassungsdimensionen	Alters- und Zielgruppe	Zeitbedarf
M. Irle & W. Allehoff (1984)	Berufs-Interessen-Test (BIT II)	Interessen in 9 Bereichen	Schüler ab dem 14./15. Lebensjahr	20-80 Minuten für forced-choice-Form, Fragebogen etwa 20-40 Minuten
E. Todt (1967, 1972 ²)	Differentieller Interessen-Test (DIT)	Interessen in 11 Richtungen und 4 Materialbereichen	15-20 Jahre	30-45 Minuten
H. Seidl & J. Turski (1975)	Schul- und Berufsinteressentest (SBIT)	20 Interessensbereiche	ab 13. Lebensjahr bzw. 8. Klasse	25-45 Minuten
A. Andre (1980)	Themenwahlverfahren für 4. bis 9. Klassen	Stärke und Richtung schulischer und außerschulischer Interessen (12 Richtungen)	4.-9. Klassen	10-30 Minuten
K. Weible & H. Bethäuser (1986)	Testverfahren für Interessen: Beruf-Schule (TIBS 1)	Berufliche vs. schulische Bildungsmotivation, Interessenschwerpunkte	Hauptschüler Ende 8./Anfang 9. Klasse	1 Schulstunde

3.4.4 Weitere Verfahren zur Erfassung nichtkognitiver Persönlichkeitsmerkmale (Moderatorvariablen)

- (1) Der **Angstfragebogen für Schüler (AFS)** von W. Wiczerkowski, H. Nickel, A. Janowski, B. Fittkau & W. Rauer (1974, 1981⁶) (TZ)

Testaufbau: Der AFS setzt sich aus 50 Items zusammen, die mit "stimmt" und "stimmt nicht" beantwortet werden. Die Items werden zu den vier Skalen **Prüfungsangst (PA, 15 Items)**, **Manifeste Angst (MA, 15 Items)**, **Schulunlust (SU, 10 Items)** und **Soziale Erwünschtheit (SE, 10 Items)** zusammengefaßt. Zusätzlich liegen dem Test Einschätzskalen für den Lehrer bei, anhand derer Lehrer- und Schülerelbsteinschätzungen verglichen werden können.

Erfassungsdimensionen: Der AFS ist als mehrfaktorieller Fragebogen konzipiert, wobei die Skala **Manifeste Angst** eine generelle Disposition zur Ängstlichkeit erfassen soll, während sich die **Prüfungsangstskala** auf den schulischen Leistungsbereich bezieht. **Schulunlust** "erfaßt die innere Abwehr von Kindern und Jugendlichen gegen die Schule" (Manual, S. 20) und den Unterricht, wohingegen Schüler mit hohen Werten in der Skala **Soziale Erwünschtheit** befürchten, von der sozialen Norm abzuweichen, und eine Neigung zu Verstellung erkennen lassen (**Kontrollskala**).

Handhabung: Der Einsatz als Individual- oder Gruppenverfahren ist möglich. Für die Durchführung müssen 10 bis 25 Minuten veranschlagt werden. Die Auswertung (etwa 5 Minuten) erfolgt mittels Schablonen und Normtabellen. Kennwerte zur Berechnung von Vertrauensintervallen und kritischen Differenzen werden in der Handanweisung mitgeteilt. Die Ergebnisse können als Profil dargestellt werden.

Gütekriterien: Die **Objektivität** ist durch die Standardisierung von Durchführung und Auswertung gegeben. Zur **Zuverlässigkeit** werden Korrelationen zwischen 0.67 und 0.85 (innere Konsistenz) und 0.67 bis 0.77 (Stabilität) berichtet. In Untersuchungen anderer Autoren wurden z.T. noch höhere Werte mitgeteilt. Die Skaleninterkorrelationen sind niedrig bis mittel, die Skala SE ist von den übrigen praktisch unabhängig. Korrelationsstudien zur **Validität** mit dem HANES (Hamburger Neurotizismus- und Extraversionsskala) und dem KAT (Kinder-Angst-Test) belegen eine relativ hohe Übereinstimmung von MA mit der Neurotizisskala des HANES und dem KAT (jeweils $r=0.61$). Die Prüfungsangstskala PA korreliert am engsten mit dem KAT ($r=0.53$), alle übrigen mitgeteilten Zusammenhänge sind niedriger. Korrelationen der AFS-Skalen mit Schulzensuren sind in den Klassen 3 bis 5 höher als in der 6. bis 9. Klasse. Die höchsten Zusammenhänge ergaben sich hierbei erwartungsgemäß für PA und SU.

SU hängt eng mit subjektiven Erfahrungen mit Lehrern zusammen. Kinder mit hohen Scores in SE akzeptieren eher autoritäres Lehrerverhalten; erhöhte negative Erfahrung mit Lehrern hängt mit stärkerer manifester und Prüfungsangst zusammen. Weiter ergaben sich Zusammenhänge zwischen der Wahrnehmung der Eltern als kontrollierend und PA sowie MA, zwischen SU und Unaufgeschlossenheit sowie

Verständnislosigkeit der Eltern in der Wahrnehmung der Kinder bzw. zwischen SE und wahrgenommener Aufgeschlossenheit der Eltern.

An der Normierung nahmen 2374 Schüler aus 83 Klassen des 3. bis 10. Schuljahrs in Schleswig-Holstein, Hamburg und Bremen teil (kleinste Teilstichproben: 238 Schüler). Wegen geringer praktischer Relevanz der Unterschiede zwischen Jungen und Mädchen, Schülern verschiedener Schularten und Herkunft aus städtischen oder ländlichen Regionen wurden keine getrennten Normen berechnet. Den Tabellen können T-Werte und Prozentränge entnommen werden.

Kritik: Trotz einiger Unzulänglichkeiten (hohe Interkorrelationen der Skalen, Mehrdimensionalität einiger Items) kann davon ausgegangen werden, daß der AFS ökonomisch, reliabel und valide das Konstrukt Angst im Sinne der Intention der Autoren des Fragebogens erfaßt (vgl. auch Rost & Haverkamp, 1979).

(2) **Das Arbeitsverhaltensinventar (AVI)** von R.D. Thiel, G. Keller & A. Binder (1979) (TZ)

Die Autoren des AVI strebten mit dem Verfahren an, nicht nur wie bei amerikanischen Vorbildern die Güte der Lern- und Arbeitstechniken von Schülern, sondern gleichzeitig auch motivationale, emotionale, personale und sozialpsychologische Determinanten des Arbeitsverhaltens zu erfassen. Was das AVI aber besonders auszeichnet, ist die Entwicklung einer darauf abgestimmten Anleitung zur Modifikation des Arbeitsverhaltens (Arbeitsverhaltensmodifikation, AVM, von Keller, Binder & Thiel, 1979). Diese Kombination von Diagnostikum und Modifikationsprogramm wirkt so überzeugend, daß das AVI trotz gravierender Mängel bei der Testkonstruktion, die z.B. Jäger (1980, 1981a u. b) heftig kritisiert hat, hier mit aufgenommen wurde. Es ist zu hoffen, daß in einer Neuauflage die methodischen Mängel behoben werden.

Testaufbau: Der AVI besteht aus 20 bipolaren Skalen mit jeweils 10 Items, die mit "trifft voll zu", "trifft nur teil- oder zeitweise zu" und "trifft nicht zu" beantwortet werden sollen. Darüber hinaus sind weitere 6 Items im Fragebogen enthalten, die keiner Skala zugeordnet sind, die aber nach Ansicht der Autoren wichtige Informationen für die Beratung liefern können.

Erfassungsdimensionen: Dem AVI liegt keine geschlossene Theorie zugrunde, vielmehr wurden aus praktischen Überlegungen heraus Skalen zusammengestellt, die aufgrund von Forschungsergebnissen und praktischen Erfahrungen für die Beratungspraxis relevant schienen. Die 20 Skalen des AVI verteilen sich wie folgt auf 5 Bereiche (die Nummern beziehen sich auf die Numerierung in der Handanweisung).

Motivationale Skalen:

- 1) **Anspruchsniveau** an schulische und berufliche Leistung.
- 2) **Bedürfnisaufschub** in bezug auf Freizeitbedürfnisse vs. Lernanforderungen.
- 3) **Erfolgsmotivation.**

4) **Lernmotiviertheit** (extrinsisch vs. intrinsisch) als relative Unabhängigkeit von kurzfristigen äußeren Verstärkern.

5) **Mißerfolgsmotivation** als Akzeptieren von leistungsmäßigen Randpositionen.
Selbstwert:

6) **Selbstwert** (leistungszentriert vs. multithematisch).

Kognitive Aspekte des Lern- und Arbeitsverhaltens:

7) **Stoffverarbeitung** (langsam vs. schnell).

8) **Aktualisierungsphase** (gestört vs. nicht gestört): Störbarkeit bei der Reproduktion von Lerninhalten.

9) **Gestaltung der Lernbedingungen** (schlecht vs. gut): Organisierung der äußeren Lernbedingungen.

10) **Denken** (impulsiv vs. reflexiv).

11) **Lernen** (faktenorientiert vs. substanzorientiert).

13) **Rezeptionsphase** (gestört vs. nicht gestört): konzentrierte Störbarkeit beim Erlernen von Stoff.

14) **Leistungskontrolle** (unfähig vs. fähig): selbständige Kontrolle des eigenen Wissensstandes.

18) **Lerntechnik** (schlecht vs. gut): zeitökonomische und erfolgreiche Lernstoffbewältigung.

Emotionale Skalen:

12) **Mißerfolgstoleranz** (gering vs. groß).

15) **Streßresistenz** (anfällig vs. resistent).

20) **Leistungsgefühle** (insuffizient vs. suffizient).

Soziale Gesichtspunkte:

16) **Lernfeldunabhängigkeit** (abhängig vs. unabhängig): Unabhängigkeit der eigenen Lernmotivation von Sympathie/Antipathie.

17) **Lernverhalten** (interessenabhängig vs. interessenunabhängig): Unabhängigkeit von fachspezifischen Interessen.

19) **Einstellung zur Schule** (negativ vs. positiv).

Handhabung: Die Schüler lesen im Einzel- oder Gruppenversuch die Fragen im Testheft und kreuzen die Antworten in einem separaten Antwortbogen an. Für die Testdurchführung sind 40 bis 45 Minuten vorgesehen. Die Auswertung ist einfach, die Objektivität ist gewährleistet. Die Rohwerte werden mittels Normtabellen in T-Werte umgewandelt und die ermittelten Normwerte in einen Profilbogen eingezeichnet. Dem AVM von Keller, Binder & Thiel (1979) können auf das AVI-Ergebnis bezogenen Modifikationshinweise entnommen werden.

Gütekriterien: Zur Kritik der berechneten Statistiken zur Prüfung von Reliabilität und Validität des Verfahrens vgl. Jäger (1980, 1981b).

Die Normierung wurde an 297 weiblichen und 235 männlichen Gymnasiasten, wegen entsprechender Unterschiede geschlechtsspezifisch, vorgenommen. Es wurden flächentransformierte T-Werte berechnet.

(3) **Der Persönlichkeitsfragebogen für Kinder zwischen 9 und 14 Jahren (PFK 9-14)** von W. Seitz & A. Rausche (1976, 1991³) (TZ)

Der PFK 9-14 stellt seit seinem Erscheinen einen der meistverwendeten Persönlichkeitstests in der Schulberatung dar. Sowohl der Fragebogen als auch das Handbuch wurden für die dritte Auflage¹⁰ wesentlich überarbeitet, so daß den Anwendern zu empfehlen ist, die alten Testmaterialien nicht mehr weiterzuverwenden. Insbesondere enthalten nun alle Skalen einheitlich 12 Items, zudem wurden unzeitgemäße Items und Fragen mit längerem Wortlauf (zur Vermeidung von Verständnisschwierigkeiten) gestrichen.

Testaufbau: Der PFK 9-14 gliedert sich in die drei Bereiche Verhaltensstile (VS, 4 Untertests), Motive (Mo, 6 Untertests) und Selbstbild (SB, 5 Untertests). Vier Faktoren 2. Ordnung wurden aufgrund von Faktorenanalysen der einzelnen Skalen bzw. Dimensionen zusammengestellt.

Erfassungsdimensionen: Die Items zu den drei Bereichen des PFK 9-14 wurden unter Berücksichtigung der einschlägigen Literatur theoriebezogen formuliert. Mit Faktorenanalysen wurde anschließend die Zuordnung der Items zu den einzelnen Skalen und Bereichen vorgenommen bzw. überprüft. Letztendlich wurden in der überarbeiteten Version folgende Skalen zusammengestellt:

Skalen des Bereichs Verhaltensstile (VS):

- VS1: Emotionale Erregbarkeit:** ängstliche Irritierbarkeit und Prüfungsangst, Ich-Schwäche;
- VS2: Fehlende Willenskontrolle:** fehlende Beherrschung und Selbstkontrolle, Über-Ich-Schwäche;
- VS3: Extravertierte Aktivität:** Initiative, Optimismus, Problembewältigung;
- VS4: Zurückhaltung und Scheu vor Sozialkontakt;**

Dimensionen des Bereichs Motive (MO):

- Mo1: Bedürfnis nach Ich-Durchsetzung, Aggression und Opposition;**
- Mo2: Bedürfnis nach Alleinsein und Selbstgenügsamkeit gegenüber Bedürfnis nach Geselligkeit;**
- Mo3: Schulischer Ehrgeiz;**
- Mo4: Bereitschaft zu sozialem Engagement;**
- Mo5: Neigung zu Gehorsam und Abhängigkeit gegenüber Erwachsenen;**
- Mo6: Maskulinität der Einstellung;**

Dimensionen des Selbstbildes (SB):

- SB1: Selbsterleben von allgemeiner (existentieller) Angst;**
- SB2: Selbstüberzeugung hinsichtlich Erfolg und Richtigkeit eigener Meinungen, Entscheidungen, Planungen und Vorhaben;**
- SB3: Selbsterleben von Impulsivität gegenüber Selbsterleben von Nachdenklichkeit und Ernst;**

¹⁰ Wir danken Herrn Prof. Dr. W. Seitz für die freundliche Vorab-Überlassung des Manuskripts zu Kapitel 1 bis 6 der überarbeiteten Neuauflage.

SB4: Egozentrische Selbstgefälligkeit (Selbstaufwertung, Selbstüberschätzung, Selbstbeschönigung);

SB5: Selbsterleben von Unterlegenheit gegenüber anderen.

Alle Skalen enthalten seit der Überarbeitung einheitlich 12 Items. Dadurch konnten Überlappungen zwischen den Dimensionen reduziert und die Homogenität der Skalen gesteigert werden. Aufgrund von Faktorenanalysen der Skalen wurden folgende **Faktoren II. Ordnung** ermittelt. Im Gegensatz zur früheren Version des Fragebogens wird den Benutzern die 4-Faktorenlösung empfohlen, die sich in einer Reihe von Studien aus inhaltlichen und methodischen Gründen als überlegen erwiesen habe:

F-II-O 1: **Derb-draufgängerische Ich-Durchsetzung** vs. sensibel-differenzierte Selbstbeherrschung (VS2, Mo1, Mo6, SB3) (48 Items);

F-II-O 2: **Emotionalität** vs. emotionale Ausgeglichenheit (VS1, VS4, SB1, SB3, SB5) (60 Items);

F-II-O 3: **Aktives Engagement** (VS3, Mo2, Mo4, SB2) (48 Items);

F-II-O 4: **Selbstgenügsame Isolierung** vs. gesellige Zuwendung (VS4, Mo2, SB4) (23 Items).

Handhabung: Für jeden der drei Bereiche des PFK 9-14 steht ein eigenes Testheft zur Verfügung, so daß die Teile auch einzeln angewendet werden können. Die Schüler haben bei jeder Frage zwischen zwei Antwortmöglichkeiten zu wählen (bis auf einzelne Items "stimmt" und "stimmt nicht"). Für den PFK 9-14 ist keine Zeitbegrenzung vorgesehen. Nach Angaben im Manual beträgt die Durchführungszeit jedes Testteils in der Regel weniger als 15 Minuten, die meisten Schüler bearbeiten den Gesamtfragebogen in ungefähr 30 Minuten. Die Auswertung erfolgt mit Hilfe von Auswertungsschablonen und Normtabellen im Handbuch. Das Manual enthält umfassende Interpretationshinweise sowie zwei ausführliche Fallbeispiele.

Gütekriterien: Die **Objektivität** des PFK 9-14 ist bei Beachtung der Durchführungs- und Auswertungsrichtlinien gegeben. Für die Itemanalysen und Berechnungen zur Reliabilität wurden aus der umfangreichen Stichprobe von 1411 Schülern 174 Probanden ausgeschieden, die mehr als 90% der Fragen in einer Richtung beantwortet hatten. Sodann wurden alle Items ausgeschieden, die von mindestens 88% bzw. 90% der Schüler der Gesamt- und verschiedener Teilstichproben in der gleichen Richtung beantwortet wurden. Nach der Absicherung der Zuordnung der Items zu den einzelnen Skalen mittels Faktorenanalyse liegen die **Trennschärfeindizes** zwischen 0.40 und 0.65 (Bereich Verhaltensstile), 0.34 und 0.73 (Bereich Motive) sowie 0.37 und 0.63 (Aspekte des Selbstbilds). Die **Zuverlässigkeit** der Skalen kann nach den im Manual mitgeteilten Halbierungs- und Konsistenzkoeffizienten als insgesamt ausreichend bis befriedigend beurteilt werden (die meisten Koeffizienten bewegen sich zwischen 0.70 und 0.75). Die entsprechenden Kennwerte für die Faktoren II. Ordnung liegen durchweg über 0.80.

Die **faktorielle Validität** des PFK 9-14 kann als gut bestätigt gelten, die Struktur der Skalen (4-Faktorenlösung) konnte nach Angaben im Manual in mehreren Untersuchungen (z.B. Kühn, 1983, zit. n. Seitz & Rausche, 1991) gut

repliziert werden, Mitte der 80er Jahre auch für eine Stichprobe aus der ehemaligen DDR (Littmann, 1986, zit. n. Seitz & Rausche, 1991).

Zur Konstruktvalidität teilen Seitz & Rausche (1991) zunächst Ergebnisse mehrerer Untersuchungen zu Unterschieden zwischen Schulleistungs- und Begabungsgruppen bezüglich der mit dem PFK erfaßten Persönlichkeitsmerkmale mit. Vor allem bezüglich SB3, aber auch im Hinblick auf SB1, SB5, VS1 und andere Skalen zeigten sich dabei signifikante Gruppenunterschiede. Weiter finden sich u.a. Vergleiche hinsichtlich der PFK-Skalen für Schülern mit unterschiedlicher Beurteilung für den Übertritt in weiterführende Schulen oder für delinquente und nicht-delinquente Schüler. Unter anderem werden Geschlechtsunterschiede, die jedoch nicht einheitlich ausfielen, und Unterschiede zwischen bundesrepublikanischen Schülern und Schülern aus der DDR berichtet. Bezüglich der Ergebnisse muß der Leser auf die ausführliche Darstellung im Manual verwiesen werden.

Im Hinblick auf die **Konstruktvalidität** finden sich im Manual Ergebnisse verschiedener Korrelationsstudien. Demnach korrelieren beispielsweise die dem Faktor II. Ordnung Emotionalität zugeordneten Skalen mit den AFS-Skalen Manifeste Angst und Prüfungsangst, die Primär-Dimension des Faktors Derb-draufgängerische Ichdurchsetzung zeigt Beziehungen zur Anstrengungsvermeidungsskala des AVT, während sich Hinweise für Beziehungen zwischen den PFK-Skalen des aktiven Engagements und der AVT-Pflichteiferskala fanden.

Die berechneten Korrelationen mit Intelligenztestleistungen und auch Schulzensuren in Mathematik, Deutsch und Sachkunde fielen bei Viertkläßlern dagegen niedriger aus (fast alle kleiner $r=0.30$). So korreliert der Faktor II. Ordnung Emotionalität mit der Durchschnittszensur in den Hauptfächern mit lediglich $r=-0.23$ am höchsten, was etwas mehr als 5% Varianzaufklärung entspricht. Für den PSB werden keine signifikanten Beziehungen berichtet, mit den AzN korreliert lediglich Mo4 nennenswert ($r=0.27$). Daneben werden u.a. Ergebnisse von Korrelationsstudien zu den Beziehungen des PFK mit verschiedenen Aspekten von Sprachleistungen und Motorik, mit Aggressionskonsum durch Fernsehen, und selbstberichteter Delinquenz angeführt. Schließlich werden Untersuchungen zum Zusammenhang der mit dem PFK erfaßten Merkmale mit Variablen des elterlichen Erziehungsstils berichtet.

Kritik: Mit der überarbeiteten Version des PFK 9-14 liegt ein Instrumentarium zur Persönlichkeitsdiagnostik bei Kindern vor, das im Hinblick auf die Sorgfalt der Testkonstruktion sowie die Testgütekriterien den Vergleich mit ähnlichen Verfahren nicht zu scheuen braucht. Besonders hervorgehoben sei auch das sehr ausführliche Manual, in dem vor allem die umfassende Aufarbeitung der Literatur zum PFK 9-14 besticht. Vielleicht hätten die Autoren hier jedoch etwas stärker auswählen können, mancher Benutzer wird sich möglicherweise mit der Fülle der Informationen und bei der Beurteilung der Qualität der vielen zitierten Untersuchungen schwer tun. Jedenfalls sollte die neue Version des PFK 9-14 bei keinem Schul- und Erziehungsberater im Testinventar fehlen.

Tabelle 16: Übersicht über Tests zur Erfassung nichtkognitiver Persönlichkeitsmerkmale

Autor und Erscheinungsjahr	Testbezeichnung	Erfassungsdimensionen	Alters- und Zielgruppe	Zeitbedarf
W. Wiczerkowski, H. Nickel, A. Janowski, B. Fittkau & W. Rauer (1974, 1981 ⁶)	Angstfragebogen für Schüler (AFS)	Manifeste und Prüfungsangst, Schulunlust, Soziale Erwünschtheit	9-16/17jährige Schüler, 3.-10. Klassen	10-25 Minuten
R. Thiel, G. Keller & A. Binder (1979)	Arbeitsverhaltensinventar (AVI)	Aspekte des Arbeitsverhaltens	Schüler ab 14 Jahren (Sekundarstufe II)	1 Schulstunde
G. Schumacher & R.B. Cattell (1977)	Deutscher HSPQ (High School Personality Questionnaire)	Persönlichkeitsfaktoren nach Cattell	12-18 Jahre	45-60 Minuten
F. Petermann & U. Petermann (1980)	Erfassungsbogen für aggressives Verhalten in konkreten Situationen (EAS)	Aggression nach dem Situation-Verhalten-Person-Konzept	9-13 Jahre	keine Angaben
V. Gärtner-Harnach (1972 und 1973)	Fragebogen für Schüler (FS 11-13 und FS 5-10)	Angst in schulischen Leistungssituationen	FS 5-10: 5.-10. Klassen FS 11-13: 11.-13. Klassen	jeweils ca. 20 Minuten
J.W.L. Wagner (1977b)	Fragebogen zum Selbstkonzept für 4. bis 6. Klassen (FSK 4-6)	Selbsteinschätzung im Vergleich zu Klassenkameraden	4.-6. Klassen	25 Minuten
F. Buggle & F. Baumgärtel (1972, 1975 ²)	Hamburger Neurotizismus- und Extraversionsskala für Kinder und Jugendliche (HANES-KJ)	Neurotizismus und Extraversion sensu Eysenck	8-16 Jahre	20-40 Minuten, je nach Alter der Probanden
H. Wagner & F. Baumgärtel (1978)	Hamburger Persönlichkeitsfragebogen für Kinder (HAPEF-K)	Persönlichkeitsstruktur (6 Skalen)	9;0 bis 13;9 Jahre	20-40 Minuten je nach Alter der Probanden
F. Thurner & U. Tewes (1969, 1972 ²)	Kinder-Angst-Test (KAT)	Manifeste Angst	9-15 Jahre	15-30 Minuten je nach Alter der Probanden

R. Jäger, E. Berbig, B. Geisel, H. Gosslar, J. Hagen, W. Liebich & R. Schafheutle (1973)	Mannheimer Biographisches Inventar (MBI)	Verhalten in 11 Inhalts- bereichen	11-18 Jahre	45-50 Minuten
W. Seitz & A. Rausche (1991 ³)	Persönlichkeits- fragebogen für Kinder zwi- schen 9 und 14 Jahren (PFK 9-14)	15 Skalen zu Verhaltens- stilen, Bedürf- nissen und Motiven sowie Selbstbild	9.-14. Klassen	ca. 30 Minuten (maximal etwa 15 Minuten pro Testteil)
L. Laux, P. Schaffner & P. Glanzmann (1981)	State-Trait Angstinventar (STAI)	Manifeste "trait"-Angst, situations- spezifische Angst	ab 15 Jahren, Normen für 19- 70jährige	3-5 Minuten

3.5 Skalen zur Erfassung des sozialen Lernumfeldes

- (1) Die Dortmunder Skala zum Lehrerverhalten (DSL) von F. Masendorf, M. Tücke, R. Kretschmann & M. Bartram (1976) (TZ)

Mit der DSL sollen Lehrer Informationen darüber erhalten, wie streng vs. unterstützend ihr Unterrichtsverhalten von den Schülern erlebt wird.

Testaufbau: Die DSL enthält eine bipolare Skala, in die 25 4stufig zu beantwortende Items eingehen.

Erfassungsdimensionen: Erfasst werden strenges vs. unterstützendes Lehrerverhalten aus der Sicht der Schüler (perzipierte Strenge vs. Unterstützung) als zwei Pole eines Kontinuums. Unterstützendes Verhalten wird als wünschenswert betrachtet.

Handhabung: Der Fragebogen kann in allen Schultypen (auch Sonderschulen) als Gruppenverfahren (Einzeluntersuchungen nicht möglich) etwa in den Klassen 5 bis 9 eingesetzt werden. Die Anonymität der Schüler ist unbedingt zu gewährleisten. Testleiter soll nicht der beurteilte Klassenlehrer sein. Für die Auswertung werden neben einem mittleren Gesamtwert über die ganze Klasse auch Mittelwerte für die einzelnen Items berechnet und mit Hilfe eines groben Klassifikationschemas ($PR \leq 25$, $25 < PR \leq 75$, $PR > 75$) interpretiert. Für die Gesamtskala stehen darüber hinaus T-Wert-Kategorien zur Verfügung, die jedoch ebenfalls nur den genannten drei Kategorien entsprechen. Das Vorgehen bei der Auswertung sowie die pädagogische Verwertung der Ergebnisse wird an Beispielen erläutert.

Gütekriterien: Die Objektivität ist durch die Standardisierung gegeben. Die Konsistenzreliabilität betrug 0.92, die Stabilität (Testwiederholung nach drei Monaten) zwischen 0.76 und 0.96. Korrelationsstudien zur **Validität** ergaben u.a., daß Angst und Anstrengungsvermeidung mit negativen Urteilen über den Lehrer ansteigen oder die DSL mit Engagement des Lehrers in mittlerer Höhe korreliert.

Die zur Normierung herangezogene Stichprobe umfaßt 95 Klassen mit 1782 Schülern. Da die Werte der DSL sich auf den Lehrer beziehen, ist die Stichprobe als sehr klein zu bezeichnen. Deswegen wurden offensichtlich auch nur grobe Richtwerte berechnet (s.o.), differenziertere Normen sind angekündigt.

Kritik: Die Konstruktion des Verfahrens leidet zwangsläufig an geringen Stichprobengrößen trotz großer Probandenzahlen. Diese können natürlich nicht beliebig vergrößert werden, da die Analyseeinheit der Lehrer und nicht der einzelne Schüler ist. Dennoch kann die DSL für bestimmte Anwendungszwecke ein hilfreiches Diagnoseinstrument darstellen.

(2) Die Hamburger Erziehungsverhaltensliste für Mütter (HAMEL) von F. Baumgärtel (1979) (Hogrefe)

Testaufbau: Die HAMEL enthält 24 Fragen nach mütterlichen Erziehungsverhaltensweisen, zu denen auf einer 6stufigen Skala die Häufigkeiten angegeben werden (0, 1, 5, 10, 20 und über 20 mal), mit denen die jeweilige Verhaltensweise in den letzten vier Wochen aufgetreten ist. Die Fragen werden zu den Skalen mütterliche Unterstützung (9 Items), Strenge (11 Items) und Zuwendung (4 Items) aufsummiert.

Erfassungsdimensionen: Unter Erziehungsstil versteht der Autor die "Gesamtheit der in Wechselwirkung stehenden komplexen Klassen von Erziehungsverhalten, -vorstellungen" und "-normen" (Baumgärtel, 1979, S. 6). Mit dem HAMEL sollen konkrete Verhaltensweisen der Mutter in bezug auf Unterstützung, Strenge und Zuwendung erfaßt werden.

Handhabung: Der Fragebogen wird von Müttern ausgefüllt. Wichtig ist der Hinweis, daß sich die Fragen auf die letzten vier Wochen beziehen. Die Auswertung erfolgt mit Hilfe von Schablonen, die Interpretation anhand von Normtabellen (C-Werte, Prozentränge). Im Manual finden sich Hilfen zur Berechnung von Vertrauensintervallen und zur Prüfung kritischer Differenzen. Als bedenklich werden Werte über Prozentrang 95 angesehen.

Gütekriterien: Die Objektivität des Verfahrens ist bei genauer Beachtung der Standardisierungsbedingungen gegeben. Die **Konsistenzreliabilität** wird im Manual für die einzelnen Skalen und verschiedene Schätzverfahren mit 0.8 bis 0.9 angegeben. Die **Konstruktvalidität** wurde mit einer Reihe von Studien untersucht, die sich auf den Zusammenhang des mütterlichen Erziehungsverhaltens mit Schichtvariablen (Unterschichtmütter sind unterstützender und strenger), Geschlecht (Jungen werden strenger erzogen), Alter (jüngere Kinder werden mehr unterstützt) und Stellung des Kindes in der Geschwisterreihe (keine Unterschiede), der Berufs-

tätigkeit (berufstätige Mütter sind unterstützender) sowie der Religion der Mutter (katholische Mütter scheinen etwas strenger zu sein) beziehen.

Korrelationsstudien mit dem HANES und dem AFS ergaben z.T. unerwartete Zusammenhänge: So wurden keine Beziehungen zwischen mütterlicher Strenge und Angstvariablen gefunden, während die HANES-Lügenskala mit der mütterlichen Unterstützung mit 0.40 korrelierte. Hingegen schildern sich Kinder überdurchschnittlich strenger Mütter erwartungsgemäß als neurotischer, leistungslabiler, aggressiver etc.

In die Normierung wurden über 3400 Mütter bzw. Kinder zwischen 9 und 14 Jahren einbezogen. Die Normen (C-Werte, Prozentränge) liegen, für Jungen und Mädchen getrennt, im jährlichen Abstand vor.

Kritik: Die HAMEL erfaßt nicht Erziehungsziele oder Einstellungen zu Erziehungsfragen, sondern konkrete Verhaltensweisen. Gerade dies dürfte das Verfahren für viele praktische Beratungsfälle interessant machen, da der Berater damit eine Möglichkeit hat, die Verhaltensweisen im Einzelfall mit einer größeren Stichprobe zu vergleichen. Allerdings ist angesichts des gefundenen Zusammenhangs der HAMEL mit der Tendenz, sich sozial erwünscht darzustellen, vor einer allzu voreiligen Interpretation der Ergebnisse zu warnen. Hierzu wären noch weitere Untersuchungen wünschenswert.

(3) Die Landauer Skalen zum Sozialklima für 4. bis 13. Klassen (LASSO 4-13) von M. v. Saldern & K.-E. Littig (1987) (Beltz)

Testaufbau: Die insgesamt 141 Items der LASSO 4-13 sind 17 Skalen zugeordnet, die sich wiederum auf drei Bereiche oder Dimensionen verteilen: Lehrer-Schüler-Beziehungen, Schüler-Schüler-Beziehungen und Merkmale des Unterrichts. Zum Bereich **Lehrer-Schüler-Beziehungen** gehören die Skalen **FUERS** Fürsorglichkeit des Lehrers (8 Items), **AGLE** Aggression gegen den Lehrer (8 Items), **ZULE** Zufriedenheit mit dem Lehrer (9 Items), **RUNG** Autoritärer Führungsstil des Lehrers (11 Items) und **BEVO** Bevorzugung und Benachteiligung durch den Lehrer (11 Items).

Die Dimension **Schüler-Schüler-Beziehungen** umfaßt **CLIQ** Ausmaß der Cliquenbildung (9 Items), **HILF** Hilfsbereitschaft der Mitschüler (8 Items), **AGMI** Aggressionen gegen Mitschüler (7 Items), **DISK** Diskriminierung von Mitschülern (6 Items), **ZUMI** Zufriedenheit von Mitschülern (9 Items) sowie **KONK** Konkurrenzverhalten von Mitschülern (7 Items).

Schließlich werden zu den allgemeinen Merkmalen des Unterrichts folgende Skalen gerechnet: **LEID** Leistungsdruck (7 Items), **ZUFU** Zufriedenheit mit dem Unterricht (5 Items), **DISZI** Disziplin und Ordnung (11 Items), **FÄHI** Fähigkeit des Lehrers zur Vermittlung von Lerninhalten (11 Items), **RESI** Resignation (6 Items) und **REDU** Reduzierte Unterrichtsteilnahme (8 Items).

Erfassungsdimensionen: Mit den LASSO soll erfaßt werden, wie Schüler ihre Lernumwelt wahrnehmen. Dabei werden zwei verhaltensrelevante Merkmale ein-

bezogen: die individuell subjektive Wahrnehmung (Einzeltest) und das Sozialklima einer Klasse (Gruppenversuch), wobei letzteres eine Aggregation über das subjektiv wahrgenommene Sozialklima aller Schüler einer Klasse darstellt. Insgesamt ist der Einsatz in drei verschiedenen Formen möglich: zur Erfassung der tatsächlichen Lernumwelt (LASSO-REAL), der gewünschten Lernumwelt (LASSO-IDEAL) und der tatsächlichen Lernumwelt aus der Sicht der Lehrer (LASSO-LEHRER) im Vergleich zu mehreren Klassen.

Handhabung: Pro Bereich sind etwa 20 Minuten Durchführungszeit einzuplanen. Die Auswertung erfolgt mittels Schablonen und Normtabellen (T-Werte, Prozentränge), die Normwertbänder enthalten. Ein Auswertungsprogramm für MSDOS-Personalcomputer liegt vor.

Gütekriterien: Die im Manual mitgeteilten Retestkoeffizienten sind relativ niedrig (zwischen 0.29 und 0.52 für einen Zeitraum von 9 Monaten), was jedoch angesichts der erfaßten Merkmale nicht verwundert. Zur Validität wird im Manual eine Reihe von Untersuchungen mitgeteilt, die u.a. belegen, daß Mädchen die Lernumwelt negativer beurteilen, wohingegen bessere Schüler eine günstigere Lernumwelt wahrnehmen. Weiter erleben Repetenten ein höheres Ausmaß an Cliquenbildung, sind resignativer und schätzen den Lehrer als weniger fähig ein. Korrelationen mit anderen Fragebogen zu Motivation und Einstellungen ergaben plausible, jedoch eher niedrige Zusammenhänge.

Die Normierung erfolgte im Jahr 1986 an insgesamt 3213 Schülern aus sieben westdeutschen Bundesländern. Die einzelnen Normierungsstichproben weisen zwischen 121 Probanden (5. Klasse Gymnasium) und 730 Probanden (5. Klasse Hauptschule) auf.

Tabelle 17: Übersicht über Tests zur Erfassung der sozialen Lernumwelt

Autor und Erscheinungsjahr	Testbezeichnung	Erfassungsdimensionen	Alters- und Zielgruppe	Zeitbedarf
F. Masendorf, M. Tücke, R. Kretschmann & M. Bartram (1976)	Dortmunder Skala zur Erfassung von Lehrerverhalten durch Schüler (DSL)	Einschätzung des Lehrers bzgl. Unterstützung und Strenge	Klassen 6-9 (Hauptschule), 5-8 (Realschule, Gymnasium), 7-9 (Sonderschule)	20 Minuten
F. Baumgärtel (1979)	Hamburger Erziehungsverhaltensliste für Mütter (HAMEL)	Mütterliche Zuwendung und Strenge	Mütter von Kindern zwischen 9;0 und 14;11 Jahren	keine Angaben

M. v. Saldern & K.-E. Littig (1987)	Landauer Skalen zum Sozialklima für 4. bis 13. Klassen (LASSO 4-13)	Tatsächliche und erwünschte Lernumwelt (3 Bereiche) aus Schülersicht, Lernumwelt aus Lehrersicht	4.-13. Klassen	20 Minuten pro Bereich
H. Petillon (1984)	Sozialfragebogen für Schüler für 4. bis 6. Klassen (SFS 4-6)	Soziales Schulumfeld aus Schülersicht (6 Skalen)	4.-6. Klassen	bis 2 Schulstunden
R. Müller (1980)	Diagnostisches Soziogramm (DSO)	Soziale Struktur einer Klasse	2.-9. Klassen	30 Minuten
H. Petillon (1980)	Soziometrischer Test für 3. bis 7. Klassen (ST 3-7)	Soziale Beziehungen in Schulklassen	3.-7. Klassen	ca. 15 Minuten

Literaturhinweise

- Brickenkamp, R. (1975). **Handbuch psychologischer und pädagogischer Tests**. Göttingen: Hogrefe.
- Brickenkamp, R. (1983). **Handbuch psychologischer und pädagogischer Tests, 1. Ergänzungsband**. Göttingen: Hogrefe.
- Groffmann, K.J. & Michel, L. (Hrsg.). (1982). **Intelligenz- und Leistungsdiagnostik (= Enzyklopädie der Psychologie, Themenbereich B, Serie II, Bd. 2)**. Göttingen: Hogrefe.
- Hiltmann, H. (1977). **Kompendium der psychodiagnostischen Tests**. Bern: Huber, 3. Aufl.
- Ingenkamp, K. & Jäger, R.S. (Hrsg.). (1990). **Tests und Trends 8 (Jb. d. Päd. Diagn.)** Weinheim: Beltz.
- PSYTKOM. **Kompendium Psychologischer und Pädagogischer Testverfahren, Volltextdatenbank**. Recherchen über die Zentralstelle für Psychologische Information und Dokumentation (ZPID) an der Universität Trier, Postfach 3825, 5500 Trier, Germany.
- Weise, G. (1975). **Psychologische Leistungstests, Bd. 1: Intelligenz, Konzentration, Spezielle Fähigkeiten**. Göttingen: Hogrefe.
- Folgenden Zeitschriften können regelmäßig Informationen zu eingeführten und neuen Testverfahren entnommen werden:
- Diagnostica**. Göttingen: Hogrefe.
- Psychologie in Erziehung und Unterricht**. München: Reinhardt.
- Zeitschrift für Differentielle und Diagnostische Psychologie**. Bern: Huber.

7. Literaturverzeichnis

- Abel, J. (1988). Untersuchung zur Reliabilität, Validität und Parallelität des "Prüfsystems für Schul- und Bildungsberatung - (PSB)". *Diagnostica*, 34, 351-363.
- Abels, D. (1974). *Konzentrations-Verlaufs-Test (KVT)*. Göttingen: Hogrefe, 3. Aufl.
- Ahrbeck, B., Lommatsch, E.M. & Schuck, K.D. (1984). Der "neue" HAWIK - ein "neues" Verfahren der sonderpädagogischen Diagnostik? *Zeitschrift für Heilpädagogik*, 35, 49-58.
- Ahrens, H.J. (1982). Meßskalen und Skalierung. In K.J. Klauer (Hrsg.), *Handbuch der Pädagogischen Diagnostik, Bd. 1* (S. 99-124). Düsseldorf: Schwann.
- Ahrens, H.J. & Stäcker, K.H. (1975). Diagnostische Urteilsbildung und sprachliche Kommunikation. *Zeitschrift für Sozialpsychologie*, 6, 129-149.
- Allehoff, W.H. (1985). *Berufswahl und berufliche Interessen*. Göttingen: Hogrefe.
- Allinger, U. & Heller, K. (1975). Automatische Klassifikation von psychologischen Untersuchungsbefunden. In Kultusministerium Baden-Württemberg (Hrsg.), *Bildungsberatung in der Praxis* (= Schriftenreihe A Nr. 29 des KM Baden-Württemberg zur Bildungsforschung) (S. 142-169). Villingen: Neckarverlag.
- Allport, G.W. & Allport, F.H. (1928). *AS Reaction Study in Personality*. Boston: Houghton Mifflin.
- Amelang, M. & Bartussek, D. (1990). *Differentielle Psychologie und Persönlichkeitsforschung*. Stuttgart: Kohlhammer, 3. Aufl.
- Amthauer, R. (1972). *Ein Test zur Untersuchung des praktisch-technischen Verständnisses (PTV)*. Göttingen: Hogrefe.
- Amthauer, R. (1973). *Intelligenz-Struktur-Test (I-S-T 70)*. Göttingen: Hogrefe, 4. Aufl.
- Anastasi, A. (1954). *Psychological testing*. New York: MacMillan, 2. Aufl (1961), 4. Aufl. (1976).
- Anderson, N.H. (1968). A simple model for information integration. In R.P. Abelson, E. Aronson, W.J. McGuire, T.M. Newcomb, M.J. Rosenberg & P.H. Tannenbaum (Eds.), *Theories of cognitive consistency: A sourcebook*. Chicago: Rand McNally.
- Andre, A. (1980). *Themenwahlverfahren für 4. bis 9. Klassen (TWV 4-9)*. Weinheim: Beltz.
- Anger, H., Bargmann, R. & Hylla, E. (1965). *Wortschatztest (WST 7-8)*. Weinheim: Beltz.
- Arnold, W. (1975). *Der Pauli-Test*. München: Barth, 4. Aufl.
- Aurin, K. (1966). *Ermittlung und Erschließung von Begabungen im ländlichen Raum. Untersuchung zur Bildungsberatung in Baden-Württemberg*. Villingen: Neckarverlag.
- Aurin, K. et al. (1968). *Gleiche Chancen im Bildungsgang* (= Schriftenreihe A Nr. 9 des KM Baden-Württemberg zur Bildungsforschung). Villingen: Neckarverlag.
- Avenarius, H. (1990). *Anwendung Diagnostischer Testverfahren in der Schule. Ein Rechtsgutachten*. Weinheim: Beltz.
- Bachmann, U. & Rauchfleisch, U. (1981). Itemanalyse zum Hamburg-Wechsler-Intelligenztest für Erwachsene (HAWIE) an einer Stichprobe psychisch Kranker. *Diagnostica*, 27, 345-359.
- Bäumler, G. (1964). Zur Faktorenstruktur der Paulitestleistung unter besonderer Berücksichtigung des sogenannten numerischen Faktors. *Diagnostica*, 10, 107-119.
- Bäumler, G. (1974). *Lern- und Gedächtnistest (LGT 3)*. Göttingen: Hogrefe.
- Baier, H. (1982). Lernbehindertenpädagogik. *Zeitschrift für Heilpädagogik*, 33, 154-159.
- Ballstaedt, S.P., Mandl, H., Schnotz, W. & Tergan, S.O. (1981). *Texte verstehen, Texte gestalten*. München: Urban & Schwarzenberg.
- Bang, R. (1965). *Untersuchungen von der Intelligenz gehörloser Kinder und Jugendlicher mit dem Progressive-Matrices-Test von J.C. Raven* (Examensarbeit). Hamburg: Sonderpädagogisches Institut.
- Bartenwerfer, H. (1964). Allgemeine Leistungstests. In R. Heiß (Hrsg.), *Psychologische Diagnostik* (= Handbuch der Psychologie, Bd. 6) (S. 385-410). Göttingen: Hogrefe.
- Bartenwerfer, H. (1983). Allgemeine Leistungsdiagnostik. In K.J. Groffmann & L. Michel (Hrsg.), *Intelligenz- und Leistungsdiagnostik* (= Enzyklopädie der Psychologie, B II, Bd. 2) (S. 482-512). Göttingen: Hogrefe.

- Bartram, M. & Rollett, B. (1984). Das Konstrukt Anstrengungsvermeidung: Faulpelz im neuen Gewand? Eine Antwort auf Joppts Replik. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 16, 81-83.
- Baud, U. & Rauchfleisch, U. (1982). Zur Diagnostik hirnrorganischer Störungen mit Hilfe des Hamburg-Wechsler-Intelligenztests für Erwachsene. Eine Untersuchung zur differentialdiagnostischen Validität des HAWIE. *Diagnostica*, 28, 248-262.
- Baudisch, W., Bröse, B. & Samski, C.S. (1987). *Hilfsschulpädagogik*. Berlin: Volk und Wissen, 3. Aufl.
- Baumert, I. (1973). *Untersuchungen zur diagnostischen Valenz des HAWIK und die Entwicklung einer Kurzform WIPKI*. Bern: Huber.
- Baumgärtel, F. (1979). *Hamburger Erziehungsverhaltensliste für Mütter (HAMEL)*. Göttingen: Hogrefe.
- Bayerisches Staatsministerium für Unterricht, Kultus, Wissenschaft und Kunst. (1991). *Aufnahme ausländischer Schüler in Schulen für Behinderte* (Schreiben an die Regierungen vom 3.5.1991).
- Bechmann, M. (1982). Eine annotierte Bibliographie zur Anwendung des Hamburg-Wechsler-Intelligenztests für Erwachsene. In J.D. Matarazzo (Hrsg.), *Die Messung und Bewertung der Intelligenz Erwachsener nach Wechsler* (S. 424-509). Bern: Huber.
- Beckmann, H.K. (Hrsg.). (1978). *Leistung in der Schule*. Braunschweig: Westermann.
- Beerman, L., Heller, K.A. & Menacher, P. (1991). *Mädchen in Mathematik, Naturwissenschaft und Technik. Analyse geschlechtsspezifischer Begabungen*. Bern: Huber.
- Berg, D. & Erlwein, M. (1991). Gute Resultate im Aufmerksamkeits-Belastungs-Test (Test d2) nur aufgrund hoher Konzentrationsfähigkeit? *Psychologie in Erziehung und Unterricht*, 38, 59-62.
- Bethäuser, H. & Reichenbecher, H. (1976). Approximative Expertenabstimmung in der Bildungsberatung. In K. Heller (Hrsg.), *Handbuch der Bildungsberatung*, Bd. 3 (S. 861-877). Stuttgart: Klett.
- Betz, D. & Breuninger, H. (1987). *Teufelskreis Lernstörungen. Theoretische Grundlegung und Standardprogramm*. München: Psychologie Verlags Union, 2. Aufl.
- Bierhoff-Alfermann, D. (1976). Die Beziehung von Noten und Schülermerkmalen bei Schülern der 9. und 10. Klasse. *Psychologie in Erziehung und Unterricht*, 23, 205-214.
- Biglmaier, F. (1987). *Richtig lesen, richtig schreiben im Grundwortschatz. Ein integriertes Diagnose- und Förderprogramm*. Weinheim: Beltz.
- Binet, A. (1898). La mesure en psychologie individuelle. *Révue Philosophique*, 46, 113-123.
- Binet, A. (1905). A propos de la mesure de l'intelligence. *Année Psychologique*, 11, 69-82.
- Binet, A. & Henry, V. (1896). La psychologie individuelle. *Année Psychologique*, 2, 411-465.
- Binet, A. & Simon, T. (1908). Le développement de l'intelligence chez les enfants. *Année Psychologique*, 14, 1-94.
- Bleidick, U. (1988). *Betrifft Integration: behinderte Schüler in allgemeinen Schulen. Konzepte der Integration: Darstellung und Ideologiekritik*. Berlin: Marhold. (a)
- Bleidick, U. (1988). Sonderschule oder Allgemeine Schule? Bildungspolitische Diskussion um den zukünftigen Lernort der pädagogischen Förderung von Behinderten am Beispiel Hamburgs. *Zeitschrift für Heilpädagogik*, 39, 541-558. (b)
- Bloom, B.S. (1976). *Taxonomie von Lernzielen im kognitiven Bereich*. Weinheim: Beltz, 5. Aufl.
- Boerner, K. (1987). *Das psychologische Gutachten. Ein praktischer Leitfaden*. München: Psychologie Verlags Union, 3. Aufl.
- Bondy, C. (Hrsg.). (1956). *Hamburg-Wechsler-Intelligenztest für Kinder (HAWIK)*. Bearbeiter: F.P. Hardesty & H.J. Priester. Bern: Huber. (a)
- Bondy, C. (Hrsg.). (1956). *Hamburg-Wechsler-Intelligenztest für Erwachsene (HAWIE)*. Bearbeiter: A. Hardesty & H. Lauber. Bern: Huber. (b)
- Bondy, C., Cohen, R., Eggert, D. & Lüer, G. (1971). *Testbatterie für geistig behinderte Kinder (TBGB)*. Weinheim: Beltz, 2. Aufl.
- Booth, J.F. (1988). Kompetenz. In R.S. Jäger (Hrsg.), *Psychologische Diagnostik. Ein Lehrbuch* (S. 104-113). München: Psychologie Verlags Union.

- Borgida, E. & Nisbett, R.E. (1977). The differential impact of abstract vs. concrete information on decisions. *Journal of Applied Social Psychology*, 7, 258-271.
- Borkowski, J.G. & Krause, A. (1983). Racial differences in intelligence: the importance of the executive system. *Intelligence*, 7, 179-395.
- Borkowski, J.G., Krause, A. & Maxwell, S. (1985). On multiple determinants of racial differences in intelligence: a reply to Jensen. *Intelligence*, 9, 41-49.
- Borkowski, J.G. & Peck, V.A. (1986). Causes and consequences of metamemory in gifted children. In R.J. Sternberg & J.E. Davidson (Eds.), *Conceptions of giftedness* (pp. 182-200). Cambridge: Cambridge University Press.
- Bortz, J. (1989). *Lehrbuch der Statistik für Sozialwissenschaftler*. Berlin: Springer, 2. Aufl.
- Bourdon, B. (1955). *Bourdon-Test*. Göttingen: Hogrefe.
- Brambring, M. (1983). Spezielle Eignungsdiagnostik. In K.J. Groffmann & L. Michel (Hrsg.), *Intelligenz- und Leistungsdiagnostik* (= Enzyklopädie der Psychologie, B II, Bd. 2) (S. 414-481). Göttingen: Hogrefe.
- Brickenkamp, R. (Hrsg.). (1975). *Handbuch psychologischer und pädagogischer Tests*. Göttingen: Hogrefe.
- Brickenkamp, R. (1981). *Aufmerksamkeitsbelastungstest (Test d2)*. Göttingen: Hogrefe, 5. Aufl.
- Brickenkamp, R. (1983). *Erster Ergänzungsband zum Handbuch psychologischer und pädagogischer Tests*. Göttingen: Hogrefe.
- Brickenkamp, R. (1991). Fehlinterpretation von Testleistungen? Anmerkungen zum Beitrag "Konzentrationsleistung ohne Konzentration?". *Diagnostica*, 37, 52-57. (a)
- Brickenkamp, R. (1991). Vermeidbare Fehlinterpretationen von Konzentrationsleistungen. Anmerkungen zu einem Beitrag von Berg und Erlwein. *Psychologie in Erziehung und Unterricht*, 38, 63-65. (b)
- Brophy, J.E. & Good, T.L. (1976). *Die Lehrer-Schüler-Interaktion*. München: Urban & Schwarzenberg.
- Brown, A.L. (1978). Knowing when, where, and how to remember: A problem of metakognition. In R. Glaser (Ed.), *Advances in instructional psychology, Vol. 1* (pp. 77-165). Hillsdale, NJ: Erlbaum & Associates.
- Brown, A.L. (1984). Metakognition, Handlungskontrolle, Selbststeuerung und andere, noch geheimnisvollere Mechanismen. In F.E. Weinert & R.H. Kluwe (Hrsg.), *Metakognition, Motivation und Lernen* (S. 60-109). Stuttgart: Kohlhammer.
- Brunswik, E. (1956). *Perception and the representative design of experiments*. Berkeley: University of California Press.
- Buggle, F. & Baumgärtel, F. (1975). *Hamburger Neurotizismus- und Extraversionskala für Kinder und Jugendliche (HANES-KJ)*. Göttingen: Hogrefe.
- Bundschuh, K. (1984). *Einführung in die sonderpädagogische Diagnostik*. München: Reinhardt, 2. Aufl.
- Bunge, M. (1967). *Scientific research, Vol. I/III*. New York: Springer.
- Burgmayer, S. (1986). Die Verwendung des Teilleistungsmodells. In U. Brack (Hrsg.), *Frühdagnostik und Frühtherapie* (S. 113-121). München: Urban & Schwarzenberg. (a)
- Burgmayer, S. (1986). Neuropsychologie: Gegenstand - diagnostische Methoden - therapeutische Konsequenzen. In U. Brack (Hrsg.), *Frühdagnostik und Frühtherapie* (S. 215-231). München: Urban & Schwarzenberg. (b)
- Burt, C. (1949). The structure of the mind: a review of the results of factor analysis. *British Journal of Educational Psychology*, 19, 100-111, 176-199.
- Butsch, Ch. & Fischer, H. (1966). *Seashore-Test für musikalische Begabung*. Bern: Huber.
- Campione, J.C. (1984). Ein Wandel in der Instruktionsforschung mit lernschwierigen Kindern: Die Berücksichtigung metakognitiver Komponenten. In F.E. Weinert & R. Kluwe (Hrsg.), *Metakognition, Motivation und Lernen* (S. 109-131). Stuttgart: Kohlhammer.
- Campione, J.C. & Brown, A.L. (1978). Toward a theory of intelligence: Contributions from research with retarded children. *Intelligence*, 2, 279-304.
- Cattell, J. McK. (1890). Mental tests and measurements. *Mind*, 15, 373-381.

- Cattell, R.B. (1944). Psychological measurement: normative, ipsative, interactive. *Psychological Review*, 51, 292-303.
- Cattell, R.B. (1957). *Personality and Motivation Structure and Measurement*. New York: World Book.
- Cattell, R.B. (1960). *Culture Fair Intelligence Test*. Champaign, Ill: IPAT.
- Cattell, R.B. (1963). Theory of Fluid and Crystallized Intelligence: A Critical Experiment. *Educational Psychology*, 54, 1-22.
- Cattell, R.B. (1965). *The Scientific Analysis of Personality*. Chicago: Penguin. - Dt. Die wissenschaftliche Erforschung der Persönlichkeit. Weinheim: Beltz (1973).
- Clancey, W.J. (1988). Acquiring, representing, and evaluating a competence model of diagnostic strategy. In M.T.H. Chi, R. Glaser & M.J. Farr (Eds.), *The nature of expertise* (pp. 343-418). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Clauß, G. & Ebner, H. (1979). *Grundlagen der Statistik für Psychologen, Pädagogen und Soziologen*. Frankfurt/M.: Deutsch, 3. Aufl.
- Cohn, S.J. (1981). What is giftedness? A multidimensional approach. In A.H. Kramer (Ed.), *Gifted Children* (pp. 33-45). New York: Trillium Press.
- Conrad, W., Baumann, E. & Mohr, V. (1980). *Mannheimer Test zur Erfassung des physikalisch-technischen Problemlösens (MTP)*. Göttingen: Hogrefe.
- Conrad, W., Eberle, G., Hornke, L., Kierdorf, B. & Nagel, B. (1976). *Mannheimer Intelligenztest für Kinder und Jugendliche (MIT-KJ)*. Weinheim: Beltz.
- Cooley, W.W. & Lohnes, P.R. (1966). *Multivariate Procedures for the Behavioral Sciences*. New York: Wiley, 4. Aufl.
- Cooley, W.W. & Lohnes, P.R. (1971). *Multivariate data analysis*. New York: Wiley.
- Cronbach, L.J. (1957). The two disciplines of scientific psychology. *American Psychologist*, 12, 671-684.
- Cronbach, L.J. & Gleser, R. (1965). *Psychological Tests and personnel decisions*. Urbana: University of Illinois Press, 2. Aufl.
- Cronbach, L.J. & Snow, R.E. (1977). *Aptitudes and instructional methods*. New York: Irvington.
- Czeschlik, T. & Rost, D.H. (1988). Hochbegabte und ihre Peers. *Zeitschrift für Pädagogische Psychologie*, 2, 1-23.
- Dahl, G. (1968). *Übereinstimmungsvalidität des HAWIE und Entwicklung einer revidierten Testform*. Meisenheim am Glan: Hain.
- Dahl, G. (1972). *WIP. Reduzierter Wechsler-Intelligenztest. Anwendung - Auswertung - statistische Analysen - Normwerte*. Meisenheim am Glan: Hain.
- Dahme, G. (1985). *Giftedness, creativity and high intelligence as seen by teachers*. Paper presented at the 6th World Conference on Gifted and Talented Children in Hamburg, Germany.
- Damm, H., Hylla, E. & Schäfer, K. (1965). *Rechtschreibtest 8+ (RST 8+)*. Neubearbeitung von H. Horn. Weinheim: Beltz.
- Daniels, J.C. (1981). *Figure Reasoning Test (FRT)*. Göttingen: Hogrefe.
- Davidson, J.E. & Sternberg, R.J. (1984). The role of insight in intellectual giftedness. *Gifted Child Quarterly*, 28, 58-64.
- Dawes, R.M. (1979). The robust beauty of improper linear models in decision making. *American Psychologist*, 34, 571-582.
- de Finetti, B. (1970). Logical foundations and measurement of subjective probability. *Acta Psychologica*, 34, 129-145.
- Deutscher Bildungsrat. (1973). *Zur pädagogischen Förderung behinderter und von Behinderung bedrohter Kinder und Jugendlicher. Empfehlungen der Bildungskommission*. Bonn-Bad Godesberg.
- DeVree, J.K. (1983). The behavioral function: An inquiry into the relation between behavior and utility. *Theory and Decision*, 15, 231-245.
- Diemand, A., Schuler, H. & Stapf, K.H. (1991). Zum Einsatz eines Lerntests bei Ingenieurstudenten - eine Pilotstudie. *Zeitschrift für Arbeits- und Organisationspsychologie*, 35, 15-22.

- Diétel, B. (1987). Sensorische Integration nach Jean Ayres. Einige kritische Anmerkungen. *Der Kinderarzt*, 18, 1360-1369.
- Diétel, B. (1988). Neuropsychologische Diagnostik und Therapie. *Behindertenpädagogik in Bayern*, 31, 183-197. (a)
- Diétel, B. (1988). Sonderpädagogische Diagnose- und Förderklassen in Bayern - Versuch eines ersten Resümées aus der Sicht der sozialwissenschaftlichen Begleitung. *Behindertenpädagogik in Bayern*, 31, 393-409. (b)
- Dieterich, R. (1977). *Psychodiagnostik: Grundlagen und Probleme*. München: Reinhardt, 2. Aufl.
- Dingel, W. (1971). Mitteilungen über Zuverlässigkeitsaspekte des Test d2. *Diagnostica*, 17, 84-86.
- Dörner, D., Kreuzig, H.W., Reither, F. & Stäudel, T. (1983). *Lohhausen: Vom Umgang mit Unbestimmtheit und Komplexität*. Bern: Huber.
- Dreesmann, H. (1986). Psychologie der Lernumwelt. In B. Weidenmann & A. Krapp et al. (Hrsg.), *Pädagogische Psychologie* (S. 447-491). München: Urban & Schwarzenberg.
- Düker, H. & Lienert, G. (1965). *Konzentrations-Leistungs-Test (KLT)*. Göttingen: Hogrefe, 2. Aufl.
- Dumke, D. (1977). Die Auswirkungen von Lehrererwartungen auf Intelligenz und Schulleistungen. *Psychologie in Erziehung und Unterricht*, 24, 93-108.
- Ebbinghaus, H. (1885). *Über das Gedächtnis*. Leipzig: Duncker & Humblot.
- Ebbinghaus, H. (1897). Über eine neue Methode zur Prüfung geistiger Fähigkeiten und ihre Anwendung bei Schulkindern. *Zeitschrift für Psychologie*, 13, 401-459.
- Edelstein, W. & Hopf, D. (Hrsg.). (1973). *Bedingungen des Bildungsprozesses*. Stuttgart: Klett.
- Edwards, W. (1961). *Behavioral decision theory* (Bd. 12, S. 473-498). In Annual Review of Psychology (Hrsg.).
- Edwards, W., Lindman, H. & Savage, L.J. (1963). Bayesian statistical inference for psychological research. *Psychological Review*, 70, 193-242.
- Eggert, D. (Hrsg.). (1972). *Zur Diagnose der Minderbegabung. Ein Handbuch und Textbuch zur Testbatterie für geistig behinderte Kinder (TBGB)*. Weinheim: Beltz.
- Eggert, D. (Hrsg.). (1978). *Hannover Wechsler Intelligenztest für das Vorschulalter (HAWTVA)*. Deutsche Bearbeitung der Wechsler Preschool and Primary Scale of Intelligence. Bern: Huber, 2. Aufl.
- Eggert, D., Liman, E. & Schirmacher, A. (1984). Vergleich des Hamburg-Wechsler-Intelligenztests für Kinder (HAWIK) mit der revidierten Version (HAWIK-R) bei sprachbehinderten Kindern. *Zeitschrift für Heilpädagogik*, 35, 54-58.
- Eggert, D. & Schuck, K.D. (1973). *Gruppenintelligenztest für lernbehinderte Sonderschüler (CMM-LB)*. Weinheim: Beltz.
- Einhorn, H.J. & Hogarth, R.M. (1978). Confidence in judgment: Persistence of the illusion of validity. *Psychological Review*, 85, 395-416.
- Einhorn, H.J. & Hogarth, R.M. (1982). Behavioral decision theory: Processes of judgment and choice. In G.R. Ungson & D.N. Braustein (Eds.), *Decision making: An interdisciplinary inquiry* (pp. 15-41). Boston, MA: Kent.
- Eisebitt, R. (1986). Eine vergleichende Validierungsstudie des KFT 4-13 und des P-S-B anhand von Fallgutachten der Schulberatung. *Psychologie in Erziehung und Unterricht*, 33, 38-45.
- Ellger-Rüttgardt, S. (Hrsg.). (1990). *Bildungs- und Sozialpolitik für Behinderte*. München: Reinhardt.
- Emmet, W.G. & Jugenkamp, K. (1966). *Bildertest 2-3 (BT 2-3)*. Weinheim: Beltz.
- Engelbrecht, W. (1975). Validierung einer Berufseignungs-Testbatterie und Verwendung der Ergebnisse für eine computerunterstützte berufsbezogene Testbefundinterpretation. *Diagnostica*, 21, 3-24, 97-106.
- Engelbrecht, W. (1978). Weiterentwicklung der maschinellen Testbefundinterpretation zur EUB-Testbatterie. *Diagnostica*, 24, 39-49.
- Eser, K.H. (1987). Reliabilitäts- und Validitätsaspekte des Aufmerksamkeits-Belastungs-Tests (Test d2) bei verhaltensgestörten Kindern und Jugendlichen. *Diagnostica*, 33, 74-80.

- Evans, L. (1966). A Comparative Study of the Wechsler Intelligence Scale for Children (Performance) and Raven's Progressive Matrices with Deaf Children. *The Teacher of the Deaf*, 69, 76-82.
- Ewert, O. (1973). Level Two - Level One ... Zero. Eine Nachprüfung der Zweistufentheorie der Intelligenz von A. Jensen. In W. Edelstein & D. Hopf (Hrsg.), *Bedingungen des Bildungsprozesses* (S. 222-230). Stuttgart: Klett.
- Eysenck, H.J. (1979). *The structure and measurement of intelligence*. Berlin: Springer.
- Eysenck, H.J. (1980). *Intelligenz. Struktur und Messung*. New York: Springer.
- Faber, E. & Nollau, W. (1969). Über einen Algorithmus zur mehrdimensionalen Diskriminanzanalyse. *Schriftenreihe des DRZ, Heft 5-5*. Darmstadt: DRZ. (a)
- Faber, E. & Nollau, W. (1969). Über ein Verfahren zur automatischen Klassifikation. *Schriftenreihe des DRZ, Heft 5-6*. Darmstadt: DRZ. (b)
- Facaoaru, C. (1985). *Kreativität in Wissenschaft und Technik*. Bern: Huber.
- Facaoaru, C. (1990). Untersuchungsbericht zum "Test der Zahlenreihen und -analogien" (TZRA). In K.A. Heller (Hrsg.), *Abschlußbericht zum Forschungsprojekt "Formen der Hochbegabung bei Kindern und Jugendlichen" an das BMBW, Anhang B 27*. München: LMU.
- Facaoaru, C. & Bittner, R. (1987). Kognitionspsychologische Ansätze der Hochbegabungsdiagnostik. *Zeitschrift für Differentielle und Diagnostische Psychologie*, 8, 193-205.
- Faßnacht, G. (1979). *Systematische Verhaltensbeobachtung*. München: Reinhardt.
- Feger, B. (1986). Selective Bibliography: 'Identification of the Intellectually Gifted'. In K.A. Heller & J.F. Feldhusen (Eds.), *Identification and Nurturing the Gifted* (pp. 163-178). Toronto: Huber.
- Feger, B. (1987). Spezialprobleme bei der Identifikation Hochbegabter aus Risikogruppen. *Zeitschrift für Differentielle und Diagnostische Psychologie*, 8, 227-233.
- Feger, B. (1988). *Hochbegabung*. Bern: Huber.
- Fehnmann, U. (1988). Recht. In R.S. Jäger (Hrsg.), *Psychologische Diagnostik. Ein Lehrbuch* (S. 96-104). München: Psychologie Verlags Union.
- Feldhusen, J.F. (1986). A Conception of Giftedness. In K.A. Heller & J.F. Feldhusen (Eds.), *Identifying and Nurturing the Gifted. An International Perspective* (pp. 33-38). Toronto: Huber.
- Festinger, L. (1957). *A theory of cognitive dissonance*. Stanford, Ca: Stanford University Press.
- Feuerstein, R. (1980). *Instrumental enrichment. An intervention program for cognitive modifiability*. Baltimore: University Park Press.
- Feuerstein, R., Hoffman, M.B., Jensen, M.R. & Rand, Y. (1985). Instrumental enrichment, an intervention program for structural cognitive modifiability: Theorie and practice. In J.W. Segal, S.F. Chipman, & R. Glaser (Eds.), *Thinking and learning skills, Vol. 1* (pp. 43-82). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Fippingger, F. (1971). *Allgemeiner Schulleistungstest für 3. Klassen (AST 3)*. Weinheim: Beltz.
- Fischer, G. (1974). *Einführung in die Theorie psychologischer Tests*. Bern: Huber.
- Fischer, G.H. (1977). Linear logistic models for the description of attitudinal and behavioral changes under the influence of mass communication. In W.H. Kempf & B.H. Repp (Eds.), *Some mathematical models for social psychology* (pp. 102-151). Bern: Huber.
- Fishburn, P.C. (1981). Subjective expected utility: A review of normative theories. *Theory and Decision*, 13, 139-199.
- Fishburn, P.C. (1986). The axioms of subjective probability. *Statistical Science*, 1, 335-358.
- Fisseni, H.J. (1982). *Persönlichkeitsbeurteilung. Zur Theorie und Praxis des psychologischen Gutachtens*. Göttingen: Hogrefe.
- Fisseni, H.J. (1990). *Lehrbuch der psychologischen Diagnostik*. Göttingen: Hogrefe.
- Flammer, A. (1978). Wechselwirkung zwischen Schülermerkmalen und Unterrichtsmethoden - eine zerronnene Hoffnung? In H. Mandl & A. Krapp (Hrsg.), *Schuleingangsdiagnose* (S. 113-120). Göttingen: Hogrefe.
- Flavell, J.H. (1971). First discussant's comments: What is memory development the development of? *Human Development*, 14, 272-278.

- Flavell, J.H. (1984). Annahmen zum Begriff Metakognition sowie zur Entwicklung von Metakognitionen. In F.E. Weinert & R.H. Kluwe (Hrsg.), *Metakognition, Motivation und Lernen* (S. 23-31). Stuttgart: Kohlhammer.
- Flavell, J.H. & Wellman, H.M. (1977). Metamemory. In R.V. Kail, Jr. & J.W. Hagen (Eds.), *Perspectives on the development of memory and cognition* (pp. 3-33). Hillsdale, NJ: Erlbaum Associates.
- Forer, B.R. (1949). The fallacy of personal validation: A classroom demonstration of gullibility. *Journal of Abnormal and Social Psychology*, 44, 118-123.
- Fried, L. (1982). Sprachdiagnostik im Vorschulalter. In K. Ingenkamp, R. Horn & R.S. Jäger (Hrsg.), *Tests und Trends 1982* (Jb. d. Päd. Diagn.) (S. 43-69). Weinheim: Beltz.
- Fried, L. (1986). Zur Diagnose des sprachlichen Entwicklungsstandes unter besonderer Berücksichtigung des Ausländerkindes. In K. Ingenkamp & R.S. Jäger (Hrsg.), *Tests und Trends 5* (Jb. d. Päd. Diagn.) (S. 63-91). Weinheim: Beltz.
- Friedrich, R. & Müller, H. (1987). Zur Kulturfairneß der Progressiven Matrizen von Raven. Anmerkungen zu Taschinskis Untersuchung mit türkischen Kindern in Deutschland. *Psychologie in Erziehung und Unterricht*, 24, 227-231.
- Fürntratt, E. (1969). *Differentieller Wissenstest (DWT)*. Göttingen: Hogrefe.
- Gadene, V. (1988). Hypothesen, Erklärungen und Prognosen in der Psychologischen Diagnostik. In R.S. Jäger (Hrsg.), *Psychologische Diagnostik. Ein Lehrbuch* (S. 276-285). München: Psychologie Verlags Union.
- Gaedike, A.K. (1976). *Untersuchungen zur Validität des Kognitiven Fähigkeits-Tests für 4. bis 13. Klassen (KFT 4-13)*. Weinheim: Beltz.
- Gaedike, A.K. (1978). Determinanten der Schulleistung. In K. Heller (Hrsg.), *Leistungsbeurteilung in der Schule* (S. 46-93). Heidelberg: Quelle & Meyer, 3. Aufl.
- Gaensslen, H. & Schubö, W. (1976). *Einfache und komplexe statistische Analyse*. München: Reinhardt, 2. Aufl.
- Gärtner-Harnach, V. (1972). *Fragebogen für Schüler (FS 11-13)*. Weinheim: Beltz.
- Gärtner-Harnach, V. (1973). *Fragebogen für Schüler (FS 5-10)*. Weinheim: Beltz.
- Gagné, F. (1985). Giftedness and talent: Reexamining a reexamination of the definitions. *Gifted Child Quarterly*, 29, 103-112.
- Galton, F. (1869). *Hereditary genius: An inquiry into its laws and consequences*. London: McMillan & Co.
- Gardner, H. (1983). *Frames of mind. The theory of multiple intelligences*. New York: Basic Books.
- Gaußmann, A., Hochhausen, R. & Schmidt-Rogge, I. (1978). Der Mehrfach-Wahl-Wortschatztest (MWT) und der Zahlen-Verbindungs-Test (ZVT) als Maße der Allgemeinen Intelligenz. *Diagnostica*, 24, 50-77.
- Geisler, H.J. (1981). *Untersuchungen zur Validität von Intelligenztests im Primarbereich. Validierungsstudie zum KFT 1-3* (Dissertation). Köln: Univ., Phil. Fak.
- Gigerenzer, G. (1981). *Messung und Modellbildung in der Psychologie*. München: Reinhardt.
- Gigerenzer, G. (1984). Messung, Modellbildung und die "Kognitive Wende". In M. Amelang & H.J. Ahrens (Hrsg.), *Brennpunkte der Persönlichkeitsforschung* (S. 49-65). Göttingen: Hogrefe.
- Gözlü, L. (1986). *Zur Anwendung von nonverbalen Intelligenztests bei türkischen Grundschulkindern* (Dissertation). Köln: Univ. Köln, Phil. Fak.
- Goldfried, M.R. & Kent, R.N. (1976). Herkömmliche gegenüber verhaltenstheoretischer Persönlichkeitsdiagnostik: ein Vergleich methodischer und theoretischer Voraussetzungen. In D. Schulte (Hrsg.), *Diagnostik in der Verhaltenstherapie* (S. 3-23). München: Urban & Schwarzenberg, 2. Aufl.
- Graumann, C.F. (1973). Grundzüge der Verhaltensbeobachtung. In C.F. Graumann & H. Heckhausen (Hrsg.), *Pädagogische Psychologie, Bd. 1* (Grundlagentexte zum Funkkolleg) (S. 14-41). Frankfurt/M.: Fischer.
- Grissemann, H. (1989). *Lernbehinderung heute. Psychologisch-anthropologische Grundlagen einer innovativen Lernbehindertenpädagogik*. Bern: Huber.

- Groffmann, K.J. (1983). Die Entwicklung der Intelligenzmessung. In K.J. Groffmann & L. Michel (Hrsg.), *Intelligenz- und Leistungsdiagnostik* (= Enzyklopädie der Psychologie, B II, Bd. 2) (S. 1-103). Göttingen: Hogrefe.
- Groffmann, K.J. & Michel, L. (Hrsg.). (1982). *Grundlagen psychologischer Diagnostik* (= Enzyklopädie der Psychologie, B II, Bd. 1). Göttingen: Hogrefe.
- Guilford, J.P. (1956). The structure of intellect. *Psychological Bulletin*, 53, 267-293.
- Guilford, J.P. (1959). *Personality*. New York: McGraw Hill. - Dt. Persönlichkeit. Weinheim: Beltz, 3. Aufl. (1965).
- Guilford, J.P. (1966). Intelligence: 1965 model. *American Psychologist*, 21, 20-26.
- Guilford, J.P. (1967). *The nature of human intelligence*. New York: McGraw Hill.
- Guilford, J.P. (1976). Aptitude for creative thinking: One or many? *Journal of Creative Behavior*, 10, 165-169.
- Guilford, J.P. & Hoepfner, R. (1971). *The Analysis of Intelligence*. New York: McGraw Hill. - Dt. Analyse der Intelligenz. Weinheim: Beltz (1976).
- Gulliksen, H. (1950). *Theory of Mental Tests*. New York: Wiley.
- Guthke, J. (1974). *Zur Diagnostik der intellektuellen Lernfähigkeit*. Berlin: Deutscher Verlag der Wissenschaften, 2. Aufl./Stuttgart: Klett (1977), 3. Aufl.
- Guthke, J. (1980). *Ist Intelligenz meßbar?* Berlin: Deutscher Verlag der Wissenschaften.
- Guthke, J. & Al-Zoubi, A. (1987). Kulturspezifische Differenzen in den Coloured Progressive Matrices (CPM) und in einer Lernstvariante der CPM. *Psychologie in Erziehung und Unterricht*, 34, 306-311.
- Guthke, J., Jäger, Ch. & Schmidt, I. (1983). *Lernstbatterie "Schlußfolgerndes Denken" (LTS)*. Berlin: Psychodiagnostisches Zentrum.
- Guthke, J., Räder, E., Caruso, M. & Schmidt, K.D. (1991). Entwicklung eines adaptiven computergestützten Lernstests auf der Basis der strukturellen Informationstheorie. *Diagnostica*, 37, 1-28.
- Guthke, J. & Wohlrab, U. (1982). Der gegenwärtige internationale Entwicklungsstand, Hauptergebnisse, Probleme und Trends in der Lernstforschung. In J. Guthke, U. Wohlrab, M. Caruso, A. Harnisch, G. Lehwald, S. Michalski & K.H. Müller (Hrsg.), *Neuere Ergebnisse der Lernstforschung - Diagnostische Programme als Lernstvariante* (S. 4-23). Leipzig: Karl-Marx-Universität, Sektion Psychologie. (a)
- Guthke, J. & Wohlrab, U. (1982). Zielstellung und Konstruktionsprinzipien diagnostischer Programme. In J. Guthke, U. Wohlrab, M. Caruso, A. Harnisch, G. Lehwald, S. Michalski & K.H. Müller (Hrsg.), *Neuere Ergebnisse der Lernstforschung - Diagnostische Programme als Lernstvariante* (S. 24-43). Leipzig: Karl-Marx-Universität, Sektion Psychologie. (b)
- Haenisch, H. & Klaghofer, R. (1979). Zusammenhänge zwischen Tests und Noten am Ende von Grundschule und Orientierungsstufe. *Psychologie in Erziehung und Unterricht*, 26, 39-43.
- Hagen, E. (1989). *Die Identifizierung Hochbegabter*. Heidelberg: Asanger.
- Halpern, D.F. (1984). *Thought and knowledge: An introduction to critical thinking*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Hampton, J.M., Moore, P.G. & Thomas, H. (1973). Subjective probability and its measurement. *Journal of the Royal Statistical Society*, A136/1, 21-42.
- Hany, E.A. (1987). *Modelle und Strategien zur Identifikation hochbegabter Schüler* (Dissertation). München: LMU, Fak. Psychol. u. Päd. (a)
- Hany, E.A. (1987). Psychometrische Probleme bei der Identifikation Hochbegabter. *Zeitschrift für Differentielle und Diagnostische Psychologie*, 8, 173-191. (b)
- Hany, E.A. (1988). Programmevaluation in der Hochbegabtenförderung. *Psychologie in Erziehung und Unterricht*, 35, 241-255.
- Hany, E.A. (1990). *The computerized Test of Spatial Design (CTSD): a new device for measuring intelligence, problem solving ability, and working style*. Paper presented at the Second ECHA Conference in Budapest, October 25-28, 1990.
- Hany, E.A. (1991). Identifikation hochbegabter Schüler. In K.A. Heller (Hrsg.), *Formen der Hochbegabung bei Kindern und Jugendlichen (Teil II)*. Göttingen: Hogrefe. (a)

- Hany, E.A. (1991). Sind Lehrkräfte bei der Identifikation hochbegabter Schüler doch besser als Tests? *Psychologie in Erziehung und Unterricht*, 38, 37-50. (b)
- Hany, E.A. & Heller, K.A. (1991). Gegenwärtiger Stand der Begabungsforschung. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 23, 241-249.
- Hany, E.A. & Rennen-Allhoff, B. (1987). Spezielle Tests zur kognitiven Entwicklung. In B. Rennen-Allhoff, P. Allhoff, E.A. Hany & U. Schmidt-Deuter (Hrsg.), *Entwicklungstests für das Stüglings-, Kleinkind- und Vorschulalter* (S. 215-304). Berlin: Springer.
- Hardesty, F.P. & Priester, H.J. (1966). *Handbuch für den Hamburg-Wechsler-Intelligenztest für Kinder (HAWIK)*. Bern: Huber, 3. Aufl.
- Hartje, W. & Orgass, B. (1972). Bewährung einer HAWIE-Kurzform (WIP nach Dahl) bei hirngeschädigten Patienten I. *Zeitschrift für experimentelle und angewandte Psychologie*, 19, 309-324.
- Hartje, W. & Orgass, B. (1974). Bewährung einer HAWIE-Kurzform (WIP nach Dahl) bei hirngeschädigten Patienten II. Numerische Übereinstimmung zwischen WIP und HAWIE. *Diagnostica*, 20, 22-30.
- Hartmann, H. (1970). *Psychologische Diagnostik. Auftrag - Testsituation - Gutachten*. Stuttgart: Kohlhammer, 2. Aufl. (1973).
- Hartmann, H.A. (1984). Zur Ethik gutachterlichen Handelns. In H. Hartmann & R. Haubl (Hrsg.), *Psychologische Begutachtung. Problembereiche und Praxisfelder* (S. 3-32). München: Urban & Schwarzenberg.
- Hasemann, K. (1964). Verhaltensbeobachtung. In R. Heiß (Hrsg.), *Psychologische Diagnostik* (= Handbuch der Psychologie, Bd. 6) (S. 807-836). Göttingen: Hogrefe.
- Hasemann, K. (1983). Verhaltensbeobachtung und Ratingverfahren. In K.J. Groffmann & L. Michel (Hrsg.), *Verhaltensdiagnostik* (= Enzyklopädie der Psychologie, B II, Bd. 4) (S. 434-488). Göttingen: Hogrefe.
- Hasselhorn, M. (1987). Kognitive Bedingungen der Leistungsdefizite lernschwacher Schüler bei Gedächtnisanforderungen. *Zeitschrift für Pädagogische Psychologie*, 1, 91-98.
- Hatch, T. C. & Gardner, H. (1986). From testing intelligence to assessing competencies: A pluralistic view of intellect. Special Issue: The IQ controversy. *Roepers Review*, 8, 147-150.
- Haubl, R. (1984). Praxeologische und epistemologische Aspekte psychologischer Begutachtung. In H. Hartmann & R. Haubl (Hrsg.), *Psychologische Begutachtung. Problembereiche und Praxisfelder* (S. 33-74). München: Urban & Schwarzenberg.
- Heckhausen, H. (1989). *Motivation und Handeln*. Berlin: Springer, 2. Aufl.
- Heck-Möhling, R. (1986). *Konzentrationstest für 3. und 4. Klassen (KT 3-4)*. Weinheim: Beltz.
- Heiß, R. (1964). Technik, Methodik und Problematik des Gutachtens. In R. Heiß (Hrsg.), *Psychologische Diagnostik* (S. 975-995). Göttingen: Hogrefe.
- Heller, K. (1967). PMT-Normen für hörgeschädigte Kinder. *Neue Blätter für Taubstummeneildung*, 21, 226-234.
- Heller, K. (1969). Zum Problem der Begabungsreserven. In H.R. Lückert (Hrsg.), *Begabungsforschung und Bildungsförderung als Gegenwartsaufgabe* (S. 352-430). München: Reinhardt.
- Heller, K. (1970). *Aktivierung der Bildungsreserven*. Bern: Huber.
- Heller, K.A. (1973). *Intelligenzmessung*. Villingen: Neckarverlag.
- Heller, K. (1975). Beratung beim Übergang von der Primarstufe zur Sekundarstufe I. In K. Heller (Hrsg.), *Handbuch der Bildungsberatung, Bd. 2* (S. 473-490). Stuttgart: Klett. (a)
- Heller, K. (1975). Untersuchung zur Schuleignungsermittlung in Mannheim. In Kultusministerium Baden-Württemberg (Hrsg.), *Bildungsberatung in der Praxis* (= Schriftenreihe A Nr. 29 des KM Baden-Württemberg zur Bildungsforschung) (S. 69-107). Villingen: Neckarverlag.
- Heller, K. (1976). Computerunterstützte Interpretation von Testbefunden in der Schullaufbahnberatung. In K. Heller (Hrsg.), *Handbuch der Bildungsberatung, Bd. 3* (S. 879-900). Stuttgart: Klett.
- Heller, K. (1976). *Intelligenz und Begabung*. München: Reinhardt.
- Heller, K.A. (1979). Beurteilung und Beratung beim Übergang in die Sekundarstufe. In D. Bolscho & C. Schwarzer (Hrsg.), *Beurteilen in der Grundschule* (S. 185-206). München: Urban & Schwarzenberg.

- Heller, K.A. (Hrsg.). (1984). *Leistungsdiagnostik in der Schule*. Bern: Huber, 4. Aufl. (a)
- Heller, K.A. (1984). Schülerbeurteilung und Schullaufbahnberatung. In K.A. Heller (Hrsg.), *Leistungsdiagnostik in der Schule* (S. 299-307). Bern: Huber, 4. Aufl. (b)
- Heller, K.A. (1986). Psychologische Diagnostik im Bereich von Schule und Erziehung. In H.D. Olbrich, H. Karlusch & J. Palme (Hrsg.), *Bericht über den 26. Kongreß des Berufsverbandes Österreichischer Psychologen* (S. 75-99). Wien: Literas.
- Heller, K.A. (Hrsg.). (1987). *Hochbegabungsdagnostik*. Bern: Huber. (a)
- Heller, K.A. (1987). Möglichkeiten und Grenzen der Diagnostik von Hochbegabung. In F.E. Weinert & H. Wagner (Hrsg.), *Die Förderung Hochbegabter in der Bundesrepublik Deutschland: Probleme, Positionen, Perspektiven* (S. 106-120). Bad Honnef: Bock. (b)
- Heller, K.A. (1987). Perspektiven einer Hochbegabungsdagnostik. *Zeitschrift für Differentielle und Diagnostische Psychologie*, 8, 159-172. (c)
- Heller, K.A. (1990). Zielsetzung, Methoden und Ergebnisse der Münchner Längsschnittstudie zur Hochbegabung. *Psychologie in Erziehung und Unterricht*, 37, 85-100.
- Heller, K.A. (Hrsg.). (1991). *Formen der Hochbegabung im Kindes- und Jugendalter*. Göttingen: Hogrefe.
- Heller, K. & Schirmer, B. (1973). *Wortschatztests für Sehbehinderte (WST 4-9)*. Weinheim: Beltz.
- Heller, K.A., Nickel, H. & Rosemann, B. (1978). *Psychologie in der Erziehungswissenschaft, Bd. 4: Beurteilen und Beraten*. Stuttgart: Klett-Cotta, 2. Aufl.
- Heller, K.A., Rosemann, B. & Steffens, K. (1978). *Prognose des Schulerfolgs. Eine Längsschnittstudie zur Schullaufbahnberatung*. Weinheim: Beltz.
- Heller, K.A., Nickel, H. & Neubauer, W. (1980). *Psychologie in der Erziehungswissenschaft, Bd. 1: Verhalten und Lernen*. Stuttgart: Klett-Cotta, 3. Aufl.
- Heller, K. & Rosemann, B. (1981). *Planung und Auswertung empirischer Untersuchungen*. Stuttgart: Klett-Cotta, 2. Aufl.
- Heller, K. & Nickel, H. (Hrsg.). (1982). *Modelle und Fallstudien zur Erziehungs- und Schulberatung*. Bern: Huber.
- Heller, K.A. & Geisler, H.J. (1983). *Kognitiver Fähigkeits-Test - Grundschulform (KFT 1-3)*. Weinheim: Beltz. (a)
- Heller, K.A. & Geisler, H.J. (1983). *Kognitiver Fähigkeits-Test - Kindergartenform (KFT-K)*. Weinheim: Beltz. (b)
- Heller, K.A. et al. (1985). Technologiemodelle als diagnostische Planungs- und Entscheidungshilfen für die psychologische Beratung. In D. Albert (Hrsg.), *Bericht über den 34. Kongreß der Deutschen Gesellschaft für Psychologie in Wien 1984, Bd. 2* (S. 713-717). Göttingen: Hogrefe.
- Heller, K., Gaedike, A.K. & Weinläder, H. (1985). *Kognitiver Fähigkeitstest (KFT 4-13+)*. Weinheim: Beltz, 2. Aufl.
- Heller, K.A., Heyse, H. et al. (1985). *Analyse von Schulleistungsschwierigkeiten* (Studienbrief 5 des Fernstudienlehrgangs "Ausbildung zum Beratungslehrer"). Tübingen: Deutsches Institut für Fernstudien. (a)
- Heller, K.A., Heyse, H. et al. (1985). *Dimensionen und Bedingungsfaktoren der Schulleistung* (Studienbrief 4 des Fernstudienlehrgangs "Ausbildung zum Beratungslehrer"). Tübingen: Deutsches Institut für Fernstudien. (b)
- Heller, K.A., Heyse, H. et al. (1985). *Pädagogisch-psychologische Diagnostik: Beratungsanlässe, Funktionen, Untersuchungsvariablen und Methoden* (Studienbrief 3 des Fernstudienlehrgangs "Ausbildung zum Beratungslehrer"). Tübingen: Deutsches Institut für Fernstudien. (c)
- Heller, K.A. & Hany, E.A. (1986). Identification, Development and Achievement Analysis of Talented and Gifted Children in West Germany. In K.A. Heller & J.F. Feldhusen (Eds.), *Identifying and Nurturing the Gifted* (pp. 67-82). Toronto: Huber.
- Helmke, A. (1990). *Leistungsbezogenes Selbstvertrauen und schulische Leistungen* (Habilitationsschrift). München: LMU, Fak. Psychol. u. Päd.

- Helmke, A., Schrader, F.W. & Lehneis-Klepper, G. (1991). Zur Rolle des Elternverhaltens für die Schulleistungsentwicklung ihrer Kinder. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 23, 1-22.
- Herbig, M. (1982). Lehrzielorientierte Tests und klassische Testtheorie. In K.J. Klauer (Hrsg.), *Handbuch der Pädagogischen Diagnostik, Bd. 1* (S. 127-136). Düsseldorf: Schwann.
- Hermans, H.J.M. (1976). *Leistungsmotivationstest für Jugendliche (LMT-J)*. Deutsche Fassung von U. Undeutsch. Amsterdam: Swets & Zeitlinger.
- Hermelin, B. & O'Connor, N. (1986). Idiot savant calculators: rules and regularities. *Psychological Medicine*, 16, 1-9.
- Hermelin, B., O'Connor, N. & Lee, S. (1987). Musical inventiveness of five idiots-savants. *Psychological Medicine*, 17, 79-90.
- Herrmann, T. (1979). Pädagogische Psychologie als psychologische Technologie. In J. Brandtstädter, G. Reinert & K.A. Schneewind (Hrsg.), *Pädagogische Psychologie: Probleme und Perspektiven* (S. 209-236). Stuttgart: Klett-Cotta.
- Hetzer, H. & Tent, L. (1971). *Weilburger Testaufgaben für Schulanfänger (WTA)*. Weinheim: Beltz.
- Hofer, M. (1970). Zur impliziten Persönlichkeitstheorie von Lehrern. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 2, 197-209.
- Hoffman, P.J. (1960). The paramorphic representation of clinical judgment. *Psychological Bulletin*, 47, 116-131.
- Holtz, K.L. (1988). Wissen, Lernen und Wissen über Lernen. *Zeitschrift für Heilpädagogik*, 41, Beiheft 14, 16-36.
- Holzcamp, K. (1966). Begutachtung als Kommunikation. In K. Holzcamp, A.O. Jäger & F. Merz (Hrsg.), *Prognose und Bewährung in der psychologischen Diagnostik* (S. 19-40). Göttingen: Hogrefe.
- Horn, R. (1984). Zum Problem der Lernzieldefinition. In K.A. Heller (Hrsg.), *Leistungsdiagnostik in der Schule* (S. 145-153). Bern: Huber.
- Horn, W. (1956). *Begabungstestsystem (BTS)*. Göttingen: Hogrefe, 2. Aufl. (1972).
- Horn, W. (1962). *Leistungsprüfsystem (L-P-S)*. Göttingen: Hogrefe, 2. Aufl. (1983).
- Horn, W. (1969). *Prüfsystem für Schul- und Bildungsberatung (P-S-B)*. Göttingen: Hogrefe.
- Horst, P. (1971). *Messung und Vorhersage*. Weinheim: Beltz.
- Hubel, S. & Schuker, G. (1990). *Metagedächtnis und Strategienutzung bei Sonderschülern mit Lernbehinderungen. Eine Interventionsstudie*. Icking und München: Unveröffentl. Manuskript.
- Hylla, E. & Kraak, B. (1968). *Aufgaben zum Nachdenken (AzN 4+)*. Weinheim: Beltz (Neueichung von R. Kühn, 1976).
- Hylla, E., Süllwold, F. & Wicht, G. (1970). *Rechtschreibtest (RST 4+)*. Neubearbeitung von E. Schwarz. Weinheim: Beltz, 2. Aufl.
- Ingenkamp, K. (Hrsg.). (1976). *Bildertest 2-3 (BT 2-3)*. Weinheim: Beltz.
- Ingenkamp, K. (1985). *Lehrbuch der Pädagogischen Diagnostik*. Weinheim: Beltz.
- Ingenkamp, K., Wolf, B., Christmann, H., Lissmann, U., Knapp, A. & Haenisch, H. (1977). *Bildungs-Beratungs-Test (konvergentes Denken) für 4. bis 6. Klassen (BBT 4-6)*. Weinheim: Beltz.
- Irle, M. (1955). *Berufs-Interessen-Test (BIT)*. Göttingen: Hogrefe.
- Irle, M. & Allehoff, W. (1984). *Berufs-Interessen-Test II (BIT II)*. Göttingen: Hogrefe.
- Jacobs, B. & Strittmatter, P. (1979). *Der schulängstliche Schüler. Eine empirische Untersuchung über mögliche Ursachen und Konsequenzen der Schulangst*. München: Urban & Schwarzenberg.
- Jäger, A.O. (1973). *Dimensionen der Intelligenz*. Göttingen: Hogrefe, 3. Aufl.
- Jäger, A.O. (1982). Mehrmodale Klassifikationen von Intelligenzleistungen: Experimentell kontrollierte Weiterentwicklung eines deskriptiven Intelligenzstrukturmodells. *Diagnostica*, 28, 195-225.
- Jäger, A.O. (1984). Intelligenzstrukturforschung: Konkurrierende Modelle, neue Entwicklungen, Perspektiven. *Psychologische Rundschau*, 35, 21-35.

- Jäger, A.O. (1986). Validität von Intelligenztests. *Diagnostica*, 32, 272-289.
- Jäger, A.O. & Althoff, K. (1983). *Der Wilde-Intelligenz-Test (WIT)*. Göttingen: Hogrefe.
- Jäger, Ch. (1987). Diagnostische Aufgaben bei der Arbeit mit chronisch auffälligen Kindern. In G. Witzlack, J. Guthke, Ch. Jäger, H. Klemm, & G. Matthes (Hrsg.), *Einführung in die Psychodiagnostik in der Schule* (S. 171-217). Berlin: Volk und Wissen, 4. Aufl.
- Jäger, R. (1978). *Differentielle Diagnostizierbarkeit in der Psychologischen Diagnostik*. Göttingen: Hogrefe.
- Jäger, R., Berbig, E., Geisel, B., Goslar, H., Hagen, J., Liebich, W. & Schafheutle, R. (1973). *Mannheimer Biographisches Inventar (MBI)*. Göttingen: Hogrefe.
- Jäger, R.S. (1980). Testbesprechung: Arbeitsverhaltensinventar von Thiel, Keller & Binder. *Psychologie in Erziehung und Unterricht*, 27, 316-319.
- Jäger, R.S. (1981). Einige Bemerkungen zum AVI sowie eine Kritik der Ausführungen von Thiel und Keller. *Diagnostica*, 27, 266-273. (a)
- Jäger, R.S. (1981). Testbesprechung: Arbeitsverhaltensinventar (AVI). In R.S. Jäger, K. Ingenkamp & G. Stark (Hrsg.), *Tests und Trends 1981* (Jb. d. Päd. Diagn.) (S. 126-128) Beltz. (b)
- Jäger, R.S. (1981). Testbesprechung: RAVEN-Matrizen test. Coloured Progressive Matrices (CPM). In R.S. Jäger, K. Ingenkamp & G. Stark (Hrsg.), *Tests und Trends 1981* (Jb. d. Päd. Diagn.) (S. 128-129). Weinheim: Beltz. (b)
- Jäger, R.S. (1983). *Der diagnostische Prozeß. Eine Diskussion psychologischer und methodischer Randbedingungen*. Göttingen: Hogrefe.
- Jäger, R.S. (Hrsg.). (1988). *Psychologische Diagnostik. Ein Lehrbuch*. München: Psychologie Verlags Union.
- Janke, W. (1982). Klassenzuordnung: Zuordnung von Personen zu vorgegebenen Klassen. In K.J. Groffmann & L. Michel (Hrsg.), *Grundlagen psychologischer Diagnostik* (= Enzyklopädie der Psychologie, B II, Bd. 1) (S. 376-466). Göttingen: Hogrefe.
- Jaspens, N. (1946). Serial correlation. *Psychometrika*, 11, 23-30.
- Jensen, A.R. (1969). How much can we boost IQ and scholastic achievement? In H. Skowronek (Hrsg.), *Umwelt und Begabung* (S. 1-123). Stuttgart: Klett.
- Jensen, A.R. (1985). The nature of the black-white difference on various psychometric tests: Spearman's hypothesis. *Behavioral and Brain Sciences*, 8, 193-219.
- Jöreskog, K.G. (1979). Statistical estimation of Structural Models in Longitudinal-Developmental Investigations. In J.R. Nesselroade & P.B. Baltes (Eds.), *Longitudinal Research in the Study of Behavior and Development* (pp. 303-351). New York: Academic Press.
- Jopt, U.J. (1982). Anstrengungsvermeidungstendenz: "Faulpelz" im neuen Gewand? *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 14, 324-336.
- Jopt, U.J. (1984). Anstrengungsvermeidungstendenz: Motiv ohne Motivation oder Ein Test ersetzt noch keine Theorie. Replik auf B. Rollett. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 16, 77-80.
- Kaemmerer, W. (1980). Allgemeine und spezielle psychosomatische Fragen und Probleme. In E. Petzold & A. Reindell (Hrsg.), *Klinische Psychosomatik* (S. 164-176). Heidelberg: Quelle & Meyer.
- Kahneman, D., Slovic, P. & Tversky, A. (Eds.). (1982). *Judgment under uncertainty: Heuristics and biases*. Cambridge: Cambridge University Press.
- Kahneman, D. & Tversky, A. (1972). Subjective probability: A judgment of representativeness. *Cognitive Psychology*, 3, 430-454.
- Kahneman, D. & Tversky, A. (1973). On the psychology of prediction. *Psychological Review*, 80, 237-251.
- Kail, R. & Pellegrino, J.W. (1985). *Human Intelligence*. New York: W.H. Freeman. - Dt. Menschliche Intelligenz. Heidelberg: Spektrum der Wissenschaft (1988).
- Kaminski, G. (1970). *Verhaltenstheorie und Verhaltensmodifikation*. Stuttgart: Klett.
- Kaminski, G. (1976). Rahmentheoretische Überlegungen zur Taxonomie psychodiagnostischer Prozesse. In K. Pawlik (Hrsg.), *Diagnose der Diagnostik* (S. 45-70). Stuttgart: Klett-Cotta.

- Kanfer, F.H. (1979). Selbstmanagement-Methoden. In F.H. Kanfer & A.P. Goldstein (Hrsg.), *Möglichkeiten der Verhaltensänderung* (S. 350-406). München: Urban & Schwarzenberg, 2. Nachdruck.
- Kant, I. (1788). *Grundlegung zur Metaphysik der Sitten*. Stuttgart: Reclam (orig. 1785).
- Kardorff, von E. (1978). Modellvorstellungen über psychische Störungen: Gesellschaftliche Entstehung, Auswirkungen, Probleme. In H. Keupp & M. Zaumseil (Hrsg.), *Die gesellschaftliche Organisation psychischen Leidens. Zum Arbeitsfeld klinischer Psychologen* (S. 539-590). Frankfurt am Main: Suhrkamp.
- Karnes, M.B. & Johnson, L.J. (1986). Early identification and programming for young gifted/talented handicapped. *Topics in Early Childhood Special Education*, 6, 50-61.
- Kastner, M. (1978). Zur Problematik von Tests zum Übergang an weiterführende Schulen. *Psychologie in Erziehung und Unterricht*, 25, 9-15.
- Kautter, H. (1982). Der Übergang zu Sonderschulen. In K. J. Klauer (Hrsg.), *Handbuch der Pädagogischen Diagnostik, Bd. 2* (S. 977-988). Düsseldorf: Schwann.
- Kautter, H. & Storz, L. (1972). *Schulleistungstestbatterie für Lernbehinderte und für schulleistungsschwache Grundschüler (SBL I)*. Weinheim: Beltz. (a)
- Kautter, H. & Storz, L. (1972). *Schulleistungstestbatterie für Lernbehinderte und für schulleistungsschwache Grundschüler (SBL II)*. Weinheim: Beltz. (b)
- Keller, G., Binder, A. & Thiel, R. (1979). *Arbeitsverhaltensmodifikation (AVM)*. Braunschweig: Westermann.
- Keßler, B.H. (1982). Biographische Diagnostik. In K.J. Groffmann & L. Michel (Hrsg.), *Persönlichkeitsdiagnostik* (= Enzyklopädie der Psychologie, B II, Bd. 3) (S. 1-55). Göttingen: Hogrefe.
- Keupp, H. (1976). *Abweichung und Alltagsroutine. Die Labeling-Perspektive in Theorie und Praxis*. Hamburg: Hoffmann und Campe.
- Kiphard, E.J. & Schilling, F. (1974). *Körperkoordinationstest (KTK)*. Weinheim: Beltz.
- Klauer, K.J. (1987). *Kriteriumsorientierte Tests*. Göttingen: Hogrefe.
- Klauer, K.J. (1989). *Denktraining für Kinder I. Ein Programm zur intellektuellen Förderung*. Göttingen: Hogrefe.
- Klayman, J. & Ha, Y.W. (1987). Confirmation, disconfirmation, and information in hypothesis testing. *Psychological Review*, 94, 211-228.
- Kleber, E.W. (1978). *Lehrbuch der sonderpädagogischen Diagnostik*. Berlin: Marhold, 3. Aufl.
- Kleber, E.W. (1982). Konzentration. In K.J. Klauer (Hrsg.), *Handbuch der Pädagogischen Diagnostik, Bd. 1* (S. 395-401). Düsseldorf: Schwann.
- Kleber, E.W. (1984). Der Hawik-R und die Umschulungsdiagnostik. *Zeitschrift für Heilpädagogik*, 35, 713-722.
- Kleber, E.W. & Kleber, G. (1974). *Differentieller Leistungstest - KE (DL-KE)*. Braunschweig: Westermann.
- Kleber, E.W. & Kleber, G. (1975). *Differentieller Leistungstest - KG (DL-KG)*. Braunschweig: Westermann.
- Kluwe, R.H. & Schiebler, K. (1984). Entwicklung exekutiver Prozesse und kognitiver Leistungen. In F.E. Weinert & R.H. Kluwe (Hrsg.), *Metakognition, Motivation und Lernen* (S. 31-59). Stuttgart: Kohlhammer.
- Knaack, R. (1978). Zur Brauchbarkeit der Coloured Progressive Matrices (CPM) von Raven bei Kindern im Vorschulalter. *Psychologie in Erziehung und Unterricht*, 25, 159-167.
- Kobi, E.E. (1985). Heilpädagogische Diagnostik. *Vierteljahresschrift für Heilpädagogik und ihre Nachbargebiete*, 54, 238-255.
- Koch, U.K. & Lotz, W. (1988). Schwerstbehinderte: Erkennung und Förderung hoher Begabung. In Stiftung zur Förderung körperbehinderter Hochbegabter (Hrsg.), *Symposiumbericht*. Zürich, Vaduz.
- Kormann, A. (Hrsg.). (1987). *Beurteilen und Fördern in der Erziehung. Orientierungshilfen bei Erziehungs- und Schulproblemen*. Salzburg: Müller. (a)

- Kormann, A. (1987). Mathematikschwierigkeiten und Möglichkeiten ihrer Behebung. In A. Kormann (Hrsg.), *Beurteilen und Fördern in der Erziehung. Orientierungshilfen bei Erziehungs- und Schulproblemen* (S. 140-160). Salzburg: Müller. (b)
- Kormann, A. (1987). Neue Intelligenz und Schulfähigkeitstests. In A. Kormann (Hrsg.), *Beurteilen und Fördern in der Erziehung. Orientierungshilfen bei Erziehungs- und Schulproblemen* (S. 57-76). Salzburg: Müller. (c)
- Kormann, A. (1988). Schullaufbahnprobleme bei ausländischen Schülern. In A. Kormann (Hrsg.), *Diagnose und Intervention bei Schullaufbahnproblemen* (S. 74-84). Weinheim: Beltz.
- Kornadt, H.J. & Zumkley, H. (1982). Thematische Apperzeptionsverfahren. In K.J. Groffmann & L. Michel (Hrsg.), *Persönlichkeitsdiagnostik* (= Enzyklopädie der Psychologie, B II, Bd. 3) (S. 258-372). Göttingen: Hogrefe.
- Kornmann, R. (1977). *Testbatterie für entwicklungsrückständige Schulanfänger (TES)*. Weinheim: Beltz.
- Kornmann, R. (1982). Förderplan für einen Schüler mit Lernbehinderungen. In K.A. Heller & H. Nickel (Hrsg.), *Modelle und Fallstudien zur Erziehungs- und Schulberatung* (S. 61-77). Bern: Huber. (a)
- Kornmann, R. (1982). Von der Auslesediagnostik zur Förderdiagnostik. *Behindertenpädagogik*, 21, 293-309. (b)
- Kornmann, R. (1983). *Diagnose von Lernbehinderungen. Strategie und Methoden im Überweisungsverfahren zur Sonderschule für Lernbehinderte*. Weinheim: Beltz, 3. Aufl.
- Kornmann, R. (1986). Förderdiagnostisch orientierte Variation der Testbedingungen bei ausgewählten Aufgaben des HAWIK. *Zeitschrift für Heilpädagogik*, 37, 674-684.
- Kornmann, R. (1990). Wie Förderdiagnostik zur Gestaltung von Übungen der Rechenfertigkeit genutzt werden kann. *Zeitschrift für Heilpädagogik*, 41, 102-108.
- Kornmann, R. (1991). Förderungsdiagnostik bei ausländischen Kindern. *Psychologie in Erziehung und Unterricht*, 38, 133-151.
- Kornmann, R., Meister, H. & Schlee, J. (1986). *Förderungsdiagnostik: Konzepte und Realisierungsmöglichkeiten*. Heidelberg: Schindele, 2. Aufl.
- Kornmann, R. & Müller, H.P. (1982). Förderdiagnostik im Schulkindergarten. In K.A. Heller & H. Nickel (Hrsg.), *Modelle und Fallstudien zur Erziehungs- und Schulberatung* (S. 89-98). Bern: Huber.
- Kornmann, R. & Zickwolf, A. (1985). Möglichkeiten qualitativer Leistungsdiagnostik durch Variation der Testbedingungen am Beispiel von HAWIK-Items. In D. Albert (Hrsg.), *Bericht über den 34. Kongreß der Deutschen Gesellschaft für Psychologie in Wien 1984* (S. 61-66). Göttingen: Hogrefe.
- Krampen, G. (1981). *IPC-Fragebogen zu Kontrollüberzeugungen*. Göttingen: Hogrefe.
- Krantz, D.H. & Tversky, A. (1971). Conjoint measurement analysis of composition rules in psychology. *Psychological Review*, 78, 151-169.
- Krapp, A. (1973). *Bedingungen des Schulerfolgs*. München: Oldenburg.
- Krapp, A. (1976). Bedingungsfaktoren der Schulleistung. *Psychologie in Erziehung und Unterricht*, 23, 91-109.
- Krapp, A. (1979). *Prognose und Entscheidung*. Weinheim: Beltz.
- Krapp, A. (1984). Forschungsergebnisse zur Bedingungsstruktur der Schulleistung. In K.A. Heller (Hrsg.), *Leistungsdiagnostik in der Schule* (S. 46-62). Bern: Huber, 4. Aufl.
- Krapp, A. (1986). Diagnose und Prognose. In B. Weidenmann, A. Krapp, M. Hofer, G.L. Huber & H. Mandl (Hrsg.), *Pädagogische Psychologie. Ein Lehrbuch* (S. 565-630). München: Psychologie Verlags Union.
- Krapp, A. & Mandl, H. (1976). Vorhersage und Erklärung der Schulleistung. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 8, 192-219.
- Krathwohl, D.R., Bloom, B.S. & Masia, B.B. (1978). *Taxonomie von Lernzielen im affektiven Bereich*. Weinheim: Beltz, 2. Aufl.
- Kratzmeier, H. (1975). *Reutlinger Test für Schulanfänger (RTS)*. Weinheim: Beltz.
- Kratzmeier, H. (1982). *Heidelberger Intelligenztest (HIT 3-4)*. Weinheim: Beltz.
- Kratzmeier, H. & Horn, R. (1980). *Advanced Progressive Matrices (APM)*. Weinheim: Beltz.

- Kratzmeier, H. & Horn, R. (1987). *Standard Progressive Matrices (SPM)*. Weinheim: Beltz, 2. Aufl.
- Kritsch, K. & Jahn, J. (1982). Eine Untersuchung zur HAWIK-Kurzform WIPKI. *Zeitschrift für Heilpädagogik*, 33, 102-104.
- Krohne, H.W. (1985). Entwicklungsbedingungen von Ängstlichkeit und Angstbewältigung: Ein Zweiprozess-Modell elterlicher Erziehungswirkung. In H.W. Krohne (Hrsg.), *Angstbewältigung in Leistungssituationen* (S. 135-160). Weinheim: VCH.
- Krüger, H. & Zietz, K. (1933). Das Verifikationsproblem. Experimentelle Untersuchungen über die psychologischen Grundlagen der Bestätigung von Charaktergutachten. *Zeitschrift für angewandte Psychologie*, 45, 140-171.
- Kruglanski, A.W., Baldwin, M.W. & Towson, S.M.J. (1985). Die Theorie der Laienepistemologie. In D. Frey & M. Irle (Hrsg.), *Theorien der Sozialpsychologie. Band III: Motivations- und Informationsverarbeitungstheorien* (S. 293-314). Bern: Huber.
- Kubinger, K.D. (1983). Verfehlte Einwände gegen den HAWIK als Differential-Diagnostikum - Eine Antwort auf Wallasch und Dony. *Diagnostica*, 29, 118-135. (a)
- Kubinger, K. (1983). *Der HAWIK - Möglichkeiten und Grenzen seiner Anwendung*. Weinheim: Beltz. (b)
- Kubinger, K. (1986). Kognitiver Fähigkeits-Test (KFT 4-13+). Testrezension. *Zeitschrift für Differentielle und Diagnostische Psychologie*, 7, 56-58.
- Kubinger, K.D. (1988). *Moderne Testtheorie. Ein Abriss samt neuesten Beiträgen*. München: Psychologie Verlags Union. (a)
- Kubinger, K.D. (1988). Testtheorie: Probabilistische Modelle. In R.S. Jäger (Hrsg.), *Psychologische Diagnostik* (S. 264-276). München: Psychologie Verlags Union. (b)
- Kubinger, K.D. & Wurst, E. (1988). *Adaptives Intelligenz Diagnostikum (AID)*. Weinheim: Beltz, 2. Aufl.
- Kühn, R. (1982). Testbesprechung: Anstrengungsvermeidungstest (AVT) von B. Rollett & M. Bartram. In K. Ingenkamp, R. Horn & R.S. Jäger (Hrsg.), *Tests und Trends 1982* (Jb. d. Päd. Diagn.) (S. 149-150). Weinheim: Beltz.
- Kühn, R. (1983). *Bedingungen für Schulerfolg. Zusammenhänge zwischen Schülermerkmalen, häuslicher Umwelt und Schulnoten*. Göttingen: Hogrefe.
- Kühn, R. (1984). Vorhersagbarkeit von Schulnoten mit Hilfe zweier Intelligenztests. *Zeitschrift für erziehungs- und sozialwissenschaftliche Forschung*, 1, 169-180.
- Kühn, R. (1985). *Zur differentiellen Vorhersagbarkeit von Schulerfolg. Eine Erkundungsstudie zur Auffindung von Moderator-effekten*. Frankfurt/M.: Deutsches Institut für Internationale Pädagogische Forschung, 2. Aufl.
- Kühn, R. (1987). Welche Vorhersage des Schulerfolgs ermöglichen Intelligenztests? Eine Analyse gebräuchlicher Verfahren. In R. Horn, K. Ingenkamp & R.S. Jäger (Hrsg.), *Tests und Trends 6* (Jb. d. Päd. Diagn.) (S. 26-64). München, Weinheim: Psychologie Verlags Union.
- Kühn, R. & Heck-Möhling, R. (1976). *Bildertest 1-2 (BT 1-2)*. Weinheim: Beltz, 3. Aufl.
- Kuhl, J. (1983). *Motivation, Konflikt und Handlungskontrolle*. Berlin: Springer.
- Kvale, S. (1972). *Prüfung und Herrschaft*. Weinheim: Beltz.
- Lambert, B. (1982). *Validierungsuntersuchungen zum Kognitiven Fähigkeits-Test (Kindergartenform)* (Unveröffentl. Staatsexamensarbeit). Köln: Univ. Köln, Phil. Fak.
- Langer, I. & Schulz v. Thun, F. (1974). *Messung komplexer Merkmale in Psychologie und Pädagogik*. München: Reinhardt.
- Langfeldt, H.P. (1977). *Determinanten von Bildungsempfehlungen* (Dissertation). Bonn: Univ., Päd. Fak.
- Langfeldt, H.P. (1984). Die klassische Testtheorie als Grundlage normorientierter (standardisierter) Schulleistungstests. In K. Heller (Hrsg.), *Leistungsdiagnostik in der Schule* (S. 40-98). Bern: Huber, 4. Aufl.
- Langfeldt, H.P. (1988). Sonderpädagogische Diagnostik: Allgemeine Grundlagen und Funktionen. *Sonderpädagogik*, 18, 67-76.

- Langfeldt, H.P. & Fingerhut, W. (1984). Empirische Ansätze zur Aufklärung des Konstruktes "Schulleistung". In K.A. Heller (Hrsg.), *Leistungsdiagnostik in der Schule* (S. 40-45). Bern: Huber, 4. Aufl.
- Langfeldt-Nagel, M. (1982). *Der Einfluß der Schulbildung auf Intelligenztestleistungen*. Bern: Lang. (a)
- Langfeldt-Nagel, M. (1982). Über die Vorhersage spezifischer Intelligenzleistungen. *Zeitschrift für Differentielle und Diagnostische Psychologie*, 3, 147-160. (b)
- Langfeldt-Nagel, M. (1982). Untersuchungen zur Konstruktvalidität der Grundintelligenztests (CFT) von Cattell und Weiss. *Diagnostica*, 28, 65-79. (c)
- Langhorst, E. (1984). Beobachtung und Beurteilung des Schülerverhaltens. In K.A. Heller (Hrsg.), *Leistungsdiagnostik in der Schule* (S. 208-228). Bern: Huber, 4. Aufl.
- Lauth, G.W. (1988). *Trainingsmanual zur Vermittlung kognitiver Fertigkeiten bei retardierten Kindern*. Oldenburg: Universität Oldenburg, Zentrum für pädagogische Berufspraxis (zu beziehen über die Deutsche Gesellschaft für Verhaltenstherapie e.V.), 2. Aufl.
- Lauth, G. & Wiedl, K.H. (1985). Zur Veränderbarkeit der Testleistung im CFT-20 durch Instruktionsintensivierung. *Diagnostica*, 31, 200-209.
- Laux, L., Schaffner, P. & Glanzmann, P. (1981). *State-Trait Angstinventar (STAI)*. Weinheim: Beltz.
- Lederle-Schenk, U. & Marschner, G. (1980). Schulübergreifende Untersuchungen mit dem Kombinierten Schultest KS 4/5. *Psychologie in Erziehung und Unterricht*, 27, 148-153.
- Lehwald, G. (1985). *Zur Diagnostik des Erkenntnisstrebens bei Schülern* (= Beiträge zur Psychologie, Bd. 20). Berlin: Volk u. Wissen.
- Lehwald, G. (1987). Theoretisch-methodologische Positionen zur Diagnostik im Kleinkind- und Vorschulalter. In U. Schaarschmidt (Hrsg.), *Neue Trends der Psychodiagnostik*. Berlin: Volk u. Wissen.
- Leichner, R. (1979). *Psychologische Diagnostik. Grundlagen, Kontroversen, Praxisprobleme*. Weinheim, Basel: Beltz.
- Leichner, R. (1983). Zum Prozeß der diagnostischen Urteilsbildung. In W.R. Minsal & R. Scheller (Hrsg.), *Brennpunkte der Klinischen Psychologie, Band V: Diagnostik* (S. 125-149). München: Kösel.
- Lenk, H. (1972). *Erklärung, Prognose, Planung. Skizzen zu Brennpunktproblemen der Wissenschaftstheorie*. Freiburg: Rombach.
- Lenz, W. (1981). *Medizinische Genetik*. Stuttgart: Thieme, 5. Aufl.
- Liebert, R.M. & Morris, L.W. (1967). Cognitive and emotional components of test anxiety: A distinction and some initial data. *Psychological Reports*, 20, 975-978.
- Lienert, G. (1964). *Form-Lege-Test (FLT)*. Göttingen: Hogrefe, 2. Aufl.
- Lienert, G.A. (1969). *Testaufbau und Testanalyse*. Weinheim: Beltz, 3. Aufl.
- Lienert, G.A. (1970). *Mechanisch-technischer-Verständnistest (MTVT)*. Göttingen: Hogrefe, 2. Aufl.
- Lienert, G.A. & Fickert, H. (1958). *Analyse des verbalen Teils des HAWIE* (unveröffentl. Manuskript). Marburg: Univ. Marburg.
- Lindquist, E.F. (Ed.). (1951). *Educational measurement*. Washington: American Council on Education, 5. Aufl. (1963).
- Lockowandt, O. (1974). *Frostigs Entwicklungstest der visuellen Wahrnehmung (FEW)*. Weinheim: Beltz.
- Löwe, A. & Heller, K. (1972). *Heidelberger Hörprüf-Bild-Test (HHBT)*. Villingen: Neckarverlag.
- Lord, F.M. & Novick, M.R. (1968). *Statistical theories of mental test scores*. Reading, Mass.: Addison-Wesley.
- Lorenz, J.H. (1990). Erscheinungsbild und Diagnose von Rechenschwächen. In K. Ingenkamp & R.S. Jäger (Hrsg.), *Tests und Trends 8* (Jb. d. Päd. Diagn.) (S. 95-127). Weinheim: Beltz.
- Lukesch, H. (1975). *Erziehungsstile, pädagogische und psychologische Konzepte*. Stuttgart: Kohlhammer.
- Magnuson, D. (1969). *Testtheorie*. Wien: Deuticke.

- Mainberger, U. (1977). *Test zum divergenten Denken für 4.-6. Klassen (TDK 4-6)*. Weinheim: Beltz.
- Mandl, H. (1975). *Kognitive Entwicklungsverläufe von Grundschulern*. München: Oldenbourg.
- Marjoribanks, K. (1973). Umwelt, soziale Schicht und Intelligenz. In C.F. Graumann & H. Heckhausen (Hrsg.), *Reader zum Funkkolleg Pädagogische Psychologie, Bd. 1* (S. 190-200). Frankfurt/M.: Fischer.
- Marschner, G. (1980). *Revisions-Test (Rev.T.)*. Göttingen: Hogrefe.
- Marschner, G. (1982). Untersuchungen zur Reliabilität und Re-Test-Stabilität des "Prüfsystems für die Schul- und Bildungsberatung (PSB)" (von Horn). *Diagnostica*, 28, 263-272.
- Marschner, G. (1985). Untersuchungen zur Re-Test-Stabilität des PSB (Horn) - Vergleich 1981/1983. *Diagnostica*, 31, 164-167.
- Masendorf, F., Tücke, M., Kretschmann, R. & Bartram, M. (1976). *Dortmunder Skala zur Erfassung von Lehrerverhalten durch Schüler (DSL)*. Braunschweig: Westermann.
- Matarazzo, J.D. (1982). *Die Messung und Bewertung der Intelligenz Erwachsener nach Wechsler*. Bern: Huber.
- Medley, D.M. & Mietzel, H.E. (1965). Measuring Classroom Behavior by Systematic Observation. In N.L. Gage, *Handbook of Research on Teaching*. Chicago: Rand McNally.
- Meehl, P.E. (1954). *Clinical versus statistical prediction: A theoretical analysis and a review of the evidence*. Minneapolis: Univ. of Minnesota Press.
- Meehl, P.E. (1956). Wanted - a good cookbook. *American Psychologist*, 11, 263-272.
- Meili, R. (1957). *Analytischer Intelligenztest (AIT)*. Bern: Huber.
- Meili, R. (1964). Die faktorenanalytische Interpretation der Intelligenz. *Schweizerische Zeitschrift für Psychologie*, 23, 135-155.
- Meili, R. & Steingrüber, H.J. (1978). *Lehrbuch der psychologischen Diagnostik*. Bern: Huber, 6. Aufl.
- Meis, R. (1967). *Kettwiger Schulreifetest (KST)*. Weinheim: Beltz.
- Meis, R. (1973). *Duisburger Vorschul- und Einschulungstest (DVET)*. Weinheim: Beltz.
- Mellone, M.A., Thomson, G.H. & Horn, H. (1967). *Bildertest 1-2 (BT 1-2)*. Weinheim: Beltz.
- Michael, W.B. (1969). Prediction. In R.L. Ebel (Ed.), *Encyclopedia of educational research* (pp. 982-993). London: MacMillan.
- Michel, L. & Conrad, W. (1982). Theoretische Grundlagen psychometrischer Tests. In K.J. Groffmann & L. Michel (Hrsg.), *Grundlagen psychologischer Diagnostik* (= Enzyklopädie der Psychologie, B II, Bd. 1) (S. 1-129). Göttingen: Hogrefe.
- Mierke, K. (1963). *Begabung, Bildung und Bildsamkeit*. Bern, Stuttgart: Huber, Klett.
- Mittenecker, E. (1982). Subjektive Tests zur Messung der Persönlichkeit. In K.J. Groffmann & L. Michel (Hrsg.), *Persönlichkeitsdiagnostik* (= Enzyklopädie der Psychologie, B II, Bd. 3) (S. 57-131). Göttingen: Hogrefe.
- Möbus, C. (1978). Zur Fairness psychologischer Intelligenztests: Ein unlösbares Trilemma zwischen Zielen von Gruppen, Individuen und Institutionen? *Diagnostica*, 24, 191-234.
- Möbus, C. (1983). Die praktische Bedeutung der Testfairness als zusätzliches Kriterium zu Reliabilität und Validität. In R. Horn, K. Ingenkamp & R.S. Jäger (Hrsg.), *Tests und Trends 3* (Jb. d. Päd. Diagn.) (S. 155-203). Weinheim: Beltz.
- Möhling, R. & Raatz, U. (1974). *Konzentrationstest für das erste Schuljahr (KT 1)*. Weinheim: Beltz.
- Mönikes, D. & Mönikes, W. (1975). Der HAWIK, ein unzulängliches Prüfinstrument in der Schule für Lernbehinderte im Vergleich zum CFT 2 - eine Voruntersuchung. *Zeitschrift für Heilpädagogik*, 8, 435-466.
- Mönks, F.J. (1987). Einzelfallanalyse in der Hochbegabungsdiagnostik. *Zeitschrift für Differentielle und Diagnostische Psychologie*, 8, 235-240.
- Mönks, F.J., Boxel, H.W. van, Roelofs, J.J.W. & Sanders, M.P.M. (1986). The identification of gifted children in secondary education and a description of their situation. In K.A. Heller & J.F. Feldhusen (Eds.), *Identifying and Nurturing the Gifted* (pp. 39-65). Toronto: Huber.

- Montgomery, H. & Svenson, O. (1976). On decision rules and information processing strategies for choices among multiattribute alternatives. *Scandinavian Journal of Psychology*, 17, 283-291.
- Moosbrugger, H. (1988). Testtheorie: Klassische Ansätze. In R.S. Jäger (Hrsg.), *Psychologische Diagnostik* (S. 253-264). München: Psychologie Verlags Union.
- Müller, A. (1984). Verkehrspsychologie: Begutachtung der Fahrtauglichkeit. In H. Hartmann & R. Haubl (Hrsg.), *Psychologische Begutachtung. Problembereiche und Praxisfelder* (S. 306-328). München: Urban & Schwarzenberg.
- Müller, H. (1990). Ist das "KLA-Modell" wirklich eine Alternative zum RASCH-Modell? In D. Frey (Hrsg.), *Bericht über den 37. Kongreß der Deutschen Gesellschaft für Psychologie in Kiel 1990, Bd. 1* (S. 373f). Göttingen: Hogrefe.
- Müller, R. (1980). *Diagnostisches Soziogramm (DSO)*. Braunschweig: Westermann.
- Munz, W. (1984). Der HAWIK-R: Ein Verfahren für die sonderpädagogische Diagnostik? *Sonderpädagogik*, 14, 145-159.
- Murken, J.D. & Cleve, H. (1979). *Humangenetik*. Stuttgart: Enke, 2. Aufl.
- Mynatt, C.R., Doherty, M.E. & Tweney, R.D. (1977). Confirmation bias in a simulated research environment: An experimental study of scientific inference. *Quarterly Journal of Experimental Psychology*, 29, 85-95.
- Nestle, W. (1976). Didaktik der Sonderpädagogik. *Zeitschrift für Heilpädagogik*, 27, 167-180.
- Neuhäuser, G. (1982). *Genetische Aspekte der Behinderung. Eine Einführung für pädagogische, medizinische und verwandte Berufe*. Berlin: Marhold.
- Neukäter, H. & Schröder, U. (1991). Metakognition bei Kindern aus Schulen für Lernbehinderte und Verhaltensgestörte im Vergleich mit Grundschulkindern. *Sonderpädagogik*, 21, 12-27.
- Nickel, H. (1982). Das Beratungsgespräch mit Eltern und Schülern. In K.A. Heller & H. Nickel (Hrsg.), *Modelle und Fallstudien zur Erziehungs- und Schulberatung* (S. 15-25). Bern: Huber.
- Nisbett, R. & Ross, L. (1980). *Human inference: Strategies and shortcomings of social judgment*. Englewood Cliffs, NJ: Prentice-Hall.
- Norden, I. (1953). *Das Binetarium - Intelligenzprüfung nach Binet-Bobertag*. Göttingen: Hogrefe.
- Nuber, F., & Riediger, B. (1985). Das Verfahren zur Überweisung ausländischer Kinder an Schulen für Lernbehinderte in der Bundesrepublik Deutschland - ein Vergleich. *Zeitschrift für Heilpädagogik*, 36, 790-803.
- Oehlschlägel, J. & Moosbrugger, H. (1991). Konzentrationsleistung ohne Konzentration? Zur Schätzung wahrer Leistungswerte im Aufmerksamkeits-Belastungs-Test d2. *Diagnostica*, 37, 42-51.
- Ort, B. (1988). Meßtheoretische Grundlagen der Diagnostik. In R.S. Jäger (Hrsg.), *Psychologische Diagnostik* (S. 233-241). München: Psychologie Verlags Union.
- Osterland, J. (1976). Verhaltensbeobachtung von Schülern mit Hilfe von Ratingskalen. In K. Heller (Hrsg.), *Handbuch der Bildungsberatung, Bd. 3: Methoden der Bildungsberatung und Bildungsforschung* (S. 805-816). Stuttgart: Klett.
- Oswald, W.D. & Roth, E. (1987). *Der Zahlen-Verbindungs-Test (ZVT). Ein sprachfreier Intelligenz-Test zur Messung der "kognitiven Leistungsgeschwindigkeit"*. Göttingen: Hogrefe, 2. Aufl.
- Otto, J. (1984). Testbesprechung: Kognitiver Fähigkeits-Test, Kindergartenform (KFT-K). *Diagnostica*, 30, 244-246.
- Parsons, T. (1951). *The social system*. Glencoe, IL: Free Press.
- Patry, J.L. & Perrez, M. (1982). Entstehungs-, Erklärungs- und Anwendungszusammenhang technologischer Regeln. In J.L. Patry (Hrsg.), *Feldforschung* (S. 389-412). Stuttgart: Huber.
- Pawlik, K. (1970). *Psychologische Diagnostik* (Vorlesungsskript). Hamburg: Univ. Hamburg.
- Pawlik, K. (1982). Modell- und Praxisdimensionen psychologischer Diagnostik. In K. Pawlik (Hrsg.), *Diagnose der Diagnostik: Beiträge zur Diskussion der psychologischen Diagnostik in der Verhaltensmodifikation* (S. 13-44). Stuttgart: Klett, 2. Aufl.
- Pegnato, C.W. & Birch, J.W. (1959). Locating gifted children in junior high schools - a comparison of methods. *Exceptional children*, 25, 300-304.

- Pellegrino, J.W. & Ingram, A.L. (1977). *Components of verbal analogy solution*. Chicago: Midwestern Psychological Association.
- Perleth, Ch. (1985). *Zur Verwendbarkeit des KFT 4-13 an Berufsschulen. Validitätsprüfungen und Normierung an einer Stichprobe bayerischer Berufsschüler* (Unveröffentl. Staatsexamensarbeit). München: LMU, Fak. Psychol. u. Päd.
- Perleth, Ch. (1991). Verfahren zur Erfassung hochbegabungsrelevanter Merkmale. In K.A. Heller (Hrsg.), *Formen der Hochbegabung bei Kindern und Jugendlichen (Teil II, Kapitel 2)*. Göttingen: Hogrefe.
- Perleth, Ch. & Geisler, H.J. (1986). Normierung und Validitätsprüfungen zum KFT 4-13 für Berufsschüler. *Psychologie in Erziehung und Unterricht*, 33, 141-152.
- Perleth, Ch., Schauer, S., Hofmann, U., Volk, H. & Wernberger, H. (1989). *Intelligence testing in a Bavarian Comprehensive School*. Paper presented at the XII. International School Psychology Colloquium, Ljubljana.
- Perleth, Ch., Schuker G. & Hubel, S. (in Druck). *Metagedächtnis und Strategienutzung bei Sonderschülern mit Lernbehinderung: Eine Interventionsstudie*. Vortrag auf dem VDS Kongress in Würzburg (Mai 1991).
- Perleth, Ch. & Sierwald, W. (1991). Entwicklungs- und Leistungsanalysen zur Hochbegabung. In K.A. Heller (Hrsg.), *Formen der Hochbegabung bei Kindern und Jugendlichen (Teil III)*. Göttingen: Hogrefe.
- Perleth, Ch., Averina, I. & Scheblanowa, H. (1991). *Measuring intelligence in gifted German and Russian students: Results from a cross-cultural project* (Unveröffentl. Institutsbericht). München: LMU, Fak. Psychol. u. Päd.
- Perrez, M., Büchel, F., Ischi, N., Patry, J.L. & Thommen, B. (1985). *Erziehungspsychologische Beratung und Intervention als Hilfe zur Selbsthilfe in Familie und Schule*. Bern: Huber.
- Perrez, M. & Patry, J.L. (1982). Nomologisches Wissen, technologisches Wissen, Tatsachenwissen - drei Ziele sozialwissenschaftlicher Forschung. In J.L. Patry (Hrsg.), *Feldforschung* (S. 45-66). Stuttgart: Huber.
- Petermann, F. & Petermann, U. (1980). *Erfassungsbogen für aggressives Verhalten in konkreten Situationen (EAS)*. Braunschweig: Westermann.
- Petillon, H. (1980). *Soziometrischer Test für 3. bis 7. Klassen (ST 3-7)*. Weinheim: Beltz.
- Petillon, H. (1984). *Sozialfragebogen für Schüler für 4. bis 6. Klassen (SFS 4-6)*. Weinheim: Beltz.
- Phillips, L.D. & Edwards, W. (1966). Conservatism in a simple probability inference task. *Journal of Experimental Psychology*, 72, 346-357.
- Pitz, G.F. & Sachs, N.J. (1984). Judgment and decision: theory and application. *Annual Review of Psychology*, 35, 139-163.
- Popper, K.R. (1959). *The logic of scientific discovery*. London: Basic Books.
- Pressley, M., Borkowski, J.G. & O'Sullivan, J. (1985). Children's Metamemory and the Teaching of Memory Strategies. In D.L. Forrest-Pressley, G.E. MacKinnon, & T.G. Waller (Eds.), *Metakognition, Cognition, and Human Performance, Vol. 1: Theoretical Perspectives* (pp. 111-153). Orlando: Academic Press.
- Pressley, M., Snyder, B.L. & Cariglia-Bull, T. (1987). How can Good Strategy Use be taught to children? Evaluation of six alternative approaches. In S.M. Cormier & J.D. Hagman (Eds.), *Transfer of learning. Contemporary research and applications* (pp. 81-120). San Diego: Academic Press.
- Priester, H.J. (1958). *Die Standardisierung des Hamburg-Wechsler-Intelligenztests für Kinder (HAWIK)*. Bern: Huber.
- Priester, H.J. (1959). Untersuchungen zum Vergleich zwischen Stanford-Intelligenztest (SIT) und Hamburg-Wechsler-Intelligenztest für Kinder (HAWIK). *Diagnostica*, 5, 49-58.
- Priester, H.J. (1964). Intelligenztests für Erwachsene. In R. Heiß (Hrsg.), *Psychologische Diagnostik* (= Handbuch der Psychologie, Bd. 6) (S. 226-259). Göttingen: Hogrefe.
- Priester, H.J. & Kerekjarto, M. (1960). Weitere Forschungsergebnisse zum HAWIE und HAWIK. *Diagnostica*, 6, 86-94.

- Prystav, G. (1985). Der Einfluß der Vorhersagbarkeit von Streßereignissen auf die Angstbewältigung. In H.W. Krohne (Hrsg.), *Angstbewältigung in Leistungssituationen* (S. 14-44). Weinheim: VCH.
- Puls, M. (1982). *Reliabilitätsuntersuchungen zum Kognitiven Fähigkeitstest für die Elementarstufe (KFT-K) und den Primarbereich (KFT 1-3)* (Unveröffentl. Diplomarbeit). Köln: Univ. Köln, Phil. Fak.
- Pulver, U., Lang, A. & Schmid, F.W. (Hrsg.). (1978). *Ist Psychodiagnostik verantwortbar?* Bern: Huber.
- Putz-Osterloh, W. & Schroiff, M. (1987). Komplexe Verhaltensmaße zur Erfassung von Hochbegabung. *Zeitschrift für Differentielle und Diagnostische Psychologie*, 8, 207-216.
- Raatz, U. & Möhling, R. (1971). *Frankfurter Tests für Fünfjährige - Konzentration - (FTF-K)*. Weinheim: Beltz.
- Räder, E. (1988). *Entwicklung eines computergestützten, fehlerorientiert-adaptiven Lerntests für schlußfolgerndes Denken im figural-anschaulichen Bereich* (Dissertation A). Leipzig: Karl-Marx-Universität.
- Rapoport, A. & Wallsten, T.S. (1972). Individual decision behavior. *Annual Review of Psychology*, 23, 131-175.
- Rathenow, R. (1973). *Rechtschreibtest für 1. Klassen (RST 1)*. Weinheim: Beltz.
- Rauchfleisch, U. (1983). Zur Reliabilität und Validität des Aufmerksamkeits-Belastungs-Tests (Test d2) bei Patienten mit hirndiffusem Psychodrom und neurotischen Störungen. *Diagnostica*, 24, 247-255.
- Raven, J.C. (1965). *Coloured Progressive Matrices (CPM)*. London: Lewis. (a)
- Raven, J.C. (1965). *Standard Progressive Matrices (SPM)*. London: Lewis. (b)
- Rebmann, W., Caspari, E., Gutke, M., Haselmann, B., Preisinger, B., Sticks, E. & Zabelt, F. (1985). *Feststellung der Sonderschulbedürftigkeit bei ausländischen Schülern - Hinweise zum Verfahren* (= Materialien zur Förderung ausländischer Kinder und Jugendlicher an allgemeinbildenden und beruflichen Schulen, Reihe C, Heft 8 des Ministerium für Kultus und Sport Baden-Württemberg). Stuttgart: Ministerium für Kultus und Sport Baden Württemberg.
- Reinartz, A. (1974). *Schulleistungstest lernbehinderter Schüler (S-L-S)*. Berlin: Marhold, 4. Aufl.
- Reitan, R. (1956). *Trail-Making-Test. Manual*. Indianapolis.
- Renzulli, J.S. (1978). What makes giftedness? Reexamining a definition. *Phi Delta Kappan*, 60, 180-184.
- Renzulli, J.S. (1986). The three-ring conception of giftedness: a developmental model for creative productivity. In R.J. Sternberg & J.E. Davidson (Eds.), *Conceptions of giftedness*. Cambridge: University Press.
- Rexilius, G. (1978). Grenzen der Testerei. In S. Grubitsch & G. Rexilius (Hrsg.), *Testtheorie - Testpraxis*. Hamburg: Rowohlt.
- Rich, C.C. (1963). *The Validity of an Adaption of Raven's Progressive Matrices Test for Use with Blind Children* (Dissertation). Texas Technological College.
- Rieder, O. (1971). *Allgemeiner Schulleistungstest für 2. Klassen (AST 2)*. Weinheim: Beltz.
- Rieder, O. (1984). *Rechtschreibtest für die 6. und 7. Klasse (RST 6-7)*. Weinheim: Beltz.
- Riegel, R.M. (1960). Faktorenanalysen des HAWIE für die Altersstufen 20-34, 35-49, 50-64 und 65 Jahre und älter. *Diagnostica*, 6, 41-66.
- Robinson, A. (1986). The identification and labeling of gifted children. What do research tell us? In K.A. Heller & J.F. Feldhusen (Eds.), *Identifying and Nurturing the Gifted* (pp. 103-109). Toronto: Huber.
- Roeder, P.M. & Treumann, K. (1974). *Dimension der Schulleistung*, 2 Bde. Stuttgart: Klett.
- Roether, D. (1983). *Vorschul-Lerntest (VLT)*. Berlin: Psychodiagnostisches Zentrum.
- Rollett, B. (1983). Anstrengungsvermeidung als Motiv. Eine Antwort auf U.-J. Jopt. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 15, 75-84.
- Rollett, B. & Bartram, M. (1981). *Anstrengungsvermeidungstest (AVT)*. Braunschweig: Westermann, 2. Aufl.

- Rosemann, B. (1975). Prognosemodell für die Schullaufbahnberatung. Ein methodologischer Beitrag zur Bildungsberatung. In K. Heller (Hrsg.), *Handbuch der Bildungsberatung, Bd. 2* (S. 429-447). Stuttgart: Klett.
- Rosemann, B. (1978). *Prognosemodelle in der Schullaufbahnberatung*. München: Reinhardt.
- Rosemann, B. (1984). Konstruktion und Auswertung informeller Schulleistungstests (Lernkontrolltests). In K. Heller (Hrsg.), *Leistungsdiagnostik in der Schule* (S. 162-197). Bern: Huber, 4. Aufl.
- Rosemann, B. & Allhoff, P. (1982). *Differentielle Prognostizierbarkeit von Schulleistung*. Opladen: Westdeutscher Verlag.
- Rosenberger, E. (1978). *Ernährungslehre*. München: BLV Verlagsgesellschaft, 3. Aufl.
- Rost, D.H. & Gebert, A. (1980). Zum Problem der Faktoreninterpretation bei Raven's Coloured Progressive Matrices - Psychologische Fakten oder methodische Artefakte? *Zeitschrift für Differentielle und Diagnostische Psychologie, 1*, 255-273.
- Rost, D.H. & Haferkamp, W. (1979). Zur Brauchbarkeit des AFS. Eine empirische Analyse und eine vergleichende Darstellung vorliegender Untersuchungen. *Zeitschrift für Empirische Pädagogik, 3*, 183-210.
- Rost, J. (1988). *Quantitative und qualitative probabilistische Testtheorie*. Bern: Huber.
- Rost, J. (1990). Rasch models in latent classes: an integration of two approaches to item analysis. *Applied Psychological Measurement, 14*, 271-282.
- Rost, J. & Spada, H. (1982). Probabilistische Testtheorie. In K.J. Klauer (Hrsg.), *Handbuch der Pädagogischen Diagnostik, Bd. 1* (S. 59-98). Düsseldorf: Schwann.
- Rüdiger, D. (1987). Der Übertritt auf weiterführende Schulen. In A. Kormann (Hrsg.), *Beurteilen und Fördern in der Erziehung* (S. 98-121). Salzburg: Müller.
- Rüppell, H., Hinnersmann, H. & Wiegand, J. (1987). Problemlösen - allgemein oder spezifisch? In H. Neber (Hrsg.), *Angewandte Problemlösepsychologie* (S. 173-192). Münster: Aschendorff.
- Saldern, M. & Littig, K.E. (1987). *Landauer Skalen zum Sozialklima 4. bis 13. Klassen (LASSO 4-13)*. Weinheim: Beltz.
- Salzgeber, J. & Stadler, M. (1990). *Familienpsychologische Begutachtung*. München: Psychologie Verlags Union.
- Sarason, I.G. (1981). Test anxiety, stress, and social support. *Journal of Personality, 49*, 101-114.
- Sarimski, K. (1990). Zum Wert neuropsychologischer Testbatterien in der Diagnostik von lern- oder geistigbehinderten Kindern. *Zeitschrift für Heilpädagogik, 41*, 88-94.
- Sawyer, J. (1966). Measurement and prediction, clinical and statistical. *Psychological Bulletin, 66*, 178-200.
- Schallberger, U. (1987). Der IQ-Verlust beim HAWIK-R und das Phänomen des IQ-gains. *Sonderpädagogik, 17*, 174-178. (a)
- Schallberger, U. (1987). HAWIK und HAWIK-R: Ein empirischer Vergleich. *Diagnostica, 33*, 1-13. (b)
- Scheerer-Neumann, G. (1979). *Intervention bei Lese-Rechtschreibschwäche. Überblick über Theorien, Methoden und Ergebnisse*. Bochum: Kamp.
- Scheerer-Neumann, G. (1984). *Rechtschreibschwierigkeiten*. Hagen: Fernuniversität Hagen.
- Scheller, R. (1973). Zur Verwendung des Hamburg-Wechsler-Intelligenztests für Erwachsene (HAWIE) und des Hamburg-Wechsler-Intelligenztests für Kinder (HAWIK) als Paralleltests bei Ober- und Sonderschülern. *Psychologie in Erziehung und Unterricht, 20*, 190-197.
- Schirmacher, A. (1990). *Lernbehinderte ausländische Schüler*. Weinheim: Deutscher Studien Verlag.
- Schlee, J. (1985). Förderdiagnostik - eine bessere Konzeption? In K. Ingenkamp & R.S. Jäger (Hrsg.), *Tests und Trends 4* (Jb. d. Päd. Diagn.) (S. 82-108). Weinheim: Beltz. (a)
- Schlee, J. (1985). Zum Dilemma der heilpädagogischen Diagnostik. *Vierteljahresschrift für Heilpädagogik und ihre Nachbargebiete, 54*, 256-279. (b)
- Schmalt, H.D. (1976). *Das LM-Gitter. Ein objektives Verfahren zur Messung des Leistungsmotivs bei Kindern*. Göttingen: Hogrefe.

- Schmalt, H.D. (1978). Leistungsthematische Kognitionen I: Kausalerklärungen für Erfolg und Mißerfolg. *Zeitschrift für Experimentelle und Angewandte Psychologie*, 25, 246-272.
- Schmalt, H.D. (1979). Leistungsthematische Kognitionen II: Kausalattributionen, Erfolgserwartungen und Affekte. *Zeitschrift für Experimentelle und Angewandte Psychologie*, 26, 509-531.
- Schmid, F.W. (1988). Ethik. In R.S. Jäger (Hrsg.), *Psychologische Diagnostik. Ein Lehrbuch* (S. 89-96). München: Psychologie Verlags Union.
- Schmidt, L.R. (1982). Diagnostische Begutachtung. In K.J. Groffmann & L. Michel (Hrsg.), *Grundlagen psychologischer Diagnostik* (S. 467-537). Göttingen: Hogrefe.
- Schmidt, L.R. (1988). Psychodiagnostisches Gutachten. In R.S. Jäger (Hrsg.), *Psychologische Diagnostik. Ein Lehrbuch* (S. 398-406). München, Weinheim: Psychologie Verlags Union.
- Schmidtko, A., Schaller, S. & Becker, P. (1980). *Coloured Progressive Matrices (CPM)*. Weinheim: Beltz.
- Schmitz, G. (1990). "Förderdiagnostik" einmal anders. Eine Einführung in die Lernfähigkeitsdiagnostik Reuven Feuersteins. *Sonderpädagogik*, 20, 1-14.
- Schneider, W. (1989). *Zur Entwicklung des Meta-Gedächtnisses bei Kindern*. Bern: Huber.
- Schneider, W. & Pressley, M. (1988). *Memory development between 2 and 20*. New York: Springer.
- Schoppe, K.J. (1975). *Verbaler Kreativitäts-Test (VKT)*. Göttingen: Hogrefe.
- Schraml, W. (1969). *Abriß der klinischen Psychologie*. Stuttgart: Klett.
- Schröder, U.J. (1990). *Grundriß der Lernbehindertenpädagogik*. Berlin: Marhold.
- Shuck, K.D., Eggert, D. & Raatz, U. (1975). *Columbia Mental Maturity Scale (CMM 1-3). Sprachfreier Gruppenintelligenztest für die Grundschule*. Weinheim: Beltz.
- Schuler, H. & Funke, U. (Hrsg.). (1991). *Eignungsdiagnostik in Forschung und Praxis*. Göttingen: Hogrefe.
- Schumacher, D. & Schumacher-Merz, I. (1982). Rechtsbedingungen psychologischer Diagnostik. In K.J. Groffmann & L. Michel (Hrsg.), *Grundlagen psychologischer Diagnostik* (= Enzyklopädie der Psychologie, B II, Bd. 1) (S. 538-567). Göttingen: Hogrefe.
- Schumacher, G. & Cattell, R.B. (1977). *Deutscher HSPQ (High School Personality Questionnaire)*. Bern: Huber.
- Schwarzer, Ch. & Schwarzer, R. (1977). *Praxis der Schülerbeurteilung*. München: Kösel.
- Schwarzer, R. (1975). *Schulangst und Lernerfolg*. Düsseldorf: Schwann.
- Scroko, J. (1983). Itemanalyse des Reduzierten Wechsler-Intelligenztests (WIP). *Diagnostica*, 29, 40-47.
- Seidl, H. & Turski, J. (1975). *Schul- und Berufsinteressentest (SBIT)*. Wien: Ketterl.
- Seitz, W. & Rausche, A. (1991). *Persönlichkeitsfragebogen für Kinder zwischen 9 und 14 (PFK 9-14)*. Braunschweig: Westermann, 3. Aufl.
- Selg, H. (1975). *Einführung in die experimentelle Psychologie*. Stuttgart: Kohlhammer, 4. Aufl.
- Shertzer, B. & Stone, S.C. (1971). *Fundamentals of counseling*. Boston: Houghton Mifflin, 2. Aufl.
- Siegler, R.S. & Kotovsky, K. (1986). Two levels of giftedness: shall ever the twain meet? In R.J. Sternberg & J.E. Davidson (Eds.), *Conceptions of giftedness* (pp. 417-435). Cambridge: Cambridge Univ. Press.
- Simons, H. & Möbus, C. (1982). Testfairness. In K.J. Klauer (Hrsg.), *Handbuch der Pädagogischen Diagnostik, Bd. 1* (S. 187-198). Düsseldorf: Schwann.
- Simonton, D.K. (1988). *Scientific genius. A psychology of science*. Cambridge: University Press.
- Slovic, P., Fischhoff, B. & Lichtenstein, S. (1977). Behavioral decision theory. *Annual Review of Psychology*, 28, 1-39.
- Slovic, P. & Lichtenstein, S. (1971). Comparison of Bayesian and regression approaches to the study of information processing in judgment. *Organizational Behavior and Human Performance*, 6, 649-744.
- Snijders, J.T. & Snijders-Oomen, N. (1978). *Snijders-Oomen nicht-verbale Intelligenztestreihe (SON)*. Groningen: Wolters-Noordhoff, 2. Aufl.

- Snijders, J.Th. & Snijders-Oomen, N. (1970). *Snijders-Oomen nicht-verbale Intelligenztestreihe (SON)*. Groningen: Wolters-Noordhoff, 4. Aufl.
- Snyder, C.R., Shenkel, R.J. & Lowery, C.R. (1977). Acceptance of personality interpretations: The "Barnum effect" and beyond. *Journal of Consulting and Clinical Psychology*, 45, 104-114.
- Sommer-Stumpfenhorst, N. (1988). *Computerunterstützte Lehr- und Lernhilfe (COLLI)*. Handbuch. Bergisch-Gladbach: Technisches Büro Dipl.-Ing. Hullmann.
- Spearman, C. (1904). "General intelligence", objectively determined and measured. *American Journal of Psychology*, 15, 201-293.
- Spearman, C. (1927). *The abilities of man*. London: MacMillan.
- Spitznagel, A. (1982). Grundlagen, Ergebnisse und Probleme der Formdeutungsverfahren. In K.J. Groffmann & L. Michel (Hrsg.), *Persönlichkeitsdiagnostik* (= Enzyklopädie der Psychologie, B II, Bd. 3) (S. 186-257). Göttingen: Hogrefe.
- Spitznagel, A. (1984). Kommunikationspsychologische Forschungsergebnisse zur Produktion und Rezeption von Gutachtentexten. In H.A. Hartmann & R. Haubl (Hrsg.), *Psychologische Begutachtung. Problembereiche und Praxisfelder* (S. 127-159). München: Urban & Schwarzenberg.
- Staatsinstitut für Schulpädagogik und Bildungsforschung. (1988). *Erstschreiben. Handreichung für Diagnose- und Förderklassen*. Würzburg: Staatsinstitut für Schulpädagogik und Bildungsforschung.
- Stahl, B. (1985). *Verhaltensfragebogen für geistig- und lernbehinderte Heimkinder (VFHK)*. Weinheim: Beltz.
- Stanley, J.C. (1986). Fostering use of mathematical talent in the USA: SMPY's rationale. In A.J. Cropley, K.K. Urban, H. Wagner & W. Wiczerkowski (Eds.), *Giftedness: A Continuing Worldwide Challenge* (pp. 227-243). New York: Trillium Press.
- Steffens, K., Hospelt, C. & Heller, K.A. (1983). Zur Faktorstruktur des KFT 4-13. Eine hypothesentestende Untersuchung unter Verwendung der konfirmatorischen Maximum Likelihood Faktorenanalyse. *Zeitschrift für Differentielle und Diagnostische Psychologie*, 4, 151-164.
- Steingrüber, H.J. (1971). *Hand-Dominanz-Test (HDT)*. Göttingen: Hogrefe.
- Stern, W. (1912). *Die psychologischen Methoden der Intelligenzprüfung und deren Anwendung an Schulkindern*. Berlin: 5. Kongr. d. Exp. Psych.
- Stern, W. (1935). *Allgemeine Psychologie auf personalistischer Grundlage*. Den Haag: Mouton, 2. Aufl. (1950).
- Sternberg, R.J. (1977). *Intelligence, information processing and analogical reasoning*. Hillsdale, NJ: Erlbaum.
- Sternberg, R.J. (1981). A componential theory of intellectual giftedness. *Gifted Child Quarterly*, 25, 86-93.
- Sternberg, R.J. (1984). Toward a triarchic theory of human intelligence. *Behavioral and Brain*, 7, 269-287.
- Sternberg, R.J. (1985). *Beyond IQ: A triarchic theory of human intelligence*. Cambridge: University Press.
- Sternberg, R.J. (1986). A triarchic theory of intellectual giftedness. In R.J. Sternberg & J.E. Davidson (Eds.), *Conceptions of giftedness* (pp. 223-243). Cambridge: University Press.
- Sternberg, R.J. (1988). *The nature of creativity. Contemporary psychological perspectives*. Cambridge: Cambridge University Press.
- Sternberg, R.J. (1990). What Constitutes a "Good" Definition of Giftedness? *Journal for the Education of the Gifted*, 14, 96-100.
- Sternberg, R.J. & Davidson, J.E. (1983). Insight in the gifted. *Educational Psychologist*, 18, 51-57.
- Sternberg, R.J. & Davidson, J.E. (Eds.). (1986). *Conceptions of Giftedness*. Cambridge: University Press.
- Sternberg, R.J. & Gardner M.K. (1983). Unities in inductive reasoning. *Journal of Experimental Psychology*, 112, 80-116.

- Sternberg, R.J. & Wagner R.K. (1986). *Practical intelligence - Nature and origin of competence in the everyday world*. Cambridge: Cambridge University Press.
- Storfer, M.D. (1990). Intelligence and Giftedness. *The Contributions of Heredity and Early Environment*. San Francisco: Jossey-Bass.
- Strack, F. (1985). Urteilsheuristiken. In D. Frey & M. Irle (Hrsg.), *Theorien der Sozialpsychologie. Band III: Motivations- und Informationsverarbeitungstheorien* (S. 239-268). Bern: Huber.
- Stumpf, H. & Fay, E. (1983). *Schlauchfiguren. Ein Test zur Beurteilung des räumlichen Vorstellungsvermögens*. Göttingen: Hogrefe.
- Sturm, W. & Büssing, A. (1982). Ein Vergleich von HAWIE und LPS bei der psychometrischen Einzelfalldiagnostik neurologischer Patienten. *Diagnostica*, 28, 348-359.
- Sührig, H. & Sührig, S. (1984). Die Bildergeschichten im Subtest Bilderordnen des HAWIK-R. Eine kritische Betrachtung. *Zeitschrift für Heilpädagogik*, 35, 725-731.
- Suhrweier, H. (1986). *Beurteilung geschädigter Kinder*. Berlin: Volk und Wissen, 6. Aufl.
- Svenson, O. (1979). Process description of decision making. *Organizational Behavior and Human Performance*, 23, 86-112.
- Tack, W.H. (1976). Diagnostik als Entscheidungshilfe. In K. Pawlik (Hrsg.), *Diagnose der Diagnostik: Beiträge zur Diskussion der psychologischen Diagnostik in der Verhaltensmodifikation* (S. 103-130). Stuttgart: Klett.
- Tannenbaum, A.J. (1983). *Gifted Children: Psychological and Educational Perspectives*. New York: Mac Millan.
- Taschinski, R. (1985). Eine Untersuchung zur Kulturfairneß der Progressiven Matrizen von Raven gegenüber türkischen Kindern in Deutschland. *Psychologie in Erziehung und Unterricht*, 32, 229-239.
- Taschinski, R. (1987). Entgegnung auf Friedrich und Müller: Zur Kulturfairneß der Progressiven Matrizen von Raven. *Psychologie in Erziehung und Unterricht*, 34, 232-233.
- Tent, L. (1969). *Die Auslese von Schülern für weiterführende Schulen*. Göttingen: Hogrefe.
- Tent, L., Fingerhut, W. & Langfeldt, H.P. (1976). *Quellen des Lehrerurteils. Untersuchungen zur Aufklärung der Varianz von Schulnoten*. Weinheim: Beltz.
- Tent, L. & Waldow, M. (1984). Pädagogische Diagnostik in der Schule für Lernbehinderte. Gruppenbezogene Leistungsmessung oder Zielerreichungstests? *Heilpädagogische Forschung*, 11, 1-29.
- Tewes, U. (1985). *Hamburg-Wechsler-Intelligenztest für Kinder. Revision 1983 (HAWIK-R)*. Bern: Huber, 3. Aufl.
- Thiel, R., Keller, G. & Binder, A. (1979). *Arbeitsverhaltensinventar (AVI)*. Braunschweig: Westermann.
- Thomae, H. (1964). Prinzipien und Formen der Gestaltung psychologischer Gutachten. In U. Undeutsch (Hrsg.), *Forensische Psychologie* (= Handbuch der Psychologie, Bd. 11) (S. 643-767). Göttingen: Hogrefe.
- Thorndike, R.L. & Hagen, E. (1971). *Cognitive Abilities Test*. Boston: Houghton-Mifflin.
- Turner, F. & Tewes, U. (1972). *Kinder-Angst-Test (KAT)*. Göttingen: Hogrefe, 2. Aufl.
- Thurstone, L.L. (1931). Multiple factor analysis. *Psychological Review*, 38, 406-427.
- Thurstone, L.L. (1938). *Primary mental abilities*. Chicago: University of Chicago Press.
- Thurstone, L.L. (1947). *Multiple factor analysis*. Chicago: University Press (1950).
- Thurstone, L.L. & Thurstone, T.G. (1941). *Factorial studies of intelligence*. Chicago: Chicago University Press.
- Tismer, K.G. (1976). Verhaltensbeobachtung bei Kindern und Jugendlichen. In K. Heller (Hrsg.), *Handbuch der Bildungsberatung, Bd. 3* (S. 817-836). Stuttgart: Klett.
- Tismer, K.G., Fisseni, H.J. & Tismer-Puschner, I. (1976). Anamnese und Exploration in der schulpсихologischen Beratung. In K. Heller (Hrsg.), *Handbuch der Bildungsberatung, Bd. 3* (S. 837-856). Stuttgart: Klett.
- Tismer-Puschner, I., Fisseni, H.J. & Tismer, K.G. (1976). Das psychologische Gutachten in der Einzelfallberatung. In K. Heller (Hrsg.), *Handbuch der Bildungsberatung, Bd. 3* (S. 901-918). Stuttgart: Klett.

- Titze, I. & Tewes, U. (1987). *Messung der Intelligenz bei Kindern mit dem HAWIK-R*. Bern: Huber, 2. Aufl.
- Todt, E. (1972). *Differenzieller Interessen-Test (DIT)*. Bern: Huber, 2. Aufl.
- Trost, G. (1975). *Vorhersage des Studienerfolgs*. Braunschweig: Westermann.
- Trost, G. (1986). Identification of the Highly Gifted Adolescents - Methods and Experiences. In K.A. Heller & J.F. Feldhusen (Eds.), *Identifying and Nurturing the Gifted. An International Perspective* (pp. 83-91). Toronto: Huber.
- Trost, G. & Bickel, H. (1979). *Studierfähigkeit und Studienerfolg*. München: Reinhardt.
- Trost, G. et al. (1987). *Test für medizinische Studiengänge (TMS), 11. Arbeitsbericht*. Bonn: Institut für Test- und Begabungsforschung.
- Trudewind, C. (1974). *Häusliche Umwelt und Motiventwicklung*. Göttingen: Hogrefe.
- Tscherner, K. (1990). Zur Frage der Übereinstimmung der Testergebnisse von Raven und HAWIK-R. *Zeitschrift für Heilpädagogik, 41*, 108-113.
- Tucker, L.R. (1964). A suggested alternative formulation in the development of Hursch, Hammond, & Hursch, and by Hammond, Hursch, & Todd. *Psychological Review, 71*, 528-530.
- Tversky, A. & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology, 5*, 207-232.
- Überla, K. (1971). *Faktorenanalyse*. Berlin: Springer, 2. Aufl.
- Verband Deutscher Sonderschulen e.V. (1987). "Förderschule" statt "Schule für Lernbehinderter". *Zeitschrift für Heilpädagogik, 38*, 907.
- Vernon, P.E. (1950). *The structure of human abilities*. London: Methuen, 2. Aufl. (1961).
- Vernon, P.E. (1965). Ability factors and environmental influences. *American Psychologist, 20*, 723-733.
- Wagner, H. (1987). Außerschulische Fördermaßnahmen. In F.E. Weinert & H. Wagner (Hrsg.), *Die Förderung Hochbegabter in der Bundesrepublik Deutschland: Probleme, Positionen, Perspektiven* (S. 63-76). Bad Honnef: Bock.
- Wagner, H. (1989). Elternrecht und Sonderschulübergang: Aspekte der Schulwahlfreiheit im Überweisungsverfahren. *Sonderpädagogik, 19*, 157-173.
- Wagner, H. & Baumgärtel, F. (1978). *Hamburger Persönlichkeitsfragebogen für Kinder (HAPEF-K)*. Göttingen: Hogrefe.
- Wagner, H., Zimmermann, B. & Stüven, N. (1986). Identifizierung und Förderung mathematisch besonders befähigter Schüler. Bericht über einen Modellversuch. In W. Wiczerkowski et al. (Hrsg.), *Hochbegabung, Gesellschaft, Schule* (S. 239-251). Bad Honnef: Bock.
- Wagner, J.W.L. (1977). *Fragebogen Einstellung zur Schule für 4. bis 6. Klassen (FES 4-6)*. Weinheim: Beltz. (a)
- Wagner, J.W.L. (1977). *Fragebogen zum Selbstkonzept für 4. bis 6. Klassen (FSK 4-6)*. Weinheim: Beltz. (b)
- Waldmann, M. & Weinert, F.E. (1990). *Intelligenz und Denken. Perspektiven der Hochbegabungsforschung*. Göttingen: Hogrefe.
- Wallasch, R. & Dony, M. (1980). Cerebralschadensdiagnostik mit dem HAWIK: zur Legende der Validität der VIQ-HIQ-Diskrepanz und der Untertestmuster. *Diagnostica, 26*, 165-185.
- Wason, P.C. (1960). On the failure to eliminate hypotheses in a conceptual task. *The Quarterly Journal of Experimental Psychology, 12*, 129-140.
- Wechsler, D. (1939). *Wechsler Bellevue Adult Intelligence Scale*. Baltimore: Williams & Wilkins.
- Wechsler, D. (1949). *Wechsler Intelligence Scale for Children (WISC)*. New York: Psychological Corporation.
- Wechsler, D. (1955). *Wechsler Adult Intelligence Scale (WAIS)*. New York: Psychological Corporation.
- Wechsler, D. (1961). *Die Messung der Intelligenz Erwachsener*. Bern: Huber, 2. Aufl.
- Wechsler, D. (1967). *Wechsler Preschool and Primary Scale of Intelligence WPPSI*. New York: Psychological Corporation.

- Wechsler, D. (1974). *Manual for the Wechsler Intelligence Scale for Children - Revised*. New York: Psychological Corporation.
- Wechsler, D. (1982). *Handanweisung zum Hamburg-Wechsler-Intelligenztest für Erwachsene (HAWIE)* (Separatdruck). Bern: Huber.
- Wegener, H. (1969). Die Minderbegabten und ihre sonderpädagogische Förderung - Sondergutachten. In Dt. Bildungsstatistik (Hrsg.), *Gutachten und Studien der Bildungskommission 4. Begabung und Lernen* (S. 505-515). Stuttgart: Klett.
- Weible, K. & Bethäuser, H. (1986). *Testverfahren für Interessen: Beruf-Schule (TIBS 1)*. Weinheim: Beltz.
- Weidenmann, B. & Krapp, A. et al. (Hrsg.). (1986). *Pädagogische Psychologie*. München: Urban & Schwarzenberg.
- Weiner, B. (1975). *Die Wirkung von Erfolg und Mißerfolg auf die Leistung*. Stuttgart: Klett.
- Weiner, B., Frieze, J., Kukla, A., Reed, L., Rests, S. & Rosenbaum, R.M. (1971). *Perceiving the causes of success and failure*. New York: General Learning Press.
- Weinert, F.E. (1974). Fähigkeits- und Kenntnisunterschiede zwischen Schülern. In F.E. Weinert et al. (Hrsg.), *Funkkolleg Pädagogische Psychologie, Bd. 2* (S. 763-793). Frankfurt/M.: Fischer.
- Weinert, F.E. (1989). *Is the past the best predictor of the future? - Short- and long-term predictability of individual differences in children's cognitive achievements*. San Francisco: Paper presented at the annual meeting of the American Educational Research Association in San Francisco, USA.
- Weinert, F.E. & Helmke, A. (1987). Schulleistungen - Leistungen der Schule oder der Kinder? *Bild der Wissenschaft, 24*, 62-73.
- Weingart, P., Kroll, J. & Bayertz, K. (1988). *Rasse, Blut und Gene: Geschichte der Eugenik und Rassenhygiene in Deutschland*. Frankfurt: Suhrkamp.
- Weisberg, R.W. (1986). *Creativity: Genius and other myths*. New York: Freeman. - Dt. Kreativität und Begabung. Heidelberg: Spektrum der Wissenschaft (1989).
- Weiß, R. (1971). *Grundintelligenztest Skala 3 (CFT 3)*. Braunschweig: Westermann.
- Weiß, R. (1987). *Grundintelligenztest Skala 2 (CFT 20)*. Göttingen: Hogrefe, 3. Aufl. (a)
- Weiß, R. (1987). *Wortschatztest (WS) und Zahlenfolgentest (ZF). Ergänzungstests zum Grundintelligenztest CFT 20*. Göttingen: Hogrefe. (b)
- Weiß, R. & Osterland, J. (1980). *Grundintelligenztest Skala 1 (CFT 1)*. Braunschweig: Westermann, 4. Aufl.
- Wendeler, J. (1981). Förderungsdiagnostik bei Schulleistungsschwächen in der Grundschule. *Psychologie in Erziehung und Unterricht, 28*, 293-305.
- Wendeler, J. (1984). Förderungsdiagnostik im Primärbereich. In K.A. Heller (Hrsg.), *Leistungsdiagnostik in der Schule* (S. 283-291). Bern: Huber, 4. Aufl.
- Westhoff, K. & Kluck, M.L. (1991). *Psychologische Gutachten schreiben und beurteilen*. Berlin: Springer.
- Westmeyer, H. (1972). *Logik der Diagnostik*. Stuttgart: Kohlhammer.
- Westmeyer, H. (1975). The diagnostic process as an statistical-causal analysis. *Theory and Decision, 6*, 57-86.
- Westmeyer, H. (1976). Grundlagenprobleme psychologischer Diagnostik. In K. Pawlik (Hrsg.), *Diagnose der Diagnostik: Beiträge zur Diskussion der psychologischen Diagnostik in der Verhaltensmodifikation* (S. 71-102). Stuttgart: Klett.
- Wewetzer, K.H. (1964). Intelligenztests für Kinder. In R. Heiß (Hrsg.), *Psychologische Diagnostik* (= Handbuch der Psychologie, Bd. 6) (S. 200-225). Göttingen: Hogrefe.
- Wewetzer, K.H. (1972). *Intelligenz und Intelligenzmessung*. Darmstadt: Wissenschaftliche Buchgesellschaft.
- Widdel, H. (1977). *Attributerungsfragebogen für Erfolg und Mißerfolg in der Schule für 5.-7. Klassen (AEM 5-7)*. Weinheim: Beltz. (a)
- Widdel, H. (1977). *Fragebogen zur Erfassung des schulischen Leistungsmotivs für 5. bis 7. Klassen (FSL 5-7)*. Weinheim: Beltz. (b)

- Wieczerkowski, W., Nickel, H., Janowski, A., Fittkau, B. & Rauer, W. (1981). *Angstfragebogen für Schüler (AFS)*. Braunschweig: Westermann, 6. Aufl.
- Wieczerkowski, W. & Schumann, M. (1982). Klassische Testtheorie. In K.J. Klauer (Hrsg.), *Handbuch der Pädagogischen Diagnostik, Bd. 1* (S. 41-58). Düsseldorf: Schwann.
- Wieczerkowski, W. & Wagner, H. (1985). Diagnostik von Hochbegabung. In R.S. Jäger et al. (Hrsg.), *Tests und Trends 4* (Jb. d. Päd. Diagn.) (S. 109-134). Weinheim: Beltz.
- Wieczerkowski, W. & zur Oeveste, H. (1982). Zuordnungs- und Entscheidungsstrategien. In K.J. Klauer (Hrsg.), *Handbuch der pädagogischen Diagnostik, Bd. 2* (S. 919-951). Düsseldorf: Schwann.
- Wiggins, J.S. (1973). *Personality and prediction: Principles personality assessment*. Reading, Mass.: Addison & Wesley.
- Witzlack, G. (1987). Theoretische Grundlagen und Methoden der pädagogisch-psychologischen Diagnostik in der Schule. In G. Witzlack, J. Guthke, Ch. Jäger, H. Klemm, & G. Matthes (Hrsg.), *Einführung in die Psychodiagnostik in der Schule* (S. 13-76). Berlin: Volk und Wissen, 4. Aufl.
- Wong, B.Y.L. (1985). Metacognition and learning disabilities. In D.L. Forrest-Pressley, G.E. MacKinnon & T.G. Waller (Eds.), *Metacognition, cognition, and human performance, Vol. 2: Instructional practices* (pp. 137-180). Orlando, FL: Academic Press.
- Wottawa, H. (1980). *Grundriß der Testtheorie*. München: Juventa.
- Wottawa, H. & Hossiep, R. (1987). *Grundlagen psychologischer Diagnostik. Eine Einführung*. Göttingen: Hogrefe.
- Wundt, W. (1913). *Grundriß der Psychologie*. Leipzig: Engelmann, 11. Aufl.
- Zielinski, W. (1980). *Lernschwierigkeiten. Verursachungsmomente, Diagnose, Behandlungsansätze*. Stuttgart: Kohlhammer.
- Zielinski, W. (1987). Lese- und Rechtschreibschwierigkeiten und Möglichkeiten ihrer Behebung. In A. Kormann (Hrsg.), *Beurteilen und Fördern in der Erziehung* (S. 122-139). Salzburg: Müller.
- Zielinski, W. & Schneider, W. (1986). Diagnostische Möglichkeiten bei Lese- und Rechtschreibschwächen - Folgerungen aus der Forschung. In K. Ingenkamp & R.S. Jäger (Hrsg.), *Tests und Trends 5* (Jb. d. Päd. Diagn.) (S. 38-62). Weinheim: Beltz.
- Zimmermann, K.W., Kornmann, R. & Lorenz, A.L. (1971). *Der HAWIK bei lernbehinderten Sonderschülern*. Oberbiel: Jarwik.

8. Personenregister

- Abel, J. 160f.
Abels, D. 188f.
Ahrbeck, B., 132
Ahrens, H.J. 48, 341, 360
Al-Zoubi, A. 84, 92
Allehoff, W.H. 193ff., 200
Allhoff, B. 226, 229
Allinger, U. 221, 224, 226ff.
Amelang, M. 25, 36
Amthauer, R. 120, 172, 177, 183
Anastasi, A. 46, 123
Anderson, N.H. 312
Andre, A. 200
Arnold, W. 184, 189
Aurin, K. 114, 157ff., 224
Avenarius, H. 109
Ayres, J. 273
Averina, I. 172
Bachmann, U. 120, 123
Baier, H. 240
Baldwin, M.W. 317
Ballstaedt, S.P., 337
Bangen, R. 112
Bartenwerfer, H. 31, 184f., 188, 233
Bartram, M. 190, 192, 208, 211
Bartussek, D. 25, 36
Baud, U. 123
Baudisch, W., 239
Baumann, E. 175f.
Baumgärtel, F. 207, 209, 211
Bäumler, G. 176, 184
Bayertz, K. 238
Bechmann, M. 115, 121
Becker, P. 110, 113, 152
Beckmann, H.K. 218
Beerman, L. 280
Berbig, E. 208
Berg, D. 186
Bethäuser, H. 197, 200, 221, 224
Betz, D. 268
Bickel, H. 226
Biglmaier, F. 268
Binder, A. 202f., 207
Binet, A. 12, 34, 89, 237f.
Birch, J.W. 287
Bittner, R. 33, 282, 285
Bleidick, U. 236
Bloom, B.S. 220
Boerner, K. 357
Bondy, C. 115f., 123, 261
Booth, J.F. 333
Borchert, J. 276
Borgida, E. 309
Borkowski, J.G. 29, 242
Bortz, J. 41
Bourdon, B. 185, 189
Brambring, M. 226
Breuninger, H. 268
Brickenkamp, R. 108, 115, 123, 185f.,
188f., 212
Brophy, J.E. 218
Bröse, B. 239
Brown, A.L. 56f., 142, 242, 267, 271
Brunswik, E. 311
Buggle, F. 207
Bundschuh, K. 237, 239f., 245, 248,
268
Bunge, M. 305
Burgmayer, S. 272f.
Burt, C. 25f.
Büssing, A. 123
Butsch, Ch. 177
Campioni, J.C. 242f.
Cariglia-Bull, T. 242f.
Caruso, M. 87
Cattell, J. 27f., 94, 141ff., 145f., 149,
152, 154, 193, 207
Cattell, J. McK. 13
Clancey, W.J. 324, 327
Clark, W.W. 124
Clauß, G. 41
Cleve, H. 238
Cohn, S.J. 288
Conrad, W. 27, 46, 172, 175f.
Cooley, W.W. 228
Cronbach, L.J. 16f., 34, 58, 223f.,
286, 300f., 333
Dahbashi, A. 276
Dahl, G. 133f.
Dahme, G. 287
Darwin, C. 11
Davidson, J.E. 21, 32, 35f.
Dawes, R.M. 306
de Finetti, B. 307
DeVree, J.K. 315
Diemand, A., 87
Dietel, B. 273f.
Dieterich, R. 68, 77, 93
Dingel, W. 186
Doherty, M.E. 317
Dony, M. 134
Dörner, D. 316

Dreesmann, H. 219
 Düker, H. 182, 189
 Dumke, D. 218
 Ebbinghaus, H. 11, 34
 Eberle, G. 172
 Ebner, H. 41
 Edelstein, W. 218
 Edwards, W. 308, 312, 315
 Eggert, D. 125, 131f., 152, 261
 Einhorn, H.J. 314, 317f.
 Eisebitt, R. 159f., 170, 214
 Ellger-Rüttgardt, S. 236
 Engelbrecht, W. 221, 226f.
 Erlwein, M. 186
 Eser, K.H. 186
 Evans, L. 114
 Ewert, O. 29
 Eysenck, H.J. 26, 207, 323
 Faber, E. 228
 Facaoaru, C. 23, 33, 35f., 282, 285
 Faßnacht, G. 102
 Fay, E. 177
 Feger, B. 280, 288, 290
 Fehnmann, U. 329
 Feldhusen, J.F. 21
 Festinger, L. 317
 Feuerstein, R. 84, 268
 Fickert, H. 119f.
 Fingerhut, W. 157, 220
 Fischer, G.H. 46, 78, 83, 139
 Fischer, H. 177
 Fischhoff, B. 316
 Fishburn, P.C. 315
 Fisseni, H.J. 108, 245, 251, 334, 338,
 347, 349, 355, 357, 360ff.
 Fittkau, B. 201, 207
 Flammer, A. 17
 Flavell, J.H. 242, 271
 Forer, B.R. 341
 Freud, S. 97
 Fried, L. 264
 Friedrich, R. 114
 Funke, U. 226
 Fürntratt, E. 176
 Gadenne, V. 319, 333
 Gaedike, A.K. 165, 169, 173, 217,
 219, 234
 Gagné, F. 21
 Galton, F. 11ff.
 Gaensslen, H. 41
 Gardner, H. 19f., 32, 84, 288
 Gärtner-Harnach, V. 207
 Gaußmann, A., 12
 Gaußmann, A., 151
 Gebert, A. 113
 Geisel, B. 208
 Geisler, H.-J. 162, 170, 173
 Gigerenzer, G. 15, 48
 Glanzmann, P. 208
 Gleser, L.J. 16f., 224, 286, 300f., 333
 Goldfried, M.R. 295
 Good, T.L. 218
 Gosslar, H. 208
 Gözülü, L. 91, 147, 149
 Graumann, C.F. 95
 Grisseemann, H. 240
 Groffmann, K.J. 12, 19, 102, 108, 212
 Guilford, J.P. 22ff., 36, 97, 167, 178,
 180, 194, 278, 282
 Gulliksen, H. 46
 Guthke, J. 38, 84ff., 92, 140
 Harnach, V. 317
 Haenisch, H. 159f.
 Hagen, E. 162, 208, 285, 290
 Halpern, D.F. 307, 310
 Hampton, J.M. 307
 Hany, E.A. 21, 33f., 165, 282f.,
 285ff., 289, 340
 Hardesty, F.P. 115, 124, 126, 151
 Hartje, W. 123
 Hartmann, H. 245, 332, 335f., 338ff.,
 360ff.
 Hasemann, K. 97, 99, 102
 Hasselhorn, M. 243f., 273
 Hatch, T. 84
 Haubl, R. 338, 362
 Haferkamp, W. 202
 Heck-Möhling, R. 153, 187, 189
 Heckhausen, H. 280, 377
 Heiß, R. 336
 Heller, K.A. 14ff., 18f., 21, 34, 41,
 102, 104, 108, 112, 114, 147,
 156ff., 162, 165, 169f., 172ff., 179,
 183, 185, 188, 193, 215, 219, 221,
 224, 226ff., 232ff., 238f., 245, 259,
 261, 265, 280, 282f., 290, 301ff.,
 378
 Helmke, A. 219
 Henry, V. 12
 Herbig, M. 78
 Hermans, H.J.M. 192
 Hermelin, B. 20
 Herrmann, T. 14
 Heyse, H. 219
 Hiltmann, H. 212
 Hochhausen, R. 151
 Hoepfner, R. 23
 Hofer, M. 97, 333

- Hoffman, M.B., 310
 Hogarth, R.M. 314, 317f.
 Holtz, K.L. 243
 Holzkamp, K. 334f.
 Hopf, D. 218
 Horn, R. 93, 110, 113, 152, 220
 Horn, W. 153ff., 158ff., 172
 Hornke, B. 172
 Horst, P. 46
 Hospelt, C. 170
 Hospelt-Renette, C. 171
 Hubel, S. 35, 274, 276
 Huber, G.L. 333
 Hylla, E. 12, 173
 Ingenkamp, K. 12, 14, 93, 153, 212
 Ingram, A.L. 35
 Irlie, M. 193f., 200
 Jacobs, B. 377f.
 Jäger, Ch. 48, 84f., 239
 Jäger, O.A. 24, 48, 145, 149, 167,
 170, 178, 282
 Jäger, R.S. 16, 48, 77f., 93, 109,
 115, 182, 202f., 208, 212, 215,
 219, 222, 245, 290, 292, 298, 305,
 322, 329f., 333f., 336
 Janke, W. 214, 228
 Janowski, A. 201, 207
 Jaspén, N. 157
 Jensen, M.R. 28f., 84
 Johnson, L.J. 280
 Jopt, U.J. 190
 Jöreskog, K.G. 169
 Kaemmerer, W. 382
 Kahneman, D. 308f., 333
 Kail, R. 20, 26, 35f., 169f., 213
 Kaminski, G. 14, 106f., 294f., 305
 Kanfer, F.H. 382
 Kant, I. 330
 Karnes, M.B. 280
 Kastner, M. 159f.
 Kautter, H. 239f., 244, 265, 276
 Keller, G. 202f., 207
 Kent, R.N. 295
 Kerekjarto, M. 122
 Keßler, B.H. 108
 Keupp, H. 331
 Kierdorf, B. 172
 Kiphard, E.J. 177
 Klaghofner, R. 159f.
 Klauer, K.J. 276
 Klauer, K.J. 48, 77f., 83, 93, 235,
 237, 268ff., 276, 305
 Klayman, J. 317
 Kleber, E.W. 133, 184, 189, 245, 263
 Kleber, G. 189
 Kluck, M.L. 344, 357, 362
 Kluwe, R.H. 271, 376, 380
 Knaack, R. 112f., 115
 Knopf-Jerchow, H. 276
 Kobi, E.E. 15
 Koch, U.K. 280
 Kormann, A. 133, 140, 235, 264, 268
 Kornadt, H.J. 94
 Kornmann, R. 122, 236, 239, 241,
 245, 258f., 261, 264f., 268, 276,
 336, 357
 Kotovsky, K. 286
 Kraak, B. 173
 Abel, J. Krampen, G. 192
 Krantz, D.H. 312
 Krapp, A. 17, 217, 219, 222, 235,
 241, 293, 295, 299ff., 319, 333,
 354, 378
 Krathwohl, D.R. 220
 Kratzmeier, H. 110, 113, 152, 172
 Krause, A. 29
 Kretschmann, R. 208, 211
 Krohne, H.W. 378, 381
 Kroll, J. 238
 Krüger, H. 341
 Kruglanski, A.W., 317
 Kubinger, K.D. 82f., 124, 126, 133ff.,
 140, 152, 171, 261
 Kuhl, J. 377
 Kühn, R. 153, 157, 173, 190, 205
 Kvale, S. 14
 Lambert, B. 165
 Langer, I. 330
 Langer, I. 101f.
 Langfeldt, H.P. 77, 157, 220, 222,
 224, 226, 237, 268, 276
 Langfeldt-Nagel, M. 147, 149, 159,
 217
 Langhorst, E. 97, 102
 Lauber, H. 115, 151
 Lauth, G.W. 149, 271f.
 Laux, L. 208
 Lederle-Schenk, U. 147
 Lee, S. 20
 Lehneis-Klepper, G. 219
 Lehwald, G. 278, 282, 288
 Leichner, R. 292ff., 333
 Lenk, H. 319
 Lenz, W. 238
 Lewin, K. 215
 Lichtenstein, S. 310f., 313, 316
 Liebert, R.M. 377
 Liebich, W. 208

- Lienert, G. 48, 63, 77, 88, 93, 119f.,
 122, 169, 173f., 176, 182, 189
 Liman, E. 131f.
 Lindman, H. 312
 Lindquist, E.F. 46
 Littig, K.E. 210, 212
 Lockowandt, O. 259
 Lohnes, P.R. 228
 Lommatzsch, E.M. 132
 Lord, F.M. 78
 Lorenz, J.H. 122, 268
 Lotz, W. 280
 Löwe, A. 259, 261
 Lowery, C.R. 341
 Lukesch, H. 218
 Magnusson, D. 46
 Mainberger, U. 180, 182
 Mandl, H. 17, 217, 219, 222, 333
 Marjoribanks, K. 218
 Marschner, G. 147, 160, 189
 Masendorf, F. 208, 211
 Matarazzo, J.D. 115f., 119ff., 123
 Maxwell, S. 29
 Medley, D.M. 96
 Meehl, P.E. 221, 306, 341
 Meili, R. 32, 302
 Meister, H. 259
 Meister, H. 236, 276
 Menacher, P. 280
 Michael, W.B. 222
 Michel, L. 46, 102, 108, 212
 Mierke, K. 18, 183
 Minsel, W.R. 333
 Mittenecker, E. 94
 Mitzel, H.E. 96
 Möbus, C. 90ff.
 Möhling, R. 187, 189
 Mohr, V. 175f.
 Mönikes, D. 149
 Mönks, F.J. 21, 282, 288
 Montgomery, H. 315
 Moore, P.G. 307
 Moosbrugger, H. 77f., 186
 Morris, L.W. 377
 Müller, H. 77, 114, 212, 259, 336
 Munz, W. 132, 263
 Murken, J.D. 238
 Murray, H. 97
 Mynatt, C.R. 317
 Nagel, B. 172
 Nestle, W. 240
 Neubauer, W. 378
 Neuhäuser, G. 238
 Neukäter, H. 243, 271
 Nickel, H. 104, 201, 207, 235, 245.,
 251, 301ff., 332, 378
 Nisbett, R.E. 308f., 359
 Nollau, W. 228
 Novick, M.R. 78
 Nuber, F. 264
 O'Connor, N. 20
 O'Sullivan, J. 242
 Oehlschlägel, J. 186
 Orgass, B. 123
 Ort, B. 48
 Osterland, J. 101, 141f., 152, 253
 Oswald, W.D. 35, 150ff.
 Otto, J. 165
 Parsons, T. 331
 Patry, J.L. 305, 354
 Pauli, R. 153, 184
 Pawlik, K. 14, 121, 134, 184, 295, 333
 Pearson, K. 40
 Peck, V.A. 242
 Pegnato, C.W. 287
 Pellegrino, J.W. 20, 26, 35f., 169f.,
 213
 Perleth, Ch. 35, 150f., 168ff., 179,
 274, 276, 285
 Perrez, M. 305, 340, 354
 Petermann, F. 207
 Petermann, U. 207
 Petillon, H. 212
 Phillips, L.D. 308
 Pitz, G.F. 316
 Popper, K.R. 316
 Pressley, M. 242f.
 Priester, H.J. 116, 118ff., 122ff., 126
 Prystav, G. 377
 Puls, M. 165
 Pulver, U. 330
 Putz-Osterloh, W. 282, 285
 Raatz, U. 187, 189
 Räder, E. 87
 Rapoport, A. 118, 316
 Rauchfleisch, U. 120, 123, 186
 Rauer, W. 201, 207
 Rausche, A. 204ff., 208
 Raven, J.C. 110
 Rebmann, W. 264
 Reichenbecher, H. 221, 224
 Reinartz, A. 267
 Reitan, R. 150
 Rennen-Allhoff, B. 165
 Renzulli, J.S. 21, 282
 Rexilius, G. 14
 Rich, C.C. 114
 Riediger, B. 264

- Riegel, R.M. 120
 Robinson, A. 287
 Roeder, P.M. 220
 Roether, D. 85
 Rollett, B. 190, 192
 Rosemann, B. 16f., 41, 104, 147, 160,
 215, 219f., 222ff., 226, 229, 235,
 301ff.
 Rosenberger, E. 346
 Ross, L. 308f., 359
 Rost, D.H. 78, 83, 113, 202
 Roth, E. 35, 150ff.
 Rüdiger, D. 224
 Ruppell, H. 285
 Sachs, N.J. 316
 Salzgeber, J. 336
 Samski, C.S. 239
 Sarason, I.G. 377
 Sarimski, K. 273
 Savage, L.J. 312
 Sawyer, J. 306
 Schaffner, P. 208
 Schafheutle, R. 208
 Schallberger, U. 131
 Schaller, S. 110, 113, 152
 Scheblanowa, H. 172
 Scheerer-Neumann, G. 268
 Scheller, R. 121, 124, 333
 Schiebler, K. 376
 Schilling, F. 177
 Schirmacher, A. 131f., 264
 Schlee, J. 14f., 236, 276, 278
 Schmalt, H.D. 192, 377
 Schmid, F.W. 330
 Schmidt, I. 84f., 87
 Schmidt, L.R. 334ff., 360
 Schmidt-Rogge, I. 151
 Schmidtke, A. 110, 113, 152
 Schmitz, G. 268
 Schneider, W. 242f., 268
 Schönpflug, W. 87
 Schoppe, K.J. 177ff., 181
 Schrader, F.W. 219
 Schraml, W. 105
 Schröder, U. 240, 243, 271
 Schroiff, M. 282, 285
 Schubö, W. 41
 Schuck, K.D. 132, 261
 Schuker, G. 35, 274, 276
 Schuler, H. 87, 226
 Schulz v. Thun, F. 101f.
 Schumacher, D. 109
 Schumacher, G. 207
 Schumacher-Merz, I. 109
 Schümann, M. 77
 Schwarzer, Ch. 98f.
 Schwarzer, R. 98f.
 Scroko, J. 123
 Seidl, H. 200
 Seitz, W. 204ff., 208
 Selg, H. 324f.
 Shenkel, R.J. 341
 Shertzer, B. 97
 Siegler, R.S. 286
 Sierwald, W. 150f., 285
 Simon, T. 12, 237f.
 Simons, H. 90ff.
 Simonton, D.K. 23
 Slovic, P. 308, 310f., 313, 316, 333
 Snow, R.E. 17, 223
 Snyder, B.L. 242f., 341
 Sommer-Stumpfenhorst, N. 268
 Spada, H. 83
 Spearman, C. 25, 27, 29f., 56f., 110,
 112, 142, 144, 267
 Spitznagel, A. 94, 334f., 337f., 342,
 361f.
 Stäcker, K.H. 341, 360
 Stadler, M. 336
 Stanley, J.C. 285
 Stapf, K.H. 87
 Steffens, K. 147, 161, 170, 215, 224,
 226
 Steingrüber, H.J. 176, 302
 Stern, W. 12, 19, 89, 237, 323
 Sternberg, R.J. 21, 24, 32ff., 242
 Stone, S.C. 97
 Storz, L. 265
 Strack, F. 308, 310
 Strittmatter, P. 377f.
 Stumpf, H. 177
 Sturm, W. 123
 Sührig, H. 133
 Sührig, S. 133
 Suhrweier, H. 239
 Svenson, O. 315f.
 Tack, W.H. 224, 300
 Tannenbaum, A.J. 21, 282
 Taschinski, R. 114
 Tent, L. 14, 31, 155ff., 233
 Terman, L. 12
 Tewes, U. 124, 126f., 129ff., 134,
 151, 207, 253, 261, 263
 Thiel, R. 202f., 207
 Thomae, H. 106, 337, 349, 354
 Thomas, H. 307
 Thorndike, R.L. 25, 30, 96, 162

- Thurstone, L.L. 20, 25, 29ff., 153f.,
 157, 159, 167, 172, 174
 Tismer, K.G. 97, 106ff., 251
 Tismer-Puschner, I. 108, 251
 Titze, I. 127, 129ff., 134, 253, 263
 Todt, E. 195ff., 200
 Towson, S.M.J. 317
 Treumann, K. 220
 Trost, G. 222, 226
 Trudewind, C. 218
 Tücke, M. 208, 211
 Tucker, L.R. 311
 Thurner, F. 207
 Turski, J. 200
 Tversky, A. 308f., 312, 333
 Tweney, R.D. 317
 Überla, K. 227
 Undeutsch, U. 192
 v. Saldern, M. 212
 v. Kardorff, E. 331
 Vernon, P.E. 25f.
 Wagner, H. 207, 248, 279, 285, 290
 Wagner, J.W.L. 192, 207
 Waldmann, M. 33, 35f., 278, 290
 Waldow, M. 14
 Wallasch, R. 134
 Wallsten, T.S. 316
 Wason, P.C. 316
 Wechsler, D. 19, 27, 88, 115f., 118,
 120ff., 133f., 152
 Wegener, H. 238
 Weible, K. 197, 200
 Weidenmann, B. 219, 235, 333
 Weiner, B. 191, 217
 Weinert, F.E. 33, 35f., 214, 218f.,
 278, 290
 Weingart, P. 238
 Weinläder, H. 165, 173, 234
 Weisberg, R.W. 24, 36
 Weise, G. 212
 Weiß, R. 141ff., 145ff., 152, 253
 Wellman, H.M. 242
 Wendeler, J. 15, 268
 Westhoff, K. 344, 357, 362
 Westmeyer, H. 298, 303ff., 320, 324,
 349
 Wewetzer, K.H. 27f., 124
 Widdel, H. 191f.
 Wiczercowski, W. 77, 102, 201, 207,
 222, 279, 290
 Wiedl, K.H. 149
 Wiggins, J.S. 220ff.
 Witzlack, G. 84
 Wohlrab, U. 84, 86f.
 Wong, B.Y.L. 244, 273
 Wottawa, H. 77
 Wundt, W. 11ff., 325
 Wurst, E. 82, 134f., 140, 152, 261
 Zielinski, W. 268
 Zietz, K. 341
 Zimmermann, B. 122
 Zumkley, H. 94
 zur Oeveste, H. 222

9. Sachregister

- Abbauindex 121
- Abbildungsfunktion 45
- Abstraktionsfähigkeit 261
- Adaptives Testen 138
- Aggressives Verhalten 268
- Ähnlichkeitsfehler 97
- Akzelerationsmodell 285
- Algorithmus 303
- Allgemeinbegabung 154
- Allgemeine Intelligenz 27, 110ff., 125ff., 142, 144, 151ff., 162, 260f.
- Alpha-Fehler 16
- Alternativhypothesen 345
- Anamnese 102f., 107, 247, 342f.
- Anforderungsprofil 228
- Angst 201
- Anschauungsgebundenes Denken 167
- Anstrengungsvermeidungstendenz 190
- Aphasiesyndrom 121
- Application (Sternberg) 35
- Approximative Expertenabstimmung 224
- Aptitude-Treatment-Interaction (ATI) 17
- Äquivalenznormen 89
- Äquivalenz von Tests 57
- Arbeitsorgfalt 184
 - verhalten 182, 202, 268
- Arithmetisches Denken 167
- Attribuierungstendenz 191
- Aufgabenunabhängigkeit 59
- Aufmerksamkeit 182
- Aufnahmediagnostik 236, 244
- Auftraggeber 293, 296, 329, 332, 334f., 338, 342f., 353ff.
- AUKL (Automatische Klassifikation) 227
- Ausländerkinder 114, 142ff., 147, 149, 261, 264, 268
- Barnum-Effekt 341, 361
- Bayes-Theorem 312f.
- Bedingungsfaktoren der Schulleistung/
des Schulerfolges 217
- Bedingungsselektion 16
- Befund 335, 337, 349f., 353
 - skizze 254f, 349ff., 356
- Begabung 18f.
- Begabungsdiagnose 215
 - potenzen 149
- Belastbarkeit 129, 184
- Beobachtungsverfahren 94f.
- Beratung 222f.
- Beratungsanlaß 362, 374
 - gespräch 103, 259
- Beschreibungsmodi 95, 98
- Beta-Fehler 16
- Beurteilungsbogen 98f.
 - maßstäbe 225
- Beziehungserkennen 162
- Bezugsnormen 76
- Biographische Analyse 283
- Bipolare Skala 102
- Bitransitivität 47
- Branched-Testing 138
- Chance-Configuration-Theorie 23
- Check-list-Verfahren 98
- Closure 154, 159
- Comparison (Sternberg) 35
- Cronbachs α 58
- Curriculum 218
- Datenmodell 45
 - quelle 347, 350
 - schutz 330, 338
 - verarbeitungsstrategie 221
- Deckeneffekt 284
- Defektologie 239
- Deklaratives Wissen 33
- Denkfähigkeit 154, 159, 261
 - inhalte 22
 - kompetenz 274
 - operationen 22
 - produkte 22
- Denkprozesse
 - divergente 23, 180
 - konvergente 23
- Determinanten des sozialen Lernumfeldes 215
- Diagnose 213, 292f., 301f., 318, 320, 324ff., 349f., 356
- Diagnose-Prognose-Ansatz 214, 216
- Diagnosebogen 98f.
- Diagnostik
 - eindimensionale 263
 - mehrdimensionale 263
 - multidimensionale 244
 - treatmentabschließende 300
 - treatmentbegleitende 300
 - treatmentbezogene 299
- Diagnostische Fragestellung 350
 - Informationsgewinnung 102
 - Kompetenz 305
 - Situation 294
 - Strategie 293
 - Zielsetzung 250, 343, 355
- Diagnostischer Prozeß 292ff.

- Algorithmus 303
- Gütekriterien 297
- Kommunikation 332
- Komponenten 293ff.
- Zielsetzungen 295
- Diagnostisches Interview 103
- Diagnostisches Programm (Guthke) 86
- Differentielle Fähigkeitstests 29, 172f.
- Differenzierungsfähigkeit eines Tests 54, 71, 74
- Diskriminanzanalyse 227
- Diskriminationsbereich 263
- Dokumente 283
- Doppelbeleg 107
- DP-Faktoren 22
- Drei-Ringe-Modell (Renzulli) 21
- Durchführung (diagn. Untersuchung) 251
- Durchhaltevermögen 129
- Effektivität 287
- Eichstichprobe 72, 88
- Eichung 88
- Eigenschaftsliste (check-list) 99
- Eignungswahrscheinlichkeit 230f.
- Einfallsreichtum 178
- Einschätzung (rating) 98
- Einsichtsprobleme 32
- Einzelbefund 301, 337, 349, 356, 360
- Einzelfalldiagnose 279
 - -gespräch 105
 - -hilfe 15, 18, 236
- Empirisches Relativ 42
- Entscheidung 64, 213, 263, 293, 310, 315, 325ff., 334f., 350f.
 - investigatorische 300
 - psychologische 91
 - terminale 300
- Entscheidungsfehler 323
 - -findung 286
 - -prozeß 300f., 338
 - -strategie 222, 301, 345, 349
- Entwicklungsdiagnostik 278
 - -hilfe (individuelle) 280
 - -störung 125
 - -verzögerung 261
- Ergebnisdarstellung 251
- Erklärungsfunktion 216
- Erklärungshypothese 248, 321ff., 328, 343ff., 353, 357, 363, 369, 376
- Erwartungswert 39
- Erziehungsstil 209
 - -verhalten 209
- Evaluation 260, 340
- Exploration 102f., 107
- Fähigkeitsmodell (Förderdiagnostik) 268
 - parameter 79
 - -profil 231
 - -tests 283ff.
- Fair-share-Konzept (Testfairneß) 92
- Faktorenanalyse 64, 69
 - -theorie (Englische Schule) 25
 - -theorien der Intelligenz 22ff.
- Familiäre Sozialisationsfaktoren 218
- Fehler, systematischer 52
 - unsystematischer 52
- Fehlerquellen bei Beobachtung und Beurteilung 96
- Fehlerwert 52f., 59
- Feststellungshypothese 248, 321, 323, 328, 344, 363, 369
- Flexibilität 271
- Flexibility of Closure (Closure 2) 32
- Flüssige Intelligenz 27
- Förderdiagnostik 14, 142, 236f., 268
- Fragebogendaten 283
 - -technik 103
- Fragestellung (diagn.) 245, 257, 355
- Frageotypen (Exploration) 104
- Fremdrating 283
- Funktionsstörung 273
- Gedächtnisstrategie 274
- Gehörlose 112, 114
- Gelegenheitsbeobachtung 97
- Geltungsbereich eines Tests 71
- General Crystallized Ability Factor (GC-Factor) 28
- General Fluid Ability Factor 27
- Generalfaktorenthorie (Spearman) 25
- Gesamtbefund 256, 302, 337, 347, 349ff., 360, 371, 381
- Gesetze (psychologische) 326f., 327
- Gesetzesaussage 324
- Gespräch (diagnostisches) 106
- Gestaltauffassung 154
- Good Strategy User Modell 242
- Graphische Schätzskaala 101
- Grenzwertmethode 263
- Grobablesung (Screening) 286
- Grundintelligenz 142, 145, 151
- Gruppenfaktoren 25
- Gültigkeit, s. Validität
- Gutachten 334ff., 353ff., 382
 - Aufbau 355
 - mündliches 336
 - schriftliches 336
 - Struktur 359
- Halo-Effekt 97
- Hauptgütekriterien 48
- Heterogenität eines Tests 68

- Hierarchische Intelligenzstrukturtheorie (Vernon) 25f
- Hirnorganische Störungen 186
- Hirnschädigung 261
- Hochbegabte Schüler 34f., 282
- Hochbegabtenidentifikation 279f., 290
- Hochbegabung 170, 172, 277, 289
- Hochbegabungsdiagnostik 277ff., 282
 - diagnostischer Prozeß (Einzelfallhilfe) 290
- Hofeffekt 97
- Homogenität eines Tests 68
- Hörgeschädigte 110ff., 148, 185, 188
- Hypothese (diagnostische) 248, 301f., 316ff., 320ff., 345f., 349ff., 357
- Hypothesenbildung 248, 257
 - entscheidung 254, 356, 380
 - prüfung 254, 369
- Idealnorm 76
- Identifikation Hochbegabter 284, 286
- Identitätskonzept (Testfairmeß) 91
- Idiot-Savant 20
- Induktives Denken 33, 112, 268f.
- Inference (Sternberg) 35
- Inferenz (Diagnostik) 98
- Inferenzstatistik 39
- Informationsverarbeiter 315
 - verarbeitung 14, 33, 308, 314, 317
 - verarbeitungsgeschwindigkeit 35, 150, 242
- Insight-Skills (Sternberg) 33
- Integrierende Hirnfunktion 129
- Intelligenz 19
 - alter 237
 - modell (multiples) 19
 - modell von Jäger 24, 145
 - quantität 135
 - quotient 237
 - strukturmodell (Guilford) 22f.
 - tests 11, 110ff., 283
- Interaktiver Beratungsprozeß 224
- Interessen 194, 196, 197
 - berufsbezogene 196
 - Freizeit- 196
 - tests 200
- Interindividuelle Differenzen 11, 13
- Interpersonale Intelligenz 21
- Interpretationsphase 96
- Interventionshypothese 257
- Intrapersonale Intelligenz 21
- Inventarisieren 14
- Investigatorische Entscheidung 224
- Irrtumswahrscheinlichkeit 62, 262
- Itemanalyse 72f.
 - bias 91
 - charakteristik 80
 - pool 71
- Items
 - binäre 58
 - polychotome 83
- Itemselektion 73
 - unabhängigkeit 79
- Justification (Sternberg) 35
- Katamnese 260
- Kausalattribution 191
- Klassifikation 16, 219, 222
- Klassische Testtheorie 46, 84
- Klinische Vorhersage 221
- Kognitive Kompetenz 21
 - Lernleistungsbedingungen 217
 - Modifikation 268
 - Persönlichkeitsmerkmale 217, 282
 - Strategie 242
 - Schülermerkmale 215
- Kognitiver Komponenten-Ansatz (Sternberg) 32
- Kommunikation 335, 354
- Komponenten-Subtheorie 33
- Konfidenzintervall 61f.
- Konsistenz
 - analyse 56
 - koeffizient 60
- Konstitutionelle Bedingungen 215
- Konstrukt 67
- Konstruktive Fähigkeiten 167
- Kontrastfehler 97
- Kontrollstrategie 271
- Konzentration 182, 184ff., 268
- Konzentrationsschwäche 147
 - tests 189
 - vermögen 129
- Körperlich-kinästhetische Intelligenz 21
- Körperliche Entwicklung 106
- Korrelationskoeffizient 40
- Kovarianz 40
- Kreative (soziale) Umwelt 24
- Kreative Persönlichkeit 24
- Kreativer Prozeß 24
- Kreatives Produkt 24
- Kreativität 24, 178, 180
- Kreativitätsforschung 22ff.
 - messung 22
 - tests 181f., 283ff.
- Kristallisierte Intelligenz 28
- Kriterien zur Hypothesenprüfung 249
- Kriterium 216, 219
- Kriteriumsleistung 219
 - orientierte Diagnostik 14

- Kritische Lebensereignisse 21
- Kritischer Diskriminationsbereich 263
- Kuder-Richardson-Formeln 56, 58
- Kulturfairneß 149
- Kurzgutachten 258, 337, 372
- Kurzzeitleerntest 85
- L-Daten 94
- Längsschnittbetrachtung 106
- Langgutachten 337
 - -zeitlerntest 85
- Legasthenie 142f., 147
- Lehrer-Schüler-Beziehungen 210
- Lehrercheckliste 283
 - -verhalten 208
- Leistungsdisposition 19
 - -motiviertheit 155
 - -störung 272
 - -vermögen unter Zeitdruck 129
- Lernbehinderte Schüler 131, 139, 142, 149, 239, 243
 - -behinderung 236, 240f.
 - -fähigkeit 84
 - -gewinn 85
 - -kompetenz 274
 - -potenz 84, 92
 - -quotient 135
 - -sensible Meßverfahren 278
 - -techniken 202
 - -test 84
- Lernumfeld 17, 218
 - soziales 18
- Lernumwelt (Tests) 211f.
 - -verhalten 268
 - -zielmodell (Förderdiagnostik) 268
- Lese-rechtschreibschwache Schüler 268
- Level I (Intelligenz) 29
- Level II (Intelligenz) 29
- Literaturhinweise 36, 41, 48, 77, 78, 83, 87, 93, 102, 108, 212, 235, 276, 290, 333, 362
- Logisch-mathematische Intelligenz 21
- Logischer Fehler 97
- Logistisches Modell 79
- Lokale stochastische Unabhängigkeit 54, 79, 86
- Mapping (Sternberg) 35
- Maximum-Likelihood-Schätzung 80
- Mechanisch-technisches Verständnis 173
- Median 39
- Mehrfachbeleg 107
- Memory (Intelligenzfaktor) 31
 - -fehler 61, 76
 - -instrument 283f.
 - -modell 45, 47
- Messung 14, 37, 41f., 45
- Meßverfahren, äquivalente 55
 - parallele 55
- Metakognition 242, 244, 271, 273
- Metakognitive Kompetenz 35
- Metakognitives Wissen 243, 274
- Metakomponenten 33
 - der kognitiven Kontrolle 282
- Methodenprobleme 213
- Milieuschädigung 261
- Minderbegabung 236, 238
- Mittelwert 38
- Mixed-Rasch-Modell 83
- Modalwert 39
- Moderatoren 215f., 230
- Modifikationsstrategie 295, 319, 322, 325, 354
- Motivation zu beruflicher Bildung 197
- Motivationstests 192
- Motive 204
- Multiple Faktoretheorie (Thurstone) 29
- Multitrait-Multimethods-Matrix 69f.
- Musikalische Intelligenz 21
- Nebengütekriterien 48, 50
- Neuropsychologie 272f.
- Neurotische Störungen 186
- Nicht-kognitive Persönlichkeitsmerkmale (Lernleistungsbedingungen) 217, 282
- Niveau-Test 28
- Nominationsverfahren 283
- Nomogramm 90
- Normierung 48, 50, 72, 88
 - beim Rasch-Modell 83
- Normorientiertheit 76
 - -tabelle 89
- Novizen-Experten-Paradigma 34
- Number Factor 31, 154
- Numerische Ratingskala 100
- Numerisches Relativ 42
- Nützlichkeit 48
 - eines Tests 50
- Oberschulsyndrom 121
- Objektivität 48, 70, 72
 - Auswertungs- 49
 - der diagnostischen Konsequenzen 49
 - Durchführung- 48
 - Interpretations- 49
 - spezifische 79
- Odd-even-Methode 57
- Ökonomie 48, 50, 71, 287
- Operationalisierung 67, 249, 296, 322ff., 345ff., 349, 363
- Organikersyndrom 121
- Overlapping-Problem 227

Pädagogisierungsphase 84
 Parallelität von Tests 57
 Paralleltestreliabilität 60
 Parameter 39
 Parametermodell 45, 78
 Perceptual Speed 31, 154
 Performanzkomponenten 33
 Persönliche Gleichung 12
 Persönlichkeitstests 207f.
 – -diagnostik 295f.
 – -fragebogen 204
 Personselektion 16
 Pflichteifer 190
 Physikalisch-techn. Problemlösen 175
 Placierungsentscheidung 17, 222, 262
 Population 39, 53
 Posttest 84
 Power-Test 28
 Prädiktoren 215f., 219, 221
 Praktische Intelligenz 173
 Prätest 84
 Primärfähigkeiten 30, 154
 Primärfaktorenmodell (Thurstone) 29f.
 Problemlösestrategie 271f.
 Produkt-Moment-Korrelationskoeffizient 40
 Profilauswertung 121f., 133f., 156f.
 Prognose 64, 213, 221, 292f., 320, 327, 350, 356
 Prognosefunktion 216
 – -gültigkeit 213
 – -kriterium 225
 – -validität (AUKL-Eignungskriterien) 229
 Projektionsfehler 97
 Protokollierung (Beobachtungsdaten) 95
 Prozedurales Wissen 33
 Prozeß, diagnostischer 223, 292ff.
 Prozeßanalyse 285
 – -diagnostik 14, 295
 – -komponenten (Informationsverarbeitung) 34
 – -merkmale 218
 – -variable 282
 Prüfungsangst 201
 Psychodiagnostisches Gespräch 102, 252
 Psychologische Gesetze 319f., 325
 Psychometrisches Paradigma 12, 34
 Psychosoziale Konzeption (Hochbegabung) 21
 Q-Daten 94
 Qualitative Unterschiede zw. Hoch- und Nichthochbegabten 285
 Range der Intelligenz 136
 Rasch-Modell 79
 Ratingverfahren 98
 Räumliche Intelligenz 21
 Räumliches Vorstellen 155
 Raumvorstellung 112, 154, 159
 Reasoning 31, 154
 Rechenfähigkeit 167
 – -fertigkeit 154
 – -schwäche 147
 Rechnerisches Denken 162
 Rechtsgrundlagen psychologischer Diagnostik 109, 329f.
 Reflexivität 271
 Registrierung von Explorationsdaten 106
 Regression 40
 Regressionskonzept (Testfairneß) 91
 Relation 42
 – -zweiwertige 47
 Reliabilität 48f., 54, 58f., 61, 67, 70, 72f., 262
 – -index 55, 67
 – -bei Lerntests 86
 – -beim Rasch-Modell 82
 – -Konsistenz- 56f.
 – -Kuder-Richardson- 56
 – -Paralleltest- 56
 – -Retest- 56
 – -Spearman-Brown- 56
 – -Split-Half- 56
 – -Wiederholungs- 56
 Residualgewinn 85
 Response 35
 Risiko erster Art (Alpha-Fehler) 286
 Schätzung 39
 Schätzverfahren 98
 Schlußfolgerndes Denken 155, 162, 261
 Schularspezifische Klassennormen 230
 Schuleignungsermittlung 214, 225
 – -kennbereiche (nach AUKL) 230ff.
 – -klassifikation 228
 – -maßstäbe 231
 – -prognose 213, 215
 Schülerbeobachtungsbogen 97
 Schulerfolg 215, 219
 Schulisches Lernumfeld 217
 Schullaufbahnberatung 15, 17
 Schulleistung 215
 Schulleistungsanalyse 214
 – -test 265, 267
 – -versagen 236
 Schulunlust 201
 Schwachsinnigensyndrom 121
 Schwellenhypothese (Guilford) 36, 278
 Schwerhörige 114

Schwierigkeit 78
 Schwierigkeitsindex 73
 – -parameter 79
 Screening 283
 Sehgeschädigte 147
 Selbstbild 204
 – -konzept 21
 – -nomination 283
 – -rating 283
 Selektion 16, 219
 Selektionsdiagnostik 14
 – -entscheidung 262, 330
 – -strategie 295, 319, 325, 354
 Separierbarkeit 75, 81
 Sequentielle Entscheidungsstrategie 225, 287
 Situationsbezogene freie Verhaltensbeobachtung 97
 Skala 42f., 99
 – Guttman- 46
 – Intervall- 44
 – Klassifikations- 43
 – Nominal- 43
 – Norm- 88
 – Ordinal- 43
 – Rational- 44
 – Verhältnis- 44, 81
 Skalenanalyse 72
 – -typ 43
 Sonderpädagogische Diagnostik 125, 132, 149, 236, 260ff., 268
 Sonderschulüberweisungsverfahren 142, 144, 149, 236, 239, 244, 262f, 265
 Soziale Erwünschtheit 201
 Soziale Intelligenz 21
 Soziale Interaktion (Gespräch) 106
 Soziales Vorurteil 97
 Sozialisationshintergrund 105
 Sozialklima 210
 Soziokulturelle Bedingungsvariablen des Lernumfeldes 282
 Space 31, 154
 Spearman-Brown-Formel 56
 Speed of Closure (Closure 1) 32
 Spezialbegabung 20
 Spezielle Fähigkeitstests 176f.
 Sprachgebundenes Denken 167
 Sprachgeschädigte 114, 148
 Sprachliche Intelligenz 21
 Sprachliche Leistungsfähigkeit 155
 Sprachverständnis 162, 167
 Stabilität 59
 Standard-Schätzskala (man-to-man) 102
 Standardabweichung 39
 – -meßfehler 61
 – -norm-Äquivalente 88
 – -normen 88
 – -schätzfehler 66
 Statistik 39
 – erschöpfende 78f.
 Statistische Vorhersage 221
 Statusdiagnostik 14, 84, 295
 Stellungnahme 349, 354ff.
 Stichprobe 38f.
 – Klumpen- 39
 – repräsentative 39
 – Zufalls- 39
 Stichprobenabhängigkeit 75
 – -unabhängigkeit 79
 Strategie 242f.
 Streuung 39
 Strukturmerkmale der Schule 217
 Systematische Verhaltensbeobachtung 97
 T-Daten 94
 Talent 22
 Talentsuche 279f.
 Technische Begabung 159
 Technologische Hilfen 226
 Teilleistungsstörungen 272
 Terminale Entscheidung 224
 Test (Definition) 11
 Testauswertung 253
 Testen 14
 Testfairneß 90f.
 – -halbierung 57
 – -konstruktion 70
 Testmodell 13
 – allgemeines logistisches 83
 – deterministisches 46
 – Klassisch Latent-Additives 65, 77
 – Klassisches 48
 – kriteriumsorientiertes 77
 – Mixed-Rasch- 83
 – nach Birnbaum 79
 – nach Rasch 79
 – probabilistisches 47, 77, 78
 – qualitatives 83
 Testtheorie 37
 Trennschärfe 73, 78
 Triarchische Intelligenztheorie 32f.
 Übertrittsberatung 215
 Umweltbedingungen 217
 Unipolare Skala 102
 Unterrichtsmerkmale 210
 Untersuchungsanlaß 245
 – -bericht 347, 355f., 360, 365
 – -hypothese 349
 – -planung 342, 346

- -variable 281, 284
- Urteil (diagnostisches) 304, 310, 313f.
- Urteiler 318
- Urteilsbildung 293
 - klinische 306
 - statistische 306
- Validität 48, 50, 63, 70, 72f., 263
 - beim Rasch-Modell 82
 - diskriminante 69
 - inhaltliche 64, 68, 76
 - Konstrukt- 64, 67
 - konvergente 69
 - Kriteriums- 64f., 67
 - Prognose- 229
 - Übereinstimmungs- 65, 228
 - von Lerntests 87
 - Vorhersage- 65
- Varianz 39
- Verarbeitungsgeschwindigkeit 145
 - -kapazität 145
- Verbal Comprehension 30
- Verbal Factor 154
- Verbale Produktivität 178
- Verbalität 159
- Vergleichbarkeit 48
- Vergleichsprozesse 35
- Verhaltensbeobachtung 283
 - -beurteilung 98
 - -diagnostik 295f.
 - -gestörte Schüler 131
 - -stile 204
- Verlagsanschriften der Testverlage 109
- Verteilung 38
- Vertrauensintervall 59, 89, 262
- Visuell-motorische Koordination 129
- Vorhersage des Schulerfolgs 218
- Vorhersagegültigkeit 229
 - -hypothese 328
 - -variablen (Prädiktoren) 220
- Vorkenntnisse 214
- Wahrnehmungstempo 142, 154, 159
 - -umfang 142
- Wahrscheinlichkeit 307f., 312f., 315, 319, 323
- Wert
 - beobachteter 50, 52
 - Fehler- 52f., 59
 - gemessener 50, 53
 - wahrer 51ff., 59, 76
- Wiederholungsmessung 51, 59
 - -reliabilität 60
- Wissenserwerbskomponenten 33
- Word Fluency 31, 154
- Wortflüssigkeit 154
- Zielpopulation eines Tests 71
- Zufallsfaktor 21
 - -größe 50
- Zugehörigkeitswahrscheinlichkeit 228
- Zusatzhypothese 256, 353, 371
- Zuverlässigkeit (Tests) 54
 - s. auch Reliabilität
- Zwei-Facetten-Subtheorie 32
- Zwei-Faktoren-Theorie (Spearman) 25
 - (Cattell) 27, 149
 - (Jensen) 29

10. Testregister

- Adaptives Intelligenz Diagnostikum (AID) 134ff., 152, 261, 264
- Advanced Progressive Matrices (APM) 110ff., 152
- Allgemeiner Schulleistungstest für 2. Klassen (AST 2) 164
- Angstfragebogen für Schüler (AFS) 148, 201f., 207, 210
- Anstrengungsvermeidungstest (AVT) 190, 192, 206
- Arbeitsverhaltensinventar (AVI) 202f., 207
- Arbeitsverhaltensmodifikation (AVM) 202f.
- Attribuierungsfragebogen für Erfolg und Mißerfolg in der Schule für 5. bis 7. Klassen (AEM 5-7) 191f.
- Aufgaben zum Nachdenken (AzN 4+) 173, 188
- Aufmerksamkeits-Belastungs-Test (Test d2) 113, 148, 176, 185ff., 188f.
- Begabungstestsystem (BTS) 186
- Berufs-Interessen-Test II (BIT II) 193ff., 200,
- Bildertest 1-2 und Bildertest 3-4 (BT 1-2 und BT 3-4) 143, 153, 164
- Bildungsberatungstest (BBT) 181
- Binetarium 113, 123f., 133
- Bourdon-Figuren-Durchstreichtest 185, 188f.
- Cognitive Abilities Tests (CAT) 162
- Coloured Progressive Matrices (CPM) 110ff., 146, 152, 269
- Columbia Mental Maturity Scale (CMM 1-3) 124, 126
- Columbia Mental Maturity Scale für lernbehinderte Sonderschüler (CMM-LB) 261
- Culture Fair Intelligence Test (CFT) 141ff. (s. auch Grundintelligenztest)
- Deutscher High School Personality Questionnaire (HSPQ) 207
- Diagnostischer Rechtschreibtest (DRT 2) 143
- Diagnostisches Soziogramm (DSO) 212
- Differentieller Interessen-Test (DIT) 179, 195ff., 200
- Differentieller Leistungstest (DL-KG) 189
- Differentieller Wissenstest (DWT) 176
- Differentieller Leistungstest (DL-KE) 189
- Dortmunder Skala zum Lehrerverhalten (DSL) 208f., 211
- Durchstreichtest 185
- Duisburger Vorschul- und Einschulungstest (DVET) 143, 164
- Erfassungsbogen für aggressives Verhalten in konkreten Situationen (EAS) 207
- Figure Reasoning Test (FRT) 176
- Figurenfolgentest 87
- Form-Lege-Test (FLT) 174
- Fragebogen Einstellung zur Schule für 4. bis 6. Klassen (FES 4-6) 192
- Fragebogen für Schüler (FS 5-10) 207
- Fragebogen für Schüler (FS 11-13) 207
- Fragebogen zur Erfassung des schulischen Leistungsmotivs für 5. bis 7. Klassen (FSL 5-7) 192
- Fragebogen zum Selbstkonzept für 4. bis 6. Klassen (FSK 4-6) 207
- Frankfurter Tests für Fünfjährige - Konzentration (FTF-K) 187ff.
- Grundintelligenztest Skala 1 (CFT 1) 141ff., 143, 152, 164, 261, 264, 269, 272
- Grundintelligenztest Skala 2 (CFT 20) 141, 143ff., 152, 167, 169, 261, 264, 272
- Grundintelligenztest Skala 3 (CFT 3) 141, 143ff., 152
- Gruppenintelligenztest für lernbehinderte Sonderschüler (CMM-LB) 261
- Hamburg-Wechsler-Intelligenztest für Erwachsene (HAWIE) 115ff., 119ff, 130, 133ff, 151, 186
- Hamburg-Wechsler-Intelligenztest für Kinder 113f, 122f, 124, 125ff, 131ff, 137, 140, 143, 149
- Hamburg-Wechsler-Intelligenztest für Kinder - Revision (HAWIK-R) 120, 124ff, 126ff., 135ff., 151, 261, 264, 269
- Hamburger Neurotizismus- und Extraversionsskala für Kinder und Jugendliche (HANES-KJ) 207
- Hamburger Erziehungsverhaltensliste für Mütter (HAMEL) 209ff.
- Hamburger Persönlichkeitsfragebogen für Kinder (HAPEF-K) 207
- Hamburger Neurotizismus- und Extraversionsskala für Kinder und Jugendliche (HANES-KJ) 201, 207, 210
- Hand-Dominanz-Test (HDT) 176

- Hannover Wechsler Intelligenztest für das Vorschulalter (HAWIVA) 125f, 133, 152, 261
- Heidelberger Hörprüf-Bild-Test (HHBT) 261
- Heidelberger Intelligenztest (HIT) 3-4 172, 182
- Intelligenz-Struktur-Test (IST 70) 120, 146, 172, 179, 183, 186
- IPC-Fragebogen zu Kontrollüberzeugungen 192
- Kinder-Angst-Test (KAT) 201, 207
- Kognitiver Fähigkeits-Test für 4. bis 13. Klassen sowie Berufsschüler und Studierende (KFT 4-13+) 141, 148, 151, 162, 165ff., 173, 230, 269, 284
- Kognitiver Fähigkeitstest für das Kindergartenalter (KFT-K) 162ff, 173, 261
- Kognitiver Fähigkeitstest für die Grundschule (KFT 1-3) 162ff., 169, 173, 261, 269, 173, 269
- Konzentrations-Verlaufstest (KVT) 186, 188f.
- Konzentrations-Leistungs-Test (KLT) 182ff., 186, 188f.
- Konzentrationstest für das erste Schuljahr (KT 1) 187, 189
- Konzentrationstest für 3. und 4. Klassen (KT 3-4) 187
- Körperkoordinationstest (KTK) 177
- Kettwiger Schulreifeftest (KST) 143
- Landauer Skalen zum Sozialklima für 4. bis 13. Klassen (LASSO 4-13) 210ff.
- Leistungsmotivationstest für Jugendliche (LMT-J) 192
- Leistungsprüfsystem (LPS) 146, 153, 158ff, 172, 174, 176, 179, 183, 230
- Lern- und Gedächtnistest (LGT 3) 176
- Lerntestbatterie Schlußfolgerndes Denken (LTS) 85f.
- LM-Gitter 192
- Mannheimer Intelligenztest für Kinder und Jugendliche (MIT-KJ) 167, 172
- Mannheimer Biographisches Inventar (MBI) 182, 208
- Mannheimer Test zur Erfassung des physikalisch-technischen Problemlösens (MTP) 170, 175f.
- Mechanical Aptitude Battery 174
- Mechanisch-technischer Verständnistest (MTVT) 173ff., 176, 179
- Pauli-Test 184f., 186, 188f.
- Persönlichkeitsfragebogen für Kinder zwischen 9 und 14 Jahren (PFK 9-14) 131, 148, 204ff., 208
- Primary Mental Abilities (PMA) 153
- Progressiver Matrizentest (PMT) 110ff., 167, 179
- Prüfsystem für Schul- und Bildungsberatung (PSB) 113, 139, 146f, 149f, 158ff., 167, 169ff, 181f., 230, 269
- Raven-Matrizen-Tests 84, 110
- Rechtschreibtest (RST) 146
- Reduzierter Wechsler Intelligenztest (WIP) 123, 133, 146
- Reutlinger Test für Schulanfänger (RTS) 143
- Revisions-Test (Rev.T.) 189
- Schlauchfiguren 177
- Schul- und Berufsinteressentest (SBIT) 200
- Schulleistungstest lernbehinderter Schüler (SLS) 267
- Schulleistungstestbatterie für Lernbehinderte und für schulleistungsschwache Grundschüler I und II (SBL I und II) 265f.
- Seashore-Test für musikalische Begabung 177
- Snijders-Oomen nicht-verbale Intelligenztestreihe (SON) 114
- Sozialfragebogen für Schüler für 4. bis 6. Klassen (SFS 4-6) 212
- Soziometrischer Test für 3. bis 7. Klassen (ST 3-7) 212
- Standard Progressive Matrices (SPM) 110ff., 152
- Stanford-Binet 124
- State-Trait Angstinventar (STAI) 208
- Test zur Untersuchung des praktisch-technischen Verständnisses (PTV) 176f
- Test zum divergenten Denken (Kreativität) für 4. bis 6. Klassen (TDK 4-6) 180ff.
- Testbatterie für geistig behinderte Kinder (TBGB) 261
- Testbatterie für Entwicklungsrückständige Schulanfänger (TES) 261
- Testverfahren für Interessen: Beruf - Schule (TIBS 1) 197ff., 200
- Themenwahlverfahren für 4. bis 9. Klassen (TWV 4-9) 200
- Trail-Making-Test (TMT) 150
- Verbaler Kreativitäts-Test (VKT) 177ff., 181
- Vorschul-Lerntest (VLT) 85f.
- Wechsler Intelligence Scale for Children (WISC) 113f., 124, 126ff.

Wechsler Intelligence Scale for Children - Revised (WISC-R) 124, 126ff.
Wechsler Preschool and Primary Scale of Intelligence (WPPSI) 125
Wechsler Adult Intelligence Scale (WAIS) 115f., 120, 124f., 133ff.
Wechsler Adult Intelligence Scale - Revised (WAIS-R) 115, 133ff.
Weilburger Testaufgaben für Schulanfänger (WTA) 143
Wilde Intelligenz-Test (WIT) 196
Wortschatztest (Ergänzungstests zum CFT 20) (WS) 143, 152
Wortschatztest (WST 7-8) 148
Zahlen-Verbindungs-Test (ZVT) 150ff., 170
Zahlenfolgentest (Ergänzungstests zum CFT 20) (ZF) 143, 152