



DIW Berlin

Deutsches Institut
für Wirtschaftsforschung

Research Notes

20

Erfassung kognitiver Kompetenzen im Vorschul- bis Jugendalter: Intelligenz, Sprache und schulische Fertigkeiten

Empfehlungen zum Ausbau des Erhebungsinstrumentariums über Kinder im Sozio-oekonomischen Panel (SOEP)

Expertise

**Sabina Pauen
Janna Pahnke
Isa Valentiner**

IMPRESSUM

DIW Berlin, 2007
Deutsches Institut für Wirtschaftsforschung
Königin-Luise-Str. 5
14195 Berlin
Tel. +49 (30) 897 89-0
Fax +49 (30) 897 89-200
www.diw.de

ISSN 1860-2185

Alle Rechte vorbehalten.
Abdruck oder vergleichbare
Verwendung von Arbeiten
des DIW Berlin ist auch in
Auszügen nur mit vorheriger
schriftlicher Genehmigung
gestattet.



Research Notes 20

Sabina Pauen*

Janna Pahnke*

Isa Valentiner*

**Erfassung kognitiver Kompetenzen im Vorschul- bis
Jugendalter:**

Intelligenz, Sprache und schulische Fertigkeiten

Empfehlungen zum Ausbau des Erhebungsinstrumentariums über
Kinder im Sozio-oekonomischen Panel (SOEP)

Expertise

Berlin, April 2007

* Universität Heidelberg, Psychologisches Institut, sabina.pauen@psychologie.uni-heidelberg.de

Inhaltsverzeichnis

1 Darstellung des aktuellen Forschungsstandes	1
1.1 Definitionen von Intelligenz	1
1.2 Strukturmodelle der Intelligenz	2
1.2.1 Das Generalfaktor-Modell.....	2
1.2.2 Das Modell mehrerer gemeinsamer Faktoren	2
1.2.3 Gruppenfaktormodelle.....	3
1.2.4 Informationsverarbeitungsmodelle.....	4
1.3 Messung von Intelligenz	5
1.3.1 Entwicklung der Intelligenzmessung.....	5
1.3.2 Der psychometrische Ansatz der Intelligenzmessung	6
2 Testverfahren zur Erfassung kognitiver Kompetenzen bei Kindern und Jugendlichen	8
2.1 Intelligenzmessung vom Kindergarten- bis ins junge Erwachsenenalter	9
2.1.1 Mehrdimensionale Intelligenztestverfahren	13
2.1.2 Eindimensionale Intelligenztestverfahren	35
2.2 Messung der Sprachkompetenz im Vorschulalter	44
2.3 Messung der Lese-Rechtschreibkompetenzen im Schulalter	48
2.4 Messung der mathematischen Kompetenzen im Schulalter	55
3 Empfehlungen zur Umsetzung geeigneter Testverfahren im Rahmen des SOEP	60
3.1 Erfassung eines Maßes der allgemeinen kognitiven Leistungsfähigkeit	60
3.2 Erfassung spezifischer kognitiver Fertigkeiten.....	62
3.2.1 Sprachliche Kompetenz.....	62
3.2.2 Lesefertigkeiten	63
3.2.3 Rechtschreibfertigkeiten.....	63
3.2.4 Mathematische Fertigkeiten	63
Literatur.....	66

Verzeichnis der Tabellen

Tab. 1 Testverfahren zur Erfassung kognitiver Kompetenzen bei Kindern und Jugendlichen	8
Tab. 2 Mehrdimensionale Intelligenztestverfahren	14
Tab. 3 K-ABC	15
Tab. 4 K-TIM	20
Tab. 5 HA WIK-III	23
Tab. 6 WIE	26
Tab. 7 AID 2	29
Tab. 8 PSB-R 4-6	31
Tab. 9 I-S-T 2000 R	34
Tab. 10 Eindimensionale Intelligenztestverfahren	36
Tab. 11 SON-R	37
Tab. 12 Grundintelligenztest	39
Tab. 13 Sprachliche Fähigkeiten	45
Tab. 14 SETK	46
Tab. 15 Lese-Rechtschreibtests	48
Tab. 16 Rechentests	55
Tab. 17 HRT 1-4	56
Tab. 18 ZAREKI	58
Tab. 19 Erfassung allgemeiner kognitiver Leistungsfähigkeit	61
Tab. 20 Erfassung von Sprachfertigkeiten	62
Tab. 21 Erfassung von Lese- und Rechtschreibkompetenzen	63
Tab. 22 Erfassung mathematischer Kompetenzen	64

Vorbemerkung

Die folgende Expertise wurde mit dem Ziel verfasst, der SOEP Arbeitsgruppe bei der Auswahl geeigneter Methoden zur Erfassung kognitiver Kompetenzen vom 4. bis 16. Lebensjahr hilfreich zu sein. Ausgehend von einer Beschreibung des aktuellen Forschungsstandes sollen konkrete Empfehlungen gegeben werden, welche Rahmenbedingungen der SOEP-Erhebung (zeitlich, räumlich, personell) berücksichtigen.

Bis heute liegt für den deutschsprachigen Raum noch kein Überblick über einschlägige Verfahren zur Erfassung der kognitiven Leistungsfähigkeit im Kindes- und Jugendalter vor, der theoretische Konzepte zur Beschreibung von Intelligenz einerseits und eine differenzierte Darstellung von Testverfahren andererseits, übersichtlich kombiniert. Diese Lücke wird durch die vorliegende Expertise geschlossen.

Zunächst folgt ein kurzer Überblick über gängige Definitionen und Strukturmodelle von Intelligenz sowie über die Entwicklung der Messung intellektueller Fähigkeiten. Wie noch zu zeigen sein wird, spielt dieses Hintergrundwissen für Entscheidungen bezüglich der Vorgehensweise innerhalb der SOEP-Erhebung eine wichtige Rolle. Im Mittelpunkt der Darstellung steht jedoch eine systematisierte Zusammenstellung von existierenden Instrumenten zur kognitiven Diagnostik und empirischen Befunden, zu deren Reliabilität und Validität. Abschließend werden ausgewählte Testverfahren im Hinblick auf ihre Anwendung innerhalb des SOEP diskutiert und Empfehlungen zur Umsetzung gegeben. Diese Empfehlungen sind gestaffelt (abhängig von der zur Verfügung stehenden Zeit) und nach verschiedenen Zielsetzungen (Erfassung eines Maßes der allgemeinen kognitiven Leistungsfähigkeit oder Erfassung spezifischer Kompetenzen) gegliedert.

1 Darstellung des aktuellen Forschungsstandes

1.1 Definitionen von Intelligenz

Den meisten Intelligenzdefinitionen ist die Vorstellung einer *allgemeinen Denk- und Problemlösefähigkeit* gemeinsam. So verstanden der Pädagoge Alfred Binet und der Arzt Theophile Simon (1905; zit. nach Amelang, 1995), welche die ersten modernen Intelligenztests entwickelten, Intelligenz als „die Art der Bewältigung einer aktuellen Situation“, genauer als „*gut urteilen, gut verstehen, gut denken*“. Der Hamburger Psychologe William Stern (1920, S. 2f) definierte Intelligenz als die „*allgemeine Fähigkeit eines Individuums, sein Denken bewusst auf neue Forderungen einzustellen; sie ist die allgemeine geistige Anpassungsfähigkeit an neue Aufgaben und Bedingungen des Lebens*“. Der amerikanische Psychologe Wechsler (1964, S. 13), der die breit verwendeten Wechsler-Tests für Kinder und Erwachsene entwickelte, konzeptualisierte Intelligenz als „*zusammengesetzte oder globale Fähigkeit eines Individuums, zweckvoll zu handeln, vernünftig zu denken und sich mit seiner Umwelt wirkungsvoll auseinanderzusetzen*“. Wenzl (1957, S. 14) sieht Intelligenz als „*die Fähigkeit zur Erfassung und Herstellung von Bedeutungen, Beziehungen und Sinnzusammenhängen*“. Einen ähnlichen Gedanken greift Hofstätter (1971, S. 197) auf, der Intelligenz bezeichnet als die „*Fähigkeit zur Auffindung von Ordnung*“.

Wesentliche Gemeinsamkeiten dieser unterschiedlichen Definitionen von Intelligenz sind die Betonung der *Anpassungsfähigkeit an neue Situationen*, das *Erfassen von Beziehungen* und das *Lösen von Aufgaben mittels Denken und Einsicht*. Das Hervorheben dieser allgemeinen Denk- und Problemlösefähigkeiten grenzt die Definition allgemeiner Intelligenz von anderen, nicht intellektuellen Begabungsbereichen wie der sozial-emotionalen, musikalischen, bildnerisch-darstellenden oder psychomotorischen Begabung ab (vgl. Stapf, 1998). Wenn im Folgenden von Intelligenz gesprochen wird, ist somit ausschließlich die *intellektuelle Begabung* einer Person gemeint. Zum *Aufbau* von Intelligenz gibt es verschiedene Theorien. Während den meisten Intelligenzkonzepten die Vorstellung eines allgemeinen Intelligenzfaktors zu Eigen ist, unterscheiden sich die Modelle hinsichtlich der postulierten Komponenten der Intelligenzstruktur. Da das theoretische Konzept maßgeblich den Aufbau des jeweiligen Intelligenztestverfahrens sowie die dadurch erfassten Fähigkeitsbereiche bestimmt, sollen im Folgenden einige zentrale Intelligenzstrukturmodelle vorgestellt werden.

1.2 Strukturmodelle der Intelligenz

1.2.1 Das Generalfaktor-Modell

Die erste präzise Modellvorstellung im Hinblick auf die Struktur der Intelligenz entwickelte Charles Spearman (1904) mit der Formulierung seiner „Zwei-Faktoren-Theorie“. Die zentrale Behauptung dieser Theorie besteht darin, dass jedes Maß für Intelligenz auf zwei Faktoren beruht, nämlich einem Anteil zu Lasten der so genannten „*general intelligence*“ (*g*) und einem Anteil, der *spezifisch* (*s*) für die jeweilige Aufgabe oder den jeweiligen Test ist. Der Zusammenhang zwischen verschiedenen Leistungsmaßen basiert entsprechend auf der gemeinsamen Komponente, nämlich *g*. Ein solcher allgemeiner Intelligenzfaktor zeigt sich empirisch immer wieder in den durchweg positiven Interkorrelationen verschiedener kognitiver Leistungen und hat sich „als die größte Quelle interindividueller Unterschiede in kognitiven Aktivitäten jedweder Art herausgestellt“ (Amelang, 1995, S. 249). Fast allen gängigen Intelligenztests liegt heute die Vorstellung von einem *g*-Faktor zugrunde, d.h. gewöhnlich wird neben den Punktwerten für die Einzelleistungen auch ein Wert für die Gesamtleistung angegeben.

1.2.2 Das Modell mehrerer gemeinsamer Faktoren

Intelligenz wird von verschiedenen Autoren als bestehend aus mehreren Faktoren definiert. Einer der bekanntesten Vertreter dieses Konzeptes ist Thurstone (1931). Thurstone nahm an, dass beim Lösen von Denkaufgaben mehrere unabhängige Gruppenfaktoren in wechselnden Gewichtungsverhältnissen beteiligt sind. Er postulierte sieben so genannte „primary abilities“: (1) *verbal comprehension*, (2) *word fluency*, (3) *number*, (4) *space*, (5) *memory*, (6) *perceptual speed* und (7) *reasoning/induction*. Alle diese Primärfaktoren stehen in Thurstones Modell gleichberechtigt nebeneinander.

Die Forderung nach völliger Unabhängigkeit der Primärfaktoren konnte jedoch nicht bestätigt werden. Thurstone fand in seiner Stichprobe zwar keine nennenswerten Zusammenhänge zwischen den einzelnen Faktoren, was jedoch hauptsächlich an der eingeschränkten Varianz der Leistungsmaße in seiner Stichprobe lag (vgl. Amelang & Bartussek, 2001). Er untersuchte nämlich eine sehr homogene Stichprobe von Collegestudenten, die alle eine ähnlich gute Leistung zeigten. In einer Stichprobe mit größerer Leistungsvarianz, die repräsentativer für die Normalverteilung von Intelligenz in der Population ist, korrelieren die Thurstoneschen

Primärfaktoren in einer Höhe von ca. $r = .35$ miteinander (Fischer, 1958). Solche Zusammenhänge erlauben aber die Durchführung von Sekundäranalysen mit der Extraktion von Faktoren zweiter Ordnung (Amelang & Bartussek, 2001). Auf dieser Methode basieren Gruppenfaktormodelle, die von einer hierarchischen Ordnung der Intelligenzfaktoren auf mehreren Ebenen ausgehen (Primär- und Sekundärfaktoren).

1.2.3 Gruppenfaktormodelle

In Vernons (1965) hierarchischem Modell der Intelligenzstruktur sind auf der untersten Ebene die spezifischen, nur den betreffenden Test kennzeichnenden Faktoren lokalisiert, darüber die „minor group factors“, dann die „major group factors“ und auf der Ebene des höchsten Allgemeingrades der g -Faktor. Die beiden größeren Gruppenfaktoren bezeichnet Vernon als $v:ed$ (*verbal educational*) und $k:m$ (*spatial and motor abilities*). Darunter gliedern sich die kleineren Gruppenfaktoren wie beispielsweise f (*fluency*), n (*numerical*) und p (*perceptual*). Die Ähnlichkeit dieser Faktoren mit den Thurstoneschen Primärfaktoren macht deutlich, dass hierarchische Modelle wie dasjenige von Vernon eine Art Kompromiss zwischen der Zwei-Faktoren-Theorie Spearman's und dem Modell mehrerer unabhängiger Faktoren von Thurstone darstellen.

Eine neuere Konzeption im Bereich der Gruppenfaktormodelle wurde von Cattell (1963; Horn & Cattell, 1966) entwickelt. Bei der Faktorisierung der miteinander korrelierenden Thurstoneschen Primärfaktoren fand Cattell zwei Sekundärfaktoren, die er als „fluid“ (g_f) und „crystallized“ (g_c) *fehlt was general intelligence* interpretierte. Die *fluide Intelligenz* spiegelt dabei die Fähigkeit wider, sich neuen Problemen und Situationen anzupassen, ohne dass besonderes Wissen oder Lernerfahrungen erforderlich sind. Hiermit sind Denkfähigkeiten wie das logische Schließen oder das Erkennen von Analogien gemeint. Die *kristalline Intelligenz* besteht aus kognitiven Fertigkeiten, die sich aus vorangegangenem Lernen herauskristallisiert haben, wie z.B. Wortschatz und Sprachverständnis. Während die fluide Intelligenz relativ kulturunabhängig ist und mehr genetisch bestimmt zu sein scheint, wird die kristalline Intelligenz in hohem Maß von gesellschaftlichen und kulturellen Erfahrungen bestimmt. Die Identifizierung von g_f und g_c gelingt empirisch immer wieder, wobei beide Komponenten keinesfalls voneinander unabhängig sind, sondern in einer Höhe von ca. $r = .50$ korrelieren (vgl. Amelang, 1995). Eine erneute Faktorisierung dieser beiden Sekundärfaktoren ergibt auf tertiä-

rer Ebene einen Faktor von noch größerem Allgemeingrad, der weitestgehend Spearman's Generalfaktor g entspricht.

Carroll (1993) versuchte in seiner Drei-Schichten-Theorie einen Konsens dieser Ansätze zu bilden: In der untersten Schicht befinden sich sehr spezialisierte Fähigkeiten, wie z.B. Gedächtnisspanne, Leseverständnis, Fremdsprachenbeherrschung, Visualisierung oder Wahrnehmungsgeschwindigkeit. In der nächsten Schicht befinden sich breiter angelegte Fähigkeiten, wie kristalline Intelligenz, fluide Intelligenz, Gedächtnis, visuelle und auditive Wahrnehmung, kognitive Geschwindigkeit und Verarbeitungsgeschwindigkeit. In der obersten Schicht ergibt sich eine Art g -Faktor der allgemeinen Intelligenz.

1.2.4 Informationsverarbeitungsmodelle

Eine Abkehr von den hierarchischen Modellvorstellungen zur Struktur der Intelligenz nahm Guilford (1967) in seinem „Structure of Intellect“-Modell vor, einem Informationsverarbeitungsmodell, in das intellektuelle Prozesse wie *Kognition*, *Gedächtnis* und *Problemlösen* mit aufgenommen wurden. Auch im Berliner Intelligenzstrukturmodell von Jäger (1984) finden sich sechs Hauptfaktoren, die sich in zwei Gruppen unterteilen lassen. Neben Inhalten wie *figural-bildhaft*, *verbal*, *numerisch*, finden sich Operationen wie *Bearbeitungsgeschwindigkeit*, *Einfallsreichtum*, *Gedächtnis* und *Verarbeitungskapazität*. Einen Versuch, die psychometrische Tradition der Strukturmodelle mit der Tradition des Informationsverarbeitungsansatzes zu verbinden, unternahm Sternberg (1985). Er entwickelte die „Triarchische Theorie der Intelligenz“, die aus drei Teiltheorien besteht: (1) Der *Komponententheorie*, die drei Arten von Prozesskomponenten beschreibt: Metakomponenten (Kontrolle von Plänen, Abläufen und Bewertungen), Performanzkomponenten (Ausführen geeigneter Problemlösestrategien) und Wissenserwerb-Komponenten (Erwerb von Problemlöse- und Informationsverarbeitungs-fähigkeiten); (2) der *Erfahrungstheorie*, die sich auf Erfahrungen und die Fähigkeit, mit neuartigen Anforderungen umzugehen und die Informationsverarbeitung zu automatisieren, bezieht; und (3) der *Kontexttheorie*, die Funktionen, die in der Alltagswelt die Anpassung an die Umgebung gewährleisten sowie soziales und praktisches Verhalten umfasst.

Zusammenfassend lässt sich sagen, dass die unterschiedlichen Forschungsergebnisse innerhalb der Intelligenzstruktur-Kontroverse (Generalfaktor vs. mehrere Faktoren) größtenteils als Folge der angewendeten Analysemethoden sowie auch der Merkmals- und Personenstichproben erklärt werden können (vgl. Amelang & Bartussek, 2001). Daher „herrscht doch mittler-

weile weitgehender Konsens darüber, dass so etwas wie ein Generalfaktor der Intelligenz und/oder ein Satz von miteinander in Beziehung stehenden Gruppenfaktoren existiert“ (Hasselhorn & Grube, 1997, S. 17), so dass den meisten Experten der *Generalfaktor* als Arbeitsdefinition für Intelligenz dient (Gottfredsen, 1999).

1.3 Messung von Intelligenz

1.3.1 Entwicklung der Intelligenzmessung

Die Entwicklung der Intelligenzmessung begann Ende des 19. Jahrhunderts mit den Versuchen von Francis Galton (einem Cousin von Charles Darwin), intellektuelle Fähigkeiten durch *Sinnes- und Gedächtnisprüfungen* zu messen. Galton richtete 1884 auf der „International Health Exhibition“ in London ein „anthropometrisches Labor“ ein, in dem Ausstellungsbesucher gegen eine Gebühr von drei Pence ihr Unterscheidungsvermögen für visuelle, akustische oder Tastreize untersuchen lassen konnten. Galton (1888; zit. nach Asendorpf, 1996) wollte überprüfen, ob und wie diese Sinnes- und Gedächtnisleistungen zueinander in Beziehung stehen und welche Zusammenhänge zum Studienerfolg bestehen, der schon damals als wichtiges Außenkriterium zur Validierung von Intelligenztests galt. Es zeigte sich jedoch, dass die einzelnen „mental tests“ nur sehr gering miteinander korrelierten und keine substantiellen Beziehungen zum Studienerfolg aufwiesen (Cattell & Farrand, 1896; Wissler, 1901; zit. nach Asendorpf, 1996). Damit hatten sich Sinnesleistungen als ungeeignet zur Erfassung intellektueller Fähigkeiten erwiesen.

Eine neue Perspektive entstand durch den Ansatz von Binet und Henri (1895; zit. nach Asendorpf, 1996), die vorschlugen, intellektuelle Fähigkeiten auf einem höheren Komplexitätsniveau zu testen. Der Auftrag durch das französische Unterrichtsministerium, ein Testverfahren zu konzipieren, mit dem lernschwache und sonderschulbedürftige Schüler zuverlässig identifiziert werden konnten (denn diese durften nur noch auf der Grundlage eines medizinisch-pädagogischen Gutachtens in die Sonderschule eingewiesen werden), war die Geburtsstunde der ersten modernen Intelligenztests. Binet und Simon (1905) entwickelten eine Reihe von Aufgaben, zu deren Lösung jeweils unterschiedliche intellektuelle Fähigkeiten nötig sind. Diese ursprünglichen *Binet-Aufgaben* erwiesen sich als erfolgreich für die Frage nach der Sonderbeschulung. Sie wurden daher in der Folge systematisch verbessert und sind die Grundlage für viele der heutigen Intelligenztests für Kinder und Jugendliche.

1.3.2 Der psychometrische Ansatz der Intelligenzmessung

Der psychometrische Ansatz besteht darin, intellektuelle Fähigkeiten mit genormten Testverfahren zu messen. Dabei wird die Leistung einer Person mit den Leistungen einer Vergleichsgruppe (meist gleichen Alters, Geschlechts oder Ausbildungsstandes) in Beziehung gesetzt. Binet und Simon (1905) als Begründer der ersten modernen Intelligenztests, verglichen die Leistung eines Kindes mit der alterstypischen Leistung. Sie verwendeten als Maß für die Intelligenz die *Differenz* zwischen dem Lebensalter und dem „Intelligenzalter“ eines Kindes, das sich zusammensetzte aus der Anzahl der im Test gelösten Aufgaben. Eine bestimmte Differenz zwischen Intelligenz- und Lebensalter konnte auf verschiedenen Altersstufen jedoch eine unterschiedliche Bedeutung haben, nämlich eine unauffällige Entwicklung oder extrem niedrige Intelligenz, je nachdem, in welchem Alter ein Kind diesen Rückstand zeigte.

Um solchen Verzerrungen vorzubeugen, entwickelte der deutsche Psychologe William Stern (1911) den bis heute als Begriff sehr populären *Intelligenzquotienten*, bei dem das Intelligenzalter durch das Lebensalter geteilt wird und dann mit 100 multipliziert wird, um ganze Zahlen zu erhalten. Dieser Quotient zeigt eine ähnliche Standardabweichung auf unterschiedlichen Altersstufen, nämlich ungefähr 15 bei den Binet-Aufgaben. Der Intelligenzquotient nach Stern gewährleistet jedoch nur bei einer absolut linearen Leistungszunahme mit dem Alter eine konstante Interpretierbarkeit von Leistungsvorsprüngen und -rückständen auf verschiedenen Altersstufen. Tatsächlich findet aber ein negativ beschleunigter Entwicklungsverlauf der Intelligenz statt, d.h. der Leistungszuwachs in Intelligenztests nimmt mit wachsendem Alter ab, bis die Leistung mit etwa 17 Jahren ein Plateau erreicht (vgl. Bayley, 1970).

Heute ist es daher üblich, den von Wechsler (1939; deutsch: 1964) eingeführten *Abweichungsquotienten* zu verwenden, bei dem der in einem Intelligenztest erzielte Punktwert x nur mit den Leistungen der entsprechenden Altersstufe verglichen wird. Ausgangspunkt ist der empirische Befund einer Leistungsstreuung jeder Altersgruppe um einen Durchschnittswert, der traditionellerweise jeweils mit 100 gleichgesetzt wird. Die individuelle Abweichung von diesem Mittelwert ($x - M$) wird auf die jeweilige Altersstreuung (SD) relativiert und dann der von den Binet-Tests bekannten Streuung ($s = 15$) durch Multiplikation mit dem Wert 15 angepasst (vgl. Amelang & Bartussek, 2001):

$$IQ = 100 + 15 \frac{x - M}{SD}$$

IQ-Werte sind damit in der Regel so normiert, dass ihr Mittelwert 100 und ihre Standardabweichung 15 beträgt, und zwar auf allen Altersstufen. Die IQ-Werte lassen sich zudem in Prozentränge (Perzentile) umrechnen, die angeben, wie viel Prozent der Normstichprobe genauso gut oder schlechter abgeschnitten haben. Ein Prozentrang (PR) von 90 bedeutet beispielsweise, dass die Person genauso gut oder besser als 90 Prozent der Stichprobe abgeschnitten hat und dass nur 10 Prozent bei diesen Aufgaben bessere Leistungen erbringen. Ein Prozentrang von 50 kennzeichnet einen durchschnittlichen Wert.

Intellektuelle Fähigkeiten sind normalverteilt, daher befinden sich im Bereich Mittelwert \pm eine Standardabweichung etwa zwei Drittel aller Personen. Dieser Bereich gilt als der Durchschnittsbereich. Folglich gelten Intelligenzwerte als normal bzw. durchschnittlich, die bis zu einer Standardabweichung unter oder über dem Mittelwert liegen, also den Skalenbereich von 85 bis 115 IQ-Punkten (PR 16 bis 84) umfassen. Als über- bzw. unterdurchschnittlich gelten Werte, die mehr als eine Standardabweichung nach oben oder nach unten vom Mittelwert abweichen. Werte ab einer Standardabweichung über dem Mittelwert, also ein $IQ \geq 116$ bzw. $PR \geq 85$, bedeuten eine überdurchschnittliche Leistung. Als intellektuell hochbegabt bezeichnet man Personen, die einen Wert von zwei Standardabweichungen über dem Mittelwert ($IQ \geq 130$ bzw. $PR \geq 97-98$) erreichen, also die obersten 2-3% der Altersgruppe ausmachen (vgl. Waldmann & Weinert, 1990). Entsprechend kennzeichnet ein Ergebnis, das mehr als eine Standardabweichung unter dem Mittelwert liegt ($IQ < 85$ bzw. $PR < 16$) ein unterdurchschnittliches Ergebnis. Ab einem Wert von zwei Standardabweichungen unter dem Mittelwert ($IQ < 70$ bzw. $PR < 3$) spricht man von geistiger Behinderung.

2 Testverfahren zur Erfassung kognitiver Kompetenzen bei Kindern und Jugendlichen

Die unten stehende Tabelle 1 gibt zunächst einen Überblick über gängige Testverfahren zur Erfassung kognitiver Kompetenzen im Kindergarten-, Schulalter und jungen Erwachsenenalter. Alle angegebenen Verfahren sind in Forschung und Anwendungsdiagnostik häufig verwendete Instrumente. In den folgenden Abschnitten wird auf die Erfassung von (1) Intelligenz (ein- und mehrdimensional), (2) Sprachfertigkeiten, (3) Lese-Rechtschreibfertigkeiten und (4) mathematischen Kompetenzen eingegangen. Dabei werden alle in der Tabelle genannten Verfahren kurz vorgestellt, jedoch nur diejenigen ausführlich erläutert, die international anerkannt und besonders verbreitet sind.

Tabelle 1
Testverfahren zur Erfassung kognitiver Kompetenzen bei Kindern und Jugendlichen

Testverfahren		Altersbereich (Jahre; Monate)	Erfasste Kompetenzen
Mehrdimensionale Intelligenztestverfahren			
Kaufmann-Assessment Battery for Children	K-ABC	2;11 – 12;6	Fluide und kristalline Intelligenz, Gedächtnis
Kaufmann-Test zur Intelligenzmessung für Jugendliche und Erwachsene	K-TIM	ab 12;0	Fluide und kristalline Intelligenz
Hamburg-Wechsler-Intelligenztests	HAWIVA	4;0 – 6;6	Allgemeine Intelligenz, verbale und handlungsbezogene Fähigkeiten
	HAWIK	6;0 – 16;11	
	WIE	ab 16;0	
Allgemeines Intelligenz Diagnostikum	AID 2	6;0 – 15;11	Allgemeine Intelligenz, verbal-akustische und manuell-visuelle Fähigkeiten
Prüfsystem für Schul- und Bildungsberatung	PSB 4-6	4. – 6. Klasse	Primärfaktoren, Allgemeine Intelligenz
	PSB 6 –13	6. – 13. Klasse	
Kognitiver Fähigkeitstest	KFT-K	4;7 – 7;11	Primärfaktoren, Denkopoperationen
	KFT 1-3	6;0 – 12;11	
	KFT 4-12+R	4.-12. Klasse	
Intelligenz-Struktur-Test	I-S-T 2000 R	ab 15;0	Fluide und kristalline Intelligenz, Primärfaktoren, Wissen
Berliner Intelligenz-Struktur-Test	BIS-HB	12;6 – 16;5	Allgemeine Intelligenz, Operative und inhaltsgebundene Intelligenzfähigkeiten
	BIS-4	ab 16;0	

Fortsetzung nächste Seite

Fortsetzung Tabelle 1

Eindimensionale Intelligenztestverfahren			
Snijders-Oomen Nonverbaler Intelligenztest	SON-R 2½-7	2;6 – 7;11	Fluide Intelligenz
Grundintelligenztest	SON-R 5½-17	5;6 – 17;11	
Raven Matrizen Test	CFT 1	5;3 – 9;5	Fluide Intelligenz
	CFT 20-R	8;5 – 19;11	
Raven Matrizen Test	CPM	3;9 – 11;8	Fluide Intelligenz
	SPM	ab 6;0	
	APM	ab 12;0	
Sprachliche Fähigkeiten			
Sprachentwicklungstest für Drei- bis Fünfjährige	SETK 3-5	3;0 – 5;11	Sprachverständnis, Sprachproduktion, Sprachgedächtnis
Heidelberger Sprachentwicklungstest	HSET	3;0 – 9;11	Sprachverständnis, Sprachproduktion
Reynell-Skalen		15 Mon. – 7 J.	Sprachverständnis, Sprachproduktion
Lese-Rechtschreibtests			
Salzburger Lese-Rechtschreibtest	SLRT	1. – 4. Klasse	Lese- und Rechtschreibfähigkeiten
Diagnostischer Rechtschreibtest	DRT 1, DRT 2, DRT 3, DRT 4, DRT 5	1. – 5. Klasse	Rechtschreibfähigkeiten
Hamburger Schreibprobe	HSP 1, HSP 2, HSP 3, HSP 4, HSP 5-9	1. - 9. Klasse	Rechtschreibfähigkeiten
ELFE-Lesetest		1. – 5. Klasse	Lesefähigkeiten
Salzburger Lesescreening	SLS 1-4	1. – 4. Klasse	Leseverständnis
	SLS 5-8	5. – 8. Klasse	
Stolperwörter			Lesefähigkeiten
Wortlisten		ab 16;0	Lesefähigkeiten
Rechentests			
Heidelberger Rechentest	HRT	1. – 4. Klasse	Mathematische Basiskompetenzen
Neuropsychologische Testbatterie für Zahlenverarbeitung und Rechnen bei Kindern	ZAREKI	7;6 – 10;11	Rechnerische Fertigkeiten und Zahlenverarbeitung
Diagnostischer Rechentest	DRE 3	3. Klasse	Rechenfertigkeit und Rechenverständnis

2.1 Intelligenzmessung vom Kindergarten- bis ins junge Erwachsenenalter

Zur Erfassung der Intelligenz auf mehreren Dimensionen liegen eine Reihe gut erprobter und standardisierter Erhebungsinstrumente wie die Kaufman- und Wechsler-Tests vor. Bei der *Kaufman-Assessment Battery for Children* (K-ABC, Melchers & Preuß, 2003 a+b) handelt es sich um die deutschsprachige Fassung eines in den Vereinigten Staaten entwickelten Ver-

fahrens (Kaufman & Kaufman, 1983), das die intellektuellen Fähigkeiten und Fertigkeiten von Kindern im Alter von 2;6 bis 12;5 Jahren erfassen soll. Der **Kaufman-Test zur Intelligenzmessung** für Jugendliche und Erwachsene (K-TIM; Melchers, Schürmann & Scholten, 2006) ist die deutsche Fassung des Kaufman Adolescent and Adult Intelligence Test (KAIT, Kaufman & Kaufman, 1993). Es handelt sich um einen Individualtest zur Erfassung der allgemeinen Intelligenz bei Jugendlichen und Erwachsenen in einem Altersbereich von 11 bis über 85 Jahren. Beide Verfahren ermöglichen eine umfassende Messung von Intelligenz und die Erfassung kognitiver Stärken und Schwächen. Sie dienen der differenzierten psychologisch-diagnostischen Untersuchung. Weiterhin finden die Kaufman-Tests Anwendung bei der Beurteilung von Kindern und Jugendlichen mit Lern- und Leistungsbeeinträchtigungen sowie anderen Auffälligkeiten, mit dem Ziel der Erstellung, Einleitung und Durchführung von Fördermaßnahmen. Aufgrund ihrer neuropsychologischen Aspekte lassen sie sich zur Abklärung kognitiver und neuropsychologischer Beeinträchtigungen einsetzen. Aufgrund ihrer weiten Verbreitung und ihres vielfältigen Einsatzes soll auf diese Verfahren unten noch genauer eingegangen werden.

Weit verbreitet sind auch die **Wechsler-Tests zur Intelligenzmessung**, die auf David Wechsler zurückgehen. Diese liegen in drei Formen vor: Für das Vorschulalter von 4;0 bis 6;6 Jahren (HAWIVA, Eggert & Schuch, 1975), für das Schulalter von 6;0 bis 16;11 (HAWIK-III, Tewes, Rossmann & Schallberger, 2002) und für das Erwachsenenalter von 15 bis 89 Jahren (WIE, Von Aster, Neubauer & Horn, 2006). Der *HAWIK-III* ist ein Individualtest zur Untersuchung der allgemeinen kognitiven Entwicklung und kann in der differenzierten Diagnostik kognitiver Fähigkeiten von Kindern und der Beschreibung von Leistungsprofilen mit speziellen Stärken und Schwächen eingesetzt werden. Im Rahmen testpsychologischer Untersuchungen lassen sich Ursachen von Lern- und Leistungsstörungen im schulischen Bereich feststellen. Beim *WIE* handelt es sich ebenfalls um einen Individualtest, der sich für eine differenzierte Messung der allgemeinen geistigen Fähigkeiten einsetzen lässt. Er ermöglicht weiterhin die Untersuchung von Lern- und Leistungsstörungen sowie die Hochbegabendiagnostik. Ebenfalls auf dem Wechsler-Testkonzept beruht das **Adaptive Intelligenzdiagnostikum 2** (AID 2) von Kubinger und Wurst (2000), welches mittlerweile in der zweiten Auflage erschienen ist und zur differenzierten, adaptiven Diagnostik bei Kindern und Jugendlichen von 6 bis 15 Jahren eingesetzt wird. Während auf HAWIK, WIE und AID 2 unten noch ausführlicher eingegangen wird, wird auf eine Beschreibung des HAWIVA verzichtet. Das Verfahren

ist bisher nur in der Erstauflage von 1975 erhältlich. Aufgrund der veralteten Normierung (Flynn-Effekt), kann der jetzige HAWIVA nicht mehr zur Anwendung empfohlen werden. Allerdings ist eine Überarbeitung des Verfahrens in Vorbereitung (HAWIVA-III) und soll demnächst erscheinen.

Das **Prüfsystem zur Schul- und Bildungsberatung** (P-S-B) ist ein schulspezifisches Intelligenztestverfahren, das auf dem Intelligenzkonzept der Thurstoneschen Primärfaktoren basiert. Das P-S-B liegt für die 4. bis 6. Klasse (PSB-R 4-6; Horn, Lukesch, Kormann und Mayrhofer, 2002) und für die 6. bis 13. Klasse vor (PSB-R 6-13; Horn, Lukesch, Mayrhofer und Kormann, 2003). Das P-S-B erfasst in erster Linie schulrelevante kognitive Fähigkeiten und wird häufig in der Schul- und Bildungsberatung eingesetzt.

Das Testsystem der **Kognitiven Fähigkeitstests** (KFT) basiert auf einem ähnlichen Intelligenzkonzept und umfasst drei Testverfahren zur Erfassung kognitiver (Lern-) Fähigkeiten vom Vorschul- bis ins Erwachsenenalter: Die Kindergartenform (KFT-K; Heller & Geisler, 1983), die Grundschulform (KFT 1-3; Heller & Geisler, 1983) und den kognitiven Fähigkeitstest für 4. bis 13. Klassen (KFT 4-13; Heller, Gaedike & Weinläder, 1976). Während für die ersten beiden Verfahren nur die ursprünglichen Normen der Ersterscheinung vorliegen, ist das Verfahren für Schüler ab der vierten Klasse mehrfach überarbeitet worden und in neu normierter Form erhältlich (KFT 4-12+ R; Heller und Perleth, 2000). Dieses dient der differentiellen Erfassung kognitiver Fähigkeitsdimensionen, die besonders für schulisches Lernen relevant sind. Der deutschsprachige Test (Dauer 2 ½ Zeitstunden oder 3 Schulstunden à 45 Minuten) basiert auf dem Cognitive Abilities Test (CAT) von Thorndike & Hagen (1971/4), einer Testbatterie zur Erfassung abstrakter Intelligenz, und wird in erster Linie für Untersuchungen im Rahmen der Bildungsberatung, Unterrichtsdifferenzierung und Bildungsforschung verwendet. Da der KFT für jüngere Kinder aufgrund der veralteten Normen nicht mehr zur Anwendung empfohlen werden kann, und der KFT 4-12+ R in der Durchführung relativ viel Zeit in Anspruch nimmt, wird als Beispiel für schulspezifische Intelligenzdiagnostik nur das Prüfsystem zur Schul- und Bildungsberatung (P-S-B) später näher erläutert.

Ein weit verwendetes Verfahren zur Intelligenzdiagnostik für Jugendliche und junge Erwachsene ist der **Intelligenz-Struktur-Test 2000 R** (I-S-T 2000 R) von Amthauer, Brocke, Liepmann und Beauducel (2001). Dieses mittlerweile in der zweiten Auflage erschienene Verfahren wird als Individual- und Gruppentest vielfältig zur differentiell-psychologischen Intelligenzdiagnostik und -forschung eingesetzt und später noch genauer beschrieben.

Der *Berliner Intelligenz-Struktur-Test* (BIS-Test) von Jäger, Süß und Beauducel (1997) basiert auf dem Berliner Intelligenzstrukturmodell von Jäger (1984) und erfasst mit 45 sehr verschiedenen Aufgaben zu operativen, inhaltsgebundenen Fähigkeiten und Kreativität eine außergewöhnliche Breite und Vielfalt von Intelligenzleistungen. Bisher liegen nur Normen für 16- bis 19-jährige Schüler vor. Seit Neuestem gibt es auch den *Berliner Intelligenzstrukturtest für Begabungs- und Hochbegabungsdiagnostik* (BIS-HB; Jäger, Holling, Preckel, Schulze, Vock, Süß & Beauducel, 2006), der bei durchschnittlich begabten und hochbegabten Jugendlichen im Alter von 12 bis 16 Jahren eingesetzt werden kann. Beide Tests sind mit Durchführungszeiten von ca. 130 und 140 Minuten (plus zwei 10- bzw. 15-minütigen Pausen) sehr aufwendig in der Umsetzung; allerdings können Kurzformen mit ca. 45 und 55 Minuten Testdauer vorgenommen werden.

Neben den genannten mehrdimensionalen Intelligenztestverfahren existieren eine Reihe weiterer Intelligenztests, die vor allem die Dimension der fluiden Intelligenz erfassen sollen und daher als eindimensionale Verfahren bezeichnet werden. Zur Unterscheidung von ein- und mehrdimensionalen Verfahren führen Holling et al. (2004) aus, dass bei mehrdimensionalen Tests aufgrund des sehr viel heterogeneren Aufgabenmaterials eine Vielzahl spezifischer intellektueller Fähigkeiten erfasst werden kann, was mit dem eher homogenen Aufgabenmaterial zur Erfassung fluiden Intelligenz nicht in diesem Maße möglich ist. Zu den eindimensionalen Verfahren können die *Snijders-Oomen Non-verbale Intelligenztests* gezählt werden. Der SON-R 2½ - 7 (Tellegen, Winkel, Wijnberg-Williams & Laros, 1998) ist ein umfassendes nonverbales Verfahren für zwei- bis siebenjährige Kinder; der SON-R 5½ - 17 (Snijders, Tellegen & Laros, 1997) kann bei Kindern und Jugendlichen zwischen 5;6 und 17 Jahren eingesetzt werden. Dem SON-R liegt kein spezifisches Intelligenzkonzept zugrunde. Er kann jedoch dem Intelligenzmodell von Cattell zugeordnet werden, da die Aufgaben hauptsächlich fluide Intelligenz erfassen sollen.

Außer diesen sehr aufwendigen, zumeist individuell zu erhebenden Verfahren liegen noch weitere eindimensionale Intelligenztests vor, die zur Anwendung als Gruppentest geeignet sind und weniger aufwendig durchzuführen sind. Der *Grundintelligenztest* basiert auf den Culture Fair Intelligence Tests (CFT) von Cattell und liegt in zwei Formen mit aktueller Normierung vor: Dem *Grundintelligenztest – Skala 1* (CFT 1) ab dem Vorschulalter (5-9 Jahre; Weiß & Osterland, 1997) und dem *Grundintelligenztest – Skala 2* (CFT 20-R) für das Schul- und Erwachsenenalter (8-60 Jahre; Weiß, 2006). Der Grundintelligenztest dient der

eindimensionalen Erfassung eines wesentlichen Aspektes intelligenten Verhaltens: der Fähigkeit, Denkprobleme in neuartigen Situationen anhand von sprachfreiem, figuralem Material zu erfassen. Das Verfahren untersucht diese Aspekte auf ökonomische Weise sprachfrei und unabhängig vom sozialen oder kulturellen Hintergrund.

Die **Raven-Matrizen-Testserie** geht auf Raven (1941) zurück und dient der ökonomischen Messung der fluiden Intelligenz bei Vorschulkindern, Schulkindern und Erwachsenen. Es liegen drei Formen vor: Die *Coloured Progressive Matrices (CPM)* für jüngere Kinder (3;9 bis 11;8 Jahre), alte oder geistig behinderte Personen (Raven, Raven & Court, 1995; deutsche Bearbeitung: Bulheller & Häcker, 2002), die *Standard Progressive Matrices (SPM)* für Kinder ab 6;0 Jahren und Erwachsene (Raven, Raven & Court, 1996; dt. Bearbeitung: Heller, Kratzmeier & Lengfelder, 1998) und die *Advanced Progressive Matrices (APM)* für überdurchschnittlich intelligente Jugendliche ab 12;0 Jahren und Erwachsene (Raven, Raven & Court, 1994; dt. Bearbeitung: Heller, Kratzmeier & Lengfelder, 1998). Sämtliche Aufgaben bestehen aus geometrischen Figuren oder Mustern, die aus jeweils sechs dargebotenen Antwortalternativen ergänzt werden sollen. Dieses häufig eingesetzte und einfach durchzuführende nonverbale Verfahren soll die fluiden Anteile des Generalfaktors der Intelligenz besonders gut erfassen. Sowohl der Grundintelligenztest als auch die Raven-Matrizen werden daher ausführlich erläutert.

2.1.1 Mehrdimensionale Intelligenztestverfahren

Auf die nachfolgend genannten mehrdimensionalen Intelligenztestverfahren soll nun näher eingegangen werden (Tabelle 2):

Tabelle 2
Mehrdimensionale Intelligenztestverfahren

Kaufmann-Assessment Battery for Children	K-ABC	2;11 – 12;6	Fluide und kristalline Intelligenz, Gedächtnis
Kaufmann-Test zur Intelligenzmessung für Jugendliche und Erwachsene	K-TIM	ab 12;0	Fluide und kristalline Intelligenz
Hamburg-Wechsler-Intelligenztest	HAWIK	6;0 – 16;11	Allgemeine Intelligenz, verbale und handlungsbezogene Fähigkeiten
Wechsler Intelligenztest für Erwachsene	WIE	ab 15;0	Allgemeine Intelligenz, verbale und handlungsbezogene Fähigkeiten
Allgemeines Intelligenz Diagnostikum	AID 2	6;0 – 15;11	Allgemeine Intelligenz, verbal-akustische und manuell-visuelle Fähigkeiten
Prüfsystem für Schul- und Bildungsberatung	PSB 4-6	4. – 6. Klasse	Primärfaktoren
	PSB 6 –13	6. – 13. Klasse	
Intelligenz-Struktur-Test	I-S-T 2000 R	ab 15;0	Fluide und kristalline Intelligenz, Primärfaktoren, Wissen

Kaufman-Assessment Battery for Children (K-ABC)

Testkonzept: Bei der Kaufman-Assessment Battery for Children (K-ABC, Melchers & Preuß, 2003 a+b) handelt es sich um die deutschsprachige Fassung eines in den Vereinigten Staaten entwickelten Verfahrens (Kaufmann & Kaufmann, 1983), das die intellektuellen Fähigkeiten und Fertigkeiten von Kindern im Alter von 2;6 bis 12;5 Jahren erfassen soll. Grundlage des Verfahrens ist die Intelligenztheorie von Cattell und Horn, die zwischen Fähigkeiten, mit unbekanntem Problemen umzugehen („fluide Intelligenz“) und den durch früheres Lernen erworbenen Fertigkeiten („kristalline Intelligenz“) unterscheidet. Allgemein werden intellektuelle Fähigkeiten von den Autoren als Fähigkeiten definiert, Probleme durch Denkopoperationen zu lösen. Im Vordergrund steht die effektive Informationsverarbeitung. Der Test gliedert sich in insgesamt 16 Aufgaben, die drei allgemeinen Skalen zuzuordnen sind: Aspekte der fluiden Intelligenz werden über die Skalen *einzelheitliches Denken* und *ganzheitliches Denken* erfasst. Beim *einzelheitlichen Denken* geht es um Operationen mit sequentiellen Anordnungen von Reizen, während beim *ganzheitlichen Denken* ein gestalthaftes und oftmals räumliches Vorgehen erforderlich ist. Große Bedeutung haben außerdem Gedächtnisfähigkeiten. Die der kristallinen Intelligenz zuzuordnende Skala *Fertigkeiten* erfasst Faktenwissen als Ausdruck erfolgten Lernens (Melchers & Preuß, 1991b).

Skalen und Untertests: Die Untertests der K-ABC sind zwei Bereichen zugeordnet: der *Skala intellektueller Fähigkeiten* und der *Fertigkeitenskala*. Die Skala intellektueller Fähigkeiten umfasst die Skala *Einzelheitliches Denken* mit drei Untertests und die Skala *Ganzheitliches Denken* mit sieben Untertests. Die Fertigkeitenskala besteht aus sechs Untertests. Es besteht weiterhin die Möglichkeit, aus fünf Untertests eine *Nonverbale Skala* zu bilden (vgl. Tabelle 3 für die Zuordnung der Untertests zu Skalen und Subskalen). Insgesamt enthält die K-ABC somit 16 Untertests, von denen je nach untersuchtem Altersbereich eine Auswahl zur Anwendung kommt: (1) *Zauberfenster*, (2) *Wiedererkennen von Gesichtern*, (3) *Handbewegungen*, (4) *Gestaltschließen*, (5) *Zahlennachsprechen*, (6) *Dreiecke*, (7) *Wortreihe*, (8) *Bildhaftes Ergänzen*, (9) *Räumliches Gedächtnis*, (10) *Fotoserie*, (11) *Wortschatz*, (12) *Gesichter und Orte*, (13) *Rechnen*, (14) *Rätsel*, (15) *Lesen/Buchstabieren*, (16) *Lesen/Verstehen*.

Tabelle 3
K-ABC

Skalen und Untertests	Aufgaben	Inhalte
Fähigkeiten		
Einzelheitliches Denken		
(3) Handbewegungen	eine Reihe gezeigter Handbewegungen soll imitiert werden	einzelheitliches Verarbeiten visuelles KZG, Organisation der Wahrnehmung, mot. Reproduktion einer Serie, räumliche Fähigkeiten, visuomotorische Koordination
(5) Zahlennachsprechen	vorgesprochene Zahlenreihen müssen nachgesprochen werden	akustisches KZG, Wiedergabe einer Vorgabe, Umgang mit Zahlen, akustisch-verbales Verständnis
(7) Wortreihe	vorgesprochene Wörter sollen anhand von Bildern erkannt und gezeigt werden	akustisches KZG, akustisch-verbales Verständnis, akustisch-visuelle Integration, frühe Sprachentwicklung
Ganzheitliches Denken		
(1) Zauberfenster	durch Drehen einer Scheibe mit Sichtschlitz sind immer nur kleine Bereiche einer Figur zu sehen, die erkannt und benannt werden soll	ganzheitliches Verarbeiten Integration sequentiell dargebotener visueller Reize, visuelles KZG, räumliche Fähigkeiten, sprachliches Ausdrucksvermögen, Verständnis Beziehung Teil/Ganzes, Detailbeachtung, LZG
(2) Wiedererkennen von Gesichtern	ein Gesicht soll wahrgenommen, eingepägt und in einer Gruppe von Personen wiedererkannt werden	Strategien zum visuellen Suchen, Wahrnehmen und Wiedererkennen eines Gesichts, visuelles KZG, Detailbeachtung
(4) Gestaltschließen	auf unvollständigen Schwarz-Weiß-Zeichnungen muss ein Gegenstand wiedererkannt und benannt werden	wahrnehmungsgebundenes Schließen und Folgern, Umwandlung abstrakter Reize in ein konkretes Objekt, LZG, Detailbeachtung, räumliche Fähigkeiten, sprachliches Ausdrucksvermögen

Fortsetzung Tabelle nächste Seite

Fortsetzung Tabelle 3

(6) Dreiecke	anhand einer Vorlage soll mit blauen und gelben Gummidreiecken eine geometrische Figur reproduziert werden	nonverbale Konzeptbildung, visuo-motorische Koordination, räumliche Fähigkeiten, Reproduktion eines Modells, logisches Denken, Verständnis Beziehung Teil/Ganzes
(8) Bildhaftes Ergänzen	anhand von zwei Figuren muss eine visuell dargebotene Analogie gebildet und ein zweites Bildpaar mit dem passenden Bild ergänzt werden	Detailbeachtung, Organisation der Wahrnehmung, logisches Denken, räumliche Fähigkeiten
(9) Räumliches Gedächtnis	in einem Koordinatensystem sollen mehrere kurz dargebotene Bilder an der richtigen Stelle wiedergegeben werden	visuelles KZG, Reproduktion einer Vorlage, visuelle Organisation
(10) Fotoserie	eine Bildergeschichte auf Fotos muss in die chronologisch richtige Reihenfolge gebracht werden	Detailbeachtung, analytische Fähigkeiten, Verständnis Beziehung Teil/Ganzes, visuelle Wahrnehmung
Fertigkeiten		kristallin
(11) Wortschatz	auf Bildern dargestellte Gegenstände sollen benannt werden	Erlernen sprachlicher Beziehungen, verfügbarer Wortschatz, LZG, Bildung von Sprachkonzepten, sprachliches Ausdrucksvermögen, visuelle Wahrnehmung
(12) Gesichter & Orte	auf Bildern gezeigte Märchenfiguren, bekannte Persönlichkeiten oder Bauwerke müssen benannt werden	Umfang des allgemeinen Faktenwissens, LZG, ganzheitliches Verarbeiten, sprachliches Ausdrucksvermögen
(13) Rechnen	anhand einer Bildergeschichte werden Rechenaufgaben vorgelesen, die im Kopf ausgerechnet werden sollen	Grundlegende mathematische Konzepte und Rechenfertigkeiten, LZG, einzelheitliches und ganzheitliches Verarbeiten, Umgang mit Zahlen, logisches Denken
(14) Rätsel	vorgelesene Informationen zu einem Gegenstand müssen behalten und daraus der Name eines Gegenstandes abgeleitet werden	Integration sequentiell dargebotener akustischer Reize, schlussfolgerndes Denken, logisches Einordnen, LZG, Konzeptbildung, Detailbeachtung, sprachliches Ausdrucksvermögen
(15) Lesen/ Buchstabieren	durch lautes Vorlesen von Buchstaben und Wörtern soll die Lesefertigkeit überprüft werden	Sprachentwicklung, Lesefertigkeit, LZG, sprachliches Ausdrucksvermögen
(16) Lesen/Verstehen	Handlungsanweisungen müssen laut vorgelesen und dann entsprechen ausgeführt werden	Bildung von Sprachkonzepten, Lesefertigkeit, LZG, visuell-motorische bzw. gestische Kommunikation

Beachte: KZG: Kurzzeitgedächtnis, LZG: Langzeitgedächtnis

Bei der Auswertung der K-ABC erhält man einen Gesamtwert für die intellektuelle Begabung eines Kindes sowie einen Gesamtwert für den Bereich der Fertigkeiten. Für die Bereiche einzelheitliches und ganzheitliches Denken lässt sich ebenfalls ein Standardwert bestimmen, und es besteht die Möglichkeit, aus den fünf sprachfreien Untertests (2, 6, 8, 9, 10) einen

nonverbalen Wert zu ermitteln. Für die einzelnen Untertests finden sich im Handbuch Informationen zu den von ihnen erfassten Fähigkeiten, da die Standardisierung jedoch für die Durchführung des Gesamttests vorgenommen wurde, sollten Aussagen über Fähigkeiten anhand einzelner Untertests mit Vorsicht vorgenommen werden.

Testkennwerte: Die deutsche Fassung der K-ABC wurde an einer Stichprobe von $N = 3098$ Kindern aus Deutschland, Österreich, Schweiz und Südtirol für elf Altersgruppen im Bereich von 2;6 bis 12;5 Jahren normiert (vgl. Melchers & Preuss, 2003 a). Die faktorenanalytische Überprüfung der Konstruktvalidität bestätigte Konzept und Aufbau der K-ABC (Melchers & Preuß, 2003 b).

Die mittlere Reliabilität der Skala intellektuelle Fähigkeiten liegt für 4;0 bis 4;11 Jahre alte Kinder bei $r = .90$, die der einzelheitlichen Skala bei $r = .88$ und die der ganzheitlichen Skala bei $r = .88$. Die Reliabilität der sprachfreien Skala liegt für diese Altersgruppe mit $r = .86$ vergleichsweise hoch. Die Split-half Reliabilitäten der einzelnen Untertests befinden sich zwischen $r = .71$ (Wiedererkennen von Gesichtern) und $r = .93$ (Dreiecke). Die Fertigkeitenskala weist eine Reliabilität von $r = .93$ auf, die Untertests eine zwischen $r = .79$ (Gesichter und Orte) und $r = .90$ (Rechnen).

Es liegen verschiedene Validitätshinweise in Form von Korrelationen zwischen Subskalen der K-ABC mit anderen Intelligenztests vor: Bei Vergleichen der K-ABC mit dem Hamburg-Wechsler-Intelligenztest für Kinder (HAWIK-R) wurde deutlich, dass die Skala intellektuelle Fähigkeiten vor allem mit dem Handlungsteil ($r = .69$) und die Fertigkeitenskala hoch mit dem Verbalteil ($r = .81$) korreliert. Für die Skala einzelheitliches Denken der K-ABC sind Korrelationen mit den Gesamtleistungen und Unterskalen des Allgemeinen Intelligenzdiagnostikums (AID) ermittelt worden, die für die Altersgruppen 8;0 bis 10;0 zwischen $r = .50$ und $r = .80$ liegen. Die sprachfreie Skala korreliert in der gleichen Altersgruppe zwischen $r = .61$ und $r = .71$ mit dem Leistungsprüfsystem (LPS). Ferner weist die Skala intellektuelle Fähigkeiten Zusammenhänge in mittlerer Höhe mit den Coloured Progressive Matrices (CPM) auf ($r = .52$ bis $r = .62$), während die entsprechenden Korrelationen zwischen der Skala Fertigkeiten und der CPM deutlich niedriger liegen ($r = .18$ bis $r = .33$). Die sprachfreie Skala korreliert ebenfalls in mittlerer Höhe ($r = .58$) mit der CPM.

Einsatzbereich: Die K-ABC dient der differenzierten und umfassenden psychologisch-diagnostischen Untersuchung im Rahmen von Einzeltestungen. Weiterhin findet sie Anwendung bei der Beurteilung von Kindern mit Lern- und Leistungsbeeinträchtigungen sowie

anderen Auffälligkeiten mit dem Ziel der Erstellung, Einleitung und Durchführung von Fördermaßnahmen. Die K-ABC kann auch als Prüfinstrument für das Vorschulalter eingesetzt werden. Aufgrund ihrer neuropsychologischen Aspekte, wie z. B. der Unterscheidung zwischen den Verarbeitungsstilen einzelheitliches und ganzheitliches Denken, kann die K-ABC als Teil neuropsychologischer Diagnostik zum Einsatz kommen. Die Durchführung der K-ABC erfordert einen kompetenten, geübten Testleiter, der die Methoden der psychologischen Intelligenztestung kennt und sich mit dem Testverfahren ausführlich vertraut gemacht hat. Der Testleiter sollte über eine Ausbildung und Erfahrung in diagnostischer Verhaltensbeobachtung und individueller Intelligenzprüfung bei Kindern verfügen.

Kaufman-Test zur Intelligenzmessung für Jugendliche und Erwachsene (K-TIM)

Testkonzept: Der Kaufman-Test zur Intelligenzmessung für Jugendliche und Erwachsene (K-TIM; Melchers, Schürmann & Scholten, 2006) ist die deutsche Fassung des Kaufman Adolescent and Adult Intelligence Test (KAIT, Kaufman & Kaufman, 1993). Es handelt sich um einen Individualtest zur Erfassung der allgemeinen Intelligenz bei Jugendlichen und Erwachsenen in einem Altersbereich von 11 bis über 85 Jahren (Dauer ca. 90 Minuten). Das Verfahren beruht auf dem Intelligenzmodell von Cattell, das fluide und kristalline Intelligenz unterscheidet. Es besteht aus zwei Skalen, der *Skala fluider Intelligenz* und der *Skala kristalliner Intelligenz*, aus denen eine *Skala Gesamtintelligenz* gebildet wird. Die *Skala fluide Intelligenz* umfasst nach Angaben der Autoren die Fähigkeit, neue Probleme zu lösen, wobei Eigenschaften wie hypothetisch-deduktives Denken, abstraktes Denken, nonverbale Konzeptbildung, Organisation der Wahrnehmung und verbales Verständnis erfasst werden. Weiterhin gehen in die fluide Intelligenz auch Gedächtnisanforderungen und Arbeitsgeschwindigkeit mit ein. Die *Skala kristalliner Intelligenz* erfasst allgemeines Wissen, verbales Verständnis und Ausdruck sowie die Fähigkeit, eigene Erfahrungen einzubringen. Der Inhalt der Aufgaben hat nach Angabe der Autoren eine enge Beziehung zu Schulbildung und Akkulturation mit Blick auf den Umgang mit grundlegenden Charakteristika der Problemstellungen und den zur Bearbeitung erforderlichen Denkopoperationen. Die *Skala Gesamtintelligenz*, mit der sich ein allgemeiner Intelligenzfaktor bilden lässt, wird weniger als Umsetzung des theoretischen Konstrukts „g-Faktor“ gesehen denn als kumulativer Wert, der eine Interpretationsgrundlage liefert. Ein zusätzlicher Bereich, der im K-TIM berücksichtigt wird, sind neuropsychologische Aspekte. Dabei lassen sich Überschneidungen zwischen Untertests des K-TIM und Anforderungen aus neuropsychologischen Testbatterien finden, wie z. B. Planungsfähigkeit oder komplexe Kon-

zeptbildung. Gezielt erfasst werden im K-TIM Gedächtnisfunktionen. Hierbei handelt es sich um die Untersuchung der verzögerten Erinnerung über den Abruf von Informationen nach einem längeren Zeitintervall. Es lässt sich ein Vergleich der unmittelbaren und der verzögerten Reproduktion durchführen.

Skalen und Untertests: Den Untertests des K-TIM sind zwei Skalen zugeordnet: vier Untertests bilden die Skala fluider Intelligenz und vier Untertests die Skala kristalliner Intelligenz. Zusätzlich finden sich zwei Untertests zum Messen des verzögerten Erinnerns (vgl. Tabelle 4 für die Zuordnung der Untertests zu Skalen und Subskalen). Insgesamt liegen somit zehn Untertests vor: (1) *Worträtsel*, (2) *Symbole lernen*, (3) *Logische Denkschritte*, (4) *Auditives Verständnis*, (5) *Zeichen entschlüsseln*, (6) *Doppelte Bedeutungen*, (7) *Symbole – Abruf nach Intervall*, (8) *Auditives Verständnis – Abruf nach Intervall*, (9) *Figurales Gedächtnis*, (10) *Persönlichkeiten*.

Bei der Auswertung des K-TIM lassen sich Standardwerte für die Gesamtintelligenz, für die Skala kristalliner und die Skala fluider Intelligenz bestimmen. Zusätzlich zu den drei Standardwerten für die Intelligenzskalen stehen auch Standardwerte für alle zehn Untertests zur Verfügung. Es besteht die Möglichkeit einer Profilinterpretation, die jedoch auf das individuelle Leistungsprofil der Testperson bezogen werden sollte und nicht auf die durchschnittlichen Leistungen in der Normierungsstichprobe. Da die Durchführung aller Untertests in vorgegebener Reihenfolge Bestandteil der Standardisierungs- und Normierungsbedingungen ist, sollten Änderungen nur im Rahmen der im Durchführungshandbuch beschriebenen Möglichkeiten vorgenommen und die Ergebnisse entsprechend vorsichtig interpretiert werden.

Tabelle 4
K-TIM

Skalen und Untertests	Aufgaben	Inhalte
Skala kristalliner Intelligenz		
(1) Worträtsel	ein Wort anhand von dargebotenen Buchstaben und einem Hinweissatz erkennen	Visuelles Schließen, Erschließen semantischer Teil-Ganzes-Beziehungen, verbale Konzeptbildung, Wortschatz
(4) Auditives Verständnis	zu vorgelesenen Nachrichten inhaltliche oder Verständnisfragen beantworten	Auditive Sequenzbildung, auditives Textverständnis, Kurzzeitgedächtnis, sprachliches Gedächtnis
(6) Doppelte Bedeutungen	zu zwei Wortpaaren, die akustisch und visuell vorgegeben werden, ein Wort mit zwei Bedeutungen suchen, das zu beiden Wortpaaren passt.	Semantische Flexibilität, Wortschatz, schlussfolgerndes Denken, erworbenes Wissen, sprachliches Verständnis und Ausdruck
(10) Persönlichkeiten	anhand von Photos und einem Hinweissatz bekannte Persönlichkeiten der Gegenwart und Vergangenheit erkennen	Ausmaß des allgemeinen Faktenwissens, verbales Gedächtnis, sprachliches Verständnis und Ausdruck
Skala fluider Intelligenz		
(2) Symbole lernen	zu einem bestimmten Symbol ein assoziiertes Wort oder Konzept lernen und dann aus diesen Symbolen gebildete Sätze lesen	Paar-Assoziationslernen, Visuelle Sequenzbildung, Kurzzeitgedächtnis, Lernfähigkeit, Speicherung und Abruf von Wörtern
(3) Logische Denkschritte	akustisch und visuell dargebotene logische Regeln zur Beantwortung von Fragen anwenden	Syllogistisches Schlussfolgern, Gewandtheit im Umgang mit Zahlen, schlussfolgerndes Denken, Planen, Verarbeitungsgeschwindigkeit
(5) Zeichen entschlüsseln	anhand eines Satzes bildhafter Reize, die mit einem bestimmten Kode versehen sind, für einen neuen bildhaften Reiz den passende Kode entschlüsseln	Planungsgeschwindigkeit, abstraktes Denken, visuelle Wahrnehmung und Verarbeitung von Reizen, Organisation der Wahrnehmung, Planen
(9) Figurales Gedächtnis	ein vorgegebenes abstraktes Muster, das für kurze Zeit gezeigt wird, aus dem Gedächtnis mit gelben und schwarzen Klötzchen nachlegen	Nonverbale Konzeptbildung, Kurzzeitgedächtnis, Organisation der Wahrnehmung, räumliches Vorstellungsvermögen, visuell-motorische Koordination
Verzögertes Erinnern		
(7) Symbole – Abruf nach Intervall	nach einem längeren Zeitintervall nochmals Sätze aus den vorher gelernten Symbolen lesen	Erinnern neuer Informationen, Vergleich von unmittelbarer und verzögerter Reproduktion
(8) Auditives Verständnis – Abruf nach Intervall	nach einem längeren Zeitintervall inhaltliche und Verständnisfragen zu den vorher vorgelesenen Nachrichten beantworten	Erinnern von auf früheren Kenntnissen beruhenden Informationen, Vergleich von unmittelbarer und verzögerter Reproduktion

Testkennwerte: Die Normen des K-TIM wurden an N = 2009 deutschen, Schweizer, österreichischen und Südtiroler Personen im Alter von 11 bis 92 Jahren für 15 Altersklassen erhoben. Die Reliabilitätsschätzungen (Split half) für die Skalen kristalliner, fluider und Gesamtintelligenz sind mit Werten von .96, .98 und .97 als sehr gut einzuschätzen. Die Untertests weisen mittlere Split-half-Reliabilitätskoeffizienten zwischen .71 und .97 auf. Bei der Überprüfung der Konstruktvalidität ergab eine explorative Faktorenanalyse für alle Altersgruppen eine Zwei-Faktoren-Lösung, die die Zuordnung der Untertests zur Skala fluider oder kristalliner Intelligenz bestätigte. Die konvergente und diskriminante Validität konnte durch Zusammenhänge mit dem HAWIE-R, dem IST-70 und der K-ABC gezeigt werden.

Einsatzbereich: Der K-TIM ist ein Instrument zur umfassenden Messung von Intelligenz und zur Erfassung kognitiver Stärken und Schwächen. Er kann im Rahmen einer differenzierten klinischen Diagnostik, bei schulpsychologischen Untersuchungen und für berufsbezogene Fragestellungen zur Anwendung kommen. Der K-TIM lässt sich auch bei neuropsychologischen Untersuchungen einsetzen, zur Abklärung kognitiver und neuropsychologischer Störungen. Die Durchführung des K-TIM erfordert einen kompetenten, geübten Testleiter, der die Methoden der psychologischen Intelligenztestung kennt und sich mit dem Testverfahren ausführlich vertraut gemacht hat. Der Testleiter sollte über eine Ausbildung und Erfahrung in diagnostischer Verhaltensbeobachtung und individueller Intelligenzprüfung bei Kindern verfügen. Es wird weiterhin darauf hingewiesen, dass die Interpretation der Testergebnisse von dem Testleiter durchgeführt werden sollte, der die Testung vorgenommen hat, so dass eine alle Aspekte berücksichtigende Beurteilung der Leistung vorgenommen werden kann.

Hamburg-Wechsler-Intelligenztest für Kinder (HAWIK-III)

Testkonzept: Der Hamburg-Wechsler-Intelligenztest für Kinder (HAWIK-III; Tewes, Rossmann & Schallberger, 2002) ist die deutschsprachige Fassung der Wechsler Intelligence Scale for Children (WISC-III; Wechsler, 1991). Es handelt sich um einen Individualtest zur Untersuchung der kognitiven Entwicklung von Kindern und Jugendlichen im Alter von 6 bis 16 Jahren (Gesamtdauer zwischen 50 und 70 Minuten). Das Verfahren beruht auf dem Intelligenzkonzept von Wechsler, der Intelligenz nicht als spezifische Fähigkeit auffasst sondern als globale Einheit, die sich aus Handlungs-, Denk- und Planungskompetenzen zusammensetzt und somit die Gesamtheit einer Gruppe von Einzelaspekten bildet. Dafür stellte Wechsler eine Reihe von Aufgaben zusammen, um unterschiedliche geistige Fähigkeiten zu erfassen, die in

ihrer Gesamtheit die allgemeine Begabung eines Kindes abbilden sollen. Die Skalen sollen daher möglichst unterschiedliche kognitive Fähigkeiten erfassen. Er berücksichtigte u.A. Aspekte wie abstraktes Denken, Gedächtnisleistungen und Wahrnehmungsfähigkeiten. Das Verfahren besteht somit aus verschiedenen Untertests, die unterschiedliche Aspekte der Intelligenz erfassen. Aus den einzelnen Aufgaben lässt sich zum einen ein Gesamtmaß der geistigen Fähigkeiten berechnen, zum anderen kann getrennt ein Handlungs- und ein Verbal-Maß gebildet werden. Während bei Aufgaben des Handlungsmaßes Wahrnehmungsfähigkeiten im Vordergrund stehen, erfordern die Aufgaben des Verbalmaßes sprachlichen Ausdruck und sprachliches Verständnis. Zwei weitere Fähigkeiten, die erfasst werden können, sind Unablenkbarkeit und Arbeitsgeschwindigkeit.

Skalen und Untertests: Der HAWIK-III besteht aus 13 Untertests, von denen sich jeweils fünf dem Verbal- bzw. Handlungsteil zuordnen lassen. Drei Untertests können optional durchgeführt werden und dienen der Ermittlung der Indexwerte Unablenkbarkeit und Arbeitsgeschwindigkeit (vgl. Tabelle 5 für die Zuordnung der Untertests zu Skalen und Indexwerten): (1) *Bilderergänzen*, (2) *Allgemeines Wissen*, (3) *Zahlen-Symbol-Test*, (4) *Gemeinsamkeiten finden*, (5) *Bilderordnen*, (6) *Rechnerisches Denken*, (7) *Mosaik-Test*, (8) *Wortschatz-Test*, (9) *Figurenlegen*, (10) *Allgemeines Verständnis*, (11) *Symbolsuche*, (12) *Zahlennachsprechen*, (13) *Labyrinth-Test*.

Bei der Auswertung des HAWIK-III lassen sich zusätzlich zum Gesamt-IQ, Verbal-IQ und Handlungs-IQ vier Index-Werte (sprachliches Verständnis, Wahrnehmungsorganisation, Unablenkbarkeit und Arbeitsgeschwindigkeit) bestimmen. Für diese sieben Skalen lassen sich IQ-Werte mit Mittelwert 100 und Standardabweichung 15 ermitteln. Die Ergebnisse der einzelnen Untertests können ebenfalls in standardisierter Form (Wertpunkte) angegeben werden, eine Interpretation sollte sich jedoch darauf beschränken, die Stärken und Schwächen der Testperson vor dem individuellen Leistungsniveau zu analysieren. Da die einzelnen Untertests zwar unterschiedliche Aspekte der Intelligenz erfassen, jedoch nicht jeder Untertest nur einen spezifischen Aspekt widerspiegelt, wird es nicht empfohlen, einzelne Teilleistungen anhand einzelner Untertests zu beschreiben.

Tabelle 5
HAWIK-III

Skalen und Untertests		
	Aufgabe	Inhalte
Verbalteil		
Sprachliches Verständnis		
Allgemeines Wissen	vorgelesene Wissensfragen zu Ereignissen, Orten oder Persönlichkeiten beantworten	Breite des erworbenen Wissens, Langzeitgedächtnis für Faktenwissen
Gemeinsamkeiten finden	zu zwei Begriffen von Alltagsgegenständen oder Konzepten einen Oberbegriff finden	verbales Schlussfolgern, sprachliche Konzeptbildung, Denken in abstrakten logischen Kategorien, Analogiebildung
Wortschatz-Test	mündlich vorgegebene Wörter definieren	Wortkenntnis, Definition von Begriffen, sprachliche Entwicklung
Allgemeines Verständnis	anhand von Fragen Lösungsmöglichkeiten zu Alltagsproblemen schildern oder das Verständnis sozialer Regeln verdeutlichen	praktisches Urteilsvermögen, Kenntnis konventioneller sozialer Regeln und ihrer Bedeutungen
Unablenkbarkeit		
Rechnerisches Denken	vorgelesene Textaufgaben im Kopf lösen und mündlich beantworten	akustische Merkfähigkeit, Arbeitsgedächtnis, Rechenfähigkeiten, Konzentrationsvermögen
Zahlennachsprechen	eine vorgeschene Zahlenreihe vorwärts und rückwärts wiedergeben	akustische Merkfähigkeit, Arbeitsgedächtnis, Aufmerksamkeit, Konzentrationsvermögen
Handlungsteil		
Wahrnehmungsorganisation		
Bilderergänzen	auf Bildern von Alltagsgegenständen oder Situationen ein fehlendes Detail erkennen	Beobachtungsgenauigkeit, Unterscheidung wesentliche/ unwesentliche Details
Bilderordnen	Bildergeschichten, die einen Handlungsablauf wiedergeben und in der falschen Reihenfolge vorgelegt werden, in die richtige Reihenfolge bringen	schlussfolgerndes Denken, Aufmerksamkeit für Details und Hintergrundinformationen, Erkennen von logischen Ereignisfolgen und Ursache-Wirkungs-Zshg.
Mosaik-Test	abstrakte Muster anhand einer Vorlage mit roten und weißen Klötzchen nachlegen	Unterscheidung Teil-Ganzes, räumliche Wahrnehmung, visuomotorische Koordination, Erkennen abstrakter visueller Muster
Figurenlegen	aus Teilen eines Puzzles eine sinnvolle Figur zusammenlegen	Erkennen Beziehung Teil-Ganzes, Wiedererkennen von vertrauten Objekten
Arbeitsgeschwindigkeit		
Zahlen-Symbol-Test	anhand einer Vorlage zusammengehöriger Zahlen und Symbole muss den Zahlen das passende Symbol zugeordnet werden	visuelles Kurzzeitgedächtnis, visuomotorische Koordination, Konzentration, Arbeitsgeschwindigkeit

Fortsetzung Tabelle nächste Seite

Fortsetzung Tabelle 5

Symbolsuche	in Reihen von abstrakten Symbolen schauen, ob zwei Zielitems in der Reihe vorhanden sind oder nicht	kognitive Verarbeitungsgeschwindigkeit, Beobachtungsgenauigkeit, Konzentration
Labyrinth-Test (nicht zugeordnet)	mit einem Bleistift eine Linie vom Zentrum eines Labyrinthes bis zum Ausgang ziehen	planerisches Denken, Wahrnehmungsorganisation, visuomotorische Koordination

Testkennwerte: Die Normierung des HAWIK-III erfolgte in den Jahren 1995 bis 1998 in Deutschland, Österreich und der Schweiz an insgesamt 1570 Kindern und Jugendlichen zwischen 6 und 16 Jahren. Die Split-half-Reliabilitäten liegen für den Gesamt-IQ, den Verbal-IQ und den Handlungs-IQ mit .96, .95 und .91 im guten bis sehr guten Bereich, auch für die Index-Werte fielen sie mit .85 bis .94 gut aus. Für die Untertests fanden sich Reliabilitätskoeffizienten zwischen .68 (Figurenlegen) und .88 (Wortschatz-Test, Zahlennachsprechen und Mosaik-Test).

Die Überprüfung der Konstruktvalidität mit einer varimax-rotierten Faktorenanalyse bestätigte die Zweifaktorenlösung Verbal- und Handlungsteil. Die faktorielle Validität der vier Index-Werte fiel für sprachliches Verständnis, Wahrnehmungsorganisation und Arbeitsgeschwindigkeit zufrieden stellend aus, das Konstrukt der Unablenkbarkeit ließ sich hingegen nicht so gut verifizieren. Zur Kriteriumsvalidität wurden Korrelationen mit Schulnoten und Lehrerurteil berechnet, die mit Werten zwischen .47 und .33 (Schulnoten) und .52 und .40 (Lehrerurteil) in einer Größenordnung liegen, die bei Intelligenztests allgemein beobachtet wird.

Einsatzbereich: Der HAWIK-III kann in der differenzierten Diagnostik der kognitiven Fähigkeiten von Kindern und der Beschreibung von Leistungsprofilen mit speziellen Stärken und Schwächen eingesetzt werden. Im Rahmen testpsychologischer Untersuchungen lassen sich Ursachen von Lern- und Leistungsstörungen im schulischen Bereich feststellen. Weiterhin finden die Wechsler-Tests international häufig im Bereich der Hochbegabtdiagnostik Anwendung. Nach Angaben der Autoren kommt der HAWIK-III häufig auch zur neuropsychologischen Diagnostik zum Einsatz, wobei darauf hingewiesen werden soll, dass mit diesem Verfahren nur ergänzende Informationen gewonnen werden können. Da es sich um einen aufwendigen und komplexen Individualtest handelt, sollte der HAWIK-III nur von entsprechend geschulten Fachkräften mit psychodiagnostischer Ausbildung und Erfahrung durchgeführt werden. Dabei ist Erfahrung in der Anleitung von Kindern bei der Durchführung von

zentraler Bedeutung sowie die perfekte Beherrschung von Material und Anweisungen sowie der Regeln zur Bewertung von Lösungen. Die Interpretation der Ergebnisse setzt ebenfalls entsprechende Sachkenntnisse voraus.

Wechsler Intelligenztest für Erwachsene (WIE)

Testkonzept: Der Wechsler Intelligenztest für Erwachsene (WIE; von Aster, Neubauer & Horn, 2006) ist die deutschsprachige Fassung der Wechsler Adult Intelligence Scale (WAIS; Wechsler, 1997). Es handelt sich um einen Individualtest zur Untersuchung der kognitiven Fähigkeiten von Jugendlichen und Erwachsenen von 15 bis 89 Jahren. Das Verfahren beruht auf dem Intelligenzkonzept von Wechsler, der Intelligenz als die Fähigkeit definiert, planvoll zu handeln, vernünftig zu denken und sich mit seiner Umwelt konstruktiv auseinanderzusetzen. Er beschrieb Intelligenz als ein allgemeines Merkmal, das sich in unterschiedlichen Formen manifestieren kann. Er entwickelte verschiedene Aufgaben, die sich auf unterschiedliche kognitive Aspekte beziehen. Dabei berücksichtigte er Fähigkeiten wie abstraktes Urteilsvermögen, Wahrnehmungsfähigkeit, sprachliche Kompetenzen oder Arbeitsgeschwindigkeit. Die einzelnen Untertests erfassen somit Teilaspekte eines übergeordneten, allgemeinen Konstrukts. Das Verfahren ermöglicht somit die Bestimmung eines Gesamtmaßes der Intelligenz und gliedert sich in zwei weitere Schwerpunkte, einen Verbal- und einen Handlungsteil, bei dem sprachliches Verständnis bzw. Wahrnehmungsorganisation im Vordergrund stehen. Weiterhin lassen sich die Fähigkeiten Arbeitsgedächtnis und Arbeitsgeschwindigkeit bestimmen.

Skalen und Untertests: Der WIE besteht insgesamt aus 14 Untertests, wobei sechs Untertests den Verbalteil bilden und fünf Untertests den Handlungsteil. Drei Untertests können optional durchgeführt werden und sind für die Bildung der Indexwerte Arbeitsgedächtnis und Arbeitsgeschwindigkeit erforderlich (vgl. Tabelle 6 für die Zuordnung der Untertests zu Skalen und Indexwerten): (1) *Bilderergänzen*, (2) *Wortschatztest*, (3) *Zahlen-Symbol-Test*, (4) *Gemeinsamkeiten finden*, (5) *Mosaik-Test*, (6) *Rechnerisches Denken*, (7) *Matrizen-Test*, (8) *Zahlen-nachsprechen*, (9) *Allgemeines Wissen*, (10) *Bilderordnen*, (11) *Allgemeines Verständnis*, (12) *Symbolsuche*, (13) *Buchstaben-Zahlen-Folgen*, (14) *Figurenlegen*.

Für die Auswertung lassen sich neben dem Gesamt-IQ, dem Verbal-IQ und dem Handlungs-IQ, Index-Werte für sprachliches Verständnis, Wahrnehmungsorganisation, Arbeitsgedächtnis und Arbeitsgeschwindigkeit bestimmen. Dabei handelt es sich um Standardwerte mit einem

Mittelwert von 100 und einer Standardabweichung von 15, die sich im Vergleich zur Altersgruppe interpretieren lassen. Die Interpretation der Untertestwerte stellt ein Problem dar, weil man den einzelnen Untertests keine nur für sie spezifischen inhaltlichen Bedeutungen zuweisen kann. Daher sollten sie nur für eine Profilinterpretation im Hinblick auf Stärken und Schwächen einer Testperson vor dem Hintergrund ihres allgemeinen Leistungsniveaus eingesetzt werden.

Tabelle 6
WIE

WIE		
Skalen und Untertests	Aufgabe	Inhalte
Verbalteil		
Sprachliches Verständnis		
Wortschatz-Test	der Testperson werden Wörter vorgegeben, die diese erklären muss	Wortkenntnis, Fähigkeit zur Definition von Begriffen, sprachliche Entwicklung
Gemeinsamkeiten finden	bei einem Wortpaar muss erkannt werden, welche Gemeinsamkeit sie aufweisen und ein Oberbegriff dafür angegeben werden	verbales Schlussfolgern, sprachliche Konzeptbildung, Denken in abstrakten logischen Kategorien, Analogiebildung
Allgemeines Wissen	anhand von Wissensfragen wird die Kenntnis bestimmter Sachverhalte, Orte oder Persönlichkeiten überprüft	Breite des erworbenen Wissens, Langzeitgedächtnis für Faktenwissen
Allgemeines Verständnis (nicht zugeordnet)	Beantwortung von Fragen zu Alltagsproblemen und sozialen Regeln	praktisches Urteilsvermögen, Kenntnis konventioneller sozialer Regeln und ihrer Bedeutungen
Arbeitsgedächtnis		
Rechnerisches Denken	Rechenaufgaben werden vorgelesen, die im Kopf gelöst und mündlich beantwortet werden müssen	akustische Merkfähigkeit, Arbeitsgedächtnis, Konzentrationsvermögen
Zahlennachsprechen	unterschiedlich lange Reihen von vorgesprochenen Zahlen müssen vorwärts oder rückwärts wiederholt werden	akustische Merkfähigkeit, Arbeitsgedächtnis, Konzentrationsvermögen und Aufmerksamkeit
Buchstaben-Zahlen-Folgen	Folgen von Buchstaben und Zahlen werden vorgelesen und müssen aufsteigend (Zahlen) und absteigend (Buchstaben) sortiert werden	akustische Merkfähigkeit, Arbeitsgedächtnis, Konzentrationsvermögen und Aufmerksamkeit

Fortsetzung Tabelle nächste Seite

Fortsetzung Tabelle 6

Handlungsteil		
Wahrnehmungsorganisation		
Bilderergänzen	auf einem Bild von einem Gegenstand soll ein fehlendes Detail erkannt werden	Beobachtungsgenauigkeit, Unterscheidung wesentliche/ unwesentliche Details
Mosaik-Test	anhand von Vorlagen sollen geometrische Muster mit roten und weißen Klötzchen nachgelegt werden	Unterscheidung Teil-Ganzes, räumliche Wahrnehmung, visuomotorische Koordination, Erkennen abstrakter visueller Muster
Matrizen-Test	unvollständige Reihen abstrakter Muster sollen durch Erkennen des Konstruktionsprinzips vervollständigt werden	visuelle Informationsverarbeitung, abstraktes Denken, induktives Denken im visuellen Bereich, Erkennen visueller Analogien
Bilderordnen (nicht zugeordnet)	Bilderserien, die einen Handlungsablauf wiedergeben, sollen in die richtige Reihenfolge gebracht werden	schlussfolgerndes Denken, Aufmerksamkeit für Details und Hintergrundinformationen, Erkennen von logischen Ereignisfolgen und Ursache-Wirkungszusammenhängen
Arbeitsgeschwindigkeit		
Zahlen-Symbol-Test	Zahlen und Symbole sind zugeordnet und in Zahlenreihen sollen die fehlenden Symbole ergänzt werden	visuelles Kurzzeitgedächtnis, visuomotorische Koordination, Konzentration, Arbeitsgeschwindigkeit
Symbolsuche	in Reihen von abstrakten Symbolen soll überprüft werden, ob zwei Zielsymbole in der Reihe vorhanden sind	kognitive Verarbeitungsgeschwindigkeit, Beobachtungsgenauigkeit, Konzentration
Figurenlegen (nicht zugeordnet)	aus Teilen eines Puzzles soll eine sinnvolle Figur zusammengesetzt werden	Erkennen von Beziehungen Teil/Ganzes, Wiedererkennen vertrauter Objekte

Testkennwerte: Die Normierung des WIE erfolgte in den Jahren 1999 bis 2005 in Deutschland, Österreich und der Schweiz an insgesamt 1897 Personen der Altersgruppe 16 bis 89 Jahre. Die Reliabilitätsbestimmung nach der Split-half-Methode ergab mit Werten von .97, .96 und .94 für die Gesamtskala, den Verbal- und den Handlungsteil gute bis sehr gute Ergebnisse. Auf Untertestebene lagen die Reliabilitätskoeffizienten zwischen .92 und .70.

Die Überprüfung der faktoriellen Validität anhand einer konfirmatorischen Faktorenanalyse ergab eine Vier-Faktoren-Lösung: verbales Verständnis, Wahrnehmungsorganisation, Arbeitsgedächtnis und Arbeitsgeschwindigkeit. Weitere Berechnungen mit Hauptkomponentenanalysen bestätigten die Zusammensetzung der einzelnen Index-Werte.

Einsatzbereich: Der WIE lässt sich für eine differenzierte Messung der allgemeinen geistigen Fähigkeiten einsetzen. Er ermöglicht weiterhin die Untersuchung von Lern- und Leistungsstörungen sowie Hochbegabtdiagnostik. Der WIE lässt sich auch in die neuropsychologische

Diagnostik integrieren, da die Bestimmung des Intelligenzniveaus bei der Untersuchung von Gedächtnis-, Sprach- und anderen kognitiven Defiziten hilfreich sein kann. Da die Testdurchführung, Ergebnisauswertung und diagnostische Beurteilung ein komplexer Prozess ist, sollte der Testleiter über eine entsprechende Ausbildung in der Durchführung und Auswertung verfügen. Die genaue Kenntnis des Tests und Erfahrung im Umgang mit Testpersonen ist dabei von zentraler Bedeutung.

Adaptives Intelligenzdiagnostikum 2 (AID 2)

Testkonzept: Das Adaptive Intelligenzdiagnostikum – AID 2 (Kubinger & Wurst, 2000) ist ein deutschsprachiger Individualtest zur Erfassung der Intelligenz von Kindern und Jugendlichen im Alter von 6;0 bis 15;11 Jahren (Dauer ca. 75 Minuten). Das Verfahren beruht auf dem Testkonzept von David Wechsler und versteht Intelligenz als eine Zusammensetzung aus vielen komplexen und basalen Informationsverarbeitungsfähigkeiten. In Anlehnung an Cattells (1987) Theorie der fluiden und kristallinen Intelligenz und dessen Investmenttheorie („Wissen ist investierte Intelligenz“), wird Intelligenz im AID 2 definiert als „das Bündel aller kognitiven Voraussetzungen, die notwendig sind, um Wissen zu erwerben und Handlungskompetenzen zu entwickeln“ (Manual, S. 30). Während die Untertests sich thematisch am HAWIK anlehnen, realisiert der AID 2 im Gegensatz zu den Wechsler-Tests das Prinzip des *adaptiven Testens* nach der probabilistischen Testtheorie (Rasch-Modell). Hierbei werden der Versuchsperson nur diejenigen Aufgaben gestellt, die ihrem Leistungsniveau entsprechen, wodurch mit einer minimalen Anzahl von Aufgaben pro Untertest eine hohe Messgenauigkeit erreicht wird.

Skalen und Untertests: Der Test besteht aus 11 Untertests (mit 13 Kennwerten), die sich in die Bereiche „*verbal-akustische*“ Fähigkeiten und „*manuell-visuelle*“ Fähigkeiten gliedern: (1) *Alltagswissen*, (2) *Realitätssicherheit*, (3) *Angewandtes Rechnen*, (4) *Soziale und Sachliche Folgerichtigkeit*, (5) *Unmittelbares Reproduzieren – numerisch*, (6) *Synonyme finden*, (7) *Kodieren und Assoziieren*, (8) *Antizipieren und Kombinieren – figural*, (9) *Funktionen abstrahieren*, (10) *Analysieren und Synthetisieren – abstrakt*, (11) *Soziales Erfassen und Sachliches Reflektieren*. Neben Leistungswerten für die einzelnen Untertests gibt der Test verschiedene Maße zur Erfassung der Gesamtleistung an. Eine Beschreibung der jeweiligen Aufgaben und erfassten Inhalte sowie der Skalenzuordnung findet sich in Tabelle 7.

Tabelle 7
AID 2

Skalen und Untertests	Aufgabe	Inhalte
Verbal-akustische Fähigkeiten		
(1) Alltagswissen	mündliche Fragen	Fähigkeit, sich Sachkenntnisse über alltägliche Inhalte der heutigen Gesellschaft anzueignen
(3) Angewandtes Rechnen	mathematische Textaufgaben	Fähigkeit, weitgehend unabhängig von schulischen Rechenfertigkeiten, alltägliche Aufgabenstellungen durch Anwendung der passenden Rechenoperation zu lösen
(5) Unmittelbares Reproduzieren - numerisch (zwei Testkennwerte)	Zahlen nachsprechen; vorwärts und rückwärts	Kapazität der seriellen Informationsverarbeitung
(6) Synonyme finden	für ein gegebenes Wort jeweils ein anderes mit gleicher Bedeutung finden	Elementares Sprachverständnis und Wortschatz
(9) Funktionen abstrahieren	Gemeinsamkeiten zweier Dinge beschreiben	Fähigkeit, durch Abstraktion zu einer Begriffsbildung zu gelangen
(11) Soziales Erfassen und Sachliches Reflektieren	Fragen zum allgemeinen gesellschaftlichen Verständnis	Verständnis von Sachzusammenhängen in der Gesellschaft und sozial angemessenen Verhaltenweisen
Manuell-visuelle Fähigkeiten		
(2) Realitätssicherheit	fehlende Details auf Bildkarten entdecken	Verständnis der Wirklichkeit um Dinge des Alltags
(4) Soziale und Sachliche Folgerichtigkeit	Bildergeschichten, die in die richtige Reihenfolgen gebracht werden müssen	Fähigkeit, soziale und sachliche Abfolgen zu verstehen
(7) Kodieren und Assoziieren (zwei Testkennwerte)	Bildern auf einem Testblatt möglichst schnell jeweils entsprechende Symbole zuordnen und von einer Vorlage abschreiben, später aus dem Gedächtnis heraus	Informationsverarbeitungsgeschwindigkeit sowie die Fähigkeit zum inzidentellen Lernen
(8) Antizipieren und Kombinieren – figural	aus vorgelegten Teilen eine Figur zusammensetzen	Schlussfolgerndes Denken und das Erkennen und Gestalten eines Ganzen aus Teilen
(10) Analysieren und Synthetisieren – abstrakt	Geometrische Muster mit den richtigen Seiten von Würfeln nachlegen	Fähigkeit, komplexe Muster durch geeignete Strukturierung zu reproduzieren

Testkennwerte: Die Normen des AID 2 beruhen auf zwei Stichproben von insgesamt N = 3121 Kindern und Jugendlichen aus Österreich und Deutschland. Die ursprüngliche Reprä-

sentativerhebung fand 1982-1983 an N = 2144 Kindern aus Österreich und der BRD statt; 1995-1997 wurde eine Neunormierung an N = 977 deutschen und österreichischen Kindern vorgenommen. Es liegen dabei Normen für zehn Altersgruppen von 6;0 bis 15;11 Jahren vor. Die Split-half-Reliabilitäten (Berechnung für neun Untertests) sind mit Werten zwischen .91 und .95 sehr hoch; die Stabilität des AID nach vier Wochen beträgt zwischen .83 und .95, nach einem Jahr zwischen .60 und .80. Die inhaltliche Validität des AID ist anhand von Expertenratings, die konvergente und diskriminante Konstruktvalidität über Extremgruppenvalidierung und in Bezug auf zahlreiche Leistungs- und Persönlichkeitstests (u. a. SPM, CFT 20, PFK 9-14) erfasst worden.

Einsatzbereich: Der AID 2 ist in erster Linie zur Individualdiagnostik durch ausgebildete Diplompsychologen in Schulpsychologie, Berufs- und Bildungsberatung und klinischer Psychologie für den normierten Altersbereich geeignet. Neben der förderungsorientierten Diagnostik durch Profilinterpretation bietet das Verfahren auch die Möglichkeit der qualitativen Beurteilung des Arbeits- und Kontaktverhaltens. Aufgrund der adaptiven Testvorgabe ist der AID auch zur Hochbegabendiagnostik bei Kindern und Jugendlichen geeignet. Während die Subtests des AID in der Differenzialdiagnostik zur Erfassung unterschiedlicher Fähigkeiten herangezogen werden können, erlaubt die Standardisierung und Normierung des Verfahrens keine Interpretation separat durchgeführter einzelner Untertests.

Prüfsystem für Schul- und Bildungsberatung (P-S-B)

Testkonzept: Das Prüfsystem für Schul- und Bildungsberatung (P-S-B) ist ein verbreitetes mehrdimensionales, schulspezifisches Intelligenztestverfahren im deutschen Sprachraum (Dauer ca. 45 Minuten; Einzel- oder Gruppentest). Es liegt für 4. bis 6. Klassen (PSB-R 4-6; Horn, Lukesch, Kormann und Mayrhofer, 2002) und für 6. bis 13. Klassen (PSB-R 6-13; Horn, Lukesch, Mayrhofer und Kormann, 2003) vor und basiert auf dem Intelligenzkonzept der Thurstoneschen Primärfaktoren (Thurstone, 1938). Neben den *primary abilities* (s. Subtests) erfasst das P-S-B auch einen *second-order general factor*, der weitestgehend Spearman's Generalfaktor *g* entspricht. Aufbau und Untertests des Verfahrens werden beispielhaft am PSB-R 4-6 beschrieben.

Skalen und Untertests: Das PSB-R 4-6 besteht aus zehn Subtests: (1) Allgemeinwissen, (2) Zahlenreihen, (3) Buchstabenreihen, (4) Figurale Reihen, (5) Wortflüssigkeit, (6) Gliede-

rungsfähigkeit, (7) Raumvorstellung, (8) Gemeinsamkeiten finden, (9) Zahlenaddition, (10) Zahlen vergleichen. Es werden Standardwerte für die Subtests und ein Gesamtwert ausgegeben. Das Verfahren liegt in zwei Parallelformen (A und B) vor. Die Aufgaben, erfassten Inhalte und zugehörige Primärfaktoren werden in Tabelle 8 beschrieben.

Tabelle 8
PSB-R 4-6

Untertests	Aufgabe	Inhalte	Primärfaktor
(1) Allgemeinwissen	Vorgabe sechsbuchstabiger Wörter, in denen ein Buchstabe falsch geschrieben ist und markiert werden muss	Bereichsspezifisches Wissen, verbale Kompetenz, Rechtschreib- und Lesefähigkeit	V: Verbal Comprehension
(2) Zahlenreihen	Reihen in ihrem Aufbau erkennen und das Zeichen, das nicht der Regel entspricht, durchstreichen	Induktive Denkfähigkeit - numerisch	R: Reasoning
(3) Buchstabenreihen	Reihen in ihrem Aufbau erkennen und das Zeichen, das nicht der Regel entspricht, durchstreichen	Induktive Denkfähigkeit - verbal	R: Reasoning
(4) Figurale Reihen	Reihen in ihrem Aufbau erkennen und das Zeichen, das nicht der Regel entspricht, durchstreichen	Induktive Denkfähigkeit - symbolisch	R: Reasoning
(5) Wortflüssigkeit	Finden von Wörtern mit gleichem Anfangsbuchstaben	Sprachgebundene Kreativität	W: Word Fluency
(6) Gliederungsfähigkeit	Zweidimensionale geometrische Gebilde in einer komplexen Zeichnung wieder finden	Räumliches Denken	Flexibility of Closure bzw. Feldabhängigkeit
(7) Raumvorstellung	Bei geometrischen Figuren die Anzahl der sie begrenzenden Flächen bestimmen	Räumliches Vorstellungsvermögen	S: Space
(8) Gemeinsamkeiten finden	Dasjenige von fünf Wörtern finden, welches keine Gemeinsamkeit mit den anderen aufweist	Begriffserfassung, sprachlogisches Denken, Rechtschreib- und Lesefähigkeit	V: Verbal Comprehension
(9) Zahlenaddition	Reihen mit zehn nebeneinander stehenden Zahlen addieren und die Endziffer ankreuzen	Konzentrationsfähigkeit mit dem Materialbereich Zahlen, Beherrschen einfacher Additionsaufgaben	N: Number
(10) Zahlen vergleichen	Vergleich von zwei Zahlenreihen und Ankreuzen der abweichenden Zahl	Kurzzeitige Konzentrationsfähigkeit	P : Perceptual Speed

Testkennwerte: Die Normen des PSB-R 4-6 beruhen auf $N = 1559$ SchülerInnen aus Bayern und Baden-Württemberg. Dabei liegen Normen für vier Klassenstufen (erstes und zweites Halbjahr der 4. Klasse; 5. Klasse; 6. Klasse) und verschiedene Schularten vor. Für das PSB-R 6-13 gibt es schulstufen- und schulartbezogene Normen für SchülerInnen der 6. bis 13. Klasse sowie vorläufige Normen für Erwachsene (Form A: $N = 3.765$, Form B: $N = 3.608$). Die Reliabilität (Cronbachs Alpha) der PSB-R 4-6 Subtests liegt zwischen .62 und .93, wobei die sieben Subtests mit relativ vielen Aufgaben die höheren Koeffizienten aufweisen. Untersuchungen zur faktoriellen Validität des Verfahrens ergaben wiederholt die beiden Faktoren „Verbalität“ (F1: Wissen und sprachliche Leistungsfähigkeit) und „Inductive Reasoning“ (F2: schlussfolgerndes Denken), sowie einen Faktor „Wahrnehmungstempo und Konzentration“ (F3) und „Feldabhängigkeit“ (F4: flexibility of closure); wobei alle Faktoren untereinander und mit dem Generalfaktor zweiter Ordnung korrelierten. Eine eigenständige Dimension des räumlichen Vorstellungsvermögens oder anderer Faktoren konnte nicht nachgewiesen werden. Die PSB-R 4-6 Ergebnisse weisen außerdem hohe Zusammenhänge zu Schulnoten, Übertrittsempfehlungen, Lehrerurteilen sowie diversen Intelligenz- und Konzentrationstests auf (CFT: .55/.63; KFT 4R: .69; KLT; d2).

Einsatzbereich: Das P-S-B ist ein schulnaher Intelligenztest, der in erster Linie schulrelevante kognitive Fähigkeiten erfassen soll. Das revidierte Verfahren wird mit seiner verbesserten Ökonomie (45-minütiger Gruppentest) vor allem in der Schul- und Bildungsberatung sowie in der psychologischen Forschung eingesetzt. Während die Subtests des P-S-B nach der Primärfaktorentheorie als eigenständige Fähigkeiten interpretiert werden können, ist die Interpretation separat durchgeführter einzelner Untertests durch die Standardisierung und Normierung des Verfahrens nicht gegeben.

Intelligenz-Struktur-Test 2000 R (I-S-T 2000 R)

Testkonzept: Der Intelligenz-Struktur-Test 2000 R (I-S-T 2000 R) von Amthauer, Brocke, Liepmann und Beauducel (2001) ist ein deutschsprachiger Intelligenztest für Jugendliche und Erwachsene im Alter von 15 bis 51 Jahren und älter (Dauer je nach Modul ca. 90 - 160 Minuten). Die Testkonstruktion basiert auf dem hierarchischen Protomodell der Intelligenzstrukturforschung (HPI), welches nach dem Prinzip der Multi-Trait-Determination von Intelligenzleistungen mindestens zwei Hierarchie-Ebenen der Intelligenzstruktur unterscheidet: Neben

den sieben *Primärfaktoren* im Sinne Thurstones (1938) wird die Hierarchieebene zweiter Ordnung durch Cattells (1963, 1987) *Generalfaktoren*, fluide und kristallisierte Intelligenz, spezifiziert. Das *Grundmodul* des I-S-T 2000 R erfasst die fünf der Primärfaktoren Thurstones, die modellübergreifend immer wieder mit hoher Konvergenz aufgewiesen wurden: verbale Intelligenz (V), numerische Intelligenz (N), figurale Intelligenz (F), Merkfähigkeit (M) und schlussfolgerndes Denken/Reasoning (SD). Das *Erweiterungsmodul* ist auf die Erfassung der Generalfaktoren fluide (gf) und kristallisierte (gc) Intelligenz ausgerichtet. Zusätzlich liegen Skalen zur Erfassung der Leistungsdimension Wissen vor.

Skalen und Untertests: Das *Grundmodul* des I-S-T besteht aus neun Aufgabengruppen, von denen jeweils drei der Erfassung der verbalen, numerischen und figuralen Intelligenz sowie zwei Aufgabengruppen zur Erfassung der Merkfähigkeit (siehe Tabelle 9) dienen. Das *schlussfolgernde Denken* (SD) als grundlegende Intelligenzfähigkeit wird als Summenwert aus den Skalen für verbale, numerische und figurale Intelligenz erfasst. Die *Grundmodul-Kurzform* besteht aus dem Grundmodul ohne die Aufgaben zur Merkfähigkeit. Das *Erweiterungsmodul* erfasst die Fähigkeit zum Wissenserwerb in unserer Kultur, indem es Aspekte des erworbenen Wissens erfragt. Aus verbal (VW), numerisch (NW) und figural (FW) kodiertem Wissen kann als Summenwert *Wissen* (W) und über Faktorwerte *Wissen/gc* erfasst werden. Auch das *schlussfolgernde Denken/gf* (Reasoning weitgehend ohne Wissensanteile) kann im Rahmen des Erweiterungsmoduls abgebildet werden. Die Kurzform des Grundmoduls sowie das Erweiterungsmodul liegen in zwei Formen (A und B) vor.

Tabelle 9
I-S-T 2000 R

Skalen und Untertests	Aufgabe	Inhalte
Grundmodul		Schlussfolgerndes Denken
Aufgabengruppe verbal		
Satzergänzung	unvollständige Sätze mit dem passenden Wort aus einer vorgegebenen Reihe vervollständigen	Verbale Intelligenz (V)
Analogien	die Relation zwischen zwei Begriffen erkennen und zu einem vorgegebenen dritten, den passenden Begriff auswählen	Verbale Intelligenz (V)
Gemeinsamkeiten	aus einer Gruppe von sechs Wörtern die zwei herausfinden, für die es einen gemeinsamen Oberbegriff gibt	Verbale Intelligenz (V)
Aufgabengruppe numerisch		
Rechenaufgaben	Rechenoperationen im Bereich der reellen Zahlen	Numerische Intelligenz (N)
Zahlenreihen	die nächstfolgende Zahl in einer nach einer bestimmten Regel gebildeten Zahlenreihe finden	Numerische Intelligenz (N)
Rechenzeichen	passende Rechenzeichen in vorgegebene Gleichungen im Bereich der rationalen Zahlen einsetzen	Numerische Intelligenz (N)
Aufgabengruppe figural		
Figurenauswahl	herausfinden, welche von jeweils zehn Figuren eine in fünf Stücke geschnittene Figur beim Zusammensetzen ergibt	Figurale Intelligenz (F)
Würfelaufgaben	in einer Reihe gemusterter Würfel einen in veränderter Lage vorgegebenen Würfel wiederfinden	Figurale Intelligenz (F)
Matrizen	nach einer bestimmten Regel, angeordnete Figuren durch Auswahl der regelkonformen Lösung ergänzen	Figurale Intelligenz (F)
Merkaufgaben		
Verbale Aufgaben	vorgegebene Wörter und deren Zuordnung zu Oberbegriffen einprägen und später aus vorgegebenen Oberbegriffen diejenigen auswählen, denen die eingepprägten Wörter zuzuordnen sind	Merkfähigkeit (M)
Figurale Aufgaben	Figurenpaare einprägen und später ein vorgegebenes Element durch Auswahl der richtigen zweiten Figur ergänzen	Merkfähigkeit (M)
Erweiterungsmodul		Wissen
Wissenstest	Fragen zu verschiedenen Wissensgebieten, die richtige Lösung ist aus fünf Alternativen auszuwählen	Wissen

Testkennwerte: Die Normen des I-S-T 2000 R wurden an N = 3484 deutschen Probanden (etwa die Hälfte davon Gymnasiasten) für zehn Altersgruppen von 15 bis 51 Jahren und älter erhoben. Die Reliabilitätsschätzungen (Split half) für die Skalen des Grundmoduls und für die Wissensskalen sind mit Werten zwischen .88 und .97 als gut bis sehr gut einzuschätzen. Eine zur Validitätsprüfung durchgeführte Dimensionsanalyse bestätigte die dreifaktorielle Struktur der inhaltsbezogenen Fähigkeiten; auch die anderen Skalen des Grundmoduls konnten durch multivariate Analysen abgesichert werden. Für das Erweiterungsmodul konnten die Faktoren für schlussfolgerndes Denken/gf und Wissen/gc fast unkontaminiert durch Inhaltskomponenten faktorenanalytisch nachgewiesen werden. Die konvergente und diskriminante Gültigkeit beider Module konnte durch Zusammenhänge mit anderen Tests (u. a. d2, MWT-B, HAWIE-R, CFT 20, Raven) sowie durch Korrelationen mit Schulnoten belegt werden.

Einsatzbereich: Der I-S-T 2000 R ist als Individual- und Gruppentest zur differentiell-psychologischen Intelligenzdiagnostik und –forschung einsetzbar. Eine fachgerechte Anwendung des Verfahrens erfordert die genaue Einhaltung der Instruktionen und der im Manual spezifizierten Bedingungen für die Durchführung, Auswertung und Interpretation, insbesondere was die Vorgabe der einzelnen Module betrifft. Die allgemeinen Qualitätsstandards zur fachlich adäquaten Testanwendung sind ausführlich in den „Standards für pädagogisches und psychologisches Testen“ (Häcker, Leutner & Amelang, 1998) beschrieben. Von der Interpretation einzelner Untertests oder Aufgabengruppen als Fähigkeiten raten die Testautoren (Handbuch I-S-T 200 R, S. 22) ausdrücklich ab:

„Es muss mit Nachdruck darauf hingewiesen werden, dass nicht einzelne Aufgabengruppen als Fähigkeiten interpretiert werden können, sondern nur die aus den Aufgabengruppen gebildeten Skalen. Die empirischen und theoretischen Grundlagen (...) fundieren nur die ausgewiesenen Skalen bzw. Faktorwerte und nicht einzelne Aufgabengruppen als Fähigkeitsmaße für die 11 mit dem I-S-T 2000 R erfassten Intelligenzfähigkeiten.“

2.1.2 Eindimensionale Intelligenztestverfahren

Bei den eindimensionalen Testverfahren finden sich drei Tests, die international anerkannt sind und häufig zum Einsatz kommen. Daher soll auf diese drei Verfahren genauer eingegangen werden (Tabelle 10). Der Snijders-Oomen Nonverbale Intelligenztest wird anhand der Form für das Vorschulalter beschrieben.

Tabelle 10
Eindimensionale Intelligenztestverfahren

Snijders-Oomen Nonverbaler Intelligenztest	SON-R 2½-7	2;6 – 7;11	Fluide Intelligenz
Grundintelligenztest	CFT 1	5;3– 9;5	Fluide Intelligenz
	CFT 20-R	8;5 – 19;11	
Raven Matrizen Test	CPM	3;9 – 11;8	Fluide Intelligenz
	SPM	ab 6;0	

Snijders-Oomen Non-verbaler Intelligenztest (SON-R 2½ - 7)

Testkonzept: Der Snijders-Oomen Non-verbale Intelligenztest (SON-R 2½ - 7; Tellegen et al., 1998) ist ein individuell einzusetzendes Verfahren für Kinder im Alter von 2;6 bis 7;0 Jahren. Dem SON-R liegt kein spezifisches Intelligenzkonzept zugrunde; die ursprüngliche Version wurde zur Untersuchung kognitiver Funktionen gehörloser Kinder für die Praxis entwickelt. Insgesamt soll der SON-R eher „fluide Intelligenz“ als „kristalline Intelligenz“ erfassen (Tellegen et al., 1998). Die verschiedenen Intelligenzfunktionen werden untersucht, ohne dabei von Sprache abhängig zu sein, da das Verfahren ohne gesprochene oder schriftliche Anweisungen durchgeführt werden kann und von den Kindern keine sprachlichen Äußerungen verlangt. Es werden zwei allgemeine Aspekte von Intelligenz unterschieden: handlungsbezogene, perzeptive Fähigkeiten, die sich an räumliches Verständnis für Formrelationen, konkretes Verständnis für Bedeutungsrelationen und visuomotorische Fähigkeiten richten (*Handlungstests*) sowie abstraktes/konkretes Denken, das sich auf die Fähigkeit bezieht, komplexe Zusammenhänge zu erkennen und logische Schlussfolgerungen zu ziehen (*Denktests*). Die letztgenannten Fähigkeiten sind stärker mit verbaler Intelligenz und Sprachfertigkeit verbunden. Bei den *Denktests* muss aus dem vorgegebenen Testmaterial ein Ordnungsprinzip abgeleitet und auf neue Aufgabenstellungen übertragen werden. In den *Handlungstests* müssen anhand von Bildmaterial Lösungsstrategien erarbeitet werden, bei denen durch visuelle Analyse einer Vorlage die Beziehung zwischen Teilen und Ganzem erfasst und die einzelnen Teile dann richtig zum Ganzen zusammenzufügen sind. Dabei wird die Fähigkeit zu ganzheitlichem Denken erfasst, sowie räumliches Vorstellungsvermögen und Raum-Lage-Orientierung.

Skalen und Untertests: Der SON-R besteht aus insgesamt sechs Untertests von denen drei als Denk- und drei als Handlungstests konzipiert sind: (1) *Mosaike*, (2) *Kategorien*, (3) *Puzzles*, (4) *Analogien*, (5) *Situationen*, (6) *Zeichenmuster* (vgl. Tabelle 11). Bei der Auswertung des SON-R können Standardwerte für die Gesamtskala sowie für die Denk- und Handlungsskala bestimmt werden.

Tabelle 11
SON-R

Skalen und Untertests	Aufgabe	Inhalte
Denktest		
(2) Kategorien	anhand von Beispielbildern müssen Oberbegriffe gebildet und Abbildungen von Gegenständen in Kategorien gruppiert werden	Abstraktes Denken, schlussfolgerndes Denken, Lösungsstrategien entwickeln, Ordnungsprinzipien erkennen und übertragen, Unterscheidung relevanter und nicht relevanter Merkmale, Umgang mit Konzepten/ Begriffen/ Kategorien
(4) Analogien	geometrische Figuren müssen nach bestimmten Form- und Farbkriterien sortiert werden	Abstraktes Denken, schlussfolgerndes Denken, Lösungsstrategien entwickeln, Zuordnungs- und Veränderungsprinzipien erkennen, Klassifikationsmerkmale unterscheiden, differenzierte visuelle Wahrnehmung, visuelle Beobachtungsfähigkeit, Umgang mit Konzepten und Analogien
(5) Situationen	Bilder müssen nach inhaltlichen Kriterien ergänzt werden.	Konkretes Denken, logisches Schlussfolgern, Erkennen komplexer Zusammenhänge, visuelle Wahrnehmung, Raum-Lage-Erkennen, Figur-Grund-Unterscheidung
Handlungstest		
(1) Mosaike	anhand einer Vorlage müssen mit roten und gelben Quadraten abstrakte Muster nachgelegt werden	Räumliches Verständnis, Handlungsplanung, Problemlösestrategien, logisch-analytisches Denken, Verständnis der Beziehung Teil/Ganzes, Unterscheidung Wesentliches/ Unwesentliches, visuomotorische Koordination, räumliche Orientierungsfähigkeit
(3) Puzzle	aus einzelnen Puzzleteilen muss ein Bild zusammengesetzt werden	Konkretes Denken, Handlungsplanung, ganzheitliches Denken, analytisches Denken, Fähigkeit Ganzes aus Teilen zusammensetzen, visuelle Wahrnehmung, Vorstellungsvermögen, Raum-Lage-Erkennen, Figur-Grund-Unterscheidung, Feinmotorik
(6) Zeichenmuster	vorgezeichnete geometrische Figuren müssen nachgezeichnet werden	Räumliches Verständnis, Handlungsplanung, Erkennen von Mustern und Strukturen, Reproduzieren einer Vorlage, differenzierte visuelle Wahrnehmung, räumliches Denken, visuomotorische Fähigkeiten, Auge-Hand-Koordination, Feinmotorik, Raum-Lage-Erkennen

Testkennwerte: Für den SON-R liegen Normen aus einer Stichprobe von N = 1100 niederländischen Kindern für die Altersgruppen 2;6 bis 7;0 Jahren vor. Faktorenanalysen der Subtests

unterstützen die Differenzierung in Denk- und Handlungstests (Tellegen et al., 1998). Die Skalen weisen insgesamt befriedigende Reliabilitäten auf (mittlerer Wert: $r = .72$). Die Korrelationen mit verschiedenen allgemeinen Intelligenztests variieren zwischen $r = .59$ und $r = .75$. Zusammenhänge mit anderen nonverbalen Tests für das Vorschulalter (z.B. Test of Memory and Learning, TOMAL; Developmental Test of Visual Perception, DTVP-2) fallen erwartungsgemäß etwas höher aus ($r = .61$) als Korrelationen mit den Verbalteilen allgemeiner Intelligenztests (z. B. Wechsler Intelligence Scale for Children, WISC-R) sowie Sprachentwicklungstests (z. B. Reynell Development Language Scales). Hier liegt die mittlere Korrelation bei $r = .55$.

Einsatzbereich: Der SON-R ermöglicht die differenzierte Erfassung kognitiver Fähigkeiten im Kindergarten- und Vorschulalter. Aufgrund seines nonverbalen Charakters ist es möglich, dieses Verfahren bei sprach- und kommunikationsgestörten Kindern sowie bei ausländischen Kindern mit eingeschränkten Sprachkenntnissen einzusetzen. Die Möglichkeit, die Testanweisungen ohne gesprochene oder schriftliche Sprache durchzuführen, erlaubt die Anwendung des SON-R bei hörgeschädigten Kindern.

Grundintelligenztests (CFT 1, CFT 20-R)

Testkonzept: Die Grundintelligenztests sind eine deutsche Adaptation der Culture-Fair Intelligence Tests von Cattell zur Bestimmung der Grundintelligenz („fluid ability“) einer Person, d.h. der Fähigkeit, in neuartigen Situationen und anhand von sprachfreiem, figuralem Material, Denkprobleme zu erfassen, Beziehungen herzustellen und Regeln zu erkennen. Der Test liegt in zwei revidierten Versionen mit aktueller Normierung vor: (1) Dem *Grundintelligenztest – Skala 1* (CFT1, Weiß & Osterland, 1997) von 5;3 – 9;5 Jahren (Dauer ca. 25 - 30 Minuten) und (2) dem *Grundintelligenztest – Skala 2* (CFT 20-R, Weiß, 2006) für Kinder und Jugendliche von 8;5 bis 19;11 Jahren sowie Erwachsene von 20 bis 60 Jahren (Dauer ca. 60 Minuten, Kurzform 25-30 Minuten). Die Grundintelligenztests basieren auf dem von Cattell entwickelten Intelligenzmodell, bei dem zwischen fluider und kristalliner Intelligenz unterschieden wird. Mit dem Ziel der Entwicklung eines „culture-fair“ Tests, also einem von Umwelt- und Lernfaktoren unabhängigen Verfahrens, konzentrierte sich Cattell beim CFT auf die Erfassung der fluiden Intelligenz, die für ihn die Fähigkeit darstellt, komplexe Beziehungen in neuartigen Situationen wahrzunehmen und zu verstehen. Das Verfahren wurde als Wahrneh-

mungstest konzipiert, eine Testart, die nach Cattell am stärksten als „kulturfrei“ angesehen werden kann. Die CFT können sowohl als Gruppenverfahren als auch als Einzeltest durchgeführt werden und liegen in zwei Parallelformen vor. Der CFT 1 besteht aus fünf Untertests mit fünf verschiedenen Aufgabenstellungen, der CFT 20 aus zwei Testteilen mit jeweils 4 Aufgaben sowie den Ergänzungstests Wortschatztest und Zahlenfolgentest.

Skalen und Untertests: Der CFT 1 für 5- bis 9-jährige Kinder umfasst 5 Untertests: (1) *Substitutionen*, (2) *Labyrinth*, (3) *Klassifikationen*, (4) *Ähnlichkeiten*, (5) *Matrizen* (siehe Tabelle 12), wobei die Untertests 1 und 2 vor allem Wahrnehmungsfähigkeiten erfassen, während die Untertests 3 bis 5 am stärksten die sprachfreie Denkkapazität (fluid ability) messen. Alle Aufgaben sind innerhalb einer vorgegebenen Zeitgrenze zu lösen. Der CFT 20-R für Kinder und Jugendliche im Alter von 8;5 bis 19;11 Jahren besteht aus zwei Testteilen, die jeweils vier Untertests umfassen: (1) *Reihenfortsetzen*, (2) *Klassifikationen*, (3) *Matrizen*, (4) *Topologische Schlussfolgerungen* (siehe Tabelle 12). Alle vier Aufgaben sind innerhalb einer vorgegebenen Zeitgrenze zu lösen und erfassen das von Sozialisationseinflüssen freie Intelligenzpotential (fluid ability, „g“-Faktor).

Tabelle 12
Grundintelligenztest

Untertests	Aufgabe	Inhalte
CFT 1		
(1) Substitutionen	Vorgegebene Bilder mit jeweils zugehörigen Symbolen versehen	Reproduktiver Aspekt der Wahrnehmung
(2) Labyrinth	Labyrinth von steigender Komplexität vom Startpunkt zum Zielpunkt durchfahren	Produktiver Aspekt der Wahrnehmung sowie visuelle Orientierung und Aufmerksamkeit
(3) Klassifikationen	In einer Reihe von fünf Bildern dasjenige herausfinden, das nicht zu den vier merkmalsähnlichen Bildern passt	Beziehungsstiftendes Denken bei figuralem Material
(4) Ähnlichkeiten	Aus einer Reihe von fünf Bildern das Bild erkennen, das zum vorgegebenem Zielitem passt	Wiedererkennen figuraler Vorgaben im Kontext ähnlicher, jedoch merkmalsveränderter Figuren
(5) Matrizen	Aus fünf vorgegebenen Bildern dasjenige auswählen, das ein dargestelltes Muster korrekt ergänzt	Fähigkeit, Regeln und Zusammenhänge bei figuralen Problemstellungen zu erkennen

Fortsetzung Tabelle nächste Seite

Fortsetzung Tabelle 12

CFT 20-R		
(1) Reihenfortsetzen	Aus fünf vorgegebenen Figuren diejenige auswählen, die eine Musterreihe korrekt ergänzt	Gesetzmäßigkeiten und Regelmäßigkeiten erkennen
(2) Klassifikationen	In einer Reihe von fünf Bildern dasjenige herausfinden, das nicht zu den vier merkmalsähnlichen Bildern passt	Beziehungstiftendes Denken
(3) Matrizen	Aus fünf vorgegebenen Bildern dasjenige auswählen, das ein dargestelltes Muster korrekt ergänzt	Gesetzmäßigkeiten und Regelmäßigkeiten erkennen
(4) Topologische Schlussfolgerungen	Aus einer Reihe von fünf Auswahlfiguren das Bild erkennen, in dem ein Punkt genauso liegen kann wie beim Zielitem	Schlussfolgerndes Denken

Testkennwerte: Die Normierung des *CFT I* fand 1995 in Gruppenuntersuchungen in Sachsen und Baden-Württemberg an insgesamt 1500 Schülern statt, 1200 Testbearbeitungen wurden in die Normberechnungen einbezogen. Es liegen Altersnormen (5;3 bis 9;5) und Klassennormen (Vorschule, 1. – 3. Klasse) für Grundschulen (1. – 4. Klasse) und für Förder-/Sonderschulen vor. Die Reliabilitätskoeffizienten (Testhalbierungsmethode) für die Untertests 3 bis 5 liegen mit Subtestwerten zwischen .65 und .88 und mit Werten zwischen .90 und .96 (je nach Normgruppe) für die Gesamtsumme dieser Untertests im zufriedenstellenden Bereich. Aufgrund des ausgeprägten Speedcharakters der ersten beiden Untertests lassen sich für diese Subtests keine Zuverlässigkeitsberechnungen erstellen. Die Validitätsbestimmung anhand von Faktorenanalysen ergab eine Zwei-Faktoren-Lösung bei der die Untertests (3) Klassifikationen, (4) Ähnlichkeiten und (5) Matrizen am stärksten grundlegende intellektuelle Fähigkeiten erfassten, während die Untertests Substitutionen und Labyrinth stärker mit Wahrnehmungsaspekten in Zusammenhang gebracht werden konnten. Korrelationen der CFT 1-Gesamtleistung mit anderen Testverfahren ergaben mit dem HAWIK-Gesamtergebnis einen Koeffizienten von .66, der als zufriedenstellend bezeichnet werden kann.

Die Normierung des *CFT 20-R* wurde zwischen Dezember 2002 und Juni 2003 bei rund 4400 Schülerinnen und Schülern vom 3. bis 13. Schuljahr durchgeführt. Für den Altersbereich 8;5 bis 19;11 liegen Normtabellen für 16 Altersgruppen vor, der Altersbereich 20 bis 60 Jahre wurde als eine Altersgruppe zusammengefasst. Weiterhin gibt es Klassenstufennormen für das 3. bis 13. Schuljahr. Die Reliabilität des CFT 20-R wurde durch Testwiederholung und Konsistenzprüfung bestimmt. Die Test-Retest-Übereinstimmung bei N = 50 Haupt-/Real- und Förderschülern lag bei einem Zeitabstand von zwei bis fünf Monaten bei .91 für die Haupt-/

Realschüler und .83 für die Förderschüler. Der Äquivalenzkoeffizient aus der Korrelation des 1. Testteils mit dem 2. Testteil lag für die Gesamtstichprobe (N = 4325) bei .95.

Zur faktoriellen Validität erbrachte eine Hauptachsen-Analyse einen Zentralfaktor (im Sinne eines „g“-Faktors), auf den alle Subtests hochsignifikant laden. Eine genauere Untersuchung der Subtestbezüge zueinander durch Varimaxrotation lässt eine inhaltliche Interpretation zu, bei der Faktor 1 als „Gesetzmäßigkeiten und Regelmäßigkeiten erkennen“ bezeichnet werden kann (UT Matrizen und Reihenfortsetzen, 50% aufgeklärte Varianz), Faktor 2 als „Schlussfolgerndes Denken“ (UT topologisches Schlussfolgern, 25% Beteiligung) und Faktor 3 als „Beziehungstiftendes Denken“ (UT Klassifikationen, 25% Beteiligung). Zur Kriteriumsvalidität wurden Korrelationen mit anderen psychodiagnostischen Verfahren und Schulleistungen untersucht. Dabei fielen die Korrelationen des CFT 20-R mit dem PSB-R 4-6 mit .56 (N = 860) eher niedrig aus, was vermutlich daran liegt, dass der PSB stärker auf verbalen und damit auch sozialisationsabhängigen Fähigkeiten beruht. Eine genauere Analyse bestätigte, dass die „konstruktnahen“ Variablen des schlussfolgernden Denkens (Reasoning-Faktor) des PSB höher mit dem CFT 20-R korrelierten als der „konstruktferne“ Verbalfaktor. Bei Korrelationen mit Schulnoten zeigte sich, dass die Ergebnisse im Fach Mathematik sehr zufriedenstellend (.49), im Fach Deutsch allerdings wie erwartet niedriger ausfallen (.35).

Einsatzbereich: Der CFT 1 dient der ökonomischen und kulturfreien Erfassung eines wesentlichen Aspektes intelligenten Verhaltens: Er erfasst die Fähigkeit, Denkprobleme in neuartigen Situationen anhand von sprachfreiem, figuralem Material zu erfassen. Der Test ist einfach anzuwenden und ermöglicht es, diese Aspekte sprachfrei und möglichst umweltunabhängig zu untersuchen. Der CFT 1 ist bei Schulschwierigkeiten sowie zur Überprüfung der Sonderschulbedürftigkeit brauchbar.

Der CFT 20-R dient, wie der CFT 1, der Bestimmung der Grundintelligenz, der grundlegenden geistigen Leistungsfähigkeit im Sinne eines g-Faktors auf sprachfreier Basis anhand von figuralen, anschaulichen Denkaufgaben. Das Verfahren findet Anwendung in der Erziehungs- und schulpsychologischen Beratung, in der Berufsberatung sowie bei Rehabilitationsmaßnahmen. Der CFT 20-R lässt sich als Einzel- oder Gruppentest in einer Lang- und einer Kurzform durchführen, wodurch die Testdurchführung flexibel gestaltet werden kann. Hinzu kommt, dass der Test aufgrund seiner Papierform und der einheitlichen Struktur mit weniger Erfahrung und Einarbeitungsaufwand durchgeführt werden kann.

Raven-Matrizen (CPM, SPM)

Testkonzept: Die Raven-Matrizen-Testserie wurde von Raven (1941) entwickelt und besteht aus drei Formen (Dauer jeweils max. 45 Minuten): den *Coloured Progressive Matrices (CPM)* für jüngere Kinder oder geistig retardierte Personen (Raven, Raven & Court, 1995; deutsche Normen von 3;9 bis 11;8 Jahren: Bulheller & Häcker, 2006), den *Standard Progressive Matrices (SPM)* für Kinder und Erwachsene (Raven, Raven & Court, 1996; dt. Normen ab 6;0 Jahre: Heller, Kratzmeier & Lengfelder, 1998) und den *Advanced Progressive Matrices (APM)* für überdurchschnittlich intelligente Jugendliche und Erwachsene (Raven, Raven & Court, 1994; dt. Normen ab 12;0 Jahre: Heller, Kratzmeier & Lengfelder, 1998). Jede Matrizie besteht aus einem Muster oder einer Figur, die nach einem bestimmten Prinzip aufgebaut ist, welches erkannt werden muss. Ein Teil des Musters ist ausgelassen und muss von der Testperson unter gegebenen Antwortvorschlägen herausgefunden werden. Die Testserie soll die gesamte kognitive Entwicklung erfassen, vom einfachen Begreifen, dass das fehlende Stück einer Figur eingesetzt werden soll, bis zur maximalen Fähigkeit eines Menschen, Vergleiche zu bilden und in Analogien zu denken. Die Gesamtleistung erfasst die *allgemeine Intelligenz* im Sinne Spearman's. Während die Erfassung von induktivem Denken, räumlicher Wahrnehmungsdifferenzierung und räumlichen Vorstellen und Denken (space factors) durch die Matrizen noch kritisch diskutiert wird, betont Carroll (1993) die gute Messqualität der Standardform des Matrizentests in Bezug auf „g“ und *fluide Intelligenz*, d.h. der von Wissen und Sozialisation weitgehend unabhängigen intellektuellen Problemlösefähigkeit. Im Folgenden wird der Matrizentest am Beispiel der SPM und CPM genauer beschrieben.

Skalen und Untertests: Die Standardform der Progressiven Matrizen (SPM; ab 6 Jahren) besteht aus einem Testheft mit fünf Aufgabensets (A bis E) von jeweils 12 Aufgaben. Jede Aufgabe besteht aus Figuren oder Mustern aus einfachen Strichzeichnungen, wobei immer ein Teil des Musters ausgelassen ist. Die Versuchsperson muss das dem Aufbau der Figur zugrunde liegende Prinzip erkennen und aus jeweils sechs Antwortalternativen die richtige Lösung zur Ergänzung des Musters herausfinden (Ankreuzen im Antwortbogen). Die Lösung der ersten Aufgaben der Sets A bis D ist unmittelbar einsichtig, während die folgenden Aufgaben zunehmend schwieriger werden. Alle Testpersonen bearbeiten den gleichen Aufgabensatz in identischer Reihenfolge. Die einzelnen Aufgaben sind ohne Unterbrechung fortlaufend zu lösen, wobei Grundschüler in der Regel die Aufgaben der Skala E nicht mehr lösen kön-

nen. Das Arbeitstempo ist dabei frei wählbar; die meisten Versuchspersonen schließen den Test ohne Zeitdruck innerhalb der zeitlichen Obergrenze von maximal 45 Minuten Dauer ab.

Die CPM für das Vorschulalter enthalten nur drei Aufgabensets (A bis C) mit jeweils 12 Aufgaben, die leichter zu lösen sind als die Items der SPM. Die CPM liegen in einer Testheftversion plus Parallelförmigkeit sowie in einer Board Form vor, die das Einfügen der Antwortmöglichkeiten nach Art eines Puzzles erlaubt. In den meisten Untersuchungen erbringen beide Durchführungsversionen praktisch die gleichen Resultate.

Testkennwerte: Für die SPM liegen Altersnormen (6 bis 19 Jahre) und klassenstufenbezogene Normen (Klasse 1 bis 12) für Schüler verschiedener Schularten (Grund-, Haupt-, Realschule und Gymnasium) vor. Die Normierung wurde an N = 2134 Schülern in Bayern und Nordrhein-Westfalen durchgeführt. Weiterhin liegen Normen für Studierende verschiedener Altersgruppen (19-60 Jahre, N = 347), für 60- bis 90jährige Personen (N = 17) sowie für lernbehinderte (14-16 Jahre, N = 156) und hörgeschädigte SchülerInnen (Klasse 2-9, N = 141) vor. Sowohl hinsichtlich der inneren Konsistenz (Split-half, Cronbachs Alpha) als auch der Retest-Reliabilität sind die SPM als gut bis sehr gut reliabel zu bezeichnen. Die Kriteriumsvalidität der SPM wurde durch Zusammenhänge mit dem Kognitiven Fähigkeitstest KFT (.66), dem Grundintelligenztest CFT (.46-.81), den Wechsler Skalen für Kinder WISC-R (.42-.70), dem Prüfungssystem für Schul- und Bildungsberatung PSB (.56) sowie mit der Informationsverarbeitungsgeschwindigkeit im Zahlen-Verbindungs-Test ZVT (.33) nachgewiesen. Weiterhin zeigten sich Zusammenhänge zwischen den SPM-Ergebnissen und Schulleistungsmaßen (.20-.60), wobei der Zusammenhang mit der Mathematiknote besonders deutlich ausgeprägt war (.40). Eine faktorielle Validierung erbrachte bei Studierenden einen Faktor mit 98,2% Varianzaufklärung (bei SchülerInnen 90,1%), der als Nachweis des angenommenen g-Faktors interpretiert wird.

Für die CPM liegen deutsche Altersnormen für Kinder von 3;9 bis 11;8 Jahren in Halbjahresabständen (N = 1.218) vor sowie Standardisierungen mit alten Menschen und aus anderen Ländern. Die CPM weisen mit einer Split-Half-Genauigkeit zwischen .85 bis .90 eine gute Reliabilität auf; die Retest-Reliabilitäten liegen zwischen .86 und .90 (Testwiederholungsabstand ein bis zwei Wochen). Die Validität der CPM konnte durch Zusammenhänge mit dem HAWIK (.43 - .73) belegt werden. Faktorenanalytische Untersuchungen bestätigten eine hohe g-Faktor Ladung, sowie die Bedeutung des visuell-räumlichen Faktors „simultane Verarbeitung“ für das Lösen der CPM-Aufgaben.

Einsatzbereich: Die Raven-Matrizen werden als Einzel- oder Gruppentest in der Schulberatung, der klinischen Diagnostik, der experimentellen Psychologie sowie in der Eignungsdiagnostik und Laufbahnberatung verwendet, und ermöglichen bei überschaubarem Durchführungsaufwand die Erfassung eines allgemeinen Intelligenzfaktors. Die SPM sind bei Kindern ab dem Schulalter und Erwachsenen einsetzbar. Genaue Anweisungen zur Standardisierung der Testsituation erhöhen die Objektivität der Testdurchführung und -auswertung.

Die CPM sind für die Anwendung bei kleinen Kindern, Geistigbehinderten, alten Menschen konstruiert worden; weiterhin sind sie zur Untersuchung von Hirnschäden und Demenz und für interkulturelle Vergleiche geeignet. Die größte Testvalidität besitzen die CPM für Personen, deren Fähigkeit, Vergleiche anzustellen und in Analogien zu denken, noch nicht voll entwickelt oder eingeschränkt ist; sobald diese Fähigkeiten vorhanden sind, bieten die SPM die größere Validität. Die CPM sind sehr einfach durchzuführen (Dauer etwa 20 bis 30 Minuten; eine Zeitbegrenzung ist nicht vorgesehen).

2.2 Messung der Sprachkompetenz im Vorschulalter

Neben Verfahren, die der Erfassung der *allgemeinen Intelligenz* dienen, gibt es inzwischen eine ganze Reihe von kognitiven Leistungstests, die sich auf *spezielle sprachliche oder mathematische Kompetenzen* beziehen und die nachfolgend ebenfalls kurz erläutert werden.

Zur Messung der Sprachkompetenz im Vorschulalter soll näher auf den Sprachentwicklungstest für drei- bis fünfjährige Kinder (SETK 3-5) eingegangen werden. Zwar stehen mit den Sprachentwicklungsskalen von Reynell (dt. Fassung; Sarimski, 1985) für Kinder von 15 Monaten bis 7 Jahren und dem Heidelberger Sprachentwicklungstest (H-S-E-T; Grimm & Schöler, 1978) für die Altersgruppe 3 bis 9 Jahre zwei alternative Verfahren zur standardisierten Erfassung der Sprachproduktion und des Sprachverständnisses zur Verfügung (siehe Tabelle 13), doch liegen für beide Tests keine aktuellen deutschen Normen vor.

Tabelle 13
Sprachliche Fähigkeiten

Sprachentwicklungstest für Drei- bis Fünfjährige	SETK 3-5	3;0 – 5;11	Sprachverständnis, Sprachproduktion, Sprachgedächtnis
Heidelberger Sprachentwicklungstest	HSET	3;0 – 9;11	Sprachverständnis, Sprachproduktion
Reynell-Skalen		15 Mon. – 7 J.	Sprachverständnis, Sprachproduktion

Der Sprachentwicklungstest für Kinder (SETK)

Testkonzept: Der Sprachentwicklungstest für drei- bis fünfjährige Kinder (SETK 3-5; Grimm, 2001; Dauer 20-30 Min.) dient der standardisierten Messung der Sprachkompetenzen im Vorschulalter. Im Rahmen dieses Verfahrens werden die Bereiche Sprachverstehen und Sprachproduktion sowie Sprachgedächtnis untersucht. Diese Fähigkeit ist laut Grimm (2001) sowohl für den Wortschatzerwerb als auch für die Grammatik-Entwicklung von Bedeutung, da die Ableitung komplexer grammatischer Regeln auf der Fähigkeit aufbaut, eine größere, komplexe Menge sprachlicher Einheiten im phonologischen Arbeitsgedächtnis zu behalten.

Skalen und Untertests: Der SETK umfasst insgesamt sechs Untertests (siehe Tabelle 14). Bei dreijährigen Kindern kommen alle sechs Untertests zum Einsatz: (1) Verstehen von Sätzen, (2) Enkodieren semantischer Relationen, (3) Morphologische Regelbildung, (4) Phonologisches Arbeitsgedächtnis für Nichtwörter, (5) Gedächtnisspanne für Wortfolgen und (6) Satzgedächtnis. Für vierjährige Kinder werden nur fünf Untertests durchgeführt, der Untertest 2 fällt weg. Alle Untertests werden einzeln ausgewertet, es kann kein Gesamtwert gebildet werden.

Tabelle 14
SETK

Untertests	Aufgabe	Inhalte
Sprachverstehen		
(1) Verstehen von Sätzen	Sätze werden vorgegeben, die anhand von Bildern oder mit Materialien gezeigt werden sollen	Erfassen von Sprachversteheleistungen bei grammatisch komplexen Strukturen
Sprachproduktion		
(2) Enkodieren semantischer Relationen	zu gezeigten Bildern sollen Sätze gebildet werden	Fähigkeit auf Bildkarten dargestellte Inhalte verbal zu enkodieren
(3) Morphologische Regelbildung	auf Bildern werden Objekte in Einzahl und Mehrzahl gezeigt und das Pluralwort soll genannt werden	Erfassen der morphologischen Kompetenzen
Sprachgedächtnis		
(4) Phonologisches Arbeitsgedächtnis für Nichtwörter	Nachsprechen von mehrsilbigen Nicht- bzw. Kunstwörtern wobei die Silbenzahl zunimmt	Genauigkeit, mit der die gehörten Spracheinheiten im phonologischen Speicher repräsentiert werden können
(5) Gedächtnis für Wortfolgen	Reihen von vorgesprochenen Wörtern müssen wiedergegeben werden	Fähigkeit bekannte, inhaltlich unverbundene Wörter zu speichern und in vorgegebener Abfolge zu reproduzieren
(6) Satzgedächtnis	sinnvolle und inhaltlich sinnlose Sätze ansteigender Länge müssen reproduziert werden	Aussagen über das Kenntnissystem der formalen linguistischen Regeln

Testkennwerte: Es liegen aktuelle SETK 3-5 Normen für fünf Altersgruppen vor (3;0 - 3;5, 3;6 - 3;11, 4;0 - 4;5, 4;6 - 4;11 und 5;0 - 5;11), die an N = 495 deutschen Kindern ermittelt wurden (Grimm, 2001). Die Reliabilitäten für die Untertests können als gut bezeichnet werden (mittlerer Wert für die Altersgruppe 4;0 - 4;5: $r = .83$).

Anhand von zwei Beispielen soll auf die Kriteriumsvalidität eingegangen werden. Untertest-Vergleiche der sprachlichen Leistungen von N = 18 dysphasischen Kindern im Alter von 4;11 bis 5;9 mit der Leistung von sprachunauffälligen Kindern haben ergeben, dass die dysphasischen Kinder in allen Untertests bis auf Verstehen von Sätzen signifikant schlechter abschnitten als die Kontrollkinder (Grimm, 2001). Im Rahmen einer weiteren Studie wurden *late talkers* untersucht. Auch hier hat sich gezeigt, dass im Vergleich zur Kontrollgruppe die N = 17 späten Wortlerner im Alter von drei Jahren bei allen Untertests des SETK 3-5 signifikant schlechtere Leistungen zeigten. Im Alter von fünf Jahren konnten sie in den Untertests Verstehen von Sätzen und Morphologische Regelbildung zwar aufholen, in allen anderen Unter-

tests erzielten sie jedoch nach wie vor deutlich unterdurchschnittliche Ergebnisse (Grimm, 2001).

Zur Bestimmung der diskriminanten Validität wurden bei einer Teilstichprobe von nur $N = 8$ Kindern im Alter von 4;5 bis 4;10 Korrelationen der Untertest des SETK 3-5 mit den Untertests der sprachfreien Skala des K-ABC berechnet (Grimm, 2001). Die Korrelationen fielen insgesamt eher gering aus (mit Handbewegungen: zwischen .05 bis .58; mit Dreiecken: zwischen -.51 und .25, mit Wiedererkennen: zwischen -.51 und .10). Dagegen berichtet die Autorin von mittleren bis hohen Korrelationen zwischen dem SETK Untertest Gedächtnisspanne für Wortfolgen und den K-ABC Untertests Wiedererkennen von Gesichtern (.44) sowie Handbewegungen (.76), und führt diese Zusammenhänge darauf zurück, dass alle Untertests Parameter des Kurzzeitgedächtnisses (hier: modalitätsübergreifend) erfassen. Angesichts der geringen Stichprobengröße sind diese Befunde jedoch wenig aussagekräftig. Sie deuten lediglich an, dass bei SETK-auffälligen Kindern eine Analyse der Leistungen in spezifischen Untertests des K-ABC lohnend erscheint, wenn man die Bedeutung des Arbeitsgedächtnisses für Sprachentwicklungsstörungen besser verstehen will.

Einsatzbereich: Der SETK 3-5 ist ein standardisiertes Verfahren zur Erfassung von rezeptiven und produktiven Sprachfähigkeiten bei drei- bis fünfjährigen Kindern. Das Verfahren ermöglicht es, Störungen der Sprachentwicklung frühzeitig zu diagnostizieren und therapeutische Maßnahmen zu planen. Aufgrund der Bedeutung der auditiven Gedächtnisfähigkeiten für den Spracherwerb werden diese mit dem SETK 3-5 umfassend untersucht.

2.3 Messung der Lese-Rechtschreibkompetenzen im Schulalter

Zur Messung der Lese-Rechtschreibkompetenzen im Schulalter stehen verschiedene Verfahren zur Verfügung (s. Tabelle 15). Ein verbreitetes Verfahren für die 1. bis 4. Klasse ist der Salzburger Lese-Rechtschreibtest, der mit einem Rechtschreib- und einem Leseteil die getrennte Überprüfung von Rechtschreib- und Lesekompetenzen ermöglicht. Weiterhin werden die Hamburger Schreibprobe (HSP), der Lesetest ELFE näher, sowie das Salzburger Lesescreening (SLS) erläutert.

Tabelle 15
Lese-Rechtschreibtests

Salzburger Lese-Rechtschreibtest	SLRT	1. – 4. Klasse	Lese- und Rechtschreibfähigkeiten
Hamburger Schreibprobe	HSP 1, HSP 2, HSP 3, HSP 4, HSP 5-9	1. - 9. Klasse	Rechtschreibfähigkeiten
ELFE-Lesetest		1. – 5. Klasse	Lesefähigkeiten
Salzburger Lesescreening	SLS 1-4 SLS 5-8	1. – 4. Klasse 5. – 8. Klasse	Leseverständnis

Salzburger Lese- und Rechtschreibtest (SLRT)

Testkonzept: Der Salzburger Lese- und Rechtschreibtest (SLRT) von Landerl, Wimmer und Moser (1997) ist ein Verfahren zur Differentialdiagnose von Störungen des Lesens und Schreibens für die 1. bis 4. Klasse. Das Verfahren basiert auf zentralen Ergebnissen der kognitions- und neuropsychologischen Schriftsprach- und Leseerwerbsforschung (Überblick bei Klicpera & Gasteiger-Klicpera, 1995). Der *Lesetest* (Dauer 10-20 Minuten) konzeptualisiert den komplexen Lesevorgang als ein Zusammenwirken von Teilfertigkeiten, die relativ unabhängig voneinander funktionieren und beeinträchtigt sein können, und prüft Störungen der *Wortlesefertigkeit*, die nicht auf eine Störung des Textverständnisses zurückgehen, sondern diese erst bedingen. Das Verfahren unterscheidet Störungen des *synthetischen Lesens* (Synthese von durch Buchstaben abgebildeten Lauten zu einer artikulatorischen Einheit wie einer Silbe oder einem Wort) und der *direkten Worterkennung* (automatische Aktivierung

einer bereits gespeicherten Gedächtnisrepräsentation für ein geschriebenes Wort). Der *Rechtschreibtest* (Dauer 20-30 Minuten) erfasst Störungen des Schriftspracherwerbs, die durch ein Defizit im Bereich der *phonologischen Verarbeitung* verursacht sind, d.h. der kognitiven Verarbeitung der Lautstruktur der gesprochenen Sprache (nicht der Sprachwahrnehmung oder -produktion), und misst Schreibstörungen in zwei Teilfertigkeiten: Störungen des *lautorientierten Schreibens* (Laute des Sprechworts in Buchstaben übersetzen) und des *orthographischen Schreibens* (orthographisch korrekte Schreibungen produzieren).

Skalen und Untertests: Der *Lesetest* enthält fünf Untertests sowie einige Übungsblätter. Die Fertigkeit der direkten Worterkennung wird durch die Subtests *Häufige Wörter* (Buch, Tier) und *Zusammengesetzte Wörter* (Filzstift, Geburtstagskuchen) erfasst. Die *Textleseaufgabe* (Lesen eines kurzen Textes) simuliert eine natürliche Lesesituation, ist jedoch nicht zur Überprüfung des Leseverständnisses geeignet. Beeinträchtigungen des synthetischen Lesens werden durch das Lesen von *Pseudowörtern* erfasst: Der Subtest *Wortunähnliche Pseudowörter* erfordert das Lesen von zwei- und dreisilbigen Pseudowörtern wie „talire“. Im Subtest *Wortähnliche Pseudowörter* müssen Wörter wie „Vaus“ (Wortanfänge der Wörter „Vater“ und „Haus“ getauscht) gelesen werden. Ausgewertet werden Lesefehler und Lesezeit.

Der *Rechtschreibtest* besteht darin, vom Testleiter gesprochene Wörter aufzuschreiben, indem diese in unvollständige Sätze eingesetzt werden. Die Erfassung von Defiziten beim lautorientierten und orthographischen Schreiben erfolgt nicht mittels verschiedener Subtests, sondern durch eine differenzierte *Fehleranalyse von Wortschreibungen* (danach, ob die Schreibung eines Wortes der Lautabfolge des gesprochenen Wortes entspricht vs. danach, ob sie orthographisch korrekt ist bzw. gegen die Groß- und Kleinschreibung verstößt).

Testkennwerte: Der SLRT wurde an über N = 2800 Kindern in Salzburg und Oberösterreich normiert. Normen liegen für fünf Klassenstufen vor (Ende 1. Klasse, erstes und zweites Halbjahr der 2. Klasse, 3. Klasse und 4. Klasse). Die Reliabilitätskoeffizienten des Lesetests (Paralleltest) sind insbesondere für die Lesezeiten sehr hoch (.84 - .99), im Vergleich zu deutlich niedrigeren Koeffizienten für die Lesefehler, weshalb die Lesezeit als das entscheidende Kriterium zur Beurteilung der Lesefertigkeiten empfohlen wird. Die logische Validität des Lesetests ist gegeben, ebenso Zusammenhänge mit der Geschwindigkeit beim leisen Lesen und mit der Zeugnisnote in Deutsch. Beim Rechtschreibtest sind besonders die orthographischen Fehler gut reliabel (.74 - .90); eine Überprüfung der Interrater-Reliabilität ergab zudem eine hohe Übereinstimmung zwischen den Fehlerbeurteilungen beider Rater (.94). Die Validierung

des Rechtschreibtests erfolgte am Lehrerurteil; dabei schnitten als rechtschreibschwach identifizierte Kinder beim Rechtschreibtest deutlich schlechter ab als unauffällige SchülerInnen.

Einsatzbereich: Der SLRT erlaubt die differenzierte Diagnostik (Erfassen von Teilfertigkeiten) von Schwächen beim Erlernen des Lesens und Rechtschreibens in der Grundschulzeit (1. bis 4. Klasse) und ist vor allem bei Kindern angezeigt, die in Bezug auf ihre Lese- und Schreibleistung bereits auffällig geworden sind. Der SLRT ist zur Diagnose von Lernbehinderungen im Sinne einer umschriebenen Entwicklungsstörung schulischer Fertigkeiten geeignet (Lese- und Rechtschreibstörung bzw. isolierte Rechtschreibstörung nach ICD-10; Dilling, Mombour, Schmidt & Schulte-Markwort, 1997). Während der Lesetest nur als Individualtest durchzuführen ist, kann der Rechtschreibtest auch als Gruppentest z.B. in Schulklassen angewendet werden. Für beide Tests liegt eine Parallelförmigkeit vor. Der SLRT kann neben der psychologisch/pädagogisch-diagnostischen Prüfung von Lese-/Rechtschreibschwierigkeiten zur Erstellung und Evaluation spezifischer Fördermaßnahmen sowie als Forschungsinstrument verwendet werden. Bei der Anwendung in Deutschland ist zu berücksichtigen, dass die Normen des SLRT aus Österreich stammen.

Hamburger Schreibprobe (HSP 1-9)

Testkonzept: Die Hamburger Schreibprobe (HSP 1-9) von May (2002; Dauer 20-30 Minuten) ist ein Verfahren zur Diagnose des Rechtschreiblernens für die Klassenstufen 1 bis 9 und liegt in sechs Versionen vor: HSP 1+ (Mitte 1. Klasse bis Mitte 2. Klasse, vier bis acht Einzelwörter und einem Satz, Auswertung auf Buchstaben- und Wortebene sowie Rechtschreibstrategien), HSP 2, HSP 3 (Mitte 3. Klasse und Ende 3. Klasse, 15 Einzelwörter und vier Sätze, Auswertung auf Buchstaben- und Wortebene sowie Rechtschreibstrategien), HSP 4-5 (Mitte 4. Klasse bis Anfang 5. Klasse, 16 Einzelwörter und fünf Sätze, Auswertung auf Buchstaben- und Wortebene sowie Rechtschreibstrategien), HSP 5-9 B (Basisanforderungen, 5. bis 9. Klasse, 14 Einzelwörter und sechs Sätze, Auswertung auf Buchstaben- und Wortebene sowie Rechtschreibstrategien) und HSP 5-9 EK (Erweiterte Kompetenz, 5. bis 9. Klasse, in einer Textvorlage mit teilweise falsch geschriebenen Wörtern die Fehler korrigieren und die Satzzeichen ergänzen). Es dient der differenzierten Erfassung von grundlegenden Rechtschreibstrategien und orthographischem Strukturwissen, wobei neben der Anzahl richtig geschriebener Wörter auch die Anzahl richtiger Grapheme (kleinste unterscheidende Einheit eines

Schriftsystems wie z.B. Diphthonge (ei, au, eu), mehrgliedrige Konsonanten (ch, sch, pf, qu), Vokale mit Längenzeichen (ie, ih, oo) oder Doppelkonsonanten als Kürzungszeichen (ll, mm, ck, tz)) und die Beherrschung von Rechtschreibstrategien (alphabetische Strategie: artikulierte Laute mit Buchstaben schriftlich festhalten, orthographische Strategie: einfache Laut-Buchstaben-Zuordnung unter Berücksichtigung orthographischer Regeln, morphematische Strategie: Herleitung der Schreibweise unter Berücksichtigung der morphematischen Struktur der Wörter, wobei die Erschließung des Wortstammes oder die Zerlegung in Wortteile erforderlich ist, wortübergreifende Strategie: Berücksichtigung sprachlicher Aspekte wie Groß-/ Kleinschreibung, Zusammen-/Getrennschreibung, Satzzeichen beim Schreiben von Sätzen) berücksichtigt wird.

Skalen und Untertests: Die Hamburger Schreibprobe besteht aus Einzelwörtern und Sätzen, die vom Testleiter vorgelesen werden und auf dafür vorgesehene Linien geschrieben werden müssen. Die Bedeutung der Wörter und Sätze wird durch Bilder neben den Linien veranschaulicht. Die Wörter können so häufig wie erforderlich wiederholt werden. Je nach Klassenstufe wird erwartet, dass die Kinder bekannte und häufig verwendete Wörter ansteigender Komplexität orthographisch richtig schreiben können. Hinzu kommt dann die Beachtung der Groß-/Kleinschreibung, Regeln der Kommasetzung und Verwendung von Satzzeichen.

Testkennwerte: Die Daten zur Entwicklung der HSP wurden in den Jahren 1987 bis 1993 erhoben und in den Jahren 2000 bis 2001 überprüft, ergänzt und neu standardisiert. Für die Klassenstufen 1 bis 4 umfassen die Stichproben zwischen 1200 und 1800 Kinder, bei der HSP 5-9 liegt die Stichprobengröße bei 1400 bis 1800 Kindern. Normen liegen für Mitte Klasse 1, Ende Klasse 1 und Mitte Klasse 2 (HSP 1+), Ende Klasse 2 (HSP 2), Mitte Klasse 3 und Ende Klasse 3 (HSP 3), Mitte Klasse 4, Ende Klasse 4 und Anfang Klasse 5 (HSP 4/5) sowie Ende Klasse 5, Ende Klasse 6, Ende Klasse 7, Ende Klasse 8 und Ende Klasse 9 (HSP 5-9) vor.

Zur Überprüfung der Reliabilität der HSP wurde die interne Konsistenz bestimmt. Die Zuverlässigkeit für die wortbezogene Auswertung (Anzahl richtig geschriebener Wörter) liegt zwischen .80 und .96, die Auswertung auf Graphemebene erreichte mit Werten von .92 bis .99 sehr hohe Zuverlässigkeitswerte. Für die Auswertung auf Strategieebene fallen die Reliabilitätskoeffizienten mit Werten zwischen .61 und .94 niedriger aus. Zur Stabilität der HSP-Ergebnisse zeigt sich eine Wiederholungszuverlässigkeit von .81 bis .93 für eine Testwiederholung nach sechs Monaten und von .84 bis .92 für eine Testwiederholung nach 12 Monaten.

Zur Bestimmung der Kriteriumsvalidität wurden Korrelationen der HSP mit der Rechtschreibnote berechnet. Dabei fanden sich für die 1. bis 5. Klasse Korrelationskoeffizienten zwischen .69 und .77, für die 7. Klasse lag der Korrelationskoeffizient immer noch bei .51. Die Höhe der Korrelationen zwischen der HSP und anderen Rechtschreibtests streut für die Gesamtergebnisse zwischen .64 und .78, d. h. dass durchaus davon auszugehen ist, dass die Testverfahren einen gemeinsamen Kern der erfassten Rechtschreibfähigkeiten aufweisen.

Einsatzbereich: Mit der HSP kann die Rechtschreibleistung aller Leistungsniveaus untersucht werden. Eine umfassende Differenzierung ist insbesondere im unteren und untersten Leistungsbereich möglich, wodurch die HSP zur Bestimmung der Förderbedürftigkeit im Rahmen einer LRS-Diagnose eingesetzt werden kann und somit die Grundlage für eine gezielte Förderung von Schülern mit Rechtschreibschwierigkeiten bietet. Weiterhin bietet sie die Möglichkeit sonderpädagogischer Leistungsüberprüfungen und der Planung sonderpädagogischer Fördermaßnahmen. Die HSP ist so konzipiert, dass mit vergleichsweise geringem Aufwand umfassende Informationen über die individuellen Lernstände der Schüler gewonnen werden können. Die HSP kann als Gruppentest oder mit einzelnen Kindern durchgeführt werden.

Ein Leseverständnistest für Erst- bis Sechstklässler (ELFE 1-6)

Testkonzept: Der Leseverständnistest für Erst- bis Sechstklässler von Lenhard und Schneider (2006) ist ein Verfahren zur Erfassung des Leseverständnisses für die 1. bis 6. Klasse. Es prüft sowohl basale Lesestrategien als auch die Fähigkeit zum Verstehen von Sätzen und Texten. Das Leseverständnis wird auf mehreren Ebenen erfasst (Wortverständnistest: 3 Minuten, Satzverständnistest: 3 Minuten, Textverständnistest: 7 Minuten), da die Autoren davon ausgehen, dass Lesen nicht nur die Dekodierung einzelner Wörter erfordert, sondern dass auch die Auseinandersetzung mit den jeweiligen Inhalten erforderlich ist. Zum Verstehen eines Textes sind somit mehrere Verarbeitungsschritte notwendig, die auf unterschiedlichen Komplexitätsniveaus stattfinden: Wortebene, Satzebene, Textebene. Auf der untersten Ebene sind Faktoren von Bedeutung, die das Leseverstehen auf Wortebene beeinflussen: Wortschatz, Dekodierung einzelner Wörter, Erfassung von Wortbedeutung und Zusammenhang von Bedeutung und Kontext. Durch die Interpretation der Wortbedeutung auf der Basis des Kontextes entsteht der Zusammenhang zur Satzebene. Auf Satzebene muss nun die grammatikalische Struktur berücksichtigt werden. Um den ganzen Text verstehen zu können, ist es

dann notwendig, die Informationen verschiedener Sätze miteinander in Beziehung zu setzen. Auf Textebene werden dann metakognitive Fähigkeiten, wie das schemageleitete Textverstehen, immer wichtiger.

Skalen und Untertests: Der ELFE umfasst in der Papier-Bleistift-Version drei Untertests, in der Computerversion kommt noch der Untertest Lesegeschwindigkeit hinzu, mit dem eine Schwellenmessung der Worterkennungsgeschwindigkeit vorgenommen werden kann. Beim Wortverständnistest werden einem Bild vier Wortalternativen gegenübergestellt, die einander möglichst graphemisch und phonetisch ähneln und von denen das Kind die richtige Alternative markieren soll. Der Test umfasst 72 Items, die aus ein- bis viersilbigen Wörtern bestehen. Beim Satzverständnistest wird ein Satz vorgegeben, bei dem es für einen Satzteil (in der Regel ein einziges Wort) mehrere Antwortalternativen gibt. Die Aufgabe besteht darin, jene Alternative zu finden, die in den Satz passt. Die fünf Auswahlalternativen gehören immer derselben Wortart an und sind graphemisch und phonetisch möglichst ähnlich. Der Test umfasst 28 Sätze bei denen das Zielwort ein Substantiv, ein Verb, ein Adjektiv, eine Konjunktion oder eine Präposition sein kann. Beim Textverständnistest wird dem Kind ein Text vorgelegt und eine Frage dazu gestellt. Die Aufgabe besteht darin, unter vier möglichen Antworten die richtige Alternative herauszufinden. Dabei variiert die Komplexität des Textverständnisses vom Auffinden isolierter Informationen bis zum satzübergreifenden Lesen.

Testkennwerte: Der ELFE wurde an 4893 Kindern aus 12 Bundesländern der BRD und Südtirol normiert. Es liegen Normen für die 1. Klasse (Schuljahresende) und für die 2. bis 6. Klasse (Schuljahresmitte und Schuljahresende) vor. Neben einer Gesamtnorm finden sich auch Normen für die drei Untertests. Zur Überprüfung der Reliabilität wurden verschiedene Maße ermittelt. Die Odd-Even-Split-Half-Reliabilität für die Untertests liegt zwischen .95 (Wortverständnis, Satzverständnis) und .89 (Textverständnis). Die Retest-Reliabilität, die im Abstand von zwei Wochen erfasst wurde, fiel mit Werten zwischen .80 und .86 etwas geringer aus. Zur Validität finden sich Angaben zur konvergenten Validität, wobei der ELFE eine hohe Übereinstimmung mit der Würzburger Leise-Leseprobe (.71) und dem Lehrerurteil für die Fertigkeiten Lesen (.70) aufweist. Die Übereinstimmung mit Knuspels Lesetest fielen geringer aus (.45). Die Koeffizienten zur diskriminanten Validität lagen im Vergleich mit dem Lehrerurteil Rechtschreibung bei .46 und mit dem Lehrerurteil Rechnen bei .49.

Einsatzbereich: Der ELFE kann für eine möglichst frühzeitige Diagnostik im Bereich Leserechtschreibstörung eingesetzt werden. Er eignet sich für die Anwendung in der medizini-

schen und psychologischen Befunderhebung sowie im schulpsychologischen Bereich. Weiterhin kann er zur Überprüfung der Effektivität von Unterrichtskonzepten und Interventionsmaßnahmen eingesetzt werden. Aufgrund seiner, insbesondere in der computergestützten Version, hochstandardisierten und ökonomischen Durchführung, ist der ELFE für die Testung großer Stichproben und zur informellen Prüfung durch weniger geschultes Testpersonal geeignet.

Salzburger Lesescreening (SLS)

Testkonzept: Das Salzburger Lese-Screening für die Klassenstufen 1 bis 4 (SLS 1-4) von Mayringer & Wimmer (2003) und für die Klassenstufen 5 – 8 (SLS 5-8) von Auer, Gruber, Mayringer & Wimmer (2005) dient der Erfassung der basalen Lesefertigkeiten (Durchführungsdauer: ca. 5 Minuten). Darunter wird das fehlerfreie, schnelle und mühelose Lesen verstanden. Es geht nicht um Textverständnis, das an Sprachkompetenz und Wissen gebunden ist, sondern um den technischen Aspekt des Lesens. Die zentrale Komponente dieser Lesefertigkeit ist das Wortlesen, da ohne fehlerfreies und müheloses Lesen der Wörter, die weitere inhaltliche Verarbeitung eines Textes beeinträchtigt ist. Eine einfach zu messende Größe, die Auskunft über die technische Lesefertigkeit gibt, ist die Lesegeschwindigkeit. Das SLS erhebt die Lesegeschwindigkeit über das Lesen von Sätzen anhand der innerhalb von drei Minuten korrekt bearbeiteten Sätze. Das Verfahren liegt in zwei Versionen, mit unterschiedlichen, hinsichtlich der Leseanforderungen vergleichbaren Sätzen vor (Satzversionen). Jede der beiden Satzversionen gibt es in zwei Varianten mit unterschiedlicher Reihenfolge der Sätze (Abfolgevarianten).

Skalen und Untertests: Das SLS besteht aus einer Liste inhaltlich sehr einfacher Sätze, die entweder wahre oder nicht wahre Aussagen wiedergeben. Am Ende jedes Satzes kann anhand von zwei Zeichen für richtig und falsch markiert werden, ob die Aussage des Satzes wahr oder nicht wahr ist. Die Testpersonen müssen diese Liste möglichst schnell lesen und den Wahrheitsgehalt jedes Satzes beurteilen.

Testkennwerte: Die Normierung des SLS erfolgte an Stichproben von 700 bis 850 Kindern für jede der Schulstufen. Die Testungen wurden in Bayern und dem Bundesland Salzburg vorgenommen. Es liegen Normen für die Klassenstufen 1 bis 8 vor, wobei sowohl Gesamtnormen als auch separate Normen für Hauptschule, Realschule und Gymnasium angegeben

sind. Die Reliabilität des Verfahrens (Paralleltest-Reliabilität) beträgt .89 und ist somit als zufriedenstellend zu bezeichnen. Die Validität des Testverfahrens, gemessen am lauten Lesen von Texten, liegt bei .78.

Einsatzbereich: Die wichtigste Anwendungsmöglichkeit des SLS besteht darin, in besonders ökonomischer Weise die Lesefertigkeit von Schülerinnen und Schülern einer Klasse zu erfassen. Aufgrund der Differenzierungseigenschaften des SLS können Rückstände oder Vorsprünge der Lesefertigkeit ermittelt werden. Für die Diagnose einer Leseschwäche ist das Screening alleine jedoch nicht ausreichend. Für Forschungszwecke können die Testrohwerte als ein valides und verlässliches Maß der basalen Lesefertigkeit herangezogen werden. Unter dem Gesichtspunkt der Ökonomie sind die kurze Bearbeitungsdauer und die Möglichkeit des Einsatzes als Gruppentest für die gesamte Klasse hervorzuheben.

2.4 Messung der mathematischen Kompetenzen im Schulalter

Zur Messung der Rechenkompetenzen im Schulalter ist der ZAREKI ein häufig eingesetztes und gut erprobtes Verfahren. Mit dem HRT steht nun ein neues, sehr umfassendes und aussagekräftiges Verfahren zur Verfügung, das immer weitere Verbreitung findet. Daher soll auf diese beiden Verfahren näher eingegangen werden (Tabelle 16).

Tabelle 16
Rechentests

Heidelberger Rechentest	HRT	1. – 4. Klasse	Mathematische Basiskompetenzen
Neuropsychologische Testbatterie für Zahlenverarbeitung und Rechnen bei Kindern	ZAREKI	7;6 – 10;11	Rechnerische Fertigkeiten und Zahlenverarbeitung

Heidelberger Rechentest (HRT 1 - 4)

Testkonzept: Der Heidelberger Rechentest für 1. bis 4. Klassen (HRT 1-4; Haffner, Baro, Parzer & Resch, 2005) dient der Erfassung mathematischer Basiskompetenzen im Grundschulalter (Dauer als Einzeltest 40- 45 Min., in Schulklassen ca. 50 - 60 Min.). Unter den theoretischen Grundlagen des Verfahrens dominiert das „Triple-Code-Modell“ von Dehaene (1992) als empirisch gestütztes, anatomisch funktionales Modell der Zahlenverarbeitung.

Danach erfordern komplexe Rechenaufgaben die Zusammenarbeit linkshemisphärischer Sprachfunktionen mit beidhemisphärisch parietalen Hirnregionen zur Verarbeitung und Integration räumlich-visueller und sprachlich-semantischer Information. Zahlen werden in drei Modulen in jeweils unterschiedlichen Kodierungen repräsentiert: der visuell-arabischen Repräsentation (Arabische Zahlen), der auditiv-sprachlichen Repräsentation (Zahlwörter) sowie der analogen Repräsentation der Mächtigkeit von Mengen (quantitativ numerische Größenrepräsentation). Weiterhin lehnt sich das Testkonzept an der Diagnosekategorie einer Rechenstörung (Dyskalkulie) nach ICD-10 (Dilling et al., 1997) an.

Skalen und Untertests: Der HRT enthält 11 Untertests sowie eine Kontrollvariable zur Erfassung der Schreibgeschwindigkeit (visuomotorisches Arbeitstempo). Es werden zwei inhaltliche Schwerpunkte geprüft. Der erste Bereich *Rechenoperationen* besteht aus sechs Untertests: (1) *Addition*, (2) *Subtraktion*, (3) *Multiplikation*, (4) *Division*, (5) *Ergänzungsaufgaben*, (6) *Größer-Kleiner-Vergleiche* (siehe Tabelle 17). Der zweite Bereich *Numerisch-logische und räumlich-visuelle Fähigkeiten* enthält fünf Untertests: (7) *Zahlenreihen*, (8) *Längenschätzen*, (9) *Würfelzählen*, (10) *Mengenzählen* und (11) *Zahlenverbinden*. Das Verfahren gibt drei Skalenwerte aus für 1. *Rechenoperationen*, 2. *räumlich-visuelle Leistung* und 3. *Gesamtleistung* aller elf Untertests.

Tabelle 17
HRT 1-4

Untertests	Aufgabe	Inhalte
Schreibgeschwindigkeit	Zahlen unter Tempobedingungen abschreiben	Visuomotorisches Arbeitstempo
Rechenoperationen		
(1) Addition	Schriftlich vorgegebene Additionsaufgaben	Grundrechnen-Plus
(2) Subtraktion	Schriftlich vorgegebene Subtraktionsaufgaben	Grundrechnen-Minus
(3) Multiplikation	Schriftlich vorgegebene Multiplikationsaufgaben	Grundrechnen-Mal
(4) Division	Schriftlich vorgegebene Divisionsaufgaben	Grundrechnen-Geteilt
(5) Ergänzungsaufgaben	Fehlende Zahlen in Gleichungen ergänzen	Rechenleistung bei variablen Gleichungs-Aufgaben
(6) Größer-Kleiner-Vergleiche	Größenvergleich von Ziffern oder Rechenergebnissen	Größenvergleiche, Überblicksrechnen, Ungleichungen

Fortsetzung nächste Seite

Fortsetzung Tabelle 17

Numerisch-logische und räumlich-visuelle Fähigkeiten		
(7) Zahlenfolgen	Zahlenreihen fortsetzen	mathematisch logisches Denken, Erkennen von Regeln
(8) Längenschätzen	Die Länge von Linien schätzen	visuelle Größenerfassung
(9) Würfelzählen	Bei aus Würfeln zusammengesetzten Figuren die Anzahl der Würfелеlemente bestimmen	Mengenerfassung unter Berücksichtigung räumlicher Vorstellung
(10) Mengenzählen	Figuren in einem Kästchen unter Tempobedingungen abzählen	Zählgeschwindigkeit, Mengenstrukturierung
(11) Zahlenverbinden	Landschaft der Zahlen von 1 bis 20 so schnell wie möglich verbinden	Wahrnehmungsgeschwindigkeit, Visuomotorik

Testkennwerte: Für den HRT liegen Normen für vier Quartale pro Schuljahr ab Ende 1. Klasse bis Ende 4. Klasse vor. Die Eichstichprobe umfasst $N = 3075$ Grundschul Kinder aus vier Bundesländern, $N = 159$ Kinder aus Sprachheilschulen und $N = 120$ Kinder aus Förderschulen. Die Retest-Reliabilität (1-2 Wochen, $N = 246$) liegt für die elf Untertests zwischen .69 und .89, für die drei Skalenwerte zwischen .87 und .93, und ist damit als zufriedenstellend zu bezeichnen. Die inhaltliche Validität ist für die Aufgaben zur Beherrschung der Grundrechenarten gegeben, die Kriteriumsvalidität des HRT ist durch die Korrelationen mit der Mathematiknote ($r = -.67$) und mit dem Mathematiktest DEMAT 4 ($r = .72$) gegeben. Die Struktur der beiden Inhaltsbereiche konnte faktorenanalytisch bestätigt werden.

Einsatzbereich: Der HRT misst mathematische Basiskompetenzen im gesamten Leistungsbereich und differenziert daher rechenschwache Kinder sowie Kinder mit besonders guter mathematischer Begabung. Das Testprofil gibt dabei Hinweise auf Förderbedarf und gezielte Interventionsmaßnahmen. Das Verfahren ist als Einzeltest zur Individualdiagnostik (Dyskalkulie, mathematische Begabung) und zur Verlaufsmessung bei Therapiemaßnahmen geeignet. Als Gruppentest dient das Instrument der Feststellung des Leistungsstands einzelner SchülerInnen sowie ganzer Schulklassen. Da der HRT mathematische Kompetenzen weitgehend sprach- und lehrplanunabhängig erfasst, eignet er sich auch für internationale Vergleichs- und Grundlagenforschung.

Neuropsychologische Testbatterie für Zahlenverarbeitung und Rechnen bei Kindern (ZAREKI)

Testkonzept: Die neuropsychologische Testbatterie für Zahlenverarbeitung und Rechnen bei Kindern (ZAREKI; von Aster & Weinhold, 2002) dient der Erfassung rechnerischer Fertigkeiten und Zahlenverarbeitung bei Grundschulkindern der zweiten bis vierten Klasse (Dauer 15 bis 30 Minuten). Als theoretische Grundlagen gingen Erkenntnisse der modernen kognitiven Entwicklungsneuropsychologie zu lateralisierten visuell-räumlichen und sprachlichen Basisfunktionen, Dehaenes (1992) „Triple-Code-Model“ zur modularen Architektur der kognitiven Zahlenverarbeitung, Andersons (1992) Theorie der „Minimalen Kognitiven Architektur“ (adaptiert für Störungen der Zahlenverarbeitung und des Rechnens) sowie das Konzept der Akalkuliebatterie für Erwachsene EC 301 (Deloche, 1995) in die Testentwicklung ein.

Skalen und Untertests: Die ZAREKI besteht aus 11 Untertests; eine genaue Beschreibung findet sich in Tabelle 18. Neben einem Wert für die Gesamttleistung gibt die ZAREKI drei Indexwerte für 1. kulturvermitteltes Zahlenwissen (UT 3, 5, 7, 9, 10, 11), 2. Rechnen (UT 2, 4) und 3. visuell-analoge Zahlenrepräsentanz (UT 6, 8) aus.

Tabelle 18
ZAREKI

Untertests	Aufgabe	Inhalte
(1) Abzählen	Abzählen von schwarzen Punkten mit dem Finger	
(2) Zählen rückwärts mündlich	Rückwärtszählen von 23 bis 1	Rechnen
(3) Zahlenschreiben	gesprochene Zahlen in ihre arabische Ziffernform kodieren	kulturvermitteltes Zahlenwissen
(4) Kopfrechnen	Additionen und Subtraktionen	Rechnen
(5) Zahlenlesen	arabische Ziffern laut lesen	kulturvermitteltes Zahlenwissen
(6) Anordnen von Zahlen auf einem Zahlenstrahl	auf einer vertikalen Linie von 0 bis 100 einer arabischen Zahl die passende von vier Markierungen zuordnen	visuell-analoge Zahlenrepräsentanz
(7) Zahlenvergleich (Worte)	die größere von zwei gesprochenen Zahlen durch eine motorische Antwortreaktion bestimmen	kulturvermitteltes Zahlenwissen
(8) Perzeptive Mengenbeurteilung	Schätzen der Anzahl von Gegenständen auf einer Vorlage	visuell-analoge Zahlenrepräsentanz

Fortsetzung Tabelle nächste Seite

Fortsetzung Tabelle 18

(9) Kognitive (kontextuelle) Mengenbeurteilung	eine durch ein Zahlwort benannte Menge im Hinblick auf einen spezifischen situativen Kontext als viel, mittel oder wenig beurteilen	kulturvermitteltes Zahlenwissen
(10) Textaufgaben	Anwenden mathematischer Prinzipien auf situative Modelle, Lösen von Additions- und Subtraktionsproblemen	kulturvermitteltes Zahlenwissen
(11) Zahlenvergleich (Ziffern)	die größere von zwei geschriebenen Zahlen markieren	kulturvermitteltes Zahlenwissen

Testkennwerte: Die Normstichprobe der ZAREKI umfasst $N = 238$ Schweizer Zweit-, Dritt- und Viertklässler in drei Altersgruppen von 7;6 bis 10;11 Jahren. Weitere $N = 40$ Kinder desselben Alters mit Rechenstörungen aus kinderpsychiatrischer Inanspruchnahmepopulationen bilden die klinische Stichprobe. Die Reliabilität des Gesamttests (interne Konsistenz) ist mit .89 bei der Eichstichprobe und .90 bei der klinischen Stichprobe für einen Leistungstest zufriedenstellend. Eine zur Prüfung der Konstruktvalidität durchgeführte Faktorenanalyse bestätigte die vor dem Hintergrund der entwicklungsneuropsychologischen Erkenntnisse erwartete Faktorenstruktur. Die Kriteriumsvalidität ist durch Übereinstimmung der ZAREKI-Testleistungen mit dem Lehrerurteil sowie mit den ICD-10 Kriterien (Dilling et al., 1997) einer Rechenstörung (Dyskalkulie) gegeben.

Einsatzbereich: Die ZAREKI ist als Individualverfahren zur Messung und Analyse zahlenverarbeitender Funktionen für sieben- bis zehnjährige Grundschul Kinder geeignet. Während das Verfahren zur Feststellung einer Dyskalkulie herangezogen werden kann, weist das Handbuch ausdrücklich darauf hin, dass die Diagnose von Lernstörungen in Mathematik immer mehrdimensional erfolgen und die störungsspezifische Entwicklungsgeschichte, Untersuchungen der Intelligenz und Sprachverarbeitung, evtl. vorliegende komorbide Störungen sowie schulische und familiäre Kontextfaktoren berücksichtigen sollten. Bezüglich der Diagnose deutscher Kinder ist zu berücksichtigen, dass die Normierung der ZAREKI an eher kleinen Stichproben in der Schweiz vorgenommen wurde.

3 Empfehlungen zur Umsetzung geeigneter Testverfahren im Rahmen des SOEP

Im Folgenden werden verschiedene Testverfahren für die Anwendung im Rahmen des SOEP empfohlen, die der Zielsetzung der Erfassung eines Maßes der *allgemeinen kognitiven Leistungsfähigkeit* (fluide Kompetenz) sowie der Erfassung *spezifischer kognitiver Fertigkeiten* (kristalline Kompetenzen) entgegenkommen und Rahmenbedingungen des SOEP vor allem hinsichtlich des erforderlichen Aufwands zur Durchführung der Verfahren berücksichtigen.

3.1 Erfassung eines Maßes der allgemeinen kognitiven Leistungsfähigkeit

Zur Erfassung eines Maßes für die allgemeine kognitive Leistungsfähigkeit soll im Folgenden ein Verfahren empfohlen werden, welches in *überschaubarer Zeit* (max. 30 Minuten), mit *einfachen Materialien*, nach *kurzer Einarbeitungsphase* der Testleiter, für den *gesamten Altersrange* der geplanten SOEP-Studie angewendet werden kann. Eine umfassende, individuelle Intelligenzdiagnostik ist im Rahmen der SOEP-Erhebung kaum zu ermöglichen, da die hierfür geeigneten Verfahren (z.B. K-ABC, HAWIK III, AID 2) einen sehr hohen Aufwand bezüglich der Erhebungsdauer, der verwendeten Materialien und vor allem hinsichtlich der fundierten psychologisch-diagnostischen Ausbildung der Testleiter erfordern. Hingegen liegt mit dem **Grundintelligenztest** (CFT1, Weiß & Osterland, 1997; CFT 20-R, Weiß, 2006) ein Verfahren vor, mit welchem sich in ökonomischer Weise ein Maß für die Grundintelligenz („general mental capacity“, „g“-Faktor) eines Kindes erfassen lässt. Da das Verfahren auf sprachfreien, figuralen Material basiert, kann das allgemeine Intelligenzpotential (fluid ability) weitgehend frei von Einflüssen des soziokulturellen, erziehungsspezifischen oder rassischen Hintergrundes gemessen werden.

Für das *Vorschulalter bis zur dritten Grundschulklasse* wird die Durchführung des **Grundintelligenztests – Skala 1** (CFT1, Weiß & Osterland, 1997: 5;3 – 9;5 Jahre) empfohlen; für Kinder *ab der vierten Klasse bis zum Jugendalter* der **Grundintelligenztest – Skala 2** (CFT 20-R, Weiß, 2006: 8;5 bis 19;11 Jahre). Beide Verfahren sind Paper-Pencil-Tests und erfordern keine weiteren Einzelmaterialien für die verschiedenen Untertests. Nach überschaubarer Einarbeitungszeit des Testleiters können beide Tests als Einzel- und als Gruppenverfahren durchgeführt werden; die Testdauer beträgt in der Einzeldurchführung jeweils 25-30 Minuten.

Beide Verfahren sind international verbreitet und finden sowohl in Forschung als auch in Praxis vielfach Anwendung. Die Validität beider Verfahren konnte faktorenanalytisch bestätigt und durch Zusammenhänge mit bewährten Intelligenztestverfahren (HAWIK, PSB) nachgewiesen werden.

Für den **CFT 1** wird die Durchführung des Gesamttests (5 Untertests, ca. 25 Min.) empfohlen. Möchte man die Testzeit noch weiter verkürzen, ist auch eine Durchführung von nur drei Untertests denkbar: (3) Klassifikationen, (4) Ähnlichkeiten und (5) Matrizen (Durchführung ca. 20 Min.). Im Gegensatz zu den Untertests 1 und 2 (Substitutionen und Labyrinth), die hauptsächlich Wahrnehmungsfähigkeiten erfassen (optische Wahrnehmung, Wahrnehmungsgeschwindigkeit, Auffassung bei einfach strukturierten Problemstellungen), laden die Untertests 3 bis 5 höher auf einem separaten Fähigkeitsfaktor („g-Sättigung“) und messen daher am stärksten die grundlegenden intellektuellen Fähigkeiten (sprachfreie Denkkapazität).

Der **CFT 20-R** besteht aus zwei Testteilen (Dauer etwa 1h), von denen jedoch bei bestimmten Diagnoseanlässen die Durchführung von Teil 1 (Kurzform) ausreicht. Für das SOEP wird hinsichtlich der knappen zur Verfügung stehenden Erhebungszeit die Durchführung der Kurzform des CFT 20-R (Dauer 25-30 Min) empfohlen (siehe Tabelle 19). Zum Erreichen einer hohen Reliabilität und Validität sollten allerdings beide Testteile durchgeführt werden.

Tabelle 19
Erfassung allgemeiner kognitiver Leistungsfähigkeit

Altersrange	Untertests	Dauer bei Einzeltestung
CFT 1 5;3 – 9;5	Substitutionen Labyrinth Klassifikationen Ähnlichkeiten Matrizen	Gesamt: ca. 25 Min. Untertests 3-5: ca. 20 Min.
CFT 20-R 8;5 – 19;11	Reihenfortsetzen Klassifikationen Matrizen Topologien	Gesamt: ca. 60 Min. Teil 1: ca. 25-30 Min.

Die Testung von Kindern *unter fünf Jahren* stellt aufgrund des Entwicklungsstands (kognitiv, motorisch, sozial, motivational) immer eine Herausforderung dar, die sehr kompetente und mit dieser Altersgruppe erfahrene Testleiter erfordert. Soll auch bei Kindern im Alter von vier

Jahren oder jünger ein Maß der allgemeinen kognitiven Leistungsfähigkeit erfasst werden, werden die **Coloured Progressive Matrices (CPM)** von Raven empfohlen (deutsche Normen von 3;9 bis 11;8 Jahren; Bulheller & Häcker, 2002), ein Paper-Pencil-Verfahren, das unkompliziert anzuwenden ist (Dauer 20-30 Minuten).

3.2 Erfassung spezifischer kognitiver Fertigkeiten

Zur Erfassung spezifischer kognitiver Fertigkeiten werden im Folgenden verschiedene Verfahren zur Messung von *sprachlichen, Lese-, Rechtschreib-* sowie *mathematischen Kompetenzen* empfohlen (vgl. Tabellen 20 bis 22).

3.2.1 Sprachliche Kompetenz

Die Erfassung sprachlicher Kompetenzen sollte für das Vorschulalter, d. h. für den Altersbereich zwischen drei und sechs Jahren vorgesehen werden. Dafür empfohlen wird der **Sprachentwicklungstest (SETK)** für drei- bis fünfjährige Kinder, der für den Altersbereich 3;0 bis 5;11 eingesetzt werden kann. Der Gesamttest, der je nach Altersgruppe fünf bis sechs Untertests umfasst, dauert 25 bis 30 Minuten. Über die einzelnen Untertests werden die Bereiche Sprachverständnis (1 Untertest), Sprachproduktion (1 bzw. 2 Untertests) und Sprachgedächtnis (3 Untertests) abgedeckt. Um als Screening eine grobe Einschätzung der sprachlichen Fähigkeiten vorzunehmen, könnte die Durchführung auf die Untertests zum Sprachverständnis und zur Sprachproduktion beschränkt werden. Diese Untertests könnten mit einem zeitlichen Aufwand von 10 bis 15 Minuten durchgeführt werden (vgl. Tabelle 20).

Tabelle 20
Erfassung von Sprachfertigkeiten

Altersrange	Untertests	Dauer bei Einzeltestung
SETK 3;0 – 5;11	(1) Verstehen von Sätzen	Gesamt: 20-30 Min.
	(2) Enkodieren semantischer Relationen	
	(3) Morphologische Regelbildung	Untertests 1-3: 10-15 Min.
	(4) Phonologisches Arbeitsgedächtnis	
	(5) Gedächtnisspanne für Wortfolgen	
	(6) Satzgedächtnis	

3.2.2 Lesefertigkeiten

Ab dem Grundschulalter besteht die Möglichkeit, die Lesefertigkeiten der Kinder zu erfassen. Dazu empfehlen wir das **Salzburger Lesescreening (SLS)**, welches mit seinen zwei Versionen SLS 1-4 und SLS 5-8 die Klassenstufen 1 bis 8 abdeckt. Somit können die Lesefertigkeiten über einen weiten Teil des Kindes- und Jugendalters mit demselben Erhebungsinstrument erfasst werden. Das SLS ist schnell und unkompliziert durchzuführen, da es sich um einen Paper-Pencil-Test handelt und keine ausführlichen Testanweisungen notwendig sind; benötigt werden nur ca. 5 Minuten Durchführungsdauer (vgl. Tabelle 21).

3.2.3 Rechtschreibfertigkeiten

Die Rechtschreibfertigkeiten können ebenfalls ab dem Grundschulalter erhoben werden. Für diesen Bereich schlagen wir die **Hamburger Schreibprobe (HSP)** vor, bei der es ebenfalls verschiedene Versionen gibt (HSP 1+, HSP 2, HSP 3, HSP 4/5, HSP 5-9), die es ermöglichen, mit demselben Verfahren die Klassenstufen 1 bis 9 abzudecken. Die HSP ist in verhältnismäßig kurzer Zeit durchzuführen, benötigt aber dennoch 20 bis 30 Minuten Zeit (vgl. Tabelle 21). Da es hier keine einzelnen Untertests gibt, besteht nicht die Möglichkeit, nur bestimmte Testteile durchzuführen. Sollte hier eine weniger zeitaufwendige Lösung notwendig sein, könnte in Betracht gezogen werden, die **Deutschnote** zu erfassen, da die HSP substantiell mit der Schulnote im Fach Deutsch zusammenhängt (.51-.77).

Tabelle 21
Erfassung von Lese- und Rechtschreibkompetenzen

	Altersrange	Dauer bei Einzeltestung
SLS	1. – 8. Klasse	ca. 5 Min.
HSP	1. – 9. Klasse	20-30 Min.
Deutschnote	3. – 13. Klasse	0 Min.

3.2.4 Mathematische Fertigkeiten

Zur Erfassung mathematischer Kompetenzen stehen je nach zur Verfügung stehender Zeit und Diagnoseanspruch verschiedene Verfahren zur Verfügung (siehe Tabelle 22). Für Grundschulkindern wird der **Heidelberger Rechentest (HRT 1-4)** für *1. bis 4. Klassen* (Haffner,

Baro, Parzer & Resch, 2005) empfohlen. Dieser Paper-Pencil-Test dient der Erfassung mathematischer Basiskompetenzen und kann als Einzeltest in ca. 45 Min. Dauer durchgeführt werden (Durchführung nur der sechs Untertests des Bereichs Rechenoperationen in 20-25 Min.). Das Ergebnis im Gesamttest korreliert hoch mit der Schulnote in Mathematik (-.67).

Die **neuropsychologische Testbatterie für Zahlenverarbeitung und Rechnen bei Kindern (ZAREKI)**; von Aster & Weinhold, 2002) für Kinder *der zweiten bis vierten Grundschulklasse* hat eine kürzere Erhebungsdauer (15 bis 30 Minuten), ist jedoch nur mit Vorsicht zum Einsatz bei deutschen Kindern zu empfehlen, da dieses Verfahren nur an einer relativ kleinen Stichprobe von 238 Schweizer Kindern normiert worden ist.

Für Kinder der *4. bis 13. Klasse* kann in nur etwa 5 Minuten der neunte **Untertest „Zahlenaddition“** des weit verbreiteten **Prüfsystems für Schul- und Bildungsberatung (P-S-B)** (PSB-R 4-6 von Horn, Lukesch, Kormann und Mayrhofer, 2002; PSB-R 6-13 von Horn, Lukesch, Mayrhofer und Kormann, 2003) angewandt werden. Dieser Untertest korreliert in substantieller Höhe (.43-.50) mit der Mathematiknote. Die Aufgabe misst das Beherrschen einfacher Additionsaufgaben unter Tempobedingungen und erfasst damit auch die Konzentrationsfähigkeit.

Da alle genannten Verfahren substantiell mit der Schulnote im Fach Mathematik zusammenhängen, wird als ein Maß für mathematische Fertigkeiten, welches fast keinen Zeit- und Durchführungsaufwand erfordert, die Erfassung der **Mathematiknote** empfohlen.

Tabelle 22

Erfassung mathematischer Kompetenzen

	Altersrange	Dauer bei Einzeltestung
HRT 1-4	1. – 4. Klasse	ca. 45 Min.
ZAREKI	2. – 4. Klasse	15-30 Min.
Zahlenaddition aus dem P-S-B	4. – 13. Klasse	5 Min.
Mathematiknote	3. – 13. Klasse	0 Min.

Alle in diesem Abschnitt empfohlenen Verfahren sind aufgrund ihrer unkomplizierten Durchführung und Materialien im Rahmen einer Panel-Studie wie dem SOEP anwendbar. Abhängig von der zur Verfügung stehenden Zeit und den zu erfassenden Zielkompetenzen können ein Verfahren zur Erfassung der allgemeinen geistigen Leistungsfähigkeit und/oder Verfahren zur

Erfassung spezifischer Fertigkeiten ausgewählt und eingesetzt werden. Zur Erfassung der Grundintelligenz sollten dabei etwa 30 Minuten eingeplant werden, zur Erfassung spezifischer Teilfertigkeiten, je nach Altersgruppe und untersuchter Leistung, zusätzlich 15 bis 45 Minuten.

Literatur

- Amelang, M. (1995). Intelligenz. In M. Amelang (Hrsg.), *Enzyklopädie der Psychologie. Differentielle Psychologie und Persönlichkeitsforschung*, Bd. 2: Verhaltens- und Leistungsunterschiede (S. 245–328). Göttingen: Hogrefe.
- Amelang, M. & Bartussek, D. (2001). *Differentielle Psychologie und Persönlichkeitsforschung* (5. Auflage). Stuttgart: Kohlhammer.
- Amthauer, R., Brocke, B., Liepmann, D. & Beauducel, A. (2001). *Intelligenz-Struktur-Test 2000 R (I-S-T 2000 R)*. Göttingen: Hogrefe.
- Anderson, M. (1992). *Intelligence and Development. A Cognitive Theory*. Blackwell, Oxford.
- Asendorpf, J. B. (1996). *Psychologie der Persönlichkeit: Grundlagen*. Berlin: Springer.
- Auer, M., Gruber, G., Mayringer, H. & Wimmer, H. (2005). *Salzburger Lese-Screening für die Klassenstufen 5-8 (SLS 5-8)*. Manual. Bern: Verlag Hans Huber.
- Bayley, N. (1970). Development of mental abilities. In P. H. Mussen (Ed.), *Carmichael's manual of child psychology*, Vol. 1 (pp. 1163-1209). New York: Wiley.
- Binet, A. & Simon, T. (1905). Méthodes nouvelles pour le diagnostic de niveau intellectuel des anormaux. *Année Psychologique*, 11, 191-244.
- Bulheller, S. & Häcker, H. (2006). *Coloured Progressive Matrices: CPM*. Manual. Dt. Bearbeitung. Frankfurt: Harcourt Test Services.
- Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. Cambridge, MA: Cambridge University Press.
- Cattell, R. B. (1963). Theory of fluid and crystallized intelligence: A critical experiment. *Journal of Educational Psychology*, 54, 1-22.
- Cattell, R. B. (1987). *Intelligence: Its Structure, Growth and Action*. Amsterdam: Elsevier Science Publishers B.V.
- Dahaene, S. (1992). Varieties of numerical abilities. *Cognition*, 44, 1 – 42.
- Deloche, G. (1995). *EC301R: Batterie standardisée d'évaluation du calcul et du traitement des nombres*. Salvador : Editora Sarah Letras.
- Dilling, H., Mombour, W., Schmidt, M. H. & Schulte-Markwort, E. (1997). *Internationale Klassifikation psychischer Störungen ICD-10 Kapitel V (F) Forschungskriterien*. Göttingen: Huber.
- Eggert, D. & Schuch, K.-H. (1975). *Hannover-Wechsler-Intelligenztest für das Vorschulalter HAWI-VA*. Bern: Verlag Hans Huber.
- Fischer, H. (1958). Ein Vergleich zwischen dem IST von Amthauer und dem PMA von Thurstone. *Diagnostica*, 4, 25-32.
- Gottfredson, L. S. (1999). Der Generalfaktor der Intelligenz. *Spektrum der Wissenschaft, Spezial 3/99: Intelligenz*, 24-30.
- Grimm, H. (2001). *Sprachentwicklungstest für drei- bis fünfjährige Kinder (SETK 3-5)*. Manual. Göttingen: Hogrefe Verlag.
- Grimm, H. & Schöler, H. (1978). *Heidelberger Sprachentwicklungstest (H-S-E-T)*. Göttingen: Hogrefe.

- Grund, M., Haug, G. & Naumann, C. L. (2004). Diagnostischer Rechtschreibtest für 4. Klassen (DRT 4). Manual. 2. Auflage. Göttingen: Beltz Test GmbH.
- Grund, M., Haug, G. & Naumann, C. L. (2004). Diagnostischer Rechtschreibtest für 5. Klassen (DRT 5). Manual. 2. Auflage. Göttingen: Beltz Test GmbH.
- Guilford, J. P. (1967). The nature of human intelligence. New York: Mac Graw Hill.
- Häcker, H., Leutner, D. & Amelang, M. (1998). Standards für pädagogisches und psychologisches Testen. Supplementum 1/1998 der Diagnostica und der Zeitschrift für Differentielle und Diagnostische Psychologie.
- Haffner, J., Baro, K., Parzer, P. & Resch, F. (2005). Heidelberger Rechentest (HRT 1-4). Manual. Göttingen: Hogrefe.
- Hasselhorn, M. & Grube, D. (1997). Entwicklung der Intelligenz und des Denkens: Literaturüberblick. In F. E. Weinert & A. Helmke (Hrsg.) (1997). Entwicklung im Grundschulalter (S. 15-26). Weinheim: Psychologie Verlags Union.
- Heller, K., Gaedike, A. K. & Weinländer, H. (1976). Kognitiver Fähigkeits-Test für 4. bis 13. Klassen (KFT 4-13). Weinheim: Beltz Test GmbH.
- Heller, K. & Geisler, H.-J. (1983). Kognitiver Fähigkeits-Test - Kindergartenform (KFT-K). Weinheim: Beltz Test GmbH.
- Heller, K. & Geisler, H.-J. (1983). Kognitiver Fähigkeits-Test für 1. bis 3. Klassen (KFT 1-3). Weinheim: Beltz Test GmbH.
- Heller, K. A., Kratzmeier, H. & Lengfelder, A. (1998). Matrizen-Test-Manual Band 1. Ein Handbuch mit deutschen Normen zu den Standard Progressive Matrices von J. C. Raven. Göttingen: Hogrefe
- Heller, K. A., Kratzmeier, H. & Lengfelder, A. (1998). Matrizen-Test-Manual Band 2. Ein Handbuch mit deutschen Normen zu den Advanced Progressive Matrices von J. C. Raven. Göttingen: Hogrefe
- Heller, K. & Perleth, Ch. (2000). Kognitiver Fähigkeits-Test für 4. bis 12. Klassen (KFT 4-12+ R). Revision. Weinheim: Beltz Test GmbH.
- Hofstätter, P. R. (1971). Differentielle Psychologie. Stuttgart: Kröner.
- Holling, H., Preckel, F. & Vock, M. (2004). Intelligenzdiagnostik. Kompendien Psychologischer Diagnostik, Band 6. Göttingen: Hogrefe Verlag.
- Horn, J. L. & Cattell, R. B. (1966). Refinement and test of the theory of fluid and crystallized ability intelligence. *Journal of educational psychology*, 57, 253-270.
- Horn, Lukesch, Kormann & Mayrhofer (2002). Prüfungssystem für Schul- und Bildungsberatung für 4. bis 6. Klassen – Revidierte Fassung (PSB-R 4-6). Göttingen: Hogrefe.
- Horn, Lukesch, Mayrhofer & Kormann (2003). Prüfungssystem für Schul- und Bildungsberatung für 6. bis 13. Klassen – Revidierte Fassung (PSB-R 6-13). Göttingen: Hogrefe.
- Jäger, A. O. (1984). Intelligenzforschung: Konkurrierende Modelle, neue Entwicklungen, Perspektiven. *Psychologische Rundschau*, 35, 21-35.
- Jäger, A. O., Süß, H.-M. & Beauducel, A. (1997). Berliner Intelligenzstrukturtest – Form 4 (BIS-4). Göttingen: Hogrefe.
- Jäger, A. O., Holling, H., Preckel, F., Schulze, R., Vock, M., Süß, H.-M. & Beauducel, A. (2006). Berliner Intelligenzstrukturtest für Begabungs- und Hochbegabungsdiagnostik (BIS-HB). Göttingen: Hogrefe.

- Kaufman, A. S. & Kaufman, N. L. (1983a). Kaufman-Assessment Battery for Children. (K-ABC) Administration and Scoring Manual. Circle Pines, MN: American Guidance Service.
- Kaufman, A. S. & Kaufman, N. L. (1983b). Kaufman-Assessment Battery for Children. (K-ABC) Interpretive Manual. Circle Pines, MN: American Guidance Service.
- Kaufman, A.S. & Kaufman, N.L. (1993). Kaufman Adolescent and Adult Intelligence Test (KAIT). Circle Pines, MN: American Guidance Service.
- Klicpera, C. & Gasteiger-Klicpera, B. (1995). Psychologie der Lese- und Schreiechwierigkeiten: Entwicklung, Ursachen, Förderung. Weinheim: Beltz.
- Kubinger, K. D. & Wurst, E. (2000). Adaptives Intelligenz Diagnostikum 2 (AID 2). Göttingen: Hogrefe.
- Landerl, K., Wimmer, H. & Moser, E. (1997). Salzburger Lese- und Rechtschreibtest (SLRT). Handbuch. Bern: Hans Huber Verlag.
- Lenhard, W. & Schneider, W. (2006). Ein Leseverständnistest für Erst- bis Sechstklässler (ELFE 1-6). Manual. Göttingen: Hogrefe.
- May, P. (2002). Hamburger Schreibprobe (HSP 1-9). Manual. 6. Auflage. Hamburg: verlag für pädagogische medien.
- Mayringer, H. & Wimmer, H. (2003). Salzburger Lese-Screening für die Klassenstufen 1-4 (SLS 1-4). Manual. Bern: Hans Huber Verlag.
- Melchers, P. & Preuß, U. (2003a). K-ABC: Kaufman-Assessment Battery for Children. Durchführungs- und Auswertungshandbuch. Deutschsprachige Fassung, 6. Auflage. Frankfurt am Main: Swets & Zeitlinger.
- Melchers, P. & Preuß, U. (2003b). K-ABC: Kaufman-Assessment Battery for Children. Interpretationshandbuch. Deutschsprachige Fassung, 6. Auflage. Frankfurt am Main: Swets & Zeitlinger.
- Melchers, P., Schürmann, S. & Scholten, S. (2006). Kaufman - Test zur Intelligenzmessung für Jugendliche und Erwachsene: K-TIM. Handbuch. Leiden: PITS Psychologische Instrumenten, Testen Services.
- Müller, R. (2004). Diagnostischer Rechtschreibtest für 1. Klassen (DRT 1). Manual. 2. Auflage. Göttingen: Beltz Test GmbH.
- Müller, R. (2004). Diagnostischer Rechtschreibtest für 2. Klassen (DRT 2). Manual. 4. Auflage. Göttingen: Beltz Test GmbH.
- Müller, R. (2004). Diagnostischer Rechtschreibtest für 3. Klassen (DRT 3). Manual. 4. Auflage. Göttingen: Beltz Test GmbH.
- Raven, J. C. (1941). Standardisation of Progressive Matrices, 1938. The British Journal of Medical Psychology, 19, 137-150.
- Raven, J. C., Raven, J. & Court, J. H. (1994). Raven's Progressive Matrices and Vocabulary Scales. Advanced Progressive Matrices. Oxford: Oxford Psychologists Press.
- Raven, J. C., Raven, J. & Court, J. H. (1995). Raven's Progressive Matrices and Vocabulary Scales. Coloured Progressive Matrices. Oxford: Oxford Psychologists Press.
- Raven, J. C., Raven, J. & Court, J. H. (1996). Raven's Progressive Matrices and Vocabulary Scales. Standard Progressive Matrices. Oxford: Oxford Psychologists Press.
- Reynell, J. (1977). Reynell Developmental Language Scales. NFER, Windsor.
- Samstag, K., Sander, A. & Schmidt, R. (1999). Diagnostischer Rechentest für 3. Klassen (DRE 3). Göttingen: Beltz Test GmbH.

- Sarimski, K. (1985). Sprachentwicklungsskalen. München: Röttger.
- Snijders, J. Th., Tellegen, P. J. & Laros, J. A. (1997). Snijders-Oomen Non-verbaler Intelligenztest: SON-R 5½ - 17, 2. Auflage. Manual. Göttingen: Hogrefe.
- Spearman, Ch. (1904). „General intelligence“, objectively determined and measured. *American Journal of Psychology*, 15, 201-293.
- Stapf, A. (1998). Hochbegabung: Was ist das? In Landesinstitut für Erziehung und Unterricht Stuttgart im Auftrag des Ministeriums für Kultus, Jugend und Sport Baden-Württemberg (Hrsg.), *Begabungen finden und fördern* (S. 12-24). Stuttgart: Klett.
- Stern, W. (1911). *Intelligenzproblem und Schule*. Leipzig: Teubner.
- Stern, W. (1920). *Die Intelligenz der Kinder und Jugendlichen*. Leipzig: Barth.
- Sternberg, R. J. (1985). *Beyond IQ: A triarchic theory of human intelligence*. New York: Cambridge University Press.
- Tellegen, P., Winkel, M., Wijnberg-Williams, B. & Laros, J. (1998). Snijders-Oomen Non-verbaler Intelligenztest: SON-R 2½ - 7, 2. Auflage. Manual. Göttingen: Hogrefe.
- Tewes, U., Rossmann, P. & Schallberger, U. (2002). *HAWIK-III: Hamburg-Wechsler-Intelligenztest für Kinder – Dritte Auflage*. Manual. Bern: Verlag Hans Huber.
- Thorndike, R. L. & Hagen, E. (1971). *Cognitive Abilities Test (CAT). Examiner's Manual. Multi-Level Edition*. Boston.
- Thorndike, R. L. & Hagen, E. (1974). *Cognitive Abilities Test (CAT). Technical Manual. Multi-Level Edition*. Boston.
- Thurstone, L. L. (1931). Multiple factor analysis. *Psychological Review*, 38, 406-427.
- Thurstone, L. L. (1938). *Primary mental abilities*. Chicago: University of Chicago Press.
- Vernon, P. E. (1965). Ability factors and environmental influences. *American Psychologist*, 20, 723-733.
- Von Aster, M. (2002). *Neuropsychologische Testbatterie für Zahlenverarbeitung und Rechnen bei Kindern (ZAREKI) – 2. Auflage*. Manual. Frankfurt: Swets & Zeitlinger.
- Von Aster, M., Neubauer, A. & Horn, R. (2006). *Wechsler Intelligenztest für Erwachsene WIE*. Manual. Frankfurt/M: Harcourt Test Services.
- Waldmann, M. & Weinert, F. E. (1990). *Intelligenz und Denken: Perspektiven der Hochbegabtenforschung*. Göttingen: Hogrefe.
- Wechsler, D. (1997). *Wechsler Adult Intelligence Scale WAIS-III – Third Edition*. San Antonio, TX: The Psychological Corporation.
- Wechsler, D. (1991). *Wechsler Adult Intelligence Scale WISC-III – Third Edition*. San Antonio, TX: The Psychological Corporation.
- Wechsler, D. (1964). *Die Messung der Intelligenz Erwachsener* (3. Aufl.). Bern: Huber.
- Wechsler, D. (1939). *The measurement of adult intelligence*. Baltimore: Williams & Wilkins.
- Wenzl, A. (1957). *Theorie der Begabung. Entwurf einer Intelligenzkunde* (2. Auflage). Heidelberg: Quelle & Meyer.
- Weiß, R. H. & Osterland, J. (1997). *Grundintelligenztest Skala 1 (CFT 1)*. 5., revidierte Auflage. Göttingen: Hogrefe.
- Weiß, R. H. (2006). *Grundintelligenztest Skala 2 (CFT 20-R). Revision*. Göttingen: Hogrefe.