

**Reliabilitäts- und Validitätsuntersuchungen zum
Kognitiven Fähigkeitstest KFT 4-12+ R
an einer Stichprobe von Fünft- und Siebtklässlern
der Willy-Brandt-Gesamtschule München**

Dissertation zur Erlangung des akademischen Grades

Doctor philosophiae (Dr. phil.)

der Philosophischen Fakultät

der Universität Rostock

Vorgelegt von

Shaza Skender, geb. am 22. Januar 1980 in Damaskus, Syrien

aus Lattakia, Syrien

Rostock, den 10. Oktober 2013

Gutachter:

1. Gutachter:

Prof. Dr. Christoph Perleth

Institut für pädagogische Psychologie, Universität Rostock

2. Gutachter:

Prof. Dr. Kurt A. Heller

Institut für pädagogische Psychologie, Universität München

3. Gutachter:

PD. Dr. Dorothee Doerfel-Baasen

Institut für pädagogische Psychologie, Universität Rostock

Datum der Einreichung: 16. Oktober 2013

Datum der Verteidigung: 11. Februar 2014

Inhaltsverzeichnis

1.	Einleitung	6
2.	Intelligenz.....	9
2.1	Definition der Intelligenz	9
2.2	Modellvorstellungen von Intelligenz	11
2.2.1	Klassische und faktorenanalytische Intelligenzmodell	11
2.2.1.1	Das Globalmodell der Intelligenz und der klassische IQ-von Binet.....	11
2.2.1.2	Das Generalfaktormodell nach Spearman.....	14
2.2.1.3	Das Primärfaktormodell von Thurstone.....	15
2.2.1.4	Das Zweifaktorenmodell von Cattell	16
2.2.1.5	Das Intelligenzstrukturmodell von Guilford.....	18
2.2.1.6	Das Berliner Intelligenzstrukturmodell.....	19
2.2.2	Moderne Intelligenzmodelle	20
2.2.2.1	Kognitionspsychologische Konzept von von Campione und Brown (1978).....	20
2.2.2.2	Das triarchische Intelligenzmodell nach Sternberg (1991).....	21
2.2.2.3	Die Theorie der multiplen Intelligenzen von Gardner	22
2.2.2.4	Das Drei Schichten-Modell der Intelligenzen (1993-2003).....	24
2.3	Zusammenfassung.....	25
3.	Intelligenzmessung.....	28
3.1	Gütekriterien psychologischer Diagnostik	28
3.1.1	Objektivität.....	28
3.1.2	Reliabilität.....	29
3.1.2.1	Stabilität bzw. Retest-Reliabilität	31
3.1.2.2	Paralleltest-Reliabilität.....	32
3.1.2.3	(Test)- Halbierungsreliabilität (Split-Half-Reliabilität)	32
3.1.2.4	Innere Konsistenz.....	34
3.1.2.5	Einflussfaktoren auf die Reliabilität	35
3.1.3	Gültigkeit oder Validität	37
3.1.3.1	Inhaltsvalidität.....	37
3.1.3.2	Kriteriumsvalidität	37
3.1.3.3	Konstruktvalidität.....	40
3.1.4	Nebengütekriterien	41

3.1.5	Zusammenfassung.....	42
3.2	Messung von Intelligenz	45
3.2.1	Individualtests	45
3.2.1.1	Hamburg-Wechsler-Intelligenztest für Kinder HAWIK-IV in der zweiten aktualisierten deutschen Fassung von Petermann und Petermann (2008)	45
3.2.1.2	Das Adaptive Intelligenz Diagnostikum (AID2) in der deutschen Bearbeitung von Kubinger und Wurst (2000)	48
3.2.1.3	Das Kaufman- Assessment Battery for Children (K-ABC) von Kaufman und Kaufmann in der deutschen Bearbeitung von Melchers und Preuß (1991)	51
3.2.2	Gruppenverfahren	54
3.2.2.1	Prüfsystem für Schul- und Bildungsberatung für 4. bis 6. Klassen (PSB-R 4-6) von Horn, Lukesch, Kormann und Mayrhofer (2002)	54
3.2.2.2	Prüfsystem für Schul- und Bildungsberatung für 6. bis 13. Klassen (PSB-R 6-13) von Horn, Lukesch, Mayrhofer und Kormann (2004)	56
3.2.2.3	Die Culture Fair Intelligence Tests (CFT) von R.B. Cattell in den deutschen Bearbeitung von Weiß und Osterland (2013) sowie Weiß & Weiß (2006)	58
3.2.2.4	Kognitiver Fähigkeits-Test (Kindergarten) KFT-K von Heller und Geisler (1983).....	63
3.2.2.5	Kognitiver Fähigkeits-Test (Grundschulform) KFT 1-3 von Heller und Geisler (1983).....	65
3.2.2.6	Kognitiver Fähigkeits-Test KFT 4-12 +R von Heller und Perleth (2000).....	67
3.2.2.7	Intelligenz-Struktur-Test (I-S-T) 2000 R von Amthauer, Brocke Liepmann und Beauducel (2001)	70
3.2.3	Zusammenfassung.....	73
4.	Zusammenhänge zwischen Intelligenz, Geschlecht, Migration und Schulleistung	75
4.1	Einflussfaktoren	75
4.1.1	Geschlechtsunterschiede	75
4.1.1.1	Biologische Faktoren	76
4.1.1.2	Psychologische Faktoren.....	77
4.1.1.3	Soziale Faktoren.....	78
4.1.1.4	Biopsychosozial Faktoren	80
4.1.2	Kulturelle Unterschiede in der Schulleistung	81
4.1.2.1	Die Erklärung durch kulturelle Defizite.....	84
4.1.2.2	Die humankapitaltheoretische Erklärung.....	87
4.2	Intelligenz als Prädikator für Erfolg in Schule und Beruf	88
4.2.1	Berufserfolg	88

4.2.2	Intelligenz und schulische Leistungen	90
4.3	Zusammenfassung.....	91
5.	Empirischer Teil: Untersuchungen zum KFT 4-12+ R.....	93
5.1	Fragestellungen und Hypothesen	93
5.2	Methode	97
5.2.1	Beschreibung der Stichprobe	97
5.2.2	Darstellung des KFT 4-12+R	101
5.2.3	Zensuren.....	104
5.2.4	Statistische Verfahren	105
5.3	Ergebnisse zur Testkonstruktion des KFT 4-12+ R.....	106
5.3.1	Itemanalysen	106
5.3.2	Reliabilität.....	148
5.3.3	Aspekte der Validität	153
5.3.3.1	Faktorielle Validität	153
5.3.3.2	Kriteriumsvalidität	159
5.4	Weitere Befunde zur Intelligenzsentwicklung, Geschlecht, Migration und Schulleistung für den KFT 4-12+ R in der vorliegenden Stichprobe	163
5.4.1	Kohorteneffekte intellektueller Kompetenzen für die Klassenstufen 5 und 7 in Bezug auf KFT4-12+R	163
5.4.1.1	Kohorteneffekte für die Klassenstufe 5	163
5.4.1.2	Kohorteneffekte für die Klassenstufe 7	168
5.4.2	Geschlechtsunterschiede im Untersuchungsraum.....	173
5.4.2.1	Klassenstufe 5	174
5.4.2.2	Klassenstufe 7	178
5.4.3	Muttersprache bzw. zu Hause überwiegend gesprochene Sprache.....	183
5.4.3.1	Klassenstufe 5	183
5.4.3.2	Klassenstufe 7	188
5.4.4	Veränderungen in den kognitiven Fähigkeiten von Klassenstufe 5 zu Klassenstufe 7	193
5.4.4.1	Veränderungen der KFT-Leistungen für unterschiedliche Kohorten (Jahrgangsgruppe)	193
5.4.4.2	Veränderungen der KFT-Leistungen für unterschiedliche Sprachgruppen	198
5.4.4.3	Geschlechtsspezifische Veränderungen der KFT-Leistungen	202

6.	Zusammenfassung und Interpretation der Untersuchungsergebnisse zum Kognitiven Fähigkeitstests KFT 4-12+ R für die Klassen 5 und 7	206
6.1	Zusammenfassung der empirischen Untersuchungsergebnisse	206
6.1.1	Itemanalysen und Verteilungsmerkmale zu den KFT-Subtests, Testdimensionen und der Gesamtleistung.....	206
6.1.2	Zur Reliabilität des KFT 4-12+ R	207
6.1.3	Zur Validität des KFT 4-12+ R.....	209
6.1.3.1	Konstruktvalidität.....	209
6.1.3.2	Kriteriumsvalidität	211
6.1.4	Einflussfaktoren auf die Intelligenzleistung	212
6.1.4.1	Geschlecht.....	212
6.1.4.2	Muttersprache.....	213
6.1.5	Veränderungen der Leistungen in den KFT-Dimensionen und der KFT-Gesamtleistung in den untersuchten Klassenstufen im letzten Jahrzehnt.....	214
6.1.6	Muster der Intelligenzentwicklung anhand der KFT-Leistungen in der Eingangsstufe der Gesamtschule	215
6.2	Interpretation	216
6.2.1	Bewertung der Qualität des KFT 4-12+ R	216
6.2.2	Einflussfaktoren	218
6.2.3	Veränderungen der Leistungen in den KFT-Dimensionen und der KFT-Gesamtleistung in den untersuchten Klassenstufen im letzten Jahrzehnt	221
6.2.4	Veränderung kognitiver Fähigkeiten von Klasse 5 zu Klasse 7	221
6.3	Folgerung und Ausblick.....	222
	Tabellenverzeichnis	224
	Abbildungsverzeichnis	229
	Literaturverzeichnis.....	231
	Zusammenfassung/Abstract	237

1. Einleitung

Intelligent ist nur der, der in einem Intelligenztest Intelligentes tut, oder es ist das, was der Intelligenztest misst, könnte man in Anlehnung an Myers (2008, S. 481) formulieren. Der erste moderne Intelligenztest in der Geschichte der psychologischen Diagnostik, der vernünftige Intelligenzmessung erst ermöglichte, wurde 1904 von Alfred Binet (1857-1911) und Theodore Simone (1873-1961) entwickelt, um zwischen behinderten von nicht behinderten Kindern genauer unterschieden werden zu können (vgl. Heller, 2000).

Intelligenztests werden häufig zur Vorhersage vom Schul- und Berufserfolg angewendet, weil Intelligenz eine wichtige Rolle für die Vorhersage der schulischen Leistung (Heller, 2008) sowie der Berufsleistung (Hülsheger, Maier, Stumpp & Muck, 2006) spielt, deren Vorläuferfertigkeiten wesentlich durch Intelligenztests erfassen werden. Man darf den Blick aber nicht nur auf die Ergebnisse der Intelligenztests richten, sondern auch darauf, ob deren Befunde überhaupt zuverlässig und gültig sind. Um den Ergebnissen von Intelligenztests vertrauen zu können, muss das jeweilige Testverfahren eine hohe diagnostische Qualität aufweisen, d.h. bestimmte diagnostische Qualitäts- oder Gütekriterien erfüllen.

In der vorliegenden Studie werden Fragen zur Bewertung der Qualität eines Tests zur Erfassung von Aspekten der Intelligenz, des kognitiven Fähigkeitstests KFT 4-12+ R von Heller und Perleth (2000), untersucht, der die deutsche Version der Cognitive Abilities Tests (CAT, Thorndike and Hagen, 1971) darstellt. Dieser Tests liefert Informationen über die verbalen, quantitativen und nonverbalen Fähigkeiten sowie über das kognitive Gesamtleistungsniveau von Schülerinnen und Schülern, die für die schulische Beratung sowie als Grundlage für Schullaufbahnempfehlungen verwendet werden.

Eine Bewertung der Testqualität erfolgt in der vorliegenden Arbeit vor allem durch die Durchführung einer sog. Itemanalyse sowie der Prüfung der Verteilungsmerkmale der Untertests, Dimensionen und der Gesamtleistung des KFT 4-12+ R. Weiter wird die Reliabilität bzw. Zuverlässigkeit oder Messgenauigkeit des KFT 4-12+R durch Einschätzungen der Reliabilität nach der Testhalbierungs- und Retest-Methode sowie durch Abschätzung der inneren Konsistenz geprüft. Außerdem werden zwei Aspekte der Validität des KFT 4-12+R (faktorielle Validität und Kriteriumsvalidität anhand von Schulzensuren) untersucht.

Für die Studie standen KFT-Daten von insgesamt 14 Kohorten (Schuleintrittsjahre 1998/99 bis 2011/12) einer Gesamtschule in München zur Verfügung, die den KFT 4-12+ R in der 5. und 7. Klasse bearbeitet hatten. Neben den oben angesprochenen Prüfungen der Skalenqualität, Reliabilität und Validität konnte für die Fünf- bzw. Siebtklässler/innen der Münchner Gesamtschule auch geprüft werden, ob deren verbale/sprachliche, quantitative und nonverbale Fähigkeiten sich in der ersten Dekade des 21. Jahrhunderts verbessert oder verschlechtert haben, indem Veränderungen der Leistungen in den KFT-Dimensionen und der KFT-Gesamtleistung über die Eingangsjahrgänge im Untersuchungszeitraum hinweg untersucht werden (Kohorteneffekte). Daneben konnten auch individuelle Veränderungen in den untersuchten Intelligenzskalen (KFT-Dimensionen) von Klasse 5 zu Klasse 7 analysiert werden (Entwicklungseffekte). Beide Effekte wurden in Abhängigkeit von Einflussfaktoren wie Geschlecht und Muttersprache überprüft.

Die vorliegende Arbeit gliedert sich in einen theoretischen und empirischen Teil. Im theoretischen Teil werden zunächst unterschiedliche Definitionen von Intelligenz behandelt sowie die wichtigste und bedeutendste Intelligenzmodelle beschrieben, die den Intelligenztests zugrunde liegen (Kapitel 2). Im Kapitel 3 werden zunächst die Gütekriterien (Haupt- und Nebengütekriterien) psychologischer Diagnostik ausführlich dargestellt erklärt. Auf der Grundlage dieses Kapitels werden auch die Fragestellungen zur Reliabilität und Validität des KFT 4-12+ R abgeleitet bzw. formuliert, die den Untersuchungen im empirischen Teil zugrunde liegen. Außerdem werden Intelligenz- und kognitive Fähigkeitstests vorgestellt, die am häufigsten von Psychologen/innen in Deutschland angewendet und in der Literatur dargestellt werden.

In Kapitel 4 werden wichtige Studien präsentiert, die den Einfluss von Faktoren wie Geschlecht und Kultur auf die Intelligenz überprüft haben, und es wird auf Befunde zur Bedeutsamkeit der Intelligenz für die Prognose der Schul- und Berufsleistung eingegangen. Diese Zusammenhänge zwischen Intelligenz auf der einen und der Schul- sowie der Berufsleistung auf der anderen Seite und ihre Abhängigkeit von Faktoren wie Geschlecht und Migration (Muttersprache) liefern die theoretische Grundlage für weitere Untersuchungen im empirischen Teil.

Das 5. Kapitel enthält den empirischen Teil bzw. die Untersuchungen zum KFT 4-12+ R im Detail. Die Fragestellungen und Hypothesen werden eingangs des Kapitels formuliert und die verwendeten Methoden genau und detailliert beschrieben. Den Hauptteil machen dann die Ergebnisse zur Testkonstruktion des KFT 4-12+ R bzw. zur Item-, Reliabilitäts- und Aspekten der Validitätsanalyse aus (faktorielle- und Kriteriumsvalidität) sowie Befunde zur Abhängigkeit der KFT-Leistungen vom Schuleintrittsjahr sowie von den Einflussfaktoren Geschlecht und

Muttersprache. All diese Befunde beziehen sich, wie oben gesagt, auf Daten aus den Klassenstufen 5 und 7 in der Willy-Brandt-Gesamtschule in München von 1998 bis 2010 (Klassenstufe 5) und von 1998 bis 2011 (Klassenstufe 7). Es konnten also sowohl Kohorteneffekte wie auch Befunde zur Veränderungen der individuellen KFT-Leistungen von Klasse 5 zu Klasse 7 analysiert und ausführlich diskutiert werden.

Gesondert sei auf die ungewöhnliche Stichprobe hingewiesen, die für die Beantwortung der erwähnten Fragestellungen zur Verfügung stand. beantworten zu können. Es lagen für die vorliegende Studie Daten aller Schülerinnen und Schüler der 5. und 7. Schulklassen der Willy-Brandt- Gesamtschule in München vor. Insgesamt konnte auf $N_5 = 2091$ Datensätze von Fünftklässlern (darunter 1113 bzw. 53,23% Jungen und 978 bzw. 46,77% Mädchen) der Jahrgänge 1998 bis 2010 zurückgegriffen werden. In der Regel von denselben Schülern lagen $N_7 = 2229$ Datensätze aus Klassenstufe 7 vor (davon 1165 oder 52,27% Jungen und 1064 oder 47,73% Mädchen; letztes Erhebungsjahr war 2011-2012). Die Differenzen ergeben sich daraus, dass einzelne Schülerinnen und Schüler bei einzelnen Tests fehlten und vor allem daraus, dass Schülerinnen und Schüler im Laufe der Unterstufe die Schule verlassen oder erst nach den Testerhebungen in Klasse 5 an sie gewechselt sind. Schülern/innen vom in Klasse 5 und bis zum Jahr 2011 in Klasse 7 gesammelt. Darüber hinaus konnten im Jahr 2011-2012 Schulzensuren (Deutsch, Englisch und Mathe) an einer Stichprobe von Klasse 7 ca. $N = 154$ Schüler/innen erhoben werden, um Aspekte der Kriteriumsvalidität zu überprüfen. Um den Einfluss des Migrationshintergrunds zumindest grob abschätzen zu können, wurden die Schuler in Sprachgruppen eingeteilt, wobei die Sprache maßgeblich war, die zu Hause am häufigsten gesprochen wird (Näheres dazu findet sich in Kapitel 5).

Im letzten, dem 6. Kapitel werden die Ergebnisse des empirischen Teils bzw. der eigenen Untersuchungen zum KFT 4-12+ R zusammengefasst. Darüber hinaus werden diese Befunde unter Rückgriff auf den theoretischen Teil interpretiert und es wird ein Ausblick zw. Empfehlungen gegeben für den zukünftigen Einsatz des KFT 4-12+ in der Willy-Brandt-Gesamtschule, besonders im Hinblick auf seinen Einsatz in der Schullaufbahnberatung. Letztlich soll auch die Intelligenzmessung mit dem KFT 4-12+ R dazu beitragen, dass die einzelnen Schülerinnen und Schüler ihren Schulerfolg in der Gesamtschule optimieren können.

2. Intelligenz

2.1 Definition von Intelligenz

Menschen sind nicht gleich. Sie unterscheiden sich nicht nur im Aussehen, sondern auch hinsichtlich ihrer Gedächtnisleistungen, ihres logischen Denkens, der Geschwindigkeit, mit der sie Informationen aufnehmen und verarbeiten, ihres Zugangs zu Inhaltsgebieten sowie vieler unterschiedlicher intellektueller Fähigkeiten. Aufgrund beobachtbarer Unterschiede auch zwischen Personen mit vergleichbaren Bildungs- und Erziehungserfahrungen ist das Konstrukt der Intelligenz einzuführen (vgl. Stern, 2001, S. 163). Intelligenz ist eins der häufigsten Merkmale, die in der Psychologie und außerhalb der Psychologie wie der Pädagogik und der Hirnforschung untersucht und kontrovers diskutiert wurde.

Der Begriff Intelligenz (vgl. Neubauer & Stern, 2007, S. 13; Rost, 2009, S. 1) stammt vom lateinischen „intelligentia“ (Einsicht, Verstand, einsehen, verstehen) und wird als „geistige Fähigkeit“, „mentale Fähigkeit“ oder auch „Begabung“ umschrieben. Die Studien und die Untersuchungen zum Thema Intelligenz sind so zahlreich, dass es viele Intelligenzdefinitionen von unterschiedlichen Intelligenzforschern gibt, d. h. es gibt keine einheitliche Definition für Intelligenz.

Dennoch wird Intelligenz meist übereinstimmend als „eine sehr allgemeine geistige Kapazität, die – unter anderem – die Fähigkeit zum schlussfolgernden Denken, zum Planen, zur Problemlösung, zum abstrakten Denken, zum Verständnis komplexer Ideen, zum schnellen Lernen und zum Lernen aus Erfahrung umfasst. Es ist nicht reines Bücherwissen, keine enge akademische Spezialbegabung, keine Testerfahrung. Vielmehr reflektiert Intelligenz ein breiteres und tieferes Vermögen, unsere Umwelt zu verstehen, ‚zu kapieren‘, ‚Sinn in Dingen zu erkennen‘, oder ‚herauszubekommen‘, was zu tun ist“ (Gottfredson, 1997, S. 13. zit. nach Rost, 2009, S. 2) beschrieben.

Im Folgenden werden mehrere wissenschaftliche Intelligenzdefinitionen genannt bzw. aufgelistet:

- Wechsler (1961) hat Intelligenz als „die zusammengesetzte oder globale Fähigkeit des Individuums, zweckvoll zu handeln, vernünftig zu denken und sich mit seiner Umgebung wirkungsvoll auseinanderzusetzen.“ definiert (S. 13).

- „Intelligenz ist die allgemeine Fähigkeit eines Individuums, sein Denken bewußt auf neue Forderungen einzustellen; sie ist allgemeine geistige Anpassungsfähigkeit an neue Aufgaben und Bedingungen des Lebens“ (Stern1912, S. 3, zit. nach Heller, 1973, S. 10).
- „Intelligenz ist die Fähigkeit des Individuums, anschaulich oder abstrakt in sprachlichen, numerischen und raumzeitlichen Beziehungen zu denken; sie ermöglicht erfolgreiche Bewältigung vieler komplexer und mit Hilfe jeweils besonderer Fähigkeitsgruppen auch ganz spezifischer Situationen und Aufgaben“. (Groffmann, 1964, S. 190, zit. nach Heller, 1973, S. 10).
- Perleth (2008) definiert Intelligenz als „allgemeine Fähigkeit zum Denken oder Problemlösen in Situationen, die für das Individuum neuartig, d.h. nicht durch Lernerfahrungen vertraut sind“ (S. 15).
- „Intelligenz kann als Potenzial eines Menschen verstanden werden, Lern- und Bildungsangebote zur Anneigung von Wissen zu nutzen.“ (Stern, 2001, S. 163).
- Intelligenz ist „die Fähigkeit zum denkgestützten Lösen von Aufgaben und Problemen in Situationen, die für die Person neu und nicht allein durch Wissensabruft erfolgreich bearbeitbar sind, die Fähigkeit zum induktiv und deduktiv-logisch schlussfolgernden Denken, die Fähigkeit zum abstrakten Denken und die Fähigkeit zu Verständnis und Einsicht zum Erkennen und zur Herstellung von Strukturen, Beziehungen, Sinnzusammenhängen und Bedeutungen. (Rindermann, 2004, S. 374)

Nach Heller (1973, S. 9) schließen die meisten der Intelligenzdefinitionen mindestens zwei Bestimmungsstücke ein:

- Die biologische versus soziale Anpassungsfähigkeit an neue Probleme oder Situationen bzw. Umweltbedingungen.
- Das Verlangen nach der Ökonomie der zur Verfügung stehenden Mittel.

Zusammenfassend kann man sagen, dass Intelligenz bei allen angeführten Definitionen als Disposition, generelle Fähigkeit oder kognitive Fähigkeit bezeichnet wird. Dies stimmt nicht mit den Theorien anderer Wissenschaftler wie Howard Gardner („multiple Intelligenz“) oder Sternberg („Erfolgsintelligenz“) überein, weil diese das Postulat einer allgemeinen Fähigkeit ablehnt und über nicht-kognitive Fähigkeiten gesprochen haben, die nach diesen Autoren zur Bewältigung des beruflichen Lebens erforderlich sind.

Wie oben gezeigt wurde, haben die Forscher unterschiedliche Meinungen darüber, ob Intelligenz ist, was der Intelligenztest misst oder ob Intelligenz eine Fähigkeit oder mehrere Fähigkeiten

darstellt. Intelligenz ist ein Komplex aus unterschiedlichen Fähigkeiten: abstraktes Denken, Problemlösen, Anpassungsfähigkeit an neue Aufgaben und an die Umwelt, und die Fähigkeit aus Erfahrungen zu lernen und Wissen zu erwerben.

2.2 Modellvorstellungen von Intelligenz

Die Intelligenzforschung beschäftigt sich mit der Frage, ob Intelligenz eine generelle, umfassende Fähigkeit ist, oder ob Intelligenz aus unabhängigen, spezifischen Fähigkeiten besteht (vgl. Kail & Pellegrino, 1989, S. 21). Anfang des 20. Jahrhunderts gab es innerhalb des psychometrischen Intelligenzparadigmas zwei Richtungen:

- Eindimensionale Intelligenzmodelle, die Intelligenz als allgemeine, umfassende Fähigkeit definieren. Diese Fähigkeit sei wesentlich verantwortlich für die Bewältigung intellektueller Aufgaben.
- Im Gegensatz dazu werden spezifische Fähigkeiten angenommen, in denen sich Menschen bei der Lösung von Aufgaben oder Tests unterscheiden. In mehrdimensionalen Intelligenzmodellen wird Intelligenz als aus mehreren Faktoren oder Fähigkeiten bestehend angesehen, die bei der Lösung von Aufgaben jeweils zusammenwirken. (vgl. Perleth, 2008, S. 15).

In diesem Kapitel werden zunächst die bedeutsamsten Intelligenzmodelle, die sich in klassische oder faktorenanalytische und moderne Intelligenzmodelle gliedern lassen, beschrieben.

2.2.1 Klassische und faktorenanalytische Intelligenzmodelle

2.2.1.1 Das Globalmodell der Intelligenz und der klassische IQ von Binet

Alfred Binet (1857-1911) wird als Begründer des ersten Intelligenztests angesehen. Der Unterrichtsminister beauftragte ihn, ein Verfahren zu entwickeln, mit dem behinderte Kinder von nicht behinderten Kindern genauer unterschieden werden können. Gemeinsam mit seinem Schüler Simon publizierte er deswegen 1905 einen Intelligenztest (Binet & Simon, 1905), um die geistigen Fähigkeiten von Kindern zu messen (vgl. Kail & Pellegrino, 1989, S. 19).

Binet behauptete, „komplexere Prozesse wie Gedächtnis, visuelle Vorstellung, Verstehen, Urteilen und die Fähigkeit zur Selbstkritik bildeten den Kern der menschlichen Intelligenz“

(Binet & Henri, 1896; Binet & Simon, 1908, zit nach Kail & Pellegrino, 1989, S. 20). Der Binet-Simon-Test besteht aus 30 verschiedenen Testaufgaben, die die höheren psychischen Funktionen wie Gedächtnis, Aufmerksamkeit, Verständnis messen. Nach Binet und Simon (1908) ist die Intelligenz nicht stabil. Sie verändert sich mit dem Alter der Kinder. Deshalb führten sie mit dem von ihnen entwickelten Test (1908) das Konzept des Intelligenzalters ein, um den intellektuellen Entwicklungsstand von Kindern vergleichen zu können. In der Berechnung des Intelligenzalters ging man ursprünglich man zuerst von dem Altersniveau aus, auf dem das Kind alle Aufgaben gelöst hat. Dann wird für jeweils fünf zusätzlich gelöste Aufgaben ein Jahr zum Ausgangsniveau addiert (vgl. Kail & Pellegrino, 1989, S. 18-21).

In den USA wurden vor dem ersten Weltkrieg zwei Revisionen für die Binet –Simon Skalen entwickelt (vgl. Grob, Hagmann-von Arx & Bodmer, 2009, S. 150; Myers, 2008, S. 482-483). Die erste Revision erarbeiteten Robert Yerkes und James Bridges (Yerkes et al., 1915, zit. nach Grob et al., 2009, S. 150), die ähnliche Aufgaben konstruierten und diese in Subtests anordneten. Außerdem transformierten sie die Leiterskala in eine Punktesskala. Die zweite Revision stammt von Lewis Terman (1916, zit. nach Grob et al., 2009, S. 150). Er war Professor an der Stanford Universität und nach dem Tod Binets im Jahr 1911 entwickelte er bis zu seinem Tod 1956 Binets Test weiter. Terman kam zu dem Ergebnis, dass die bei französischen Schülern erhobenen Altersmaßstäbe auf die kalifornischen Schüler nicht anwendbar waren. Aus diesem Grund legte er neue Altersnormen fest, erweiterte den Altersbereich bis ins Erwachsenenalter und ersetzte das Entwicklungsalter durch den Intelligenzquotienten (IQ), der von dem deutschen Psychologen William Stern (1912) formuliert wurde und der „als das Verhältnis von Intelligenzalter (IA) zum Lebensalter (LA) multipliziert mit 100, nach der Formel

$$IQ = IA/LA * 100$$

definiert ist (Myers, 2008, S. 483). Der Wert IQ = 100 bezeichnet durchschnittliche Intelligenz, da in diesem Fall das Intelligenzalter gleich dem Lebensalter eines Kindes ist (vgl. Myers, 2008, S. 482-483).

Dieser Intelligenzquotient (vgl. Holling & Kannig, 1999, S. 25-26; Schweizer, 2006, S. 4) ist im Kindes- und Jugendalter passend, jedoch nicht bei Erwachsenen, da das Lebensalter ununterbrochen zunimmt, das Intelligenzalter ab dem jungen Erwachsenenalter jedoch stabil bleibt. Aus diesem Grund hat Wechsler (1939, zit. nach Schweizer, 2006, S. 4) den IQ als Abweichungs-

koeffizient definiert, der die Position der einzelnen Person in der Populationsverteilung durch die folgende Formel angibt:

$$IQ = 100 + 15 \cdot \frac{(X - M)}{SD}$$

Dabei sind

- X: der Testscore,
- M: der Mittelwert der Normstichprobe,
- SD¹: die Standardabweichung der Normstichprobe.

Dieser IQ folgt der Normalverteilung mit Mittelwert 100 und Standardabweichung 15. Durch den Abweichungs-IQ können sowohl die Leistungen von Kindern als auch die Leistungen von Erwachsenen unterschiedlichen Alters jeweils untereinander verglichen werden. Außerdem kann der IQ-Wert in einen Prozentrangwert umgerechnet werden. IQ = 100 passt zu einem Prozentrang 50%, d. h. dass das betreffende Individuum ein gleiches gutes oder besseres Ergebnis wie 50% seiner Altersgruppe erreicht. IQ = 130 entspricht einem Prozentrang von 97,6% d. h. dass das betreffende Individuum ein gleiches gutes oder besseres Ergebnis wie 97,6% seiner Altersgruppe erzielt.

Obwohl Binet wegen seines Begriffs Intelligenzalter kritisiert wurde, wird der Binet-Simon-Test trotzdem als erster und guter Intelligenztest für Kinder bezeichnet. Außerdem stellt er eine eindeutige Abkehr von dem vorher von Galton und Cattell vertretenen Forschungsansatz im Bereich der Intelligenz und der individuellen Unterschiede dar, die sagten, dass einfache Prozesse wie die sensorischen und motorischen Prozesse die grundlegenden Elemente der Intelligenz ausbilden und nicht die komplexen Prozesse wie Gedächtnis und visuelle Vorstellung, wie Binet behauptete (vgl. Kail & Pellegrino, 1989, S. 20).

Folgende Kritikpunkte ergeben sich bei der Intelligenzmessung von Binet und Simon: Einerseits sind die Tests von Binet und Simon sprach- und bildungsabhängig, d.h. Kinder mit besseren sozioökonomischen Bedingungen haben mehr Erfolg als Kinder mit schlechteren sozioökonomischen Bedingungen. Andererseits wurde die Form der Berechnung der Vorsprünge und Rückstände in der Intelligenzsentwicklung kritisiert, weil eine bestimmte Differenz zwischen (IA) und

¹ Die Standardabweichung wird hier in Übereinstimmung mit der zitierten bzw. verwendeten Literatur mit „SD“ bezeichnet. In dieser Arbeit wird daneben auch „s“ bzw. „StdAbw“ bzw. „StandAbw“ zur Bezeichnung der Standardabweichung verwendet.

(LA) in unterschiedlichen Altersbereichen unterschiedliche Bedeutung hat (vgl. Neubauer & Stern, 2007, S. 29).

2.2.1.2 Das Generalfaktormodell nach Spearman (1904)

Bei der Faktorenanalyse der Englischen Schule sind als prominenteste Vertreter Spearman (1904, 1927), Burt (1949) und Vernon (1961) zu nennen (Heller, 1973, S. 14). Spearman, der als Begründer faktorenanalytischer Untersuchungen geistigen Verhaltens gilt, konnte aufgrund seiner empirischen Untersuchungen eine einheitliche Fähigkeit der Intelligenz oder den allgemeinen generellen Faktor, „g-Faktor“ bestimmen, der für Erfolg bei intellektuellen Aufgaben erforderlich ist. Außerdem gibt es nach Spearman spezifische „s-Faktoren“, die bei jeweiliger Aufgabe zusammen mit dem „g“ Faktor wirksam werden (vgl. Schön-Gaedike, 1978, S. 25-26; Perleth, 2008, S.15-16; siehe auch Abbildung 1). Das Zwei-Faktoren-Modell gibt an, warum einige Menschen bessere geistige Leistungen als andere erbringen können, wenn der „g-Faktor deutlich höher liegt. Bei den spezifischen Begabungen ist dies der Fall, wenn bestimmte“ s- Faktoren“ stark ausgeprägt sind (vgl. Holling & Kännig, 1999, S. 26).

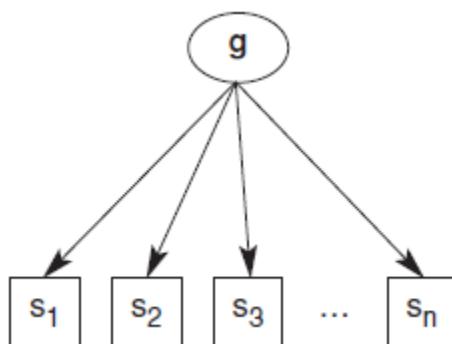


Abbildung 2.1: Spearmans Modell, zit. nach Perleth (2008, S. 16).

Das Modell von Spearman gilt als Grundlage für viele Intelligenztests wie z.B. den Hamburg-Wechsler-Intelligenztest und ist bis heute ein sehr nützliches Konstrukt. Kritisch bei diesem Modell ist anzumerken, dass Spearman die Generelle Fähigkeit „Intelligenz“ nur durch eine Zahl, den IQ, ausgedrückt hat.

2.2.1.3 Das Primärfaktormodell von Thurstone (1938)

Mit Thurstone (1938) hat die Entwicklung der mehrfaktoriellen Intelligenzforschung begonnen. Seiner Auffassung nach besteht Intelligenz aus sieben unterschiedlichen, gleichberechtigten und unabhängigen Primärfaktoren, die aufgrund der Faktorenanalyse von Intelligenztests bei amerikanischen Studenten entdeckt wurden. Bei diesen Fähigkeiten handelt es sich um:

- Wahrnehmungsgeschwindigkeit: Die Fähigkeit schnell wahrzunehmen und Informationen zu verarbeiten.
- Wort und Sprachverständnis: die Fähigkeit mit verbalen Begriffen umzugehen, sprachliche Bedeutungen und Beziehungen zu verstehen.
- Wortflüssigkeit: Die Fähigkeit ein Synonym eines Worts zu finden.
- Gedächtnis: Die Fähigkeit paarweise Assoziationen zu behalten und zu erinnern.
- Schlussfolgerndes Denken: Die Fähigkeit Regeln zu erkennen und anzuwenden.
- Numerische Fähigkeit: Die Fähigkeit einfache Rechenoperationen ausführen zu können und schnell zu rechnen.
- Räumliches Vorstellungsvermögen: Die Fähigkeit zum räumlichen Vorstellen und orientieren.

Nach Thurstone unterscheiden sich Menschen nicht nur in ihrem Intelligenzniveau, sondern auch in ihrer Intelligenzstruktur (vgl. Rindermann, 2004, S. 379). Die Primärfaktoren von Thurstone zeigen allerdings auch gewisse korrelative Zusammenhänge hingewiesen, die nach seiner Ansicht auf Faktoren zweiter Ordnung zurückgehen und die den allgemeinen generellen Faktor ausdrücken. Dieser Faktor hat einen Einfluss auf die Aktivitäten der Primärfaktoren (vgl. Ingenkamp & Lissman, 2008, S. 228).

Als Vorzug dieses Modells kann man festhalten, dass das Primärfaktormodell als Forschungsinstrument und Grundlage für verschiedene Tests gilt, die für die Intelligenzdiagnostik im schulischen Umfeld verwendet werden wie das Prüfsystem für Schul- und Bildungsberatung PSB R 4- 6 und PSB R 6-13 von Horn et al. (2002; 2004). Zusätzlich ist Intelligenz nach Thurstone (1938) ein Komplex von verschiedenen unabhängigen Faktoren und nicht wie bei Spearman nur von zwei Faktoren. Durch die Faktorenanalyse wurde ein Zusammenhang zwischen den Primärfaktoren nachgewiesen und diese Faktoren werden von dem generellen Faktor beeinflusst d. h., dass Thurstone letztlich zu dem gleichen Ergebnis wie Spearman kommt.

2.2.1.4 Das Zweifaktorenmodell von Cattell (1963)

Cattell war Schüler von Spearman. Er entwickelte das Zweifaktorenmodell der Intelligenz von seinem Lehrer weiter und unterschied zwischen der flüssigen und kristallisierten Intelligenz. Die flüssige Intelligenz gilt als die allgemeine Fähigkeit „in neuartigen Situationen und anhand von sprachfreiem, figuralem Material, Denkprobleme zu erfassen, Beziehungen herzustellen, Regeln zu erkennen, Merkmale zu identifizieren und rasch wahrzunehmen“ (Weiß & Osterland, 1980, S. 4). Die fluide Intelligenz ist nach Cattel angeboren und analog zu Spearman stellt diese die allgemeine generelle Fähigkeit oder den „g-Faktor“ dar. Die Ausprägung der fluiden Intelligenz ist unabhängig von der soziokulturellen Umwelt.

Die kristallisierte Intelligenz ist ein Komplex aus der fluiden Intelligenz und der sozialen Umwelt. Darüber hinaus enthält sie Erfahrung in Form von erworbenem Wissen. Aus diesem Grund ist die kristallisierte Intelligenz abhängig von den soziokulturellen Einflüssen und sie ist nicht angeboren. Nach Cattel erwirbt man kristallisierte Intelligenz, wenn man in einer guten Lernumwelt lebt und über fluide Intelligenz verfügt (vgl. Perleth, 2008, S. 16-17). Je älter der Mensch wird, desto mehr verfügt er über kristallisierte Intelligenz, wohingegen die fluide Intelligenz im frühen Erwachsenenalter ihren Höhepunkt erreicht und danach zuerst leicht, dann stärker abfällt (Neubauer & Stern, 2007, S. 62).

Cattell hat in seiner Theorie die Anlage mit der Lernumwelt verknüpft. Allerdings erscheint mir die Entwicklung der fluiden Intelligenz im Sinne Cattel nicht unabhängig von der Lernumwelt zu verlaufen, was auch mit den neuen Modellen übereinstimmt (siehe z.B. die Modelle von Gardner oder Campione und Brown in den Abschnitten 1.2.2.1 und 1.2.2.3 unten).

Auf Grundlage der Konzeption der fluiden und kristallisierten Intelligenz nach Cattell (1971) und Horn (1982, 1989) sind Zweikomponentenmodelle der Intelligenz entstanden (vgl. Lindenberger & Schaefer, 2008, S. 377-378). Dieses Modell unterscheidet zwischen der fluiden (mechanischen) kognitiven Fähigkeit und der kristallisierten (pragmatischen) kognitiven Fähigkeit. Während die Mechanik die biologische Komponente der intellektuellen Leistungsfähigkeit darstellt, repräsentiert die Pragmatik die kulturelle Komponente der intellektuellen Fähigkeit.

Die Mechanik der intellektuellen Fähigkeit zeigt Fähigkeiten, die sich auf Schnelligkeit, Genauigkeit elementarer kognitiver Prozesse (Wahrnehmungsgeschwindigkeit, Denkfähigkeit, Merkfähigkeit, usw.) beziehen. Diese Fähigkeiten wachsen im Kindes- und Jugendalter und sie

nehmen im höheren Erwachsenenalter ab. Die Gründe für die Zunahme der Mechanik oder der mechanischen Fähigkeiten in der frühen Kindheit liegen im Aufbau der neuronalen Strukturen (Bjorklund, 1997; Elman et al., 1996; Wellman & Gelman, 1992, zit. nach Lindenberger & Schaefer, 2008, S. 377). Die Abnahme der mechanischen Fähigkeiten im höheren Erwachsenenalter hingegen sind auf das Nachlassen des phylogenetischen Selektionsdrucks und altersbezogene Dysfunktionen zurückzuführen.

Die Pragmatik umfasst nach Lindenberger und Schaefer, (2008, S. 378) die kulturelle Dimension der kognitiven Entwicklung durch das erworbene Wissen, das nicht nur internal wie z.B. in semantischen Netzwerken, sondern auch external z. B. in Büchern repräsentiert wird. Die Entwicklung der Pragmatik erfolgt durch den Erwerb des deklarativen und prozeduralen Wissens, das in normativen Gesellschaften z. B. durch Schulpflicht oder in universellen Gesellschaften z.B. durch informelle Unterweisung durch Mentoren usw. vermittelt wird. Die individuellen Unterschiede in den pragmatischen Fähigkeiten stehen mit Bildungschancen und sozialer Ungleichheit in Beziehung (vgl. Lindenberger & Schaefer, 2008, S. 378). Im Gegensatz dazu korrelieren die mechanischen Fähigkeiten bzw. die Wahrnehmungsgeschwindigkeit stärker mit sensorischsensomotorischen Kompetenzen als mit pragmatischen Fähigkeiten im Sinne von Wissen (Wortschatz) (Lindenberger und Baltes, 1997, zit. nach Lindenberger & Schaefer, 2008, S. 379).

Die kognitive Entwicklung erfolgt durch Interaktionen zwischen der Pragmatik und Mechanik, die sich im Laufe der Ontogenese verändern. Einerseits stellt die Pragmatik das Medium und die Voraussetzung für den Aufbau der Mechanik in der frühen Kindheit dar. Andererseits bestimmen reifungs- und altersbedingte Veränderungen hinsichtlich Mechanik die Fähigkeit zum Erwerb des pragmatischen Wissens und die Wahrscheinlichkeit, mit der einmal erworbenes Wissen erhalten und in bestimmten Kontexten verwendet werden kann. In dem Alter, in dem beispielsweise Experten ihre höchsten Leistungen erzielen, können Unterschiede zwischen dem Fertigkeitsniveau auf Wechselwirkungen zwischen dem Alter, in dem mit dem Fertigkeitserwerb begonnen wurde, und dem altersbedingten Nachlassen der Mechanik zurückgeführt werden (vgl. Lindenberger & Schaefer, 2008, S. 380-381).

2.2.1.5 Das Intelligenzstrukturmodell von Guilford

Dieses Modell gilt als das komplexeste der Intelligenzforschung (vgl. Heinze, 2005, S. 17; Neubauer & Stern; 2007, S. 64-65). Guilford unterscheidet zwischen drei Dimensionen der Intelligenz: Denkinhalte, Denkprodukte und Denkoperation. Diese Dimensionen wurden in 120 Intelligenzfaktoren untergliedert (siehe Abbildung 2). Denkinhalte bestehen aus vier unterschiedlichen Inhalten: symbolische, semantische, figurale und verhaltensmäßige Inhalte. Im Hinblick auf die Operationen unterscheidet Guilford zwischen Erkenntnisvermögen, Gedächtnis, divergente und konvergente Produktion sowie Evaluation. Bezuglich der Denkprodukte unterscheidet Guilford zwischen Einheiten, Klassen, Beziehungen, Systemen, Transformationen und Implikationen. Dieses Modell ist auf Kritik gestoßen, da es sehr kompliziert ist und noch nicht alle Faktoren empirisch gesichert sind, d.h es ist praktisch nicht anwendbar, obwohl die meisten neuen Studien auf die Mehrdimensionalität der Intelligenz hinweisen. Außerdem sei dieses Modell nach Heller geeigneter zur Messung von Kreativität als von Intelligenz (Heller, 2000, S. 26).

VORGANG:

Evaluation

konvergente Produktion

divergente Produktion

Gedächtnis

Erkenntnisvermögen

Einheiten

Klassen

Beziehungen

Systeme

Transformationen

Implikationen

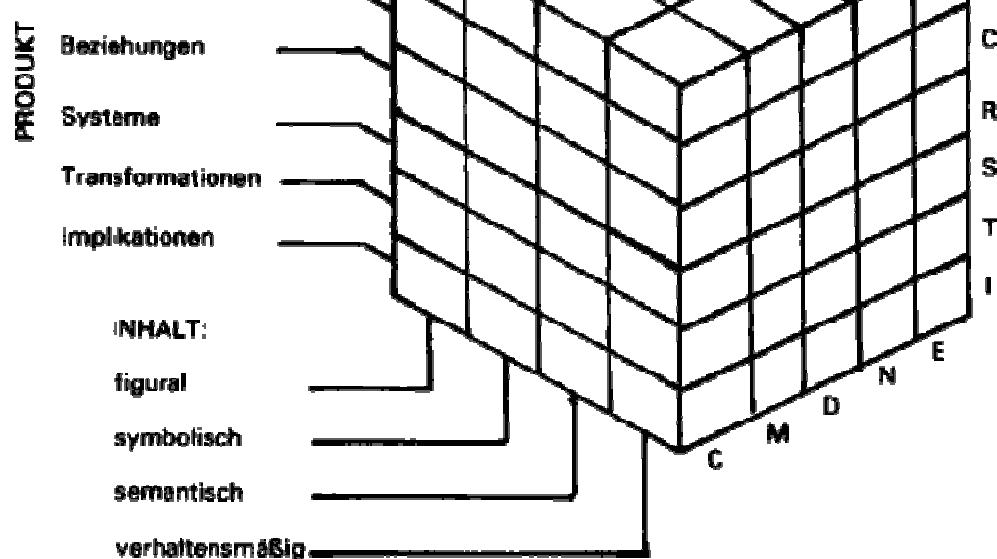


Abbildung 2.2: Intelligenzstrukturmodell von Guilford (vgl. Neubauer & Stern, 2007, S. 64).

2.2.1.6 Das Berliner Intelligenzstrukturmodell

Das Modell von Jäger (1984) gilt als das beste empirische psychometrische Intelligenzmodell zumindest in der deutschen Sprache (vgl. Perleth, 2008, S. 18). In diesem Modell geht es um ein integratives Modell, in dem sich die älteren Modelle der Intelligenz als Teilmodelle wiederfinden. Außerdem stellt dieses Modell eine Verbindung zwischen dem Modell von Spearmans „g“ Faktor, der alle Intelligenzleistungen beeinflusst, und Thurstones Modell „unterschiedlicher Fähigkeiten“ dar (vgl. Heinze, 2005, S. 17-18).

Das Berliner Intelligenzstrukturmodell (siehe Abbildung 3) enthält sieben Hauptkomponenten, die in zwei Modalitäten „Inhalte“ und „Operation“ untergliedert werden und die nach Jäger als Fähigkeitsbündel bezeichnet werden. Die Inhaltsdimensionen umfassen drei Faktoren bzw. Fähigkeitsbündel, V (verbal), N (numerisch) und F (figural) sowie die Operationsdimensionen bestehend aus vier Fähigkeitsbündeln, B (Bearbeitungsgeschwindigkeit), K (Verarbeitungskapazität), G (Gedächtnis) und E (Einfallsreichtum). Außerdem spielt der „g“ Faktor eine integrierende Rolle über alle Komponenten.

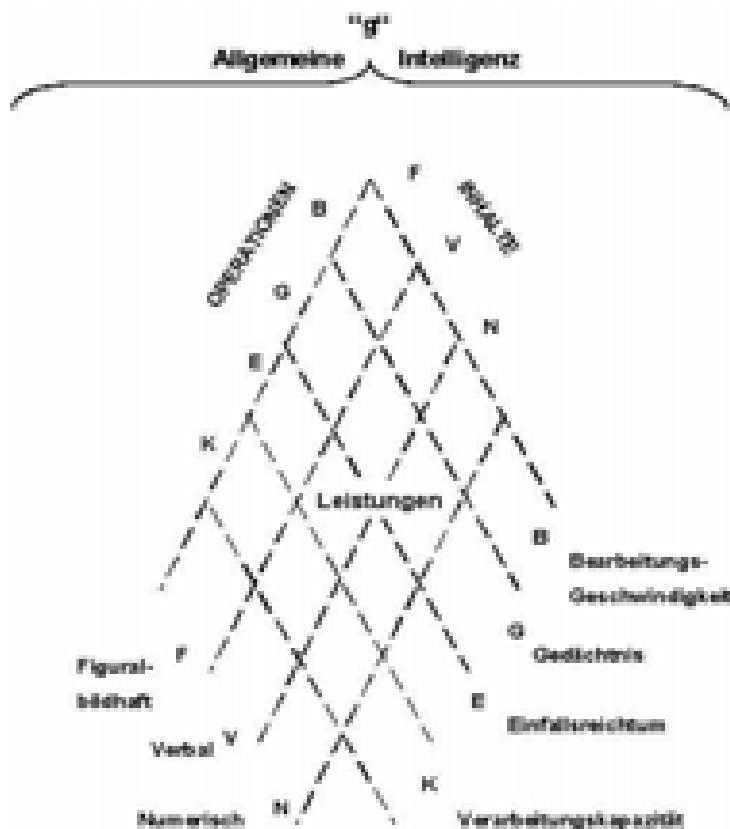


Abbildung 2.3: Berliner Intelligenzstrukturmodell von Jäger (zit. nach Heller, 2000, S. 28)

Empirisch hat Jäger (1984) 48 Variablen, die er aus seinem Itempool ausgewählt hat, zu 16 operationshomogenen Bündeln und zu 12 inhaltshomogenen Bündeln zusammengefasst. Anhand der Faktorenanalyse dieser Bündel konnte er die vier Faktoren von Operationen und drei Faktoren von Inhalten bestätigen (vgl. Perleth, 2000, S. 37).

Das Modell von Jäger gilt als Grundlage für den BIS (Berliner Intelligenzstruktur-Test). Darüber hinaus haben Perleth und Sierwald (1996, zit. nach Perleth, 2000, S. 37) ein allgemeines, mehrdimensionales State-Trait-Modell in Anlehnung an das Berliner Intelligenzstrukturmodell für den kognitiven Fähigkeiten-Test 4-13+ von Heller, Gaedike und Weinläder (1985) konzipiert. Dem State-Trait-Modell liegt das „Latente Personen-Methoden-Zeitpunkte-Modell“ (LPZM) von Sierwald (1996) zugrunde. Die Dimensionen sind im Falle des Modells für den KFT nach Perleth und Sierwald (1996): Zeit, Material und Operation.

Das Berliner Intelligenzstrukturmodell ist einfacher als das Modell von Guilford und hat praktische Implikationen. Aber wie Guilford in Bezug auf Kreativität kritisiert wurde, kann man auch Jäger kritisieren, da er ebenfalls eine Komponente Einfallsreichtum konzipierte.

2.2.2 Moderne Intelligenzmodelle

2.2.2.1 Das kognitionspychologische Konzept von Campione und Brown (1978)

In diesem Modell unterscheiden Campione und Brown (1978) zwischen zwei Elementen oder Ebenen der Intelligenz. Die erste stellt die „architecture-Ebene“ oder Hardware des kognitiven Apparates dar. Dazu gehört das Dreispeicher-Gedächtnismodell (Sensorische Register, Langzeitgedächtnis, und Kurzzeitgedächtnis). Dieses umfasst die drei Komponenten:

- Kapazität (Speichergröße),
- Speicherdauer,
- Effizienz (Informationsverarbeitungsgeschwindigkeit).

Nur die Geschwindigkeit der Verarbeitung und des Abrufs der Informationen steht mit Intelligenz in Zusammenhang. Die „architecture Ebene“ ist stabil und unveränderbar.

Die „exekutive Ebene“ besteht aus drei Komponenten:

- der Wissensbasis, in der Wissen organisiert und repräsentiert wird;

- Regeln und Strategien, die das Wissen lenken;
- metakognitivem Wissen und metakognitiven Kontrollprozessen, welche Strategien steuern, die beim Problemlösen oder bei Verstehensprozessen benötigt werden.

Die Komponenten dieser Ebene sind trainierbar (vgl. Perleth, 2008, S. 20).

Bei diesem Modell verbinden die Autoren die angeborene Intelligenz repräsentiert durch die „architecture Ebene“ mit der erworbenen Intelligenz, dem Wissen durch die exekutive Ebene. Außerdem stellt dieses Modell eine Brücke zu der Expertiseforschung dar.

2.2.2.2 Das triarchische Intelligenzmodell nach Sternberg (1991)

Sternberg definiert Intelligenz als „die Fähigkeit, kontextuell angemessenes Verhalten in neuen Situationen oder während der Automatisierung des Umgangs mit bekannten Situationen zu zeigen. Intelligentes Verhalten wird aufgefasst als Funktion von Meta-, Ausführungs- und Wissenserwerbskomponenten“ (Asendorpf, 2004, S. 197).

Sternberg versuchte in seinem Modell, elementare kognitive Prozesse zu identifizieren. Nach Sternberg hängt der Schwierigkeitsgrad einer Aufgabe davon ab, welche und wie viele wesentliche Informationsverarbeitungsschritte zur Lösung erforderlich sind. Dementsprechend hängt die Fähigkeit eines Individuums zur Lösung einer Aufgabe davon ab, ob das intellektuelle System des Individuums die elementaren Prozeduren besitzt oder nicht (vgl. Holling & Kannig; 1999, S. 28).

Er unterscheidet dabei zwischen drei Subtheorien:

- Kontext-Subtheorie: Die Kontext Subtheorie bedeutet die Fähigkeit mit der sozialen Umwelt durch die soziokulturelle Intelligenz umzugehen. Die Kontext-Subtheorie basiert auf der praktischen Intelligenz.
- Zwei-Facetten-Subtheorie: In dieser Subtheorie wird einerseits die Fähigkeit thematisiert, mit neuen Aufgaben gut umzugehen. Andererseits verfügen gute Problemlöser über wachsende Automatisierung der Informationsverarbeitung bzw. mehr Kapazität für neue und schwierige Probleme.
- Komponenten-Subtheorie: Diese Theorie beschäftigt sich besonders mit analytischer Intelligenz. Sternberg unterscheidet hier zwischen drei Komponenten:

- Performanzkomponenten: Basisoperationen wie Kombinieren und vergleichen, Verknüpfen neuer Informationen mit altem Wissen usw. gelten als Performanzkomponenten, die zur Lösung aller Aufgaben des induktiven Denkens notwendig sind.
- Metakomponenten: Anhand der Metakomponenten kontrolliert und steuert die Person die Informationsverarbeitung.
- Wissenserwerbskomponenten sind wichtig für die Ausprägung der Intelligenz wie die Kombination von Information und die Unterscheidung zwischen bedeutsamen und unbedeutsamen Informationen (vgl. Perleth, 2008; S. 21).

Zusammenfassend kann man sagen, dass Intelligenz nach Sternberg auf dem Zusammenhang zwischen Umwelt, Erfahrungen oder Wissen und den kognitiven Intelligenzkomponenten beruht. Jede Ebene hat eine eigene Struktur. Aus diesem Grund ist das Modell von Sternberg sehr komplex und schwer empirisch zu prüfen. Der Vorteil dieses Modells ist, dass es die praktische Intelligenz berücksichtigt.

2.2.2.3 Die Theorie der multiplen Intelligenzen von Gardner

Intelligenz nach Gardner ist die Begabung, Probleme zu lösen oder Ergebnisse zu produzieren, die in einer Kultur oder mehreren Kulturen als wertvoll angesehen werden“ (Gardner & Hatch, 1989, zit. nach Tücke, 2005, S. 199). Gardner (1994, zit. nach Perleth, 2000, S. 40) nimmt mehrere unabhängige Intelligenzfaktoren an und gründet seine Annahme von „unabhängigen Intelligenzen“ auf folgenden Kriterien:

- Neurologisches Kriterium: Zentrale Funktionseinheiten einer Intelligenzdomäne sollten sich gehirnphysiologisch lokalisieren lassen.
- Phylogenetisches Kriterium: Die Intelligenzdomäne muss in der Entwicklung der Menschheit eine eigene Rolle spielen.
- Ontogenetisches Kriterium: Die Intelligenzdomäne sollte auch in der individuellen Entwicklung klar definiert werden.
- Die Intelligenzdomäne ist als unabhängig zu erkennen, wenn ihr ein autonomes Symbol- system zugeordnet wird, über das die Kommunikation in dieser Domäne stattfinden kann.
- Darüber hinaus bezieht er sich beim Nachweis, dass zwei Intelligenzbereiche eigenständig sind, auf experimentelle Ergebnisse, die zeigen, dass die Bearbeitung zweier Aufgaben aus gleichen Bereichen wegen Interferenzeffekten schwerer fallen als bei Aufgaben aus verschiedenen Bereichen.

Gardner (1994) unterscheidet zwischen sieben Arten von Intelligenzen:

- Die linguistische Intelligenz: Die Fähigkeit andere Sprachen einzusetzen, um andere Menschen zu verstehen, das Gefühl für Phonetik und den Klang der Wörter zu besitzen, Sensibilität für die Bedeutung und die Regeln der Wörter.
- Die musikalische Intelligenz: Die Fähigkeit Musik zu komponieren und zu spielen, sich mit Tonhöhe und Rhythmen zu befassen und dann das besondere Gefühl für Klang, Harmonie und für Rhythmus zu haben.
- Die logisch-mathematische Intelligenz: Die Fähigkeit mit Zahlen umzugehen, mathematische Operationen durchzuführen, abstrakt zu denken, schwierige Probleme zu erkennen und zu lösen.
- Die räumliche Intelligenz: Die Fähigkeit die visuelle Welt richtig wahrzunehmen, die ursprüngliche Wahrnehmung zu transformieren und zu modifizieren, die Ähnlichkeit zwischen unterschiedlichen Domänen zu erkennen, komplexe Formen in mehreren Dimensionen zu manipulieren, eine mentale Vorstellung zu produzieren und im Kopf zu verändern.
- Die körperlich- kinästhetische Intelligenz: Die Fähigkeit die Körperbewegung zu beherrschen und zu kontrollieren, mit Objekten umzugehen, den Körper geschickt sowohl ausdrucksvoll (Ballett) als auch zielgerichtet (Sport) einzusetzen.
- Die personale Intelligenz: Unter diesem Begriff unterscheidet Gardner zwischen intrapersonaler Intelligenz und interpersonaler Intelligenz. Während sich die intrapersonale Intelligenz mit den eigenen komplexen Gefühlen der Person beschäftigt, richtet sich die interpersonale Intelligenz nach außen, auf das Verhalten, die Absichten, Motive und Wünsche von den anderen Menschen.

Diese Intelligenzen sind nach Gardner vollkommen unabhängig. Obwohl Gardner einige Beweise oder Argumentationen für seine Theorie vorgelegt hat, fehlen ihr trotzdem die empirischen Belege. Er hat nicht über die Messung dieser Intelligenzen durch Testverfahren gesprochen. Darüber hinaus hat er die Bedeutung der kognitiven Fähigkeiten ignoriert, obwohl sie in den Bereichen wie z. B: Sprachen oder Mathematik eine große Rolle spielen.

2.2.2.4 Das Drei-Schichten-Modell der Intelligenzen (1993, 2003)

Das Drei-Schichten-Modell der Intelligenzen (vgl. Rost, 2009, S. 59-63) wurde von Carroll (1993) konzipiert. Es kann heute als das prominenteste und meist akzeptierte Modell kognitiver Fähigkeiten angesehen werden. Carroll hat sein hierarchisches Modell auf der Faktorenanalyse von 461 Datensätzen begründet, die zwischen 1925 und 1987 erhoben wurden. Daraus resultierte sein „Drei-Schichten-Modell der Intelligenzen“, bestehend aus den drei folgenden Ebenen oder Schichten:

Die erste, unterste Schicht umfasst ca. 70 spezifische Fähigkeiten oder Primärfaktoren (die nicht mit den Primärfaktoren Thurstones gleichgesetzt werden können!). In der zweiten Schicht befinden sich acht General-Sekundärfaktoren (Carroll, 1993, S. 624-625, zit. nach Rost, 2009, S. 60-61):

- 2GF-Fluide Intelligenz: grundlegende Prozesse schlussfolgernden Denkens wie Klassifikation und Begriffsbildung, unabhängig von Lernerfahrungen und Akkulturation.
- 2GC-Kristalline Intelligenz: Zur kristallisierten Intelligenz gehört das deklarative Denken, das sich durch schulische und berufliche Aufgaben sowie Erfahrungen erwerben lässt.
- 2GH-Ein Faktor, der die fluide Intelligenz mit der kristallinen Intelligenz verbindet. Nach Carroll (1993, S. 628, zit. nach Rost, 2009, S. 60) wurde dieser Faktor eingeführt, weil er in einigen Fällen als ein Sekundärfaktor aufgetreten wäre. Laut Carroll unterscheidet sich dieser Faktor von den Faktoren 2GF und 2GC nicht. Deswegen fehlt dieser Faktor in Carrolls Modellabbildung (Carroll, 2003, S. 627, zit. nach Rost, 2009, S. 61).
- 2GY-Allgemeine Gedächtnisfähigkeit: Dieser Faktor ist sehr wichtig bei jeder Aufgabe, die das Lernen oder Erinnern neuer Inhalte verlangt.
- 2GV-Breite visuelle Wahrnehmung: Dieser Faktor ist vorhanden bei jeder Aufgabe, welche die Wahrnehmung von visuellen Formen und nicht von Sprachformen erfordert.
- 2GU-Breite auditive Wahrnehmung: Dieser Faktor spielt eine Rolle bei jeder Leistung, die die Differenzierung von auditiven Mustern, Tönen oder von gesprochener Sprache verlangt.
- 2GR-Breite Abruffähigkeit: Dieser Faktor ist involviert bei jeder Leistung, die den Abruf von Konzepten und Elementen aus dem Langzeitgedächtnis erfordert.
- 2GS-Breite kognitive Schnelligkeit: Dieser Faktor ist vorhanden bei jeder Aufgabe, die von einer schnellen Informationsverarbeitung abhängig ist.
- 2GT-Verarbeitungsgeschwindigkeit-Entscheidungsgeschwindigkeit bei Reaktionszeiten: Reaktionsgeschwindigkeit bei unterschiedlichen Reaktionszeitaufgaben.

In der dritten, obersten Schicht gibt es den Generalfaktor „g“, der Spearmans „g“ ähnelt.

Der große Vorzug des Modells von Carroll ist, dass es das analytische Ergebnis vieler Daten und Studien darstellt. Außerdem integriert es unterschiedliche Theorien wie die Zweifaktorentheorie von Spearman, die multiple Faktorentheorie von Thurstone und die fluide, kristalline Intelligenz von Cattell.

Eine Erweiterung des Modells von Carroll stellt die Vereinigung des Modells mit der Theorie von Cattell-Horn (CH) mit 10 Sekundärfaktoren und 75 Primärfaktoren dar. In der Anwendung bildet die das Cattell-Horn-Carroll-Modell (CHC-Modell) die Grundlage für Intelligenztests wie z.B. den Woodcock- Johnson III-Test (Rost, 2009, S. 63; Rohrmann, 2010, S. 158).

2.3 Zusammenfassung

In diesem Kapitel geht es um das kontroverse Thema „Intelligenz“, zu dem es keine einheitliche Definition gibt, sondern unterschiedliche Meinungen, die auf verschiedenen Forschungsansätzen wie dem psychometrischen Ansatz oder dem kognitivistischen Informationsverarbeitungsansatz beruhen. Die meisten der Intelligenzforscher definieren Intelligenz als Disposition oder allgemeine Fähigkeit, die aus mehreren Fähigkeiten oder Faktoren wie z. B. der Fähigkeit, Probleme zu lösen, der Lernfähigkeit und Anpassungsfähigkeit besteht. Andere Wissenschaftler wie Gardner („multiple Intelligenz“) und Sternberg („Erfolgsintelligenz“) nehmen nicht an, dass Intelligenz eine generelle Fähigkeit ist, sondern mehrere Fähigkeiten umfasst, die erforderlich für Erfolg im beruflichen Leben sind.

Um die Struktur der Intelligenz zu analysieren und die Auswirkungen von Intelligenz zu beschreiben, wurden unterschiedliche Intelligenzmodelle mithilfe statistischer Methoden wie z.B. der Faktorenanalyse entwickelt und in diesem Kapitel vorgestellt.

Faktorenanalysen. Faktorenanalysen beruhen auf der statistischen Technik der Faktorenanalyse, die die Ergebnisse von vielen Testaufgaben zu wenigen Faktoren zusammenfassen. In diesen Modellen wird Intelligenz entweder als eine generelle Fähigkeit bezeichnet oder aus mehreren Faktoren bestehend beschrieben. Spearman stellte in seinem *Generalfaktormodell* Intelligenz als eine allgemeine Fähigkeit dar, die alle kognitiven Fähigkeiten beeinflusst. Außerdem gibt es spezifische Fähigkeiten, die mit der allgemeinen Fähigkeit bei jeder Aufgabe zusammenwirken.

Cattell hat das *Zweifaktorenmodell* in Anlehnung an bzw. als Weiterentwicklung von Spearmans Modell entwickelt. Cattell postulierte eine fluide und eine kristallisierte Intelligenz. Fluide Intelligenz nach Cattell stellt die allgemeine Fähigkeit dar und repräsentiert den „g“-Faktor bei Spearman. Die Entwicklung der fluiden Intelligenz ist unabhängig von der soziokulturellen Umwelt. Kristallisierte Intelligenz setzt sich aus der fluiden Intelligenz, sozialen Umwelt und Wissen zusammen. Die Entwicklung der kristallisierten Intelligenz ist abhängig von der soziokulturellen Umwelt. d. h. man verfügt über kristallisierte Intelligenz, wenn man in guter Lernumwelt lebt und fluide Intelligenz hat.

Laut Thurstones *Primärfaktormodell* besteht Intelligenz aus sieben unabhängigen Fähigkeiten (Memory, Number, Space, Verbal Comprehension, Reasoning, Word Fluency und Perceptual Speed). Aber anhand der Faktorenanalyse wurden auch Zusammenhänge zwischen diesen Faktoren bestätigt. Die Primärfaktoren werden demnach von der einheitlichen Fähigkeit oder vom „g“-Faktor beeinflusst. Im *Intelligenzstrukturmodell* von Guilford setzt Intelligenz sich aus drei Dimensionen zusammen: Denkinhalte (symbolische, semantische, bildliche und soziale Verhältnisinhalte), Denkprodukte (Einheiten, Klassen, Systeme) und Denkoperation (Erkenntnis, Gedächtnis, divergente und konvergente Produktion). Diese Dimensionen umfassen 120 Intelligenzfaktoren.

Kognitionspsychologisch orientierte Intelligenzmodelle. Diese Modelle beschreiben das Zusammenspiel von Gedächtnis und Denkfunktionen, die beim Lösen von Testaufgaben zusammenwirken. Beispiele sind die Konzeption von Campione und Brown (1978) und das triarchische Intelligenzmodell von Sternberg. *Campione und Brown (1978)* unterscheiden in ihrem Konzept zwischen der angeborenen Intelligenz (architecture Ebene) und erworbenen Intelligenz (exekutive Ebene). Die architecture Ebene besteht aus dem Dreispeicher-Gedächtnismodell (Kapazität, Speicherdauer, Effizient). Nach den Autoren steht lediglich Effizient mit Intelligenz in Zusammenhang. Die Exekutive Ebene setzt sich aus drei Komponenten (Wissensbasis, Regeln und Strategien, metakognitives Wissen und metakognitive Kontrollprozesse) zusammen. Diese Ebene ist nach Campione und Brown trainierbar.

Laut Sternbergs *triarchischem Intelligenzmodell* ist Intelligenz ein Komplex aus drei Subtheorien: Kontext-Subtheorie (Umwelt), Zwei Facetten-Subtheorie (Erfahrungen) und Komponenten-Subtheorie (Wissen). Jede Subtheorie hat ihre eigene Struktur. Deswegen ist dieses Modell sehr kompliziert.

Gardner setzt in seinem Modell der *multiplen Intelligenzen* Intelligenz mit dem gleich, was sonst auch als Begabung oder Talent bezeichnet wird. Er ignorierte übergreifende, allgemeine intellektuelle Fähigkeit und postulierte sieben Arten von unabhängigen Intelligenzen (linguistische, mathematische, räumliche, körperliche, musikalische, interpersonale und intrapersonale Intelligenz). Eingehende empirische Belege für dieses Modell fehlen bisher.

Ein modernes Beispiel für faktorielles Modell ist das *Berliner Intelligenzstrukturmodell* von Jäger, das als das am besten empirisch belegte psychometrische Intelligenzmodell im deutschen Sprachraum gilt. Dieses Modell verbindet Spearmans Modell mit dem Primärfaktorenmodell von Thurstone. Nach diesem Modell besteht Intelligenz aus zwei Modalitäten oder Fähigkeitsbündeln (Inhalte, Operation). Die Inhaltsdimensionen umfassen die folgenden Fähigkeiten: V (Verbal), N (numerisch), F (Figural). Die Operationsdimensionen bestehen aus vier Fähigkeiten: B (Bearbeitungsgeschwindigkeit), K (Verarbeitungskapazität), M (Merkfähigkeit), E (Einfallsreichtum): In diesem Modell spielt der „g“ Faktor eine integrierende Rolle über alle Komponenten.

Das *Drei-Schichten-Modell der Intelligenzen* von Carroll integriert verschiedene Theorien, z. B. die Zweifaktorentheorie von Spearman, die multiple Theorie von Thurstone sowie die fluide und kristallisierte Intelligenz von Cattell. Deswegen besteht es aus drei Schichten: die unterste Schicht mit 70 spezifischen Faktoren wie Wahrnehmungsgescheindigkeit, die zweite Schicht mit acht General-Sekundärfaktoren wie fluide und kristalline Intelligenz sowie die dritte Schicht mit dem Generalfaktor „g“.

3. Intelligenzmessung

3.1. Gütekriterien psychologischer Diagnostik

Bevor in diesem Kapitel auf Möglichkeiten der Intelligenzmessung eingegangen wird, soll beschrieben werden, welchen Kriterien psychologische Tests genügen müssen. Damit Intelligenztests zuverlässig sind und man mit ihnen gute Ergebnisse ermitteln kann, müssen sie bestimmte Bedingungen erfüllen, die als Gütekriterien der psychologischen Diagnostik bezeichnet werden. Die Gütekriterien eines Tests werden in Haupt- und Nebengütekriterien unterteilt (vgl. Lienert & Raatz, 1998; Perleth & Sierwald, 2000; Holling & Kannig, 1999; Ingenkamp, 1997; Bühner, 2006; Bühner, 2004), die hier in Anlehnung an die klassische Testtheorie vorgestellt werden. Zu den Hauptgütekriterien zählen Objektivität, Reliabilität und Validität.

3.1.1 Objektivität

Objektivität zeigt den Grad an, in dem die Testergebnisse nicht vom Untersucher beeinflusst sind. Ein Test wird als objektiv bezeichnet, wenn Testdurchführung und Auswertung durch verschiedene Personen bei dem gleichen Probanden zum gleichen Ergebnis führen. Verschiedene Aspekte von Objektivität wie die Durchführungsobjektivität, Auswertungsobjektivität und Interpretationsobjektivität werden im Folgendem dargestellt.

Durchführungsobjektivität. Die Testergebnisse eines Tests müssen unabhängig von den Untersuchungssituationen und vom Testleiter sein. Bei der Durchführungsobjektivität wird erläutert, wie und unter welchen Voraussetzungen ein Test oder Verfahren durchgeführt werden muss. Um die Durchführungsobjektivität zu erreichen, muss die Untersuchungssituation (Aufgabestellung, Bearbeitungszeit, zulässige Hilfsmittel, die Erklärung der Aufgaben) standardisiert werden und alle Personen müssen unter gleichen Bedingungen arbeiten. Der Testleiter muss die Instruktion des Tests wörtlich einhalten.

Um zu wissen, ob die Durchführungsobjektivität gewährleistet ist, werden die Aufgabenstellungen, die Testinstruktionen für die Probanden und die Testanweisungen für die Testleiter durch Experten begutachtet. Zusätzlich können unterschiedliche Testleiter bei der Testanwendung beobachtet werden, um herauszufinden, welche Differenzen auf nicht ausreichende Anweisungen zurückzuführen sind (vgl. Ingenkamp, 1997, S. 35, Bühner, 2006, S. 34).

Unter **Auswertungsobjektivität** versteht man, dass verschiedene Testleiter denselben Test bewerten und das gleiche Ergebnis ermitteln. Die Auswertungsobjektivität wird bei Leistungstests und bei den Fragebögen durch festgelegte Antworten gesichert, wohingegen die Auswertungsobjektivität bei Leistungstests und Fragebögen, deren Aufgaben frei beantwortet werden müssen, schwierig zu verwirklichen ist (vgl. Lienert, Raatz, 1998, S. 8). Um die Auswertungsobjektivität bei dieser Art von Tests zu erreichen, müssen die Aufgaben so entwickelt werden, dass, wenn eine offene Frage ohne festgelegte Antwort gestellt wird, die Antwort genau definiert ist. Bei den meisten der heute gebräuchlichen Testverfahren wird anhand der Auswertungsschablonen ein Rohwert gesichert, die durch Anzahl der richtigen Antworten berechnet wird. Bei manchen Testverfahren kann die Auswertungsobjektivität durch maschinenlesbare Antwortbögen und computerisierte Auswertung ermittelt werden (vgl. Perleth & Sierwald, 2000, S. 52).

Interpretationsobjektivität ist verwirklicht, wenn mehrere Auswerter die Testergebnisse gleich beurteilen oder interpretieren. Der Test-Rohwert ist allein nicht ausreichend, die Position des Probanden entlang der Testskala festzulegen. Es ist sehr wichtig bei den Intelligenztests, dass man weiß, ob seine gemessene Leistung überdurchschnittlich oder unterdurchschnittlich ist oder eine Hochbegabung vorliegt. Deswegen findet man in den Handbüchern der psychometrischen Testverfahren Normtabellen, die Vergleichswerte vorlegen. Die Test-Rohwerte werden anhand dieser Normtabellen in Werte umgesetzt, die mit bestimmten Populationen oder Bezugsgruppen verglichen werden (vgl. Perleth & Sierwald, 2000, S. 53).

3.1.2 Reliabilität

Unter Reliabilität oder Zuverlässigkeit eines Tests versteht man „Den Grad der Genauigkeit, mit dem ein Test ein bestimmtes Merkmal misst, unabhängig davon, ob er dieses Merkmal auch zu messen beansprucht“ (Bühner, 2006, S. 35). Der Grad der Zuverlässigkeit wird durch den Reliabilitätskoeffizienten geschätzt. Dieser gibt an, „in welchem Maße unter gleichen Bedingungen gewonnene Messwerte über ein und denselben Pbn übereinstimmen, in welchem Maße also das Testergebnis reproduzierbar ist“ (Lienert & Raatz, 1998, S. 9). In der klassischen Testtheorie zeigt die Reliabilität, wie genau ein Test unterschiedlich ausgeprägte Merkmale voneinander unterscheiden kann. In der Mathematik (vgl. Perleth & Sierwald, 2000, S. 58-59) wird die Reliabilität als der Anteil der Varianz von den wahren Werten an der Varianz der gemessenen Werte durch die folgende Formel definiert.

$$Rel(X) = Rel_X = \frac{\sigma_T^2}{\sigma_X^2}$$

Da die Varianz der wahren Werte immer entweder kleiner oder gleich der Varianz der gemessenen Werte ist, muss die Reliabilität zwischen 0 und 1 liegen:

$$0 \leq Rel_X \leq 1$$

Reliabilität kann auch als der Anteil der Fehlervarianz an der Varianz der gemessenen Werte definiert werden:

$$Rel_X = \frac{\sigma_T^2}{\sigma_X^2} = \frac{\sigma_X^2 - \sigma_e^2}{\sigma_X^2} = 1 - \frac{\sigma_e^2}{\sigma_X^2}$$

Die Reliabilität ergibt sich aus dem Quadrat der Korrelation zwischen den wahren und den gemessenen Werten: Die Korrelation wird als Reliabilitätsindex bezeichnet.

$$Rel_X = r_{XT}^2, \text{Reliabilitätsindex} = \sqrt{Rel_X} = r_{XT}$$

Die Reliabilität gibt an, ob die Schwankungen der Messergebnisse auf die Messfehler oder auf die Differenzen der wahren Werte zurückgehen. Reliabilität $Rel_X = 1$, wenn es keinen Messfehler gibt und $Rel_X = 0$, wenn die gesamte gemessene Varianz auf dem Meßfehler beruht, d.h. der Test misst kein Merkmal mehr.

Die Varianzen des wahren Wertes und des gemessenen Wertes können nicht unmittelbar ermittelt werden. Deswegen sind obige Formeln der Reliabilität nicht unmittelbar anwendbar. Die Reliabilität eines Tests kann zwar mathematisch definiert werden, muss aber immer empirisch-statistisch geschätzt werden. Wenn man einen parallelen und lokal stochastisch unabhängigen Test zur Verfügung hat, kann dass die Reliabilität beispielsweise über die Korrelation r_{XX} der beiden parallelen Tests abschätzen:

$$\text{Rel}_X = \text{Rel}_{X'} = r_{XX'},$$

Zur praktischen Bestimmung oder Abschätzung der Reliabilität werden unterschiedliche Aspekte oder Methoden wie Wiederholungsmethode oder Retest-Reliabilität, Paralleltestmethode bzw. Paralleltest-Reliabilität oder die (Test-)Halbierungsmethode sowie die Methode der (inneren) Konsistenz-Reliabilität verwendet, die im Rahmen der Testkonstruktion besonders häufig zur Anwendung kommen.

3.1.2.1 Stabilität bzw. Retest-Reliabilität

Bei dieser Methode wird der gleiche Test zweimal in einem zeitlichen Abstand bei gleichen Personen angewendet. Dann wird die Korrelation zwischen den Ergebnissen der Durchführungen berechnet, die den Reliabilitätskoeffizienten darstellt. Die Nachteile dieser Methode bestehen darin, dass die Reliabilität sowohl überschätzt, beispielsweise durch Testungseffekte (Erinnerung an die Lösungen der Aufgaben, besonders, wenn das Zeitintervall zwischen den Anwendungen kurz ist und wenn der Test wenige Aufgaben mit interessanten Inhalten enthält), als auch unterschätzt werden kann, wenn der Zeitabstand nach der ersten Durchführung groß ist. Darüber hinaus ist diese Methode nicht geeignet für Tests, die schwankende Merkmale messen, weil sich diese Merkmale über die Zeit verändern können. Außerdem ist diese Methode sehr aufwendig (zeitlicher und technischer Aufwand; die Stichprobe muss beim zweiten Mal wieder motiviert werden, noch einmal mitzumachen).

Diese Methode empfiehlt sich nach Bühnen (2004) vor allem bei den Schnelligkeitstests und Persönlichkeitsfragebogen für stabile Merkmale, ist aber nach Perleth & Sierwald (2000) auch bei Intelligenz- und Leistungstests wichtig, da die Stabilität der damit gemessenen Merkmale vor allem für die Prognose ein wichtiges Merkmal dieser Verfahren darstellt. Zu Ermittlung der Retest-Reliabilität kann man die folgende Formel benutzen (Lienert & Raatz, 1998, S. 181):

$$r_{tt} = \frac{N \sum_i X_{1i} X_{2i} - \sum_i X_{1i} \sum_i X_{2i}}{\sqrt{\left[N \sum_i X_{1i}^2 - \left(\sum_i X_{1i} \right)^2 \right] \left[N \sum_i X_{2i}^2 - \left(\sum_i X_{2i} \right)^2 \right]}}$$

Dabei sind:

X_{1i} : Die Rohwerte der ersten Durchführung.

X_{2i} : Die Rohwerte der zweiten Durchführung.

3.1.2.2 Paralleltest-Reliabilität

Bei dieser Methode werden zwei Tests an gleichen Personen durchgeführt. Die beiden Verfahren müssen dabei so ähnlich wie möglich sein. Die Anwendung der Parallelformen erfolgt entweder direkt nacheinander oder zu zwei verschiedenen Zeitpunkten. Dann wird die Paralleltest-Reliabilität durch die Korrelation aus den Ergebnissen der Testformen berechnet.

Aber es ist schwierig, für diese Methode zwei Tests zu konstruieren, die homogene Itemmengen enthalten und die in ihrer Trennschärfe und Schwierigkeit vergleichbar sind. Aus diesem Grund ist die Paralleltestreliabilität empirisch nicht exakt ermittelbar. Außerdem ist hier die Unabhängigkeit der Messung nicht gesichert. Bei der Durchführung der Paralleltest-Reliabilität muss berücksichtigt werden, dass eine Hälfte der Stichprobe am Anfang mit dem Test A und die andere Hälfte mit dem Test B beginnt und bei der zweiten Anwendung in umgekehrter Reihung. Der Zeitabstand zwischen den beiden Durchführungen soll höchstens wenige Tage sein. Besser ist es, wenn man sofort mit der zweiten Anwendung beginnt. Diese Methode wird häufig bei Schnelligkeitstests und Powertests verwendet.

3.1.2.3 (Test-)Halbierungsreliabilität (Split-Half- Reliabilität)

Hier wird der Test nur einmal an einer Stichprobe durchgeführt. Danach wird der Test in zwei Testhälften geteilt, die so parallel bzw. vergleichbar wie möglich sein müssen. Dann wird der Korrelationskoeffizient aus den Ergebnissen der beiden Testteile berechnet. Um die Halbierungsreliabilität zu ermitteln, können unterschiedliche Aspekte bei der Teilung des Tests verwendet werden (vgl. Bühner, 2004, S. 117):

- *Odd-Even*: Bei dieser Methode umfasst die erste Testhälfte die Aufgaben mit geraden Reihungsnummern und die zweite Testhälfte besteht aus Items mit ungeraden Reihungsnummern.

- *Zufall-Methode*: Hier werden die Aufgaben oder Items in den beiden Testhälften zufällig ausgewählt.
- *Testzeit-Methode*: Nach dieser Methode wird die ganze Testzeit in zwei Hälften geteilt. Dann gibt der Testleiter nach Ablauf der ersten Hälfte der Zeit ein Zeichen und danach machen die Probanden weiter. Diese Methode wird bei den Schnelligkeitstests verwendet.
- *Analysedaten-Methode*: Es ist hier wichtig, Aufgabenpaare zu bilden, die ähnliche Schwierigkeit und Trennschärfe haben. Dann werden die Paare isoliert und zufällig in jeder Hälfte ausgewählt.

Die Methode, die bei der Halbierung des Tests angewendet wird, hat einen Einfluss auf die berechnete Reliabilität. Wenn die Testteile gleiche Anzahl von Items enthalten und sie parallel sind, wird die folgende Formel von *Spearman und Brown* zur Berechnung der Halbierungsreliabilität verwendet (vgl. Bühner, 2004, S. 120):

$$r_{tt} = \frac{2 \cdot r_{12}}{1 + r_{12}}$$

Dabei sind:

r_{tt} : Die Reliabilität.

r_{12} : Die Korrelation aus der Testhälften 1 und der Testhälften 2.

Man kann auch zur Berechnung der Halbierungsreliabilität die einfache Formel von Guttman (aus dem Jahr 1945) benutzen, die lautet (vgl. Bühner, 2004, S. 120):

$$r_{tt} = 2 \cdot \left[1 - \frac{S_1^2 + S_2^2}{S_x^2} \right]$$

Dabei sind:

r_{tt} : Die Reliabilität.

S_x : Die Gesamtstreuung des Tests.

S_1 : Die Streuung der Testhälften 1.

S_2 : Die Streuung der Testhälften 2.

Wenn die Größe der Stichprobe klein ist, unterschätzen diese beiden Formeln (Spearman- Brown und Guttman) die Reliabilität. Deswegen wird die Formel von Kristof (aus dem Jahr 1963) verwendet (vgl. Bühner; 2004, S. 121):

$$r_{tt} = \frac{2}{N-1} + \frac{N-3}{N-1} \left[\frac{4 \cdot S_1 \cdot S_2 \cdot r_{12}}{S_1^2 + S_2^2 + 2 \cdot S_1 \cdot S_2 \cdot r_{12}} \right]$$

Dabei ist:

N: Die Anzahl der Probanden

r_{12} : Die Korrelation aus den Testhälften 1 & 2

S_1 : Streuung Testhälft 1

S_2 : Streuung Testhälft 2

3.1.2.4 Innere Konsistenz

Bei dieser Methode (vgl. Perleth & Sierwald, 2000, S. 61-62) wird der Test in so viele Teile aufgeteilt, wie der Test Aufgaben besitzt. Jede Aufgabe wird als eigenständiger Testteil angesehen. Dann wird die Reliabilität der Aufgaben durch unterschiedliche Formeln geschätzt und für den Gesamttest aufgewertet.

Die Formel von Cronbach (aus dem Jahr 1951) wird zur Ermittlung der inneren Konsistenz benutzt. Aber sie setzt eigentlich äquivalente Items voraus. Wenn diese Voraussetzung nicht erfüllt ist, legt der Koeffizient α von Cronbach eine „untere Grenze“ der Reliabilität vor. Die Formel lautet:

$$\alpha = \frac{n}{n-1} \left(\frac{S_X^2 - \sum S_{xi}^2}{S_X^2} \right)$$

Dabei sind:

N: Die Anzahl der Testitems.

S_X^2 : Die Varianz der Gesamtrohwerte.

S_{xi}^2 : Die Varianz der Testitems.

Für dichtome Aufgaben kann die Formel von *Kuder-Richardson* verwendet werden, die lautet:

$$r_{tt} = \frac{n}{n-1} \cdot \left(\frac{S_X^2 - \sum p_i q_i}{S_X^2} \right)$$

Dabei ist:

p_i : Schwierigkeit der einzelnen Testaufgaben

q_i : $1 - p_i$

n: Anzahl der Items

S_X^2 : Varianz der Testrohwerte

3.1.2.5 Einflussfaktoren auf die Reliabilität

Wie oben angegeben wurde, ist liefern die genannten Reliabilitätskoeffizienten nicht immer ein genaues Maß für die Reliabilität, sondern sie hängt von einigen Faktoren ab, durch welche die Reliabilität über- oder unterschätzt werden kann. Diese Faktoren sind (Bühner, 2004; Perleth & Sierwald, 2000):

Homogenität und Heterogenität des Tests. Während homogene Tests stets reliabel sind, weil sie ähnliche oder gleiche Aufgaben umfassen, sind heterogene Tests beschränkt reliabel, weil sie Items unterschiedlicher Charakteristik beinhalten. Darüber hinaus hängt die Höhe der inneren Reliabilität nicht nur von Homogenität der Items ab, sondern auch von der Länge des Tests. Die Reliabilität nach Cronbachs Alpha fällt höher aus, wenn mehr homogene Aufgaben einer Skala hinzugefügt werden, d.h. je homogener der Test ist, desto höher fällt die Reliabilität aus (vgl. Bühner, 2004, S. 126).

Streuung der Testkennwerte. Die Streuung der Testkennwerte übt einen wichtigen Einfluss auf die Reliabilität aus. Während eine hohe Streuung normalerweise zu hohen Reliabilitäten führt, lässt eine geringe Streuung keine hohen Reliabilitäten zu. Die Streuung ist aber auch von der Itemschwierigkeit abhängig. Die Itemstreuung bei dichtomen Aufgaben ist rechnerisch vollkommen durch die Itemschwierigkeit bestimmt, aber das gilt nicht für intervallskalierten Aufgaben wegen möglicher Boden- oder Deckeneffekte, die geringe Trennschärfe zur Konsequenz

haben (vgl. Bühner, 2004, S. 91). Aber es kann trotzdem aus einer geringen oder hohen Schwierigkeit eine hohe Streuung resultieren, weil die erhöhte Streuung durch einen oder mehrere Probanden, die eine Aufgabe anders als die übrigen Probanden bearbeitet haben, zustande kommen oder weil die Stichprobe sehr heterogen ist (vgl. Bühner, 2004, S. 127).

Meßfehler. Laut Schmidt und Hunter (1999, zit. nach Bühner, 2004, S. 127-128) werden verschiedene Meßfehler unterschieden:

- *Zufällige Antwortmessfehler* innerhalb einer Durchführung wegen Ermüdung der Testpersonen oder der Verschiebung der Aufmerksamkeit auf etwas anders.
- *Systematische Fehler*, die z. B. bei Wiederholung einer Messung über unterschiedliche Messzeitpunkte (Übungs- und Lerneffekte, Antworttendenzen usw.) auftreten.
- *Spezifische Fehler*: Wenn die Probanden gleiche Items unterschiedlich begreifen.
- *Vorübergehende Fehler*, die einen variierenden Messfehler über verschiedenen Messungen zeigen.

Die Methode der Reliabilitätsberechnung. Einige Methoden der Reliabilitätsberechnung benötigen zwei Messungen wie bei der Stabilität bzw. Retest-Reliabilität, eine Messung wie bei der Konsistenz-Reliabilität oder bei der Verwendung zweier Testformen, die je einmal durchgeführt werden (Paralleltest-Reliabilität). Das Zeitintervall über verschiedene Messzeitpunkte und Testformen stellen Fehlerquellen dar, die einen Einfluss auf die berechnete Reliabilität ausüben können. Perleth und Sierwald (2000, S. 63) unterscheiden folgende Fehlervarianzquellen bei jeder Methode: Untersuchungssituation, Probandenfaktoren und Auswertungsfehler bei Wiederholungs-Reliabilität, Paralleltest-Reliabilität und Konsistenz-Reliabilität.

Darüber hinaus gibt es eigene Fehlervarianzquellen bei jeder Methode, z. B. Wiederholungseffekt bei der Retestreliabilität, mangelnde Parallelität der Testformen bei der Paralleltest-Reliabilität und mangelnde Homogenität des Meßverfahrens (bzw. der Items) bei der Konsistenz-Reliabilität. Diese Fehlerquellen beeinflussen unsystematisch die beiden Messungen bei der Wiederholungsreliabilität und der Paralleltestreliabilität sowie im Zuge der einzigen Messung im Falle der Konsistenzreliabilität. Aber diese Effekte wirken zwischen zwei Testungen stärker als während einer Testung. Deshalb fällt der Reliabilitätskoeffizient bei der Konsistenz-Reliabilität in der Regel höher als bei der Wiederholungs-Reliabilität und der Paralleltest-Reliabilität aus (vgl. Perleth & Sierwald, 2000, S. 63-64).

3.1.3 Gültigkeit oder Validität

Validität ist das wichtigste Hauptgütekriterium bei Intelligenztests. Unter Validität versteht man „den Grad der Genauigkeit, mit dem dieser Test dasjenige Persönlichkeitsmerkmal oder diejenige Verhaltensweise, das (die) er messen soll oder zu messen vorgibt, tatsächlich misst“ (Lienert, 1969, S. 16). Es gibt aber keine mathematische Definition von Validität wie bei der Reliabilität. Perleth & Sierwald (2000) wie viele andere Autoren differenzieren zwischen drei Validitätsarten: Inhaltsvalidität, Konstruktvalidität und Kriteriumsvalidität.

3.1.3.1 Inhaltsvalidität

Unter Inhaltsvalidität versteht man, dass ein Test oder eine Testaufgabe das zu messende Merkmal wirklich erfasst. Der Ausgangspunkt der Inhaltsvalidität ist, dass die Testaufgaben eine repräsentative Auswahl von dem Universum der Testaufgaben darstellen müssen. Die Inhaltsvalidität wird nach Perleth & Sierwald (2000, S. 68) aufgrund logischer Überlegung und nicht auf mathematische Weise ermittelt. Deshalb gibt es keinen Koeffizienten zur Inhaltsvalidität, sondern es werden die Beurteilungen von unterschiedlichen Experten erhoben und miteinander verglichen. Murphy und Davidshofer (2001, zit. nach Bühner, 2004, S. 31) erwähnen ein hilfreiches Vorgehen, wie man Inhaltsvalidität bestimmen kann:

Zuerst wird die Inhaltsebene des gemessenen Gegenstands wie Fähigkeiten und Eigenschaften geschildert. Dann werden für jede Ebene oder für jeden Bereich die entsprechenden Items festgestellt. Danach wird die Teststruktur mit den Inhaltsebenen des Konstruktks verglichen. Die logische Überlegung ist allein nicht genug für die Testkonstruktion, es folgt eine statistische Absicherung. Bei der Testkonstruktion wird das Konstrukt in verschiedene Facetten untergliedert. Dann werden für jede Facette viele Items konstruiert. Danach und anhand der Faktorenanalyse verbleiben nur jene Items, die inhaltlich zu einem Faktor passen, die unpassenden Items werden ausgeschieden (vgl. Bühner, 2006, S. 36-38).

3.1.3.2 Kriteriumsvalidität

Die Kriteriumsvalidität bezeichnet den Zusammenhang oder die Korrelation zwischen den Testergebnissen des Messinstruments und einem Kriterium und setzt voraus, dass das Kriterium das

gemessene Konstrukt erfassen oder den Testzweck darstellen soll (vgl. Perleth & Sierwald, 2000, S. 68). Die Autoren geben an, wie man die Kriteriumsvalidität durch das folgende Vorgehen ermitteln kann:

Zunächst werden die Testwerte (X) an einer Stichprobe aus der Population erhoben. Dann werden auch die Testwerte des Kriteriums (Y) an den gleichen Personen erhoben. Danach wird die Korrelation berechnet. Der Korrelationskoeffizient $r(X, Y)=r_{XY}$ stellt die Kriteriumsvalidität von X bezüglich Y dar.

Bei der Kriteriumsvalidität werden vier Aspekte unterschieden (vgl. Bühner, 2006, S. 38):

- *Vorhersagevalidität:* Bei der Vorhersagevalidität wird die Korrelation zwischen dem Testergebnis und einem Kriterium berechnet. Die Testergebnisse und das Kriterium werden jedoch nicht zur gleichen Zeit erhoben, sondern das Kriterium wird später erhoben. Beispielsweise wird ein Intelligenztest vor Beginn des Studiums mit dem Kriterium am Ende des Studiums (Abschlussnoten) korreliert.
- *Übereinstimmungsvalidität:* Hier wird die Korrelation zwischen dem Testergebnis und dem Kriterium fast zugleich berechnet, wie z.B. die Korrelation zwischen dem Ergebnis des Rechentests und dem Lehrerurteil.
- *Retrospektive Validität:* Bei der Retrospektiven Validität wird die Korrelation mit einem vorher vorhandenen Kriterium berechnet. Beispielsweise wird ein Intelligenztest während des Studiums angewendet und mit Schulnoten als Kriterium korreliert.
- *Inkrementelle Validität:* Bei dieser Methode wird ein Test zur Verbesserung der Prognose eines Kriteriums durch einen anderen zusätzlichen Test oder eine andere Methode benutzt. Beispielsweise werden Intelligenztests zur zusätzlichen, besseren Vorhersage des Berufserfolgs benutzt. Eine weitere Methode wie z. B. das Interview kann noch etwas zur Prognose von Berufserfolg beitragen. Diese Methode kann durch Regressionsanalysen ermittelt werden.

Laut Perleth & Sierwald (2000, S. 68-70) bestehen die Probleme der Kriteriumsvalidität darin, dass sie von der Reliabilität des Messinstruments, des Kriteriums und vom Ausmaß, in dem das Messinstrument und das Kriterium etwas Gemeinsames erfassen, abhängt. Je höher die Reliabilität des Tests und des Kriteriums ist, umso höher kann der Validitätskoeffizient ausfallen. Aus diesem und den Reliabilitäten kann man die korrigierte Korrelation der wahren Werte durch die folgende Formel berechnen:

$$r_{korr} = r_{T_X T_Y} = \frac{r_{XY}}{\sqrt{\text{Rel}_X \cdot \text{Rel}_Y}}$$

Außerdem kann man auch die maximale Kriteriumsvalidität aus dem Ergebnis der Reliabilität eines Tests (falls diese angegeben wurde) durch die Formel des *Reliabilitätsindex* berechnen:

$$\text{Reliabilität sin dex} = \max(r_{XY}) = \sqrt{\text{Rel}_X}$$

Nach Ansicht von Perleth und Sierwald (2000, S. 69) ergibt sich bei der Durchführung eines Tests auch das Problem, wie und mit welcher Zuverlässigkeit aus dem Testwert eines Probanden der dazugehörige Kriteriumswert berechnet werden kann. Zu diesem Zweck muss vor allem der Schätzwert \hat{y}_i für das Kriterium aus dem Messwert x_i durch die Regressionsformel berechnet werden:

$$\hat{y}_i = r_{XY} \frac{S_Y}{S_X} (x_i - M_X) + M_Y$$

Dabei ist:

r_{XY} : Der Validitätskoeffizient.

Nach der Berechnung des Schätzwerts des Kriteriums muss das Vertrauensintervall des Kriteriums abgeschätzt werden. Dies setzt voraus, dass die Schätzfehler normalverteilt sind und der Standardfehler für alle Probanden gleich bleibt. Das Vertrauensintervall wird durch die folgende Formel angegeben:

$$\hat{y}_i - z_{\alpha} S_E \leq y \leq \hat{y}_i + z_{\alpha} S_E$$

Dabei ist:

S_E : Der Standardschätzfehler, der nach folgender Formel berechnet wird:

$$S_E = S_y \sqrt{1 - r_{XY}^2}$$

Die Formel des Vertrauensintervalls ist so konstruiert, dass die „wahren“ Kriteriumswerte mit der Irrtumswahrscheinlichkeit α innerhalb der Intervallgrenzen liegen (vgl. Perleth & Sierwald, 2000, S. 69-70).

3.1.3.3 Konstruktvalidität

Unter Konstruktvalidität versteht man, dass die Messungen das Konstrukt erfassen, das erfasst werden soll. Nach Lienert und Raatz (1998) ist die Konstruktvalidierung „weniger pragmatisch als theoretisch orientiert. Ihre Bedeutung liegt daher nicht in einer unmittelbaren praktisch-diagnostischen Verwertbarkeit, sondern in der theoretischen Klärung dessen, was der betreffende Test misst“ (S. 226).

Ein Konstrukt wird als eine theoretische Variable bezeichnet, die nicht operational fassbar ist, sondern durch eine Theorie (vgl. Perleth & Sierwald, 2000, S. 71). Zum Beispiel kann man Intelligenz nicht durch eine Menge von Verhaltensweisen definieren, sondern durch eine psychologische Theorie. Die Konstrukte können durch empirische Indikatoren wie z. B. Tests, Fragebogen usw. gemessen werden. Aus einer Theorie können Hypothesen über die Beziehungen zwischen einem Konstrukt und anderen Konstrukten aufgestellt und diese Hypothese mit Hilfe der empirischen Indikatoren überprüft werden.

Bei der Beurteilung der Konstruktvalidität formuliert man a priori konkrete Erwartungen über den Zusammenhang des vorliegenden Tests mit konstruktverwandten konvergenten und diskriminanten Tests (vgl. Bühner, 2004, S. 32-33). Es wird hier also zwischen der konvergenten Validität und der diskriminanten Validität unterschieden:

- *Konvergente Validität.* Hier wird die Korrelation mit Tests, die das gleiche oder ähnliche Merkmal messen, berechnet wie z. B. die Korrelation zwischen einem neuen Intelligenztest und einem bereits etablierten Meßverfahren z. B. dem HAWIE-R (Tewes, 1991). Bei diesem Aspekt muss die Korrelation hoch sein.
- *Diskriminante/divergente Validität.* Hier wird die Korrelation mit Tests, die andere Konstrukte oder andere Merkmale erfassen, ermittelt, wie z. B. die Korrelation eines Konzentrationstests mit einem Arbeitsgedächtnisleistungstest. Hier muss und soll der Konzentrationstest gerade nicht die Eigenschaft „Arbeitsgedächtnisleistung“ messen, sondern eben die Eigenschaft „Konzentration“. Deswegen kann man erwarten, dass die Korrelation hier niedrig ausfällt.

Um die Konstruktvalidität zu bestimmen, werden drei unterschiedliche Methoden angewandt (vgl. Bühner, 2004, S. 32-33):

- **Korrelationsmethode:** Bei dieser Methode werden die Zusammenhänge von einem Test mit konstruktnahen und konstruktfernen Verfahren erläutert, d. h. es wird überprüft mit welchen Variablen der Test korreliert und mit welchen nicht.
- **Faktorielle Validität:** Diese Methode untersucht die Zusammenhänge von unterschiedlichen Tests anhand der Faktorenanalyse. Die Faktorenanalyse fasst die konstruktnahen Variablen, homogenen Tests oder die Tests die einen positiven Zusammenhang zeigen, zusammen, welche folglich auf einem Faktor laden. Außerdem trennt sie diese von konstruktfernen Tests, mit denen sich kein oder ein lediglich geringer Zusammenhang zeigt.
- **Multitrait-Multimethod Ansatz:** Bei dieser Methode werden unterschiedliche Matrizen erzeugt:
 - **Monotrait-Monomethoden-Matrix:** Diese Methode zeigt die Korrelation zwischen der Messung eines Merkmals (z. B. Intelligenz) mit der gleichen Methode (z. B. Intelligenztests). Hier sollte die Korrelation hoch sein.
 - **Monotrait-Heteromethoden-Matrix:** Diese Methode zeigt die Korrelation zwischen der Messung einer Eigenschaft (z. B. Intelligenz) mit unterschiedlichen Methoden (z. B. Intelligenztest und Intelligenzbeobachtung). Hier sollte die Korrelation zwischen Indikatoren unterschiedlicher Informationsquellen niedriger ausfallen.
 - **Heterotrait-Monomethode- Matrix:** Hier wird die Korrelation zwischen verschiedenen Eigenschaften (z. B. Intelligenz und Konzentration) mit gleicher Methode (z. B. Tests) berechnet. Die Korrelation sollte hier ebenfalls niedrig ausfallen.
 - **Heterotrait-Heteromethoden-Matrix:** Die Korrelation wird zwischen unterschiedlichen Eigenschaften (z. B. Intelligenz und Konzentration) mit verschiedenen Methoden (z. B. Test und Verhaltenbeobachtung) ermittelt. Diese Korrelationen sollten am niedrigsten ausfallen.

3.1.4 Nebengütekriterien

Neben Hauptgütekriterien gibt es andere Voraussetzung für einen psychometrischen Test und diese Voraussetzungen werden *Nebengütekriterien* genannt. Man unterscheidet hier zwischen den vier Nebengütekriterien: Normierung, Vergleichbarkeit, Ökonomie und Nützlichkeit (vgl. Bühner, 2004, S. 34; Lienert & Raatz, 1998, S. 11-13).

Normierung. Jeder Test soll über Angaben oder Normtabellen verfügen, die als Bezugssystem für die Einordnung des individuellen Testwertes dienen können und Informationen darüber liefern, ob ein Proband unterdurchschnittlich oder überdurchschnittlich im Vergleich zu seiner Altersgruppe abgeschnitten hat.

Vergleichbarkeit. Ein Test gilt als vergleichbar, wenn eine oder mehrere Parallelformen, die einen Vergleich des Tests mit sich selbst erlaubt, vorliegen. Beide Parallelformen besitzen gleiche Gütekriterien.

Ökonomie. Ein Test wird als ökonomisch angesehen, wenn

- die Durchführung des Tests nicht viel Zeit kostet und nicht viel Material verbraucht.
- der Test als Gruppentest angewendet wird.
- er einfach zu handhaben ist.
- die Auswertung des Tests schnell und bequem ist.

Der Test ist sehr ökonomisch, wenn er alle Bedingungen oder die Wichtigsten erfüllt. Treffen nur eine oder keine der Bedingungen zu, handelt es sich um kein ökonomisches Verfahren.

Nützlichkeit. Ein Test gilt als nützlich, wenn ein Persönlichkeitsmerkmal oder eine Verhaltensweise erfasst oder vorhergesagt wird und für die Untersuchung von diesem Merkmal ein praktisches Bedürfnis besteht.

3.1.5 Zusammenfassung

In diesem Kapitel werden Gütekriterien von psychodiagnostischen Messverfahren behandelt, die als Voraussetzungen für eine gute Messmethode angesehen werden. Gütekriterien werden in *Haupt- und Nebengütekriterien* unterteilt. Hauptgütekriterien sind: Objektivität, Reliabilität und Validität.

- **Objektivität** bedeutet, dass die Testwerte unabhängig vom Testleiter und Testauswerter sind. Hier wird zwischen Durchführungsobjektivität, Auswertungsobjektivität und Interpretationsobjektivität unterschieden:
 - *Durchführungsobjektivität* bedeutet, dass die Testergebnisse nicht vom Testleiter abhängen. Für eine hohe Durchführungsobjektivität muss die Untersuchungssituation standardisiert werden

- Die *Auswertungsobjektivität* ist gegeben, wenn zwei oder mehrere Personen den gleichen Test bewerten und zum gleichen Ergebnis kommen.
- *Interpretationsobjektivität* ist hoch, wenn mehrere Auswerter die Testergebnisse eines Tests gleich beurteilen und interpretieren.
- **Reliabilität** zeigt, wie genau ein Test das Merkmal misst, das er messen soll. Bei der Reliabilität werden die Messfehler berücksichtigt, weil die Messung eines Persönlichkeitsmerkmals von Messfehlern abhängt. Reliabilität wird mathematisch als den Anteil der Varianz der wahren Werte an den gemessenen Werten definiert. In der Statistik wird die Reliabilität als Korrelationskoeffizient bestimmt, der von unterschiedlichen Variablen wie z. B. der Homogenität und Heterogenität des Tests, der Methode der Reliabilitätbestimmung usw. abhängt. Bei der Reliabilität werden vier Arten unterschieden:
 - *Innere Konsistenz und Testhalbierung*: Diese Methoden ähneln sich, weil der Test bei den beiden nur einmal an einer Gruppe durchgeführt wird. Außerdem wird der Test bei diesen Arten in zwei (bei der Testhalbierung) oder mehrere Teile (bei der inneren Konsistenzanalyse) aufgeteilt. Danach wird die Reliabilität aus der/den Korrelation/en der Ergebnisse der beiden oder mehreren Testteile berechnet, die dann noch anhand einer Korrekturformel aufzuwerten ist.
 - *Stabilität/Retest-Reliabilität*: Hier wird das gleiche Messverfahren in einem Zeitabstand an der gleichen Stichprobe durchgeführt. Danach wird die Retest-Reliabilität durch die Korrelation zwischen den Ergebnissen der beiden Durchführungen bestimmt.
 - *Paralleltest-Reliabilität*: Diese Methode bedeutet, dass zwei Verfahren (Testformen) an den gleichen Probanden angewendet werden. Aber sie setzt voraus, dass sich die beiden Testformen so ähnlich wie möglich sind. Dann wird die Korrelation zwischen den Ergebnissen der beiden Testformen berechnet, die die Paralleltest-Reliabilität ausdrückt.
- **Gültigkeit/Validität** ist ein wichtiges Kriterium zu Beurteilung der Gültigkeit eines Testverfahrens und der Messergebnisse. Insbesondere erklärt die Validität, ob ein Verfahren wirklich das Merkmal misst, was es zu messen vorgibt, d. h. die Validität bezieht sich auf den Inhalt der Messung und nicht auf die Genauigkeit der Messung wie bei der Reliabilität. Hier unterscheidet man zwischen drei Validitätsaspekten:
 - *Inhaltsvalidität* liegt vor, wenn ein Verfahren oder Testitem ein bestimmtes Merkmal in all seinen Facetten tatsächlich erfasst, die Aufgaben also eine repräsentative Menge aus dem Universum der Testitems darstellen. Inhaltsvalidität kann nicht auf numerische oder mathematische Weise bestimmt werden, sondern durch Expertenurteil oder logische Überlegungen.

- *Kriteriumsvalidität* überprüft die Korrelation des Testergebnisses mit einem Kriterium. Hier werden nach dem Zeitpunkt der Erhebung des Kriteriums die folgenden Arten von Kriteriumsvalidität unterschieden:
 - *Vorhersagevalidität*: Bei diesem Aspekt wird das Kriterium später (in der Zukunft) erhoben und nicht gleichzeitig mit dem Verfahren.
 - *Übereinstimmungsvalidität*: Die Erhebung des Kriteriums findet zur gleichen Zeit mit der Erhebung des Tests statt.
 - *Retrospektive Validität*: Hier erfolgt die Erhebung des Kriteriums vor der Erhebung des Testergebnisses.
 - Außerdem gibt es noch einen anderen Aspekt der Kriteriumsvalidität. Dieser Aspekt ist die *inkrementelle Validität*, welche die Verstärkung der Vorhersage eines Kriteriums über einen anderen zusätzlichen Test hinaus kennzeichnet.
- *Konstruktvalidität* ist der letzte Aspekt der Validität (Gültigkeit), die angibt, inwieweit ein Messinstrument ein bestimmtes Konstrukt misst, das gemessen werden soll. Das Konstrukt wird als eine theoretische Variable definiert. Um Konstruktvalidität zu überprüfen, muss das Konstrukt von einer Theorie ausgehen und anhand empirischer Indikatoren untersucht werden. Deswegen wird zwischen konvergenter und diskriminanter Validität unterschieden:
 - *Konvergente Validität* bezeichnet die Übereinstimmung zwischen den Testergebnissen zweier Verfahren, die das gleiche oder ähnliche Konstrukt erfassen.
 - *Diskriminante Validität* beinhaltet, dass es im Idealfall keine Übereinstimmung zwischen den Testergebnissen zweier Tests gibt, die unterschiedliche Konstrukte erfassen.
- **Nebengütekriterien** zur Beurteilung der Qualität des psychologischen Tests sind: Normierung, Vergleichbarkeit, Ökonomie, und Nützlichkeit:
 - *Normierung* ermöglicht Leistungen anhand von Normentabellen mit jenen anderer Personen zu vergleichen.
 - *Vergleichbarkeit*: Der Test sollte mehrere Parallelformen besitzen.
 - *Ökonomie*: Ein Test sollte ökonomisch (Zeit, Durchführung, Auswertung, usw.) sein.
 - *Nützlichkeit*: Wenn der Test ein praktisches Bedürfnis befriedigen kann.

3.2 Messung von Intelligenz

Nach der Darstellung von Intelligenzmodellen und den Intelligenztheorien im 2. Kapitel, die den Intelligenztests zugrunde liegen, werden in diesem Kapitel Intelligenz- und kognitive Fähigkeitstests vorgestellt. Die Auswahl dieser Testverfahren berücksichtigt die am häufigsten von Psychologen/innen in Deutschland und dem deutschsprachigen Raum angewendeten Intelligenz- und kognitiven Fähigkeitstests sowie die regelmäßig in der Literatur dargestellten Verfahren (vgl. etwa Roth, Schmidt & Herzberg, 2010; Perleth & Heller, 1995; Heller & Perleth, 2000; Holling, Preckel & Vock, 2004; Preckel, 2010). Diese Verfahren werden für die folgende Darstellung in zwei Gruppen, nämlich Individualtests und Gruppentests, eingeteilt.

3.2.1. Individualtests

3.2.1.1 *Hamburg-Wechsler-Intelligenztest für Kinder (HAWIK-IV) in der zweiten aktualisierten deutschen Fassung von Petermann und Petermann (2008)*

Testart. Der Hamburg-Wechsler- Intelligenztest (HAWIK-IV) von Petermann und Petermann (2008) ist ein Intelligenztest und geeignet für Kinder im Alter von 6;0 bis 16;11 Jahren. Der HAWIK-IV dient nach Aussage der Autoren auch zur Bestimmung der kognitiven Fähigkeiten bei hochbegabten Kindern, bei Kindern mit Aufmerksamkeitsdefiziten, bei Kindern mit einer leichten Intelligenzminderung, bei Kindern mit Lernstörungen und bei Kindern mit anderer Sprache.

Testaufbau. Der Test HAWIK-IV umfasst 15 Untertests bzw. zehn Kerntests, die aus dem HAWIK-III übernommen wurden, sowie fünf neue optionale Untertests. Beim HAWIK-IV kann ein Gesamt-IQ als Maß für die kognitiven Fähigkeiten eines Kindes und einzelne IQ-Werte für die vier Indices Sprachverständnis (SV), Wahrnehmungsgebundenes logisches Denken (WLD), Arbeitsgedächtnis (AGD) und Verarbeitungsgeschwindigkeit (VG) berechnet werden. Der Gesamt-IQ ergibt sich aus der Berechnung der Kerntests und nicht aus den optionalen Tests. Die Kerntests und die optionalen Subtests sind in Tabelle 3.1 zusammengestellt.

Der HAWIK-IV wird als Einzeltest verwendet. Er muss standardisiert angewendet werden, um seine Zuverlässigkeit zu sichern. Die Testanwendung dauert 65 bis 90 Minuten. Die Untertests werden in einer bestimmten Reihenfolge durchgeführt (vgl. Tabelle 3.1).

Tabelle 3.1: Skalen des HAWIK-IV (Numerierung zeigt die Reihenfolge der Durchführung an)

Kerntests	Optionale Untertests
2. Gemeinsamkeiten finden	13. Allgemeines Wissen
6. Wortschatz-Test	15. Begriffe erkennen
9. Allgemeines Verständnis	11. Bilder ergänzen
1. Mosaik-Test	14. Rechnerisches Denken
4. Bildkonzept	12. Durchstreich-Test
8. Matrizen-Test	
3. Zahlen nachsprechen	
7. Buchstaben-Zahlen-Folgen	
5. Zahlen-Symbol-Test	
10. Symbol-Suche	

Die Auswertung des HAWIK-IV erfolgt anhand eines Protokollbogens, auf dem das Lebensalter des Kindes bestimmt wird und die Bewertungen der einzelnen Untertestaufgaben zu Rohwerten summiert werden. Anschließend werden die Rohwerte in Wertpunkte umgerechnet. Die Summe der Wertpunkte können in die Index-Werte und den Gesamt-IQ umgewandelt werden. Die Wertpunkte der Untertests, die Index-Werte und der Gesamt-IQ-Wert können auch grafisch dargestellt werden. Bei der Interpretation der Skalen des HAWIK-IV dürfen die einzelnen Ergebnisse der Skalen, die wichtigen Informationen über die intellektuelle Leistungsfähigkeit eines Kindes liefern, nicht isoliert betrachtet werden.

Erfassungsdimensionen. Der HAWIK-IV bezieht sich auf die theoretische Grundlage (vgl. Petermann & Petermann, 2008, S. 21), dass Intelligenz sowohl ein globales Konstrukt ist, da Intelligenz das Verhalten einer Person als Ganzes bestimmt, als auch ein spezifisches, da Intelligenz aus Faktoren besteht, in denen sich die Menschen unterscheiden. Dieser Test erfasst somit die allgemeine Intelligenz in Form von spezifischen Fähigkeiten: Sprachverständnis (SV), Wahrnehmungsgebundenes logisches Denken (WLD), Arbeitsgedächtnis (AGD) und Verarbeitungsgeschwindigkeit (VG).

Testgütekriterien. In Bezug auf die *Objektivität* des HAWIK-IV ist die **Durchführungsobjektivität** durch allgemeine Durchführungsanweisungen und spezielle Anweisungen für jeden Untertest zu Start-, Umkehr- und Abbruchregeln, sowie Antwortbeispielen im Manual gewährleistet.

leistet. Die Auswertungsobjektivität ist durch den Protokollbogen, auf dem die Antworten des Kindes bewertet werden, gesichert.

Die **Reliabilität** der Untertests, Indizes und der Gesamtleistung des HAWIK-IV (vgl. Petermann & Petermann, 2008, S. 113-115) wurde nach der Split-Half-Methode berechnet, mit Ausnahme der Untertests zur Verarbeitungsgeschwindigkeit (Zahlen-Symbol-Test, Symbol-Suche und Durchstreich-Test), weil für diese Untertests die Stabilität durch die Korrelationen zwischen den Testwerten in zwei Testungen von 103 Kindern aus drei Altersgruppen, 6-7, 11-12 und 15-16, bestimmt wurde. Die mittleren Reliabilitäten im Manual zeigen, dass die Koeffizienten der einzelnen Untertests ($r = .76$ bis $r = .91$) niedriger als die Reliabilitäten der Index-Werte ($r = .87$ bis $r = .94$) und des Gesamt-IQ ($r = .97$) ausfallen.

Zur **Validität** des HAWIK-IV (vgl. Petermann & Petermann, 2008, S. 125-144) wurde die Konstruktvalidität mit Hilfe der Faktorenanalyse überprüft. Es gibt Angaben zur Faktorenanalyse für die Kerntests über alle Altersgruppen und hinsichtlich separater Altersgruppen: 6-7, 8-10, 11-13 und 14-16, sowie Angaben über alle 15 Untertests. Die Befunde der Faktorenanalyse bestätigen die vorgeschlagene Vier-Faktoren-Struktur in Sprachverständnis, Wahrnehmungsgebundenes logisches Denken, Arbeitsgedächtnis und Verarbeitungsgeschwindigkeit.

Bezüglich der Kriteriumsvalidität wurde der HAWIK-IV mit dem HAWIK-III verglichen. Zum Vergleich der beiden Testverfahren wurden diese einer Stichprobe von 100 Kindern und Jugendlichen in ausbalancierter Reihenfolge zur Bearbeitung vorgelegt. Der Abstand zwischen den beiden Testungen lag bei 35 Tagen. Bei 54 Kindern, die zunächst den HAWIK-IV bearbeiteten, variierten die Korrelationen der Untertests von $r = .34$ bis $r = .79$. Außerdem ergaben sich Korrelationen in den IQ-Werten und den Indizes zwischen $r = .50$ und $r = .97$, sowie für den Gesamt-IQ von $r = .73$. Bei 46 Kindern, die zuerst den HAWIK-III bearbeiteten, ergaben die Korrelationen der beiden Testverfahren in den Subtests Werte von $r = .43$ bis $r = .72$ und in den IQ-Werten sowie den Indizes von $r = .35$ bis $r = .80$. Darüber hinaus ergibt sich eine Korrelation bei dem Gesamt-IQ von $r = .63$.

Die **Normierung** des HAWIK-IV wurde an einer Stichprobe von 1650 Kindern und Jugendlichen im Alter von 6 bis 16 Jahren vorgenommen (vgl. Petermann & Petermann, 2008, S. 48-49). Die Daten wurden zwischen März 2005 und Mai 2006 in Deutschland, Österreich und der Schweiz erhoben und unter Angaben des statistischen Bundesamtes der BRD nach unterschiedlichen Merkmalen wie Länder, Geschlecht, Schultyp usw. geschichtet.

3.2.1.2 Das Adaptive Intelligenz Diagnostikum (AID2) in der deutschen Bearbeitung von Kubinger und Wurst (2000)

Testart. Das Adaptive Intelligenz Diagnostikum (AID 2) von Kubinger und Wurst (2000) ist ein Intelligenztest für Kinder und Jugendliche im Alter von 6 bis 15;11 Jahren.

Testaufbau. Der Test AID 2 besteht aus 11 Untertests, die in verbal-akustische und manuell-visuelle Fähigkeiten unterteilt sind. Außerdem gibt es noch drei neue Zusatztests. Die Subtests des AID2 sind in Tabelle 3.2 zusammengestellt. Die Nummern der Subtests des AID 2 in der Tabelle stimmen mit der Reihenfolge bei der Testdurchführung überein:

Tabelle 3.2: Subtests des AID2

Verbal-akustische Fähigkeiten	Gemessene Fähigkeiten/Fertigkeiten
1. Alltagswissen	Sachkenntnisse über alltägliche und gesellschaftliche Inhalte
3. Angewandtes Rechnen	Anwendung der passenden Rechenoperationen durch entsprechende Schlussfolgerungen bei der Problemlösung der alltäglichen Aufgabenstellungen
5. Unmittelbares Reproduzieren-numerisch	Die Kapazität der seriellen Informationsverarbeitung im verbal-akustischen Bereich
6. Synonyme Finden	Elementares Sprachverständnis, Erfassung der Bedeutung der sprachgebundenen Begriffe bzw. über einen Wortschatz verfügen, der solche Begriffe alternativ ausdrücken lässt
9. Funktionen Abstrahieren	Die Fähigkeit, durch Abstraktion zu einer Begriffsbildung zu gelangen
11. Soziales Erfassen und Sachliches Reflektieren	Erfassung der Sachzusammenhänge von der gesellschaftlichen Umwelt, Wissen des Bescheides über sozial angepasste Verhaltensweisen und gesellschaftliche Bedingungen
Manuelle-visuelle Fähigkeiten	
2. Realitätssicherheit	Das Verstehen der Wirklichkeit um Dinge des Alltags
4. Soziale und Sachliche Folgerichtigkeit	Die Fähigkeit, die Abfolge sozialen Geschehens bzw. alltäglicher Sachgegebenheiten zu verstehen und zu kontrollieren
7. Kodieren und Assoziieren	Erfassung zweier unabhängiger Fähigkeiten, der Informationsverarbeitungsschnelligkeit und der Fähigkeit zum inzidentellen Lernen
8. Antizipieren und Kombinieren-figural	Teile eines konkreten Ganzen erkennen und dieses Ganze gestalten durch das schlussfolgernde Denken
10. Analysieren und Synthesieren-abstrakt	Die Fähigkeit, komplexe (abstrakte) Gestalten durch eine geeignete Strukturierung reproduzieren zu können
Zusatztests	
5a. Unmittelbares Reproduzieren-figural/abstrakt	Die Kapazität der seriellen Informationsverarbeitung im visumotorischen Bereich
5b. Merken und Einprägen	Die Behaltenskapazität durch eine einmalige Wiederholung der Reizdarbietung
10a. Strukturieren (visumotorisch)	Die Fähigkeit, komplexe (abstrakte) Gestalten in elementare Teilkomponenten zerlegen zu können

Das AID 2 wird als Einzeltest durchgeführt. Die Anwendung der standardmäßigen Vorgabe (ohne Zusatztests) dauert ca. 75 Minuten und (mit den Zusatztests) noch zusätzlich 15 Minuten. Der Testleiter muss die Testinstruktionen aller Subtests gut beherrschen, die wörtlich vorzugeben und die am besten auswendig zu lernen sind. Die Vorgabe der meisten Untertests des AID 2 erfolgt adaptiv, d.h. die Anzahl der gelösten vorangegangenen Aufgaben bestimmt die Auswahl der nächsten Aufgaben, mit Ausnahme der Subtests 2, 5, 7, 5a, 5b und 10a. Außerdem kann eine Kurzform des AID 2 in den Untertests 1, 3, 6, 9, und 11 sowie die Parallelformen in den Subtests 1, 3, 4, 6, 9, 10 und 11 durchgeführt werden. Darüber hinaus gibt es sprachfreie Testinstruktionen für die Untertests 2, 4, 7, 8, 10, 5a und 10a.

Die Auswertung des AID 2 kann entweder manuell oder durch das Computer- Auswertungsprogramm AIDScore erfolgen, das zusätzlich zu den normierten Testwerten auch pro Subtest das resultierende Konfidenzintervall für die Interpretation der Testergebnisse liefert.

Erfassungsdimensionen. AID 2 ist die Revision und die Neunormierung des AID, dessen Untertests inhaltlich an Wechslers Intelligenzkonzept orientiert sind und methodisch auf der probabilistischen Testtheorie beruhen. Die pragmatische Intelligenztheorie ist die Grundlage für das AID2, um „viele (komplexe und basale) Fähigkeiten, die für „intelligentes“ Verhalten verantwortlich scheinen zu erfassen“ (Kubinger & Wurst, 2000, S. 14). In Anlehnung an den pragmatischen Ansatz, an Cattell (1987) und seine Investmenttheorie (Wissen ist investierte Intelligenz) kann Intelligenz im AID 2 als „das Bündel aller kognitiven Voraussetzungen, die notwendig sind, um Wissen zu erwerben und Handlungskompetenzen zu entwickeln“ (Kubinger & Wurst, 2000, S. 30) definiert werden. Das Ziel des AID2 ist förderungsorientierte Diagnostik und es wurde als Screening-Verfahren zur Messung ausgewählter Teilleistungsstörungen verwendet. Darüber hinaus „erteilt der AID2 der Generalfaktortheorie im Sinne Charles E. Spearman endgültig eine Absage: Ein Intelligenzquotient, definiert als Durchschnitt aller geprüften Fähigkeiten, ist intelligenztheoretisch nicht mehr vertretbar und förderungsorientiert nicht zielführend“ (Kubinger & Wurst, 2000, S. 14).

Testgütekriterien. In Bezug auf das erste Kriterium **Objektivität** (vgl. Kubinger & Wurst, 2000, S. 20- 21) sind Durchführungs- und Auswertungsobjektivität eingeschränkt. Es wurden Testleitereffekte in zwei Untertests, in dem Untertest *Synonyme Finden* und in dem Untertest *Funktionen Abstrahieren* durch eine Studie beim AID nachgewiesen. Außerdem gibt es eine gewisse Ermessensfreiheit vom Testleiter bei den Subtests 6 *Synonyme Finden*, 9 *Funktionen*

Abstrahieren und 11 *Soziales Erfassen und Sachliches Reflektieren* sowie 10a *Strukturieren*. Die Interpretationsobjektivität ist gegeben.

Zur **Reliabilität** des AID2 (vgl. Kubinger & Wurst, 2000, S. 21-23) sind die meisten der Untertests vom AID2 nach dem Rasch-Modell konstruiert. Das ermöglicht die Berechnung der Zuverlässigkeit pro Person unabhängig von allen getesteten Personen mit Berücksichtigung der Fähigkeitsparameter, das gilt nicht für die Verfahren, die auf der klassischen Testtheorie basieren. Die Standardschätzfehler der Fähigkeitsparameter und die minimalen Messfehler für das AID2 (außer der Untertests 5 und 7) wurden in der Kurz- und Parallelform bestimmt. Darüber hinaus gibt es im Manual Angaben zur Split-Half-Reliabilität für das ursprüngliche AID mit Ausnahme der Subtests 5 *Unmittelbares Reproduzieren* und 7 *Kodieren und Assoziieren* zwischen $r = .70$ (Realitätssicherheit) und $r = .95$ (Alltagswissen), sowie zur Retest-Reliabilität nach vier Wochen zwischen $r = .67$ und $r = .95$, nach mindestens einem Jahr zwischen $r = .39$ und $r = .80$. Zusätzlich wurde die Retest-Reliabilität für einige Untertests des AID 2 (Untertests 5 und 7, alle Zusatztests) nach vier Wochen bestimmt, deren Koeffizienten zwischen $r = .57$ und $r = .89$ lagen.

Bezüglich der **Validität** des AID 2 (vgl. Kubinger & Wurst, 2000, S. 23-27) wurde die faktorielle Validität durch die Faktorenanalyse nach Hauptachsenmethode mit Rotation (Varimax-Methode) bestimmt. Die Faktorenanalysen vom AID und AID 2 haben vier Faktoren erbracht. Die extrahierten Faktoren werden als Informationsverarbeitung in der gesellschaftlichen Umwelt, Informationsverarbeitung neuer Inhalte, Auffassungskapazität und (Re)Produktionsfähigkeit durch Strukturieren interpretiert. Zusätzlich wurde die Konstruktvalidität des AID durch die Extremgruppenvalidierung und die konvergente sowie diskriminante Validität berechnet. Um die Extremgruppenvalidierung des AID zu bestimmen, wurden 40 lernbehinderte Kinder mit 22 Kindern von Elternteilen, die Mitglieder des internationalen Vereins „MENSA“ waren, verglichen. Der Vergleich erbrachte signifikante Ergebnisse. „Während die Gruppe der Kinder von MENSA-Mitgliedern in sämtlichen Untertests kaum auf T-Werte unter 45 kam, erreichte die Gruppe der Sonderschüler kaum T-Werte über 50“ (Kubinger & Wurst, 2000, S. 24). Die diskriminante Validität wurde durch die Korrelationen des AID mit Leistungstests und Persönlichkeitsfragebogen an einer kleinen Stichprobe von 37 Kindern im Alter von 9-11 Jahren ermittelt. Außerdem gibt es im Manual Angaben über die Korrelationen der AID2 Zusatztests mit dem CFT 20. Diese Korrelationen lagen zwischen $r = .29$ und $r = .59$.

Die inhaltliche Validität des AID 2 ist durch Expertenratings gewährleistet. In Bezug auf die Kriteriumsvalidität wurde das AID mit den Schulnoten in den Fächern Deutsch und Mathematik korreliert. Die Ergebnisse wiesen auf gute Zusammenhänge mit den AID Untertests 1, 3, 6 und 11 hin.

Nach Kubinger (1996a; S. 63. zit. nach Kubinger & Wurst, 2000, S. 20) erfüllt ein Test das Gütekriterium **Skalierung** „wenn die laut Verrechnungsvorschriften resultierenden Testwerte die empirischen Verhaltensrelationen adäquat abbilden“. Dieses Gütekriterium wurde im AID 2 durch die einzelnen Verrechnungsvorschriften gesichert.

Die **Normierung** (vgl. Kubinger & Wurst, 2000, S. 66-71) erfolgte an einer Stichprobe von $N = 977$ Kindern und Jugendlichen im Alter von 6 bis 15;11 Jahren, die nach Geschlecht, 495 Jungen (50,7%) und 482 Mädchen (49,3%), Ländern Österreich (67,6%) sowie deutschen alten und neuen Bundesländern (32,4%), Alter, unterschiedlichen Schultypen und nach der Sozialschicht aufgeteilt wurde. Außerdem wurden die Testwerte (Fähigkeitsparameter) pro Altersstufe in T-Werte im jährlichen Abstand transformiert. Diese Transformation ist im Fall der Normalverteilung linear, im anderen Fall wurde eine Flächentransformation vorgenommen. Darüber hinaus werden die Normwerte der Intelligenzquantität, der zweitniedrigsten Untertestleistung und des Range der Intelligenz in Prozenträngen angegeben.

3.2.1.3 Die Kaufman Assessment Battery for Children (K-ABC) von Kaufman und Kaufman in der deutschen Bearbeitung von Melchers und Preuß (1991;1997)

Testart. Die Kaufman Assessment Battery (K-ABC) ist ein Individualtestverfahren zur Messung von Intelligenz und erworbenen Fertigkeiten bei Kindern im Alter von 2;6 bis 12;5 Jahren. Die K-ABC kann daher im Kindergarten, in der Vorschule und Grundschule sowie in den ersten Jahren der Sekundarschule angewendet werden (vgl. Melchers & Preuß, 1997, S. 7).

Testaufbau. Die K-ABC enthält 16 Untertests, die sich in vier Skalen unterteilen: Skala des einzelheitlichen und ganzheitlichen Denkens, die zur Skala *intellektuelle Fähigkeiten* zusammengefasst wird, die *Fertigkeitsskala* und die *Sprachfreie Skala*, die bestimmte Subtests umfasst, um die Intelligenz bei sprachbehinderten und fremdsprachigen Kindern zu erfassen. Die Untertests der intellektuellen Fähigkeiten, der Fertigkeitsskala und der Sprachfreien Skala sind in Tabelle 3.3 angegeben (vgl. Melchers & Preuß, 1991; Melchers & Preuß, 1997):

Tabelle 3.3: Skalen und Untertests des K-ABC

Skala intellektueller Fähigkeiten		Fertigkeitsskala	Sprachfreie Skala
Skala des einzelheitlichen Denkens	Skala des ganzheitlichen Denkens		
Handbewegungen	Zauberfenster	Wortschatz	Handbewegungen
Zahlen nachsprechen	Wiederkennen von Gesichtern	Gesichter und Orte	Wiederkennen von Gesichtern
Wortreihe	Gestaltschließen	Rechnen	Dreiecke
	Dreiecke	Rätsel	Bildhaftes Ergänzen
	Bildhaftes Ergänzen	Lesen/Buchstabieren	Räumliches Gedächtnis
	Räumliches Gedächtnis	Lesen/Verstehen	Fotoserie
	Fotoserie		

Die Durchführungsdauer der K-ABC als Individualtest hängt vom Alter des Kindes ab. Die Durchführungsdauer der K-ABC beträgt 30 Minuten bei Kindern im zweiten Lebensjahr (nur 7 Subtests bei der Anwendung) und 90 Minuten bei den Kindern im Alter von 12 Jahren (13 Untertests bei der Anwendung). Das Verfahren sollte nur von Psychologen/innen angewendet werden. Die Durchführung und die Auswertung des Tests erfordern gute Erfahrungen in den Bereichen Entwicklungspsychologie, Statistik und gute Kenntnisse in diagnostischer Verhaltensbeobachtung und individueller Intelligenzprüfung. Die Auswertung der K-ABC erfolgt anhand der Ermittlung der Rohwerte für jeden Subtest der K-ABC, die in Skalenwerte für jeden Untertest der Skala intellektueller Fähigkeiten mit einem Mittelwert von 10 und Standardabweichung von 3 umgewandelt werden.

Die Standardwerte für jeden Subtest der Skala *Fertigkeiten* sowie für die vier Gesamtskalenwerte werden in Skalenwerte mit einem Mittelwert von 100 und Standardabweichung von 15 umgerechnet. Außerdem können die Skalenwerte der Skala *intellektueller Fähigkeiten* und die Standardwerte der Fertigkeitensubtests sowie der Gesamtskalen in Prozentränge und in andere skalierte Werte transformiert werden. Zusätzlich werden das Konfidenzintervall für die Standardwerte der Skala *Fertigkeiten* und der Gesamtskalen sowie Stärken und Schwächen der Versuchsperson für die Untertests der Skala *intellektueller Fähigkeiten* und für die Fertigkeitenuntersets angegeben.

Erfassungsdimensionen. Die theoretische Grundlage des K-ABC bezieht sich auf die Definition der Intelligenz „als die Art und Weise, in der ein Individuum Probleme löst und Informationen verarbeitet“ (Melchers & Preuß, 1997, S. 7), die auf den Erkenntnissen in der Neuropsychologie

und der kognitiven Psychologie basiert. Die K-ABC erfasst mit der Skala der *intellektuellen Fähigkeiten* ein Gesamtmaß für Intelligenz, die als Problemlösen interpretiert wird. Die Skala der *Fertigkeiten* misst Faktenwissen, das das erworbene Lernen darstellt. Die Autoren betonen in ihrem Ansatz die Unterscheidung zwischen Problemlösen und Faktenwissen. Dieser Ansatz entspricht der Intelligenztheorie von Cattell und Horn (vgl. Melchers & Preuß, 1997, S. 8).

Testgütekriterien. Bezuglich der *Objektivität* der K-ABC sind Durchführungs- und Auswertungsobjektivität durch genaue Instruktionen und Auswertungsrichtlinien gegeben. Zur *Reliabilität* (vgl. Melchers & Preuß, 1997, S. 91-96) beinhaltet das Manual Angaben zur Split-Half-Reliabilität für jede Alterstufe und für jeden Subtest. Die Split-Half-Koeffizienten für die Skala der *intellektuellen Fähigkeiten* liegen zwischen $r = .70$ und $r = .88$ mit Ausnahme des Untertests Gestaltschließen, dessen Reliabilität mit $r = .58$ geringer ausfällt. Die Split-Half-Koeffizienten der *Fertigkeitsskala* sind etwas höher und für die Gesamtskalen zwischen $r = .83$ und $r = .98$. Außerdem wurde die Stabilität an einer Stichprobe von $N = 24$ Kindern im Alter von 9;0 bis 12;5 Jahren bestimmt. Die Koeffizienten der Stabilität für die einzelnen Untertests der K-ABC variieren zwischen $r = .57$ und $r = .96$, sowie für die Gesamtskalen zwischen $r = .84$ und $r = .97$.

In Bezug auf die *Validität* der K-ABC (vgl. Melchers & Preuß, 1997, S. 105-121) wurde die Faktorenanalyse zur Bestimmung der Konstruktvalidität durchgeführt, die sowohl für die Subtests der Skala *intellektueller Fähigkeiten* zwei Faktoren wie *einzelheitliches Denkens* und *ganzheitliches Denkens*, als auch für alle Untertests der K-ABC drei Faktoren: *einzelheitliches Denkens*, *ganzheitliches Denken* und *Fertigkeiten* bestätigt hat. Darüber hinaus wurde die Übereinstimmungsvalidität durch die Korrelationen der K-ABC Gesamtskalenstandardwerte mit anderen Tests wie z. B. HAWIK-R und AID ermittelt. Die Korrelationen der K-ABC Gesamtskalen mit der Gesamtleistung des HAWIK-R lagen zwischen $r = .57$ und $r = .79$, mit dem AID zwischen $r = .50$ und $r = .86$ und mit dem Begabungssystem (BTS) zwischen $r = .54$ und $r = .84$.

Normierung. Die K-ABC wurde an einer Stichprobe von $N = 3098$ Kindern aus Deutschland, der Schweiz, Österreich und Südtirol in den Jahren von 1986 bis 1989 für elf Altersgruppen von 2;6 bis 12;5 Jahren normiert. Die Repräsentativität der Normierungsstichprobe könnte von daher eingeschränkt sein, weil zwar über 90% der Kinder per Zufall ausgewählt wurden, die übrigen jedoch aus Beratungsstellen u.ä. Einrichtungen stammten und die beiden Gruppen ohne weitere Korrektur zu den Normstichproben zusammengefasst wurden (vgl. Melchers & Preuß, 1997, S. 79-83).

3.2.2 Gruppenverfahren

3.2.2.1 *Prüfsystem für Schul- und Bildungsberatung für 4. bis 6. Klassen (PSB-R 4-6) von Horn, Lukesch, Kormann und Mayrhofer (2002)*

Testart. Das Prüfsystem für Schul- und Bildungsberatung für 4. bis 6. Klassen (PSB-R 4-6) ist ein Intelligenztest und kann als Einzel- oder Gruppentest durchgeführt werden. Er ist geeignet für Schüler/innen der 4. bis 6. Klasse.

Testaufbau. Das Verfahren besteht aus 10 Subtests: (1) Allgemeinwissen, (2) Zahlenreihen, (3) Buchstabenreihen, (4) Figurale Reihen, (5) Wortflüssigkeit, (6) Gliederungsfähigkeit, (7) Raumvorstellung, (8) Gemeinsamkeiten finden, (9) Zahlenaddition und (10) Zahlenvergleich. Der Test liegt in zwei Parallelformen vor. Die Testdurchführung dauert ca. 44 Minuten.

Die Auswertung des Tests erfolgt anhand von Schablonen für beide Testformen A und B. Die Rohwerte für jeden Untertest werden berechnet, deren Summationen die Gesamtleistung bildet. Außerdem werden die Rohwerte des Tests mithilfe von Normtabellen in Prozentränge (PR) und Standardwerte (SW) transformiert. Darüber hinaus können die Standardwerte in IQ Werte umgewandelt werden.

Erfassungsdimensionen. Der PSB-R 4-6 bezieht sich auf die Grundlage von Thurstones Theorie „Primary Mental Abilities“ (Thurstone, 1938), sowie auf das Intelligenzmodell von Jäger (1982) (vgl. Horn, et al. 2002, S. 24-27). Die Subtests (1) Allgemeinwissen und (8) Gemeinsamkeiten finden erfassen den Faktor V (Verbal comprehension nach Thurstone). Die Subtests (2) Zahlenreihen, (3) Buchstabenreihen und (4) Figurale Reihen messen das induktive numerische, verbale und figurale Denken (R: Faktor Reasoning-Kompetenzen nach Jäger), der Untertest (5) Wortflüssigkeit misst den Faktor W (Word Fluency nach Thurstone), der Subtest (6) Gliederungsfähigkeit erfasst das räumliche Denken (F: Faktor Flexibility of Closure nach Thurstone), der Subtest (7) Raumvorstellung misst die Fähigkeit des räumlichen Vorstellungsvermögens (S: Faktor Space nach Thurstone). Der Subtest (9) Zahlenaddition erfasst die Konzentrationsfähigkeit mit dem Materialbereich Zahlen (N: Faktor Number nach Thurstone) und der Untertest (10) Zahlen vergleichen erfasst die Konzentrationsfähigkeit (P: Faktor Perceptual Speed nach Thurstone).

Gütekriterien. Zur **Objektivität** des PSB-R 4-6 sind die Durchführungs- und Auswertungsobjektivität durch die genauen und klaren Testinstruktionen gewährleistet. Das gilt auch für die Interpretationsobjektivität, die durch Anweisungen und Normtabellen gesichert ist. In Bezug auf die **Reliabilität** des PSB-R 4-6 (vgl. Horn et al. 2002, S. 47-48) wurde die innere Reliabilität der Subtests nach Cronbach Alpha bestimmt, deren Koeffizienten zwischen $r = .72$ und $r = .94$ liegen.

Bezüglich der **Validität** des PSB- R 4-6 (vgl. Horn et al. 2002, S. 49-60) wurde die Konstruktvalidität anhand einiger Faktorenanalysen zum PSB- R 4-6 ermittelt, die unterschiedliche Faktorenlösungen bzw. Drei- oder Vier-Faktoren-Lösungen erbracht haben. Aus diesen Analysen kann man schließen, dass der PSB-R 4-6 zwei Dimensionen der kognitiven Leistungsfähigkeit wie den Faktor des „Wissens und der sprachlichen Leistungsfähigkeit“ (F1) und den Faktor des „Schlussfolgernden Denkens“ (F2) gut misst. Zusätzlich erfasst das Verfahren den Faktor „Wahrnehmungstempo und Konzentration“ (F3) sowie den Faktor „flexibility of closure“ und „Feldabhängigkeit“ (F4). Nur der Faktor „räumliche Vorstellungsvermögen“ ist nicht nachweisbar (vgl. Horn et al., 2002, S. 53). Außerdem wurde die konvergente Validität durch die Korrelationen der PSB-R 4-6 Gesamtleistung mit anderen Intelligenztests wie mit dem CFT 20 geprüft. Hier lagen die Werte zwischen $r = .55$ und $r = .63$. Die Korrelation der Gesamtleistungen beider Tests mit der Kurzform des KFT 4-12 R betrug $r = .69$.

Um die Kriteriumsvalidität zu bestimmen, wurde die PSB-R 4-6 Gesamtleistung an einer Stichprobe von $N = 152$ Schülern/innen aus der 4. Klassenstufe (Schätz, 2000, zit. nach Horn et al. 2002, S. 53) mit der Deutschnote ($r = .62$), mit der Mathematiknote ($r = .61$) und mit der Note in Heimat- und Sachkunde ($r = .47$) korreliert. Außerdem wiesen die Korrelationen der PSB-R 4-6 Leistungen mit den Lehrerurteilen auf hohe Zusammenhänge hin.

Normierung. Der PSB-R 4-6 (vgl. Horn et al. 2002, S. 62) wurde an einer Stichprobe von $N = 1559$ Schülern/innen aus Bayern und Baden-Württemberg im Jahr 1999/2000 normiert. Es wurden getrennte Normtabellen nach Testformen A und B sowohl für die Schulstufen (4. Jahrgangsstufen erstes und zweites Schulhalbjahr, 5. und 6. Jahrgangsstufe), als auch für schulspezifische Gesamtleistungswerte (Haupt-, Realschule und Gymnasium) für die 5. und 6. Jahrgangsstufe ausgegeben. Zusätzlich gibt es im Manual Umrechnungstabelle von den Standardwerten (SW) in IQ-Werte. Darüber hinaus wurden die Standardwertkorrekturen für die Schüler/innen

mit nicht-deutscher Muttersprache berechnet, um die Leistungsdifferenzen mit den deutschen Schülern/innen vergleichen zu können.

3.2.2.2 Prüfsystem für Schul- und Bildungsberatung für 6. bis 13. Klassen (PSB-R 6-13) von Horn, Lukesch, Mayrhofer und Kormann (2004)

Testart. Das Verfahren PSB-R 6-13 ist ein Intelligenztest für Schüler/innen der 6. bis 13. Schulstufe.

Testaufbau. Der PSB-R 6-13 besteht aus neun Subtests, denen die drei Faktoren Verbal, Reasoning, und Konzentrationsfaktor zugeordnet werden. Die Subtests sind in Tabelle 3.4 zusammengestellt:

Tabelle 3.4: Untertests des PSB-R 6-13

Untertests des PSB-R 6-13		
Verbalfaktor (V)	Reasoningfaktor (R)	Konzentrationsfaktor (K)
(1) Allgemeinwissen	(2) Zahlenreihen	(8) Zahlenaddition
(5) Wortflüssigkeit	(3) Buchstabenreihen	(9) Zahlenvergleich
(7) Gemeinsamkeiten finden	(4) Figurale Reihen	
	(6) Raumvorstellung	

Der PSB-R 6-13 liegt in zwei Parallelversionen vor. Der Test kann als Einzel- oder Gruppentest angewendet werden. Die Durchführung des Verfahrens beträgt inkl. Testinstruktionen etwa 45 Minuten. Für die Auswertung des Verfahrens gibt es Schablonen für beide Testformen A und B, auf denen die Rohwerte für jeden Untertest berechnet werden. Danach werden die Rohwerte für jeden Faktor summiert. Die Summation aller Unterleistungen bildet den Gesamtleistungswert. Anhand der Normtabellen können die Rohwerte in Prozentränge sowie in Standardwerte umgerechnet werden. Zusätzlich können die Standardwerte in IQ-Werte transformiert werden.

Erfassungsdimensionen. Das Verfahren PSB-R 6-13 stellt die revidierte Fassung des PSB von Horn (1969) dar (Horn et al., 2004, S. 5). Bei der Neubearbeitung des PSB wurde gezeigt, dass die Aufgaben des ursprünglichen PSB für ältere Testpersonen mit einem gymnasialen Bildungshintergrund zu leicht waren. Aus diesem Grund wurden neue Items für die meisten der Subtests des PSB produziert, die nach Berechnung ihres Schwierigkeitsgrades in zwei Tests, dem PSB-R

4-6 für jüngere Probanden und dem PSB-R 6-13 für ältere oder auch erwachsene Probanden aufgeteilt wurden.

Dem PSB-R 6-13 (vgl. Horn, et al., 2004, S. 7-9) liegt das Primärfaktormodell von Thurstone (1938) zugrunde (siehe Kapitel 1). Die Subtests (1) Allgemeinwissen, (5) Wortflüssigkeit und (7) Gemeinsamkeiten finden erfassen die verbale Fähigkeit (V: Verbalfaktor nach Thurstone). Die Untertests (2) Zahlenreihen, (3) Buchstabenreihen, (4) Figurale Reihen und (6) Raumvorstellung sind in Anlehnung an das Intelligenzmodell von Jäger (1982) (siehe Kapitel 1) konzipiert und erfassen die Reasoning-Kompetenzen (R: Resoningfaktor nach Jäger). Die Untertests (8) Zahlenaddition und (9) Zahlenvergleich messen den Konzentrationsfaktor (K).

Gütekriterien. Bezuglich der ***Objektivität*** des PSB-R 6-13 sind die Durchführungs- und Auswertungsobjektivität durch genaue Testanweisungen und Schablonen gewährleistet (vgl. Horn, et al., 2004, S. 16-21). Das gilt auch für die Interpretationsobjektivität aufgrund der Interpretationshinweise, Normtabellen und Fallbeispiele. Zur ***Reliabilität*** des PSB-R 6-13 wurde die innere Konsistenz nach Cronbach Alpha geschätzt, deren Koeffizienten für die Subtests und für die beiden Testformen des PSB-R 6-13 außer der Untertests (8) Zahlenaddition und (9) Zahlenvergleich zwischen $r = .55$ und $r = .94$ variieren, für die Faktorenskalen zwischen $r = .89$ und $r = .92$.

In Bezug auf die ***Validität*** des PSB-R 6-13 (vgl. Horn, et al., 2004, S. 24-37) wurde die Konstruktvalidität anhand der Faktorenanalyse berechnet und bestätigte die Drei-Faktoren-Lösungen. Diese Faktoren klären 69% bis 73% der Varianz auf. Der erste Faktor wird als Dimension des verbalen Denkens (V) bezeichnet, der zweite Faktor ist als Faktor des induktiven Denkens (R) zu erkennen und der dritte Faktor wird als Faktor der Anstrengungsbereitschaft und Konzentrationsfähigkeit (K) identifiziert.

Zusätzlich gibt es im Manual Angaben über die Kriteriumsvalidität, die in verschiedenen Studien (Frohnholzer, 2001; Moser, 2003; zit. nach Horn, et al., 2004, S. 28-34) anhand der Korrelation zwischen den PSB-R 6-13-Ergebnissen und den Schulnoten von Hauptschülern geprüft wurde. Die Korrelationsbefunde zeigten, dass signifikante Zusammenhänge zwischen dem PSB –R 6-13- Gesamtwert und den Schulzensuren für die Schulstufe 6 bis 8 gesichert werden können. In der Studie von Frohnholzer (2001) (zit. nach Horn, et al., 2004, S. 29) weisen die Korrelationsbefunde der PSB-R 6-13 mit den Subtests Allgemeinwissen, Wortflüssigkeit und Gemeinsamkeiten auf signifikante Zusammenhänge mit der Deutschnote hin, sowie der PSB-R 6-13

Untertest-Reasoning mit der Mathematiknote. Die Konzentrationssubtests zeigen keine große Bedeutung für die Prognose von Schulzensuren.

In einer Untersuchung zur Ermittlung der Übereinstimmungsvalidität korreliert der PSB-R 6-13 Gesamtwert signifikant mit dem CFT 20 Gesamtwert mit $r = .41$ (Form A) bzw. $r = .31$ (Form B), jedoch nicht besonders hoch (Burgstaller, 2003 zit. nach Horn, et al., 2004, S. 37).

Normierung. Die Normierung des PSB-R 6-13 (vgl. Horn, et al., 2004, S. 41 ff) wurde an einer Stichprobe von $N = 7373$ bayerischen Schülern/innen durchgeführt. Es wurden getrennte Normtabellen nach Testformen A und B für die Schulstufen 6 bis 13, die Schultypen: Haupt-, Real-, Wirtschafts- und Berufsschule sowie Normtabellen für Erwachsene erstellt. Außerdem gibt es Umrechnungstabellen für die Standardwerte in IQ-Werte. Darüber hinaus können für die ausländischen Schüler die Korrekturwerte bestimmt werden.

3.2.2.3 *Die Culture Fair Intelligence Tests (CFT) von R.B. Cattell in den deutschen Bearbeitungen von Weiß und Osterland (2013) sowie Weiß & Weiß (2006)*

(a) Grundintelligenztest Skala 1 (CFT 1-R) von Weiß und Osterland (2013)

Testart. Der CFT 1-R ist ein Intelligenztest und kann als Einzel- oder Gruppentest angewendet werden. Er ist geeignet für Kinder im Alter von 5;4 bis 9;11 Jahren. Dieses Verfahren kann in schulischen und vorschulischen Bereichen bzw. in Kindergärten/Vorschulen, Grundschulen (Klassen 1 bis 3) sowie Förder- und Sonderschulen (Klasse 1 bis 4) angewendet werden. Der CFT 1-R ist in der Bildungs- Erziehungs- und Schulberatung, Praxis, Forschung, bei Diagnose von Lese-Rechtschreib-Schwäche und Dyskalkulie sowie bei Schulschwierigkeiten bzw. Rechenschwäche und bei Überprüfung der Sonderschulbedürftigkeit nützlich (vgl. Weiß & Osterland, 2013, S. 10-12).

Testaufbau. Der CFT 1-R besteht aus sechs Untertests, die sich in zwei Teile gliedern. Während der erste Teil die Subtests Substitutionen, Labyrinthe und Ähnlichkeiten umfasst, bilden die Untertests Reihenfortsetzen, Klassifikationen und Matrizen den zweiten Teil. Es existiert keine Parallelform. Die Subtests des CFT 1-R sind in Tabelle 3.5 zusammengestellt:

Tabelle 3.5: Die Skalen des CFT 1-R

Untertests des CFT 1-R	Gemessene Fähigkeiten
1. Substitutionen	die Fähigkeit, schnell die entsprechenden Symbole zu erkennen und den Darstellungen richtig zuzuordnen
2. Labyrinthe	Erfassung des optischen Wahrnehmungsumfangs und der Wahrnehmungsgeschwindigkeit, sowie der visuellen Orientierung und Aufmerksamkeit
3. Ähnlichkeiten	Messung des detailgenauen Wiedererkennens bei figuralen Vorgaben
4. Reihenfortsetzen	Erfassung des Erkennens von Regeln bei einfachen bis stufenweise komplexen figuralen Aufgabenstellungen, sowie des beziehungsstiftenden Denkens mit figuralem Material
5. Klassifikationen	Messung des Klassifizierens bzw. des beziehungsstiftenden Denkens bei figuralem Material
6. Matrizen	Die Fähigkeit, Regeln und Zusammenhänge bei figuralen Problemstellungen zu erkennen

Der CFT 1-R kann als Kurzform oder Langform angewendet werden. Die Testdurchführung dauert bei der Kurzform ohne Pausen 13 Min 10 Sek und bei der Langform 17 Min 30 Sek. Der Zeitbedarf inkl. Pausen liegt bei ca. 45 Minuten in der Grundschule Klasse 2 und 3, ca. 50 Minuten in der Grundschule Klasse 1, bis 60 Minuten in Kindergarten/Vorschule, Förder- und Sonderschule sowie in der Einzeltestung bei ca. 40 Minuten. Folgende Gruppengrößen im Zuge der Testung sollten nicht überschritten werden: Vorschule 6, Förder- und Sonderschule 1. Klasse 4 bis 5, 2. Klasse 5 bis 10, 3. und 4. Klasse 10 bis 12, 1. Klasse Grundschule 10, 2. und 3. Klasse bis 15 Kinder. Es ist möglich diese Zahl zu überschreiten, wenn es zwei Testleiter gibt. Für die Durchführung des Tests werden Testhefte, in denen die Beantwortungen eingetragen werden, 2 Stifte (ein Bleistift und ein roter Farbstift) für jede Testperson, Wandtafel für Übungsbeispiele, Stoppuhr und Beiheft mit Anweisungen und Lösungszeiten benötigt. Für die Auswertung werden die Rohwerte für jeden Untertest berechnet, die in T-Werte, IQ-Werte und in Prozentränge transformiert werden können.

Erfassungsdimensionen. Die Intelligenztheorie von Cattell liegt dem CFT 1-R zugrunde. Cattell geht in seiner Theorie davon aus, „dass der Bereich der allgemeinen intellektuellen Leistungsfähigkeit („general ability“ abgekürzt mit „g“) sich in zwei allgemeine Intelligenzformen aufgliedern lässt, nämlich in einen „flüssigen Intelligenzfaktor“ und in einen „kristallisierten Intelligenzfaktor“ (Weiß & Osterland, 2013, S. 20).

Der CFT 1-R misst die Grundintelligenz, d.h. die Fähigkeit „eines Kindes, in neuartigen Situationen und anhand von sprachfreiem, figuralem Material, Denkprobleme zu erfassen,

Beziehungen herzustellen, Regeln zu erkennen, Merkmale zu identifizieren und rasch wahrzunehmen“ (Weiß & Osterland, 2013, S. 10). Darüber hinaus liefert der Test Informationen darüber, bis zu welchem Komplexitätsgrad ein Kind bereits in der Lage ist, nonverbale Problemstellungen zu erfassen und zu lösen.

Testgütekriterien. Die **Objektivität** des CFT 1-R ist durch genaue Anweisungen zur Durchführung und Auswertung gegeben. Zur **Reliabilität** des CFT 1-R (vgl. Weiß & Osterland, 2013, S. 28) gibt es Angaben über die Retest-Reliabilität nach vier Wochen (Zeitintervall), deren Koeffizienten $r = .88$ für den 1. Teil, $r = .94$ für den 2. Teil und $r = .95$ für den Gesamttest betragen, sowie über die Konsistenzkoeffizienten (Kuder- Richardson Formula 8), die von $r = .94$ für den 1. Teil, $r = .95$ für den 2. Teil bis $r = .97$ für den Gesamttest variieren.

Bezüglich der **Validität** des CFT 1-R (vgl. Weiß & Osterland, 2013, S. 36-41) wurde die faktorielle Validität anhand der Hauptkomponentenanalyse und Varimax-Rotationen untersucht. Bei der Hauptkomponentenanalyse wiesen alle sechs Untertests auf hohe Ladungen auf dem Faktor General Fluid Ability „gf“ bei Förder- und Grundschülern und bei den Migratenkindern hin. In Bezug auf die Varimax-Rotationen ergeben sich zwei Faktoren, der Faktor Wahrnehmung/Speed, auf dem der erste Teil die höchste Ladung hat sowie der Faktor figurales Denken, auf dem der zweite Teil die höchste Ladung hat.

Darüber hinaus wurde die Kriteriumsvalidität durch die prognostische Validität und Übereinstimmungsvalidität bestimmt. Zur Überprüfung der prognostischen Validität wurde der CFT 1-R bei den dritten und vierten Klassen der Förderschüler durchgeführt und die Schulzensuren zwei Jahre später im 5. und 6. Schuljahr erhoben. Während der erste Teil des CFT 1-R mit der Mathematiknote mit $r = .23$ korreliert, weist die Korrelation des zweiten Teils des CFT 1-R auf einen guten prognostischen Wert von $r = .43$ bis $r = .60$ mit der Mathematiknote hin.

In Bezug auf die Übereinstimmungsvalidität gibt es im Manual Angaben über die Korrelationen des CFT 1 mit dem HAWIK für Kinder. Der CFT 1 korreliert mit dem HAWIK- Handlungsteil höher als mit dem Verbalteil. Hinsichtlich der Angaben über die Korrelation des CFT 1 mit dem Coloured Progressive Matrizen-Test (CPM), liegen die Korrelationen mit dem Untertest Matrizen des CFT 1 am höchsten ($r = .75$).

Normierung. Die Normierung des CFT 1-R (vgl. Weiß & Osterland, 2013, S. 48-50) erfolgte an einer Stichprobe von ca. 4700 Kindern aus Vorschulen, Grundschulen und Förderschulen aus

sieben deutschen Bundesländern: Baden Württemberg, Bayern, Berlin, Hamburg, Niedersachsen, Brandenburg und Sachsen. Es wurden getrennte Normtabellen (Altersnormen) in Form von T-Werten für die beiden Testteile sowie den Gesamttest, sowohl für die Kurzform als auch für die Langform für alle Altersgruppen in Vierteljahresschritten und Halbjahresschritten erstellt. Zusätzlich gibt es im Manual Angaben über Klassennormen für die Grundschule Klasse 1 bis 3 und eigene Normen für die Förderschule bzw. Altersnormen für 6;6 bis 11;11 Jahre sowie Quartil- und Klassennormen Klasse 1 bis 4. Darüber hinaus gibt es Hilfstabellen zur Transformation der T-Werte in IQ-Werte, Standardwerte und Prozentränge.

(b) Grundintelligenztest Skala 2 (CFT 20-R) von Weiß und Weiß (2006)

Testart. Das Testverfahren CFT 20-R ist ein Intelligenztest und stellt die Weiterentwicklung des CFT 20 dar. Der CFT 20-R ist geeignet für Kinder und Jugendliche von 8;5 bis 19 Jahren sowie für Erwachsene von 20 bis 60 Jahren. CFT 20 R kann in der Schulpsychologischen Beratung (Schullaufbahnberatung, Fördermaßnahmen bei Lernschwierigkeiten in Grundschule 3. und 4. Klasse, Hauptschule, Realschule, Gymnasium, Gesamtschule und Berufsschule), in der Erziehungsberatung, in der Berufsberatung und in der Forschung eingesetzt werden (vgl. Weiß & Weiß, 2006, S. 11-15).

Testaufbau. Der CFT 20 R enthält zwei Teile mit je vier Subtests, dem Reihenfortsetzen, Klassifikationen, Matrizen und Topologische Schlussfolgerungen. Der erste Testteil umfasst 65 Items, der zweite Testteil 45 Items, die nach Schweregrad geordnet sind. Der CFT 20 R kann als Einzel- und Gruppentest, in Langform (beide Testteile) sowie Kurzform (erster Testteil) angewendet werden. Die Durchführungszeit des CFT 20 R (beide Testteile) inkl. Instruktionen beträgt etwa 60 Minuten, bei der Durchführung der Kurzform ca. 35-40 Minuten. Testzeitverlängerung ist sinnvoll bei Testpersonen, die geringe Testerfahrung und schwieriges Anweisungsverständnis (z.B. Migrantenkinder) haben, sowie im Sonder- und Förderschulbereich in allen Alterstufen und bei Grundschülern.

Für die Testdurchführung werden als Testmaterial die Testhefte und die Antwortbögen, auf denen die Antworten eingetragen werden müssen, benötigt. Die Auswertung des CFT 20 R erfolgt anhand des Antwortbogens/Durchschreibbogens, auf denen die Antworten übertragen werden. Es werden die Rohwerte durch die Zahl der richtigen Antworten (Punkt für jede richtige Antwort) in jedem Untertest ermittelt, danach werden die Rohwerte für die vier Untertests in

beiden Testteilen summiert und der Gesamtrohwert aus der Summe der Einzellrohwerte der beiden Testteile gebildet. Die Rohwerte können mithilfe der Normtabellen in IQ-Werte; T Werte, Standardwerte oder Prozentränge umgewandelt werden und es kann ein Testprofil erstellt werden. Darüber hinaus kann die Auswertung des CFT 20 R anhand eines PC Programms geschehen, das eine schnelle Auswertung mit Profildarstellung und wichtigen statistischen Parametern wie Profildifferenzen und Kurzbeschreibung zur Interpretation ermöglicht.

Erfassungsdimensionen. Der CFT 20 R erfasst „wesentliche Teile des allgemeinen intellektuellen Niveaus (Grundintelligenz) im Sinne der general fluid ability nach Cattell. Diese kann interpretiert werden als eine Fähigkeit, komplexe Beziehungen in neuartigen Situationen wahrnehmen und erfassen zu können“ (Weiß & Weiß, 2006, S. 30). Zusätzlich können durch die Ergänzungstests, dem Wortschatztest (WS) und dem Zahlenfolgentest (ZF) die Elemente des Faktors Verarbeitungskapazität von Jäger gemessen werden. Nach Jäger (1982) gilt die Verarbeitungskapazität als „Verarbeitung komplexer Informationen bei Aufgaben, die nicht auf Anhieb zu lösen sind, sondern Heranziehen, Verfügbarhalten, vielfältiges Beziehungsstiften, formallogisch exaktes Denken und sachgerechtes Beurteilen von Informationen erfordern“ (Jäger, 1982; S. 213. zit. nach Weiß & Weiß, 2006, S. 31).

Testgütekriterien. In Bezug auf die **Objektivität** des CFT 20-R wird die Durchführungsobjektivität durch klare Testanweisungen gewährleistet. Das gilt auch für die Auswertungsobjektivität durch den Durchschreibbogen und das PC Auswertungsprogramm, sowie für die Interpretationsobjektivität durch die Normtabellen.

Zur **Reliabilität** des CFT 20 R (vgl. Weiß & Weiß, 2006, S. 48-49) wurde die Stabilität nach zwei bis fünf Monaten berechnet. Die Stabilitätskoeffizienten des CFT 20 R nach zwei Monaten betrugen $r = .85$ im ersten Teil, $r = .82$ im zweiten Teil, $r = .91$ im Gesamttest, nach fünf Monaten $r = .69$ im ersten Teil, $r = .86$ im zweiten Teil und $r = .83$ im Gesamttest. Darüber hinaus wurde die Konsistenzreliabilität durch die Korrelation zwischen dem ersten Teil Kurzform und Langform mit dem zweiten Teil bestimmt, deren Koeffizienten zwischen .80 und .82 variieren.

Bezüglich der **Validität** des CFT 20-R (vgl. Weiß & Weiß, 2006, S. 80-89) wurde die faktorielle Validität überprüft. Die Hauptkomponenten-Analyse mit drei Faktoren erbrachte eine Varianzaufklärung von 69%. Bei der Hauptachsen-Analyse ergab sich ein Zentralfaktor „General Fluid Ability“ mit einem Varianzanteil von 50%, auf dem alle vier Untertests hohe Ladungen hatten.

Die Faktorenstruktur des CFT 20-R nach der Varimaxrotation besteht aus drei Faktoren: (1) Gesetzmäßigkeiten und Regelhaftigkeiten erkennen (Reasoning- Faktor), der den größten Varianzanteil 50% besitzt und der von den Untertests Matrizen und Reihenfortsetzen repräsentiert wird, (2) Schlussfolgerndes Denken mit 25 % Varianzanteil, der von dem Subtest Topologien repräsentiert wird und (3) Beziehungsstiftendes Denken mit 25 % Varianzanteil, dieser wird vom Untertest Klassifikationen repräsentiert.

Außerdem gibt es im Manual Angaben sowohl über die Korrelationen zwischen dem CFT 20-R mit anderen Intelligenztests. Beispielsweise liegen die Korrelationen des CFT 20-R Gesamttest (Kurzform) mit den Ergänzungstests bzw. Wortschatztest und Zahlenfolgen zwischen $r = .53$ und $r = .61$ und mit der CFT 20-R (Langform) zwischen $r = .55$ und $.63$. Die Korrelationen des CFT 20-R mit dem Gesamtwert des PSB-R 4-6 lagen in einer Berliner Stichprobe von $N = 860$ Schülern aus den 4. Grundschulklassen im Jahr 2005 bei $r = .56$. Im Hinblick auf Korrelationen des CFT 20-R mit Schulzensuren zeigen die Ergebnisse, dass der CFT 20-R immer mit der Mathematiknote über alle Klassen, Schularten und Schulstufen sehr zufriedenstellend mit $r = .49$ korreliert. Das gilt nicht für die Deutschnote, deren Korrelation mit CFT 20-R mit $r = .35$ niedriger ausfällt. Zum CFT 20-R liegen noch keine prognostischen Validitätsstudien vor.

Normierung. Der CFT 20-R (vgl. Weiß & Weiß, 2006, S. 55-58) wurde im Jahr 2002/2003 an einer Stichprobe von ca. 4400 Schülern/innen aus verschiedenen Schultypen und sechs deutschen Bundesländern normiert. Es wurden schulartrrepräsentative Standardnormen für die Altersgruppen von 8;5 bis 19 Jahren erstellt. Diese Altersnormen sind nach Teil 1, Gesamtwert (mit verschiedenen Testzeiten) und Teil 2 getrennt. Die Altersnormen von 8;5 bis 15;0 Jahren sind in Form von Halbjahresschritten, und ab 15;1 bis 19 Jahre in Form von Jahresschritten angegeben. Darüber hinaus gibt es Altersnormen für Personen im Alter von 20 bis 60 Jahren nur für den ersten Teil in Fünfjahresschritten. Neben den Altersnormen wurden auch Klassennormen für 3-10/13. Schuljahr berechnet. Die Rohwerte können anhand der Normtabellen in IQ-Werte, T-Werte, Standardwerte und Prozentränge umgewandelt werden.

3.2.2.4 Kognitiver Fähigkeits-Test (Kindergarten) KFT-K von Heller und Geisler (1983)

Testart. Der Kognitive Fähigkeits-Test (Kindergarten) (KFT-K) von Heller und Geisler (1983) ist ein Intelligenztest zur Erfassung der kognitiven Lernfähigkeiten von Kindern im Alter von 5 bis 6 Jahren. Dieses Verfahren kann in der Schuleingangsdagnostik und in der Einzelfallhilfe,

bei der individuellen Begabungs- und Bildungsförderung in Kindergärten und Vorschulklassen (Förderungsdiagnostik) sowie bei wissenschaftlichen Untersuchungen verwendet werden (vgl. Heller & Geisler, 1983, S. 6).

Testaufbau. Der Test besteht aus den vier Subtests, Sprachverständnis, Beziehungserkennen, Schlussfolgerndes Denken und Rechnerisches Denken. Er kann als Einzel- und Gruppentest durchgeführt werden. Es gibt keine Zeitbegrenzung. Die Durchführung des Tests erfolgt an mehreren Tagen, ein Testteil pro Tag. Der Test sollte von Psychologen/innen und Pädagogen/innen angewendet werden. Für die Auswertung des Verfahrens wurden die Rohwerte für jeden Testteil berechnet, deren Summe den Gesamtwert bildet. Außerdem wurden die Rohwerte in T-Wertnormen transformiert.

Erfassungsdimensionen. Der KFT-K bezieht sich auf den Cognitive Abilities Tests von Thorndike, Hagen und Lorge bzw. auf die Parallelformen des Primary-Tests I der Cognitive Abilities Tests (Thorndike & Hagen, 1977, zit. nach Heller & Geisler, 1983, S. 6). Das Verfahren KFT-K misst die allgemeine Intelligenz oder das kognitive Fähigkeitsniveau (vgl. Heller & Geisler, 1983, S.3).

Gütekriterien. In Bezug auf die **Objektivität** des KFT-K (vgl. Heller & Geisler, 1983, S. 7-8) sind die Durchführungs- und Auswertungsobjektivität durch die detaillierten Testanweisungen und die numerischen Lösungsschlüssel gesichert. Die Interpretationsobjektivität ist auch durch Normtabellen und Fallsbeispiele gewährleistet.

Die **Reliabilität** des KFT-K (vgl. Heller & Geisler, 1983, S.8-9) wurde durch die innere Konsistenz nach Kuder- Richardson-Formel 20 ermittelt. Der Reliabilitätskoeffizient der KFT-K Gesamtleistung beträgt $r = .90$ in den Altersgruppen ab 5. Jahren, für die Untertests liegen die Koeffizienten zwischen $r = .53$ und $r = .88$. Die Konsistenzkoeffizienten des KFT-K weisen in den jüngsten Altersgruppen mit 3 bis 4 Jahren auf keine hinreichende Zuverlässigkeit hin. Außerdem wurde die Retest- Reliabilität nach vier bis fünf Wochen berechnet, deren Koeffizienten bei den Kindern im Alter von 5 bis 6. Jahren um $r = .80$ für die einzelnen Subtests und $r = .93$ für die KFT-K Gesamtleistung variieren.

Zur Überprüfung der **Validität** des KFT- K (vgl. Heller & Geisler, 1983, S. 10-17) wurde die Übereinstimmungsvalidität durch die Korrelation der KFT-K Gesamtleistung mit dem CFT 1 Gesamtwert mit $r=.64$ belegt. Darüber hinaus wurde auch die Konstruktvalidität des KFT-K

anhand der faktoriellen Validität sowie Mittelwertsvergleiche bestimmt. Um die Faktorenanstruktur des KFT-K zu ermitteln, wurden Hauptkomponentenanalysen mit Varimax-Rotation durchgeführt. Die Befunde zeigten, dass alle Untertests des KFT-K hohe Ladung auf nur einem Faktor haben. Deswegen wird der Test als eindimensionales Verfahren bezeichnet. Bei der Überprüfung der Gruppenunterschiede ergab sich, dass nicht nur die Testrohwerte mit dem zunehmenden Alter ansteigen, sondern auch die Differenzen zwischen den sozialen Schichten.

Die **Normierung** des KFT-K (vgl. Heller & Geisler, 1983, S. 18 ff) erfolgte an einer Stichprobe von $N = 509$ Kindern im Alter von 5;1 bis 7;0 Jahren. Normtabellen bzw. Altersnormen in T-Werten für die KFT-K Subtests und KFT-K Gesamtleistung unter Berücksichtigung der Testtermine September bis Februar und März bis August stehen zur Verfügung. Darüber hinaus können die T-Werte in Prozentränge oder in IQ-Werte umgewandelt werden.

3.2.2.5 Kognitiver Fähigkeits-Test (Grundschulform) KFT 1-3 von Heller und Geisler (1983)

Testart. Der Kognitive Fähigkeits-Test (Grundschulform) KFT 1-3 von Heller und Geisler (1983) wird als ein differentieller Intelligenztest zur Messung der intellektuellen Lernfähigkeiten bei Schülern der 1. bis 3. Grundschulklassen bezeichnet. Der Einsatz dieses Tests ist wie bei KFT-K in der Schullaufbahnberatung, Einzelfallhilfe und bei der individuellen Begabungs- und Bildungsförderungen in der Grundschule sowie in der wissenschaftlichen Untersuchung (vgl. Heller & Geisler, 1983, S. 6).

Testaufbau. Das Verfahren KFT 1-3 umfasst vier Untertests: Sprachverständnis, Beziehungs erkennen, Schlussfolgerndes Denken und Rechnerisches Denken. Der KFT 1-3 kann als Einzel- und Gruppentest angewendet werden und es gibt keine Zeitbegrenzung. Bei der Durchführung des Tests wird eine Pause von ca. 5 Minuten benötigt. Im ersten Schuljahr kann man den Test in zwei Hälften an zwei Tagen durchführen oder zwischen dem zweiten und dritten Untertest eine größere Pause von etwa 15 Minuten machen. Die durchschnittliche Bearbeitungszeit beträgt 45 bis 60 Minuten. Die Auswertung des Verfahrens erfolgt anhand des Lösungsschlüssels. Die Rohwerte werden für jeden Subtest ermittelt. Aus der Summe der Rohwerte aller Subtests ergibt sich der Gesamtwert. Zusätzlich können die Rohwerte in T-Werte umgewandelt werden.

Erfassungsdimensionen. Dem KFT 1-3 liegen die Cognitive Abilities Tests (CAT) von Thorndike, Hagen und Lorge bzw. die Parallelformen der Primary-Tests I und II der Cognitive

Abilities Tests zugrunde (Thorndike & Hagen, 1977, zit. nach Heller & Geisler, 1983, S. 6). Mit dem KFT 1-3 werden die schulische Lernfähigkeit/Denkfähigkeit und das kognitive Fähigkeitsniveau erfasst (vgl. Heller & Geisler, 1983, S. 3).

Gütekriterien. Bezuglich der ***Objektivität*** des KFT 1-3 (vgl. Heller & Geisler, 1983, S. 7-8) sind die Durchführungs- und Auswertungsobjektivität durch die detaillierten Testinstruktionen und Lösungsschlüssel gewährleistet. Das gilt auch für die Interpretationsobjektivität, die anhand der Interpretationsbeispiele und Normtabellen gesichert ist.

Zur Überprüfung der ***Reliabilität*** des KFT 1-3 (vgl. Heller & Geisler, 1983, S. 8) wurde die innere Konsistenz nach Kuder- Richardson-Formel 20 berechnet, deren Koeffizienten für die Untertests zwischen $r = .56$ und $r = .77$ variieren. Dazu kommt, dass der Konsistenzkoeffizient der KFT 1-3-Gesamtleistung höher als die Reliabilitätskoeffizienten der KFT 1-3 Untertests ausfällt. Außerdem wurde auch die Retest- Reliabilität nach fünf Wochen bestimmt. Die Wiederholungskoeffizienten der KFT 1-3 Untertests liegen zwischen $r = .58$ und $r = .80$ und für die KFT 1-3-Gesamtleistung beträgt der Reliabilitätskoeffizient ca. $r = .80$.

Bezuglich der ***Validität*** des KFT 1-3 (vgl. Heller & Geisler, 1983, S. 10-17) wurde die Konstruktvalidität anhand der Faktorenanalyse mit der Hauptachsenmethode und Varimax- Rotation ermittelt. Die Ergebnisse der Faktorenanalysen bestätigten die zwei Faktoren, den Schulleistungsfaktor und Intelligenzfaktor. Während die Untertests 1. Sprachverständnis und Untertest 2. Beziehungserkennen signifikant mit dem Schulleistungsfaktor korrelieren, stehen die Subtests 3. Schlussfolgerndes Denken und der Untertests 4. Rechnerisches Denken mit dem Intelligenzfaktor in Zusammenhang.

Zur Überprüfung der Übereinstimmungsvalidität wurde die Korrelation der KFT 1-3 Gesamtleistung mit den CFT 1 Untertests von Weiss & Osterland, 1977 errechnet, deren Koeffizienten zwischen $r = .16$ und $r = .55$ liegen, sowie mit den BT 2-3 Subtests von Ingenkamp, 1976, deren Korrelationskoeffizienten von $r = .40$ bis $r = .61$ reichten (vgl. Heller & Geisler, 1983, S. 10). Darüber hinaus korrelierte die KFT 1-3-Gesamtleistung mit der mündlichen Deutschleistung mit $r = .53$, mit der Mathematiknote mit $r = .44$ und mit dem Sachunterricht mit $r = .53$. Außerdem wurde die prognostische Validität durch die Korrelation zwischen den KFT 1-3-Ergebnissen und den Noten, die nach einem Jahr der Testdurchführung in der dritten Klasse erhoben wurde, bestimmt. Die Korrelationsbefunde zeigten, dass die KFT 1-3 Gesamtleistung am höchsten mit der Mathematikzensur ($r = .53$) und am niedrigsten mit der Sportnote ($r = .18$) korreliert. Dazu

kommt, dass die Untertests Beziehungserkennen und Rechnerisches Denken am meisten den Schulerfolg vorhersagen können.

Normierung. Die Normierung des KFT 1-3 (vgl. Heller & Geisler, 1983, S. 19) wurde an einer Stichprobe von $N = 4351$ Schülern/innen im Alter von 6;0 bis 12;0 im Jahr 1978-1979 durchgeführt. Normtabellen bzw. Altersnormen und Klassennormen der KFT 1-3 Subtests und Gesamtleistung wurden erstellt. Diese Normen sind nach Geschlecht getrennt. Zusätzlich gibt es im Manual eine Tabelle für die kritischen Differenzen zum Vergleich zwischen zwei Testergebnissen, sowie eine Tabelle für die Transformation von T-Werten in IQ, Prozentränge und andere Standardwerte.

3.2.2.6 *Kognitiver Fähigkeits-Test KFT 4-12 +R von Heller und Perleth (2000)*

Testart. Der Kognitive Fähigkeits-Test (KFT 4-12+ R) von Heller und Perleth (2000) ist ein differentieller Intelligenztest, und liefert Informationen über die intellektuellen Fähigkeiten wie sprachliches, quantitatives und figural-nonverbales Denken sowie das Gesamtleistungsniveau von Schülern der 4. bis 12. Klasse (vgl. Heller & Perleth, 2000, S. 8).

Testaufbau. Der KFT 4-12+ R besteht aus drei Teilen: verbaler, quantitativer und nonverbaler-figuraler Teil. Jeder Teil umfasst drei Untertests mit 12 bis 25 Items pro Klassenstufe. Darüber hinaus muss jeder Subtest in bestimmter Zeit bearbeitet werden. Deswegen muss der KFT 4-12+ R als ein Power-Speed-Test angesehen werden. Testaufbau und Zeitbedarf des KFT 4-12+ R finden sich ausführlich in Abschnitt 5.2.2 (Tabelle 5.2.2.1). Der KFT 4-12+ R liegt in zwei parallelen Formen, Form A und Form B, vor. Außerdem gibt es eine Kurzform, die aus den folgenden Untertests besteht: V1 Wortschatztest, V3 Wortanalogien, Q1 Mengenvergleiche, Q2 Zahlenreihen, N1 Figurenklassifikation und N2 Figurenanalogien.

Das Verfahren KFT 4-12+ R kann als Einzel- sowie Gruppentest angewendet werden. Die Testdurchführung der Normalform (inkl. Pausen) dauert 130 Minuten (drei Unterrichtsstunden) und bei der Kurzform 90 Minuten (zwei Unterrichtsstunden). Der KFT 4-12 +R kann von Psychologen/innen sowie von Klassenlehrer/innen durchgeführt werden. Die Auswertung des KFT 4-12 + R kann per Hand durch die Auswertungsschablone erfolgen, es werden die Rohwerte der einzelnen Subtests V1, V2 usw. ermittelt, danach werden die Rohwerte für die drei Testteile V, Q und N summiert. In jedem Teil werden die Rohwerte der einzelnen Subtests

zusammengerechnet ($V = V1 + V2 + V3$, $Q = Q1 + Q2 + Q3$ usw.), dann wird die KFT-Gesamtleistung als Summe der Testteile $GL = V + Q + N$ bestimmt. Durch das Computer-Auswertungsprogramm anhand einer Spezialsoftware und eines Scanners können die Antwortbögen eingelesen werden.

Erfassungsdimensionen. Der KFT 4-12 +R stellt die deutsche Version des Cognitive Abilities Tests (CAT) von Thorndike und Hagen (1971, 1993) und die Revision der KFT 4-13 Subtests von Heller, Gaedike und Weinläder (1974) dar (vgl. Heller & Perleth, 2000, S. 12). Er misst die folgenden Fähigkeitsbereiche: Sprachverständnis, sprachgebundenes Denken, arithmetisches Denken, Rechenfähigkeit, anschauungsgebundenes Denken, konstruktive Fähigkeiten sowie das allgemeine intellektuelle Gesamtleistungs niveau (vgl. Heller & Perleth, 2000, S. 3).

Gütekriterien. Zur *Objektivität* des KFT 4-12 +R sind die Durchführungs- und Auswertungsobjektivität durch die Testanweisungen, Auswertungsschablonen und das Auswertungsprogramm gewährleistet. Das gilt auch für die Interpretationsobjektivität durch Normtabellen des KFT 4-12+ R für Jahrgang und Schultyp.

Bezüglich der *Reliabilität* des KFT 4-12+ R (vgl. Heller & Perleth, 2000, S. 20-22) gibt es im Manual Angaben zu den Konsistenzkoeffizienten der Skalen des KFT 4-12+ R für beide Formen (A und B): Es ergibt sich über alle Subtests und über beide Testformen ein Median von $Rel = .79$, über alle Subtests außer Q1 und über alle Klassenstufen sowie die beiden Testformen ein Median von $Rel = .77$, für die drei Testteile ein Median von $Rel = .88$ und für die Gesamtleistung beider Testformen über alle Klassenstufen ein Median von $Rel = .95$. Die Stabilität des KFT 4-12+ R über zwei Retestungszeitpunkte betrug für die beiden Testformen A und B über mehrere Schuljahrgänge nach einem Jahr (erste Re-Testung bzw. 5/6 6/7 Klasse) zwischen $Rel = .83$ und $Rel = .87$ sowie nach zwei Jahren (zweite Re-Testung bzw. 5/7 Klasse) $Rel = .83$.

In Bezug auf die *Validität* des KFT 4-12+ R (vgl. Heller & Perleth, 2000, S. 28-47) wurde die Kriteriumsvalidität durch die Korrelationen der KFT-Subtests, der KFT-Testteile und der KFT-Gesamtleistung mit den Schulzensuren bzw. Deutsch-, Englisch- und Mathematiknoten berechnet. Die Ergebnisse zeigten, dass die Skalen der V Dimension auf mittlere bis höhere Beziehungen zu den Schulnoten hinweisen, insbesondere mit der Deutschnote. Zusätzlich zeigten die Untertests des Q Teils mittlere Zusammenhänge mit der Mathenote. Darüber hinaus zeigten auch die Untertests des N Teils mittlere bis hohe Beziehungen zu der Mathematiknote.

Zur Überprüfung der Konstruktvalidität wurde die Korrelation des KFT 4-12+ R mit anderen Intelligenztests wie PSB von Horn (1969) und CFT 20 von Weiss (1971) berechnet. Die Ergebnisse zeigten, dass die verbalen Skalen des KFT 4-12 gute Zusammenhänge zu den verbalen Skalen von PSB und CFT 20 haben. Das bestätigt, dass die verbalen Untertests des KFT 4-12+ R das sprachgebundene, logische Denken erfassen. Dazu kommt, dass die Reasoninganteile der quantitativen Skalen gewährleistet sind. Während der Q1-Test den Schulleistungsaspekt erfasst, messen die übrigen Tests Q2 und Q3 zahlengebundenes logisches Denken. Außerdem korrelierte der N-Teil hoch mit anderen Reasoningtests.

Es wurde auch die faktorielle Validität durch exploratorische Faktorenanalyse ermittelt, die drei Faktoren erbrachte. Die extrahierten Faktoren klären von 63,9 bis 73,5% der Gesamtvarianz auf. Außerdem wurde anhand der Strukturgleichungsmodelle von Perleth et al. (1994) und Sierwald (1996) (vgl. Heller & Perleth, 2000, S. 46-47) mit dem KFT 4-13 zu Strukturunterschieden der Testbatterien zwischen KFT 4-13 und KFT 4-12+ R konnten ein verbaler Faktor (15% der erklärten Varianz) und ein quantitativer Faktor (40 % der erklärten Varianz) nachgewiesen werden. Die Skalen des N-Teils bildeten keinen gemeinsamen Faktor, weil die nonverbalen Untertests unterschiedliche Primärfaktoren von Thurstone erfassen. N1 und N2 erfassen induktives Denken (reasoning) mit figuralem Material, N3 misst räumliches Denken: Der Großteil der aufgeklärten Varianz zwischen 34% und 88% lässt sich auf die allgemeine Intelligenz zurückführen. Der KFT 4-12 + R ist als differentieller Intelligenztest konzipiert, aber trotz dieser Konzeption erfassen der KFT 4-13 und der KFT 4-12+ R bei jüngeren Sekundarstufenschülern einen übergeordneten Faktor der allgemeinen Intelligenz.

Normierung. Der KFT 4-12+ R (vgl. Heller & Perleth, 2000, S. 48-49) wurde an einer Stichprobe von $N = 6765$ Schülern/innen normiert, die nach Testform A und B, Geschlecht und Schultyp (Grundschule, deren Schüler/innen aus Bayern sind, Hauptschule, Realschule und Gymnasium, deren Schülern/innen aus Baden-Württemberg sowie Bayern sind) gegliedert war. Es stehen Normtabellen (T-Werte und Prozentränge) sowohl jahrgangsweise als auch schultyp-spezifisch zur Verfügung. Darüber hinaus enthält das Manual Angaben über die Verteilungsmerkmale der KFT-Subtests, KFT-Teile und KFT-Gesamtleistung beider Testformen bzw. Mittelwerte, Standardabweichung, Schiefe, Exzeß und Quartilen, sowie kritische Differenzwerte für die KFT-Profilanalyse.

3.2.2.7 Intelligenz-Struktur-Test (I-S-T) 2000 R von Amthauer, Brocke Liepmann und Beauducel (2001)

Testart. Der I-S-T Intelligenz-Struktur-Test 2000 R ist ein Intelligenztest und anwendbar im Alter von 15 bis 60 Jahren.

Testaufbau. Der I-S-T 2000 R umfasst drei Module: 1) Das Grundmodul, mit dem die folgenden Intelligenzbereiche gemessen werden sollen: verbale (V), numerische (N) und figurale Intelligenz (F). Darüber hinaus werden in diesem Modul sowohl Merkfähigkeit (M) durch zwei weitere Aufgaben gemessen, als auch Schlussfolgerndes Denken (SD), das sich als Summenwert aus den Subtests der verbalen, numerischen und figuralen Intelligenz ergibt. 2) Grundmodul-Kurzform, welches das Grundmodul ohne die beiden Aufgabengruppen für Merkfähigkeit darstellt. 3) Erweiterungsmodul, das verbales (VW), numerisches (NW) und figural kodiertes Wissen (FW), Wissen (W), das sich aus den Skalen für verbales, numerisches und figural kodiertes Wissens (FW) ergibt, schlussfolgerndes Denken (gf) ohne Wissensanteile und Wissen (gc) ohne Anteile schlussfolgernden Denkens enthält. Außerdem besitzt der I-S-T 2000 R zwei Testformen A und B, mit gleichen Aufgaben, aber in anderer Reihenfolge. Die Module des Intelligenz Struktur Tests I-S-T 2000 R sind in Tabelle 3.6 zusammengestellt:

Tabelle 3.6: Die Untertests des I-S-T 2000 R

Grundmodul	
Intelligenzbereiche	gemessene Fähigkeiten
1. verbale Intelligenz	V die Fähigkeit mit sprachlichem Material im Rahmen des schlussfolgernden Denkens umzugehen
2. numerische Intelligenz	N die Rechenfähigkeit und die Fähigkeit logische Beziehungen herzustellen
3. figurale Intelligenz	F die Fähigkeit mit figural-bildhaftem Material umzugehen
4. Merkfähigkeit	M die Fähigkeit zum Einprägen und kurzfristigen Wiedererkennen von Informationen
5. schlussfolgerndes Denken	SD es wird das schlussfolgernde Denken erfasst
Erweiterungsmodul	
Wissenstest	gemessene Fähigkeiten
6. verbal kodiertes Wissen	VW Erfassung des erworbenen verbalen Wissens
7. numerisch kodiertes Wissen	NW Messung des erworbenen numerischen Wissens
8. figural kodiertes Wissen	FW Erfassung des erworbenen figuralen Wissens
9. Wissen	W Wissen mit Anteilen schlussfolgernden Denkens, Erfassung der Ausschnitte des Wissens bzw. verbales, numerisches und figurales Wissen
10. schlussfolgerndes Denken (gf)	gf. Erfassung des Generalfaktors „fluide Intelligenz“
11. Wissen (gc)	gc. Messung des Generalfaktors „kristallisierte Intelligenz“

Der I-S-T 2000 R kann als Einzel- oder Gruppentest angewendet werden. Bei der Durchführung des Tests als Grundmodul-Kurzform ohne Merkaufgaben beträgt die Bearbeitungszeit 77 Minuten, vollständig 84 Minuten, die Bearbeitungszeit bei dem Erweiterungsmodul liegt bei 40 Minuten. Die Auswertung des Tests findet mit Hilfe von Schablonen und Normtabellen statt. In Bezug auf die Grundmodul-Kurzform wurden drei Skalenwerte für die verbale, numerische und figurale Intelligenz und ein Gesamtwert oder eine Rohwertsumme aus allen drei Skalen für das schlussfolgernde Denken sowie für die Merkfähigkeit gebildet. Darüber hinaus wurden die Rohwerte in Standardwerte, Prozentränge und in das I-S-T 2000 Profil umgewandelt.

Bezüglich des Erweiterungsmoduls kann wie im Grundmodul eine Gesamtwissensskala gebildet werden, welche die drei Skalen, verbale, numerisch und figural kodiertes Wissen, umfasst. Zusätzlich werden die Rohwerte von Schlussfolgerndem Denken (gf) und für Wissen (gc) zu Punktwerten umgewandelt und diese Punktwerte für schlussfolgerndes Denken (gf) und für Wissen (gc) in Standardwerten umgerechnet. Der I-S-T 2000 R wird anhand seiner Skalen interpretiert, weil nur die Skalen theoretisch begründet und empirisch fundiert sind. Deswegen ist der Einsatz von Normen und die Interpretation der einzelnen Aufgabengruppen nicht zulässig.

Erfassungsdimensionen. Der I-S-T 2000 R (vgl. Amthauer, et al., 2001, S.10-17) bezieht sich auf das hierarchische Rahmen- bzw. Protomodul der Intelligenzstrukturforschung (HPI), das die inhaltlichen und strukturellen Konvergenzen bzw. Primärfaktoren von Thurstone (1938) sowie fluide und kristallisierte Intelligenz von Cattell (1963, 1987) umfasst. Der I-S-T 2000 R erfasst 11 Fähigkeiten, die in Grund- und Erweiterungsmodulen geteilt wurden (Tabelle 3.6).

Gütekriterien. In Bezug auf die **Objektivität** des I-S-T 2000 R sind die Durchführungsobjektivität durch die Testinstruktionen, die Auswertungsobjektivität durch die Schablonen und die Interpretationsobjektivität durch Normtabellen gewährleistet. Zur **Reliabilität** des I-S-T 2000 R (vgl. Amthauer, et al., 2001, S. 25) wurde die innere Reliabilität nach Cronbachs Alpha und Split-Half für die Skalen der beiden Module bestimmt. Bezuglich der Skalen des Grundmoduls betragen die Reliabilitätskoeffizienten zwischen $r = .87$ und $r = .97$, für das Erweiterungsmodul oder die Wissensskalen (vgl. Amthauer, et al., 2001, S. 74) variieren die Reliabilitätskoeffizienten zwischen $r = .82$ und $r = .93$.

In Bezug auf die **Validität** des I-S-T 2000 R (vgl. Amthauer, et al., 2001, S. 39-43) wurde die Konstruktvalidität des Grundmoduls anhand der Dimensionsanalyse und der konfirmatorischen

Faktorenanalyse der Normierungsstichprobe $N = 2208$ belegt, welche die drei Faktoren verbal, numerisch und figural erbracht haben. Außerdem gibt es im Manual Angaben zur diskriminanten Validität über die die geringen Beziehungen zum Verfahren d2 (Brickenkamp, 1994); am stärksten ist hier der Zusammenhang mit der numerischen Intelligenz $r = .24$. Im Hinblick auf die konvergente Validität ergaben sich mittleren und hohen Korrelationen mit anderen Tests wie dem HAWIK-R-Wissensteil nach Tewes (1991), die höchsten Zusammenhang ergaben sich hierbei zwischen der verbalen Intelligenz ($r = .48$) sowie dem Schlussfolgerndem Denken ($r = .46$). Mit den CFT 20-Matrizen nach Weiß (1997) ergab sich der höchste Zusammenhang mit dem IST-Gesamtwert ($r = .63$), der figuralen Intelligenz ($r = .55$), der verbale Intelligenz ($r = .47$) und mit der numerischen Intelligenz ($r = .45$). Zum Raven-Test bestanden die höchsten Zusammenhänge mit $r = .69$ mit dem schlussfolgerndem Denken, der figuralen Intelligenz ($r = .50$) und mit der verbalen Intelligenz ($r = .54$).

Darüber hinaus wurde die Kriteriumsvalidität anhand von Korrelationen mit Schulnoten in Deutsch, Englisch, Mathematik, Chemie und Physik berechnet. Die Ergebnisse zeigen, dass nur der I-S-T 2000 R im Verbalteil gute Beziehungen zu Englisch- und Deutschzensuren hat. Zusätzlich ergeben sich die stärksten Zusammenhänge für die numerische Intelligenz und das schlussfolgernde Denken mit Mathematik-, Chemie-, und Physiknoten.

Bezüglich des I-S-T 2000 R Erweiterungsmoduls (vgl. Amthauer, et al., 2001, S. 58-72) wurden die Werte zur Konstruktvalidität beim Wissenstest sowie beim schlussfolgernden Denken (gf) und Wissen (gc) durch die multidimensionale Ähnlichkeitsstruktur-Analyse und die konfirmatorischen Faktorenanalyse bestimmt. Die Struktur des Wissenstests und Faktoren der fluiden und kristallisierten Intelligenz wurden bestätigt.

Die Kriteriumsvalidität des Erweiterungsmoduls wurde durch die Korrelationen mit dem alten I-S-T 2000 zwischen den aktuellen und den alten Faktorwerten der fluiden und kristallisierten Intelligenz bestimmt, die für schlussfolgerndes Denken (gf) $r = .98$ und Wissen (gc) $r = .92$ betragen, und mit den Schulnoten in Geschichte, Erdkunde, Mathematik und Physik belegt. Die Noten in Geschichte und Erdkunde korrelieren mit Wissen (gc) höher als mit schlussfolgerndem Denken (gf). Das gilt nicht für die Mathematik- und Physiknoten, deren Korrelationen mit dem schlussfolgernden Denken (gf) am höchsten waren. Zusätzlich wurde die konvergente und diskriminante Validität des Erweiterungsmoduls an einer Stichprobe von 206 Hamburger Schülern durch die Korrelationen mit anderen Tests bestimmt. Geringe Korrelationen zeigten

sich mit dem Ergebnis aus dem d2: $r = .09$ (gc) und $r = .25$ (gf), sowie hohe Korrelationen mit HAWIK-R Wissen: $r = .68$ (gc) und $r = .31$ (gf).

Normierung. Die Normierung des I-S-T 2000 R Grundmoduls (vgl. Amthauer, et al. 2001, S. 44-45) erfolgte an einer Stichprobe von $N = 3484$ Probanden im Alter von 15 bis über 51 Jahren. Davon besuchten 1892 (54,3%) der Stichprobe ein Gymnasium oder hatten ein Abitur, 1592 (45,7%) der Stichprobe waren Schüler/innen, Auszubildende, Studenten/innen usw. Die Gymnasiasten/innen wurden in acht Altersgruppen zwischen 15 und älter als 50 Jahre unterteilt, Nicht Gymnasiasten/innen in fünf Altersgruppen zwischen 15 und älter als 40 Jahre. Normtabellen wurden für die Gymnasiasten, nicht Gymnasiasten und für die Gesamtstichprobe vorgelegt. Die Rohwerte wurden in Prozentränge und IQ Werte umgerechnet. Das I-S-T Erweiterungsmodul (vgl. Amthauer, et al., 2001, S. 55-56) wurde anhand einer Stichprobe von $N = 661$ Probanden im Alter von 15 bis 60 Jahren normiert. Es wurden Normtabellen für Wissensskalen, schlussfolgerndes Denken (gf) und Wissen (gc) von drei Altersgruppen: 15-25, 26-35 und von 36 bis 60 Jahren, sowie von Gymnasiasten, nicht Gymnasiasten und von der Gesamtstichprobe erstellt. Um diese Normen zu erstellen, wurden die Faktorwerte in Prozenträngen umwandelt und diese in Z-Werte umgerechnet. Diese Z- Werte wurden in Standardwerte transformiert.

3.2.3 Zusammenfassung

In zweiten Teil dieses Kapitels wurden kurze Beschreibung der am häufigsten durchgeführten Intelligenz- und kognitiven Fähigkeitstests in Deutschland gegeben, untergliedert in Einzel- und Gruppentests. Tabelle 3.7 gibt einen hierzu einen zusammenfassenden Überblick.

Tabelle 3.7: Übersicht über Testverfahren zur Messung der Intelligenz und kognitiver Fähigkeiten

Autor und Erscheinungsja hr	Testbezeichnung	Erfassungsdimensionen	Ziel- gruppen
Petermann & Petermann (2008)	Hamburg- Wechsler-Intelligenztests HAWIK-IV	- allgemeine Intelligenz - spezifische Fähigkeiten: Sprachverständnis, Wahrnehmungsgeb. logisches Denken, Arbeitungsgedächtnis, Verarbeitungsgeschwindigk.	6;0 bis 16;11 Jahre
Kubinger & Wurst (2000)	Adaptives Intelligenz Diagnostikum 2 AID 2	- allgemeine Intelligenz - verbal-akustische Fähigkeiten - manuell-visuelle Fähigkeiten	6;0 bis 15;11 Jahre
Melchers & Preuss (1991)	Kaufman-Assessment Battery for Children K-ABC	- intellektuelle Fähigkeiten (flüssige Intelligenz) - erworbene Fertigkeiten (kristallisierte Intelligenz) - Gedächtnis	2;6 bis 12;5 Jahre
Amthauer, Brock, Liepmann & Beauducel (2001)	Intelligenz- Struktur-Test I-S-T 2000 R	- fünf Primärfaktoren wie verbale, numerische und figurale Intelligenz, Merkfähigkeiten und Schlussfolgerndes Denken - erworbenes Wissen - die Generalfaktoren der fluiden und kristallisierten Intelligenz	15 bis 60 Jahre
Horn, Lukesch, Kormann & Mayrhofer (2002)	Prüfsystem für Schul- und Bildungsberatung PSB- R 4-6	- allgemeine Intelligenz - Primärfaktoren	gilt für die 4. bis 6. Klassen
Horn, Lukesch, Mayrhofer, & Kormann (2004)	Prüfsystem für Schul- und Bildungsberatung PSB- R 6-13	- allgemeine Intelligenz - Primärfaktoren	gilt für die 6. bis 13. Klassen
Weiβ & Osterland (2013)	Grundintelligenztest Skala 1 CFT 1-R	Flüssige Intelligenz nach Cattell	5;4 bis 19;11 Jahre
Weiβ & Weiβ (2006)	Grundintelligenztest Skala 2 CFT 20-R	Flüssige Intelligenz nach Cattell	8;5 bis 19;0 Jahre; 20 bis 60 Jahre
Heller & Geisler (1983)	Kognitiver Fähigkeitstest (Kinder- gartenform) KFT-K	allgemeine Intelligenz / kognitives Fähigkeitsniveau	5 bis 6 Jahre
Heller & Geisler (1983)	Kognitiver Fähigkeitstest (Grund- schulform) KFT 1-3	- schulische Lernfähigkeit /Denkfähigkeit - das kognitive Fähigkeitsniveau	gilt für die 1. bis 3. Klassen
Heller & Perleth (2000)	Kognitiver Fähigkeitstest KFT 4-12 + R	- Sprachverständnis - Sprachgebundenes Denken - arithmetisches Denken - Rechenfähigkeit - anschauungsgebundenes Denken - konstruktive Fähigkeiten - die allgemeine intellektuelle Intelligenz	gilt für die 4. bis 12. Klassen

4. Zusammenhänge zwischen Intelligenz, Geschlecht, Migration und Schulleistung

In diesem Kapitel werden Studien vorgestellt, die den Einfluss von Faktoren wie Geschlecht und Kultur auf die Intelligenz untersucht haben. Außerdem wird die Bedeutung der Intelligenz zur Vorhersage von schulischer Leistung und Berufserfolg behandelt.

4.1 Einflussfaktoren

4.1.1 Geschlechtsunterschiede

Untersuchungen über Geschlechtsunterschiede in Bezug auf die allgemeine Intelligenz sind zahlreich. Lynn (1994, zit. nach Hirnstein & Hausmann, 2010, S. 70-71) hat festgestellt, dass Männer in Intelligenztests im Durchschnitt 4 IQ Punkte besser als Frauen abschneiden. Lynn führt diesen Unterschied auf den schwachen Zusammenhang zwischen der Hirngröße und den IQ Punkten bei einer Hirngrößendifferenz von Männern und Frauen von ca. 100cm³ zurück. Die 4 IQ-Punkte Differenz sind nach Hirnstein und Hausmann (2010) im verwendeten Intelligenztest (Wechsler-Intelligenztest) begründet, für andere Tests gebe es keine Unterschiede oder die Differenzen fielen zugunsten der Frauen aus. Diese Ergebnisse zeigen bereits die Schwierigkeiten bei der Erforschung von Geschlechtsunterschieden bezüglich der allgemeinen Intelligenz, da die allgemeine Intelligenz ein psychologisches Konstrukt ist, dessen Messung mit Annahmen und Operationalisierungen verbunden ist (vgl. Hirnstein & Hausmann, 2010, S. 70-71).

In manchen Untertests, die spezifische Fähigkeiten messen, zeigten sich regelmäßig Geschlechtsunterschiede. Während Mädchen im Vergleich zu Jungen in der Grundschule bis in die höhere Schule besser hinsichtlich Lese- und Schreibleistungen abschneiden (Berk, 2005; Halpern, 2000, zit. nach Woolfolk, 2008, S. 224), erbringen Jungen bessere Leistungen in Tests für höhere Mathematik. Aber dieser Unterschied besteht nicht immer zugunsten der Jungen. Im Jahr 2001 gab es zwar doppelt so viele Testergebnisse von Jungen mit Werten über 700 Punkten im Mathematiktest SAT (S Standardabweichungen über dem Mittelwert), jedoch hatten auch mehr Jungen Misserfolg in dem Test als Mädchen (Angier & Chang, 2005, zit. nach Woolfolk, 2008, S. 224). Die Testwerte bei Jungen schwanken also stärker als bei Mädchen, d.h. die

Jungen zeigen im Vergleich zu Mädchen häufiger extrem hohe und niedrige Testwerte (Berk, 2005; Willingham & Cole, 1997, zit. nach Woolfolk, 2008, S. 224).

Halpern (2004, zit. nach Woolfolk, 2008, S. 224) fand in ihrer Untersuchung Geschlechtsunterschiede in einem kognitiven Fähigkeitstest. Sie kam auch zu dem Ergebnis, dass Mädchen bessere Zensuren in der Schule bekommen. Darüber hinaus erreichten Mädchen bessere Testwerte in Schreib- und Wissenstests, die sich inhaltlich stärker auf die Inhalte des schulischen Curriculums, und mehr Universitätsabschlüsse als Jungen. Jungen hatten aber bessere Noten in Mathematik- und Naturwissenschaftstests, die nicht das in der Schule gelernte Wissen überprüfen. Außerdem erbrachten die Jungen bessere Leistungen in den Fachbereichen Politik und Geographie. Andere Studien haben wiederum festgestellt, dass es keine Geschlechtsunterschiede in Mathematik gibt, wie die TIMSS-Studie (Third International Mathematics and Science Study), deren Ergebnisse für Deutschland zeigen, dass es in der Altersgruppe Primar- und Sekundarstufe 1 keine Geschlechtsunterschiede in Mathematik gibt, sich aber in der Sekundarstufe 11 Unterschiede zugunsten der Jungen zeigen (Baumert et al., 1997; Baumert, Bos & Waterman, 1998, zit. nach Woolfolk, 2008, S. 225).

Zusammenfassend kann man sagen, dass die Frage, ob es Geschlechtunterschiede in Bezug auf die kognitive Fähigkeit gibt, umstritten ist. Aber die meisten Studien, die Geschlechtunterschiede untersucht haben, kamen zu der Feststellung, dass Jungen besser im Fach Mathematik und Naturwissenschaft abschneiden und die Mädchen bessere Leistungen in den Bereichen Sprache und geistig-kulturelle Gebiete erbrachten.

Die Ursachen für diese Unterschiede sind komplex und werden durch assoziierte a) biologische, b) psychologische, c) soziale und d) biopsychosoziale Faktoren angegeben.

4.1.1.1 Biologische Faktoren

Die Sexualhormone werden für die biologische Basis von Geschlechtsunterschieden in kognitiven Fähigkeiten gehalten, weil sie an den Differenzen zwischen männlichem und weiblichem Gehirn beteiligt sind. Sie beeinflussen die Geschlechtsdifferenzen durch organisierende Effekte, d.h. Sexualhormone verändern prä- bzw. neonatal die neuronale Organisation des Gehirns und verändern durch aktivierende Effekte, d.h. aktuelle Hormonschwankungen, die Informationsverarbeitung (Beerman, Heller & Menacher, 1992; Hirnstein & Hausmann, 2010).

Viele Studien haben den Einfluss organisierender¹ Effekte bei Menschen untersucht, die aufgrund von Entwicklungsstörungen unnormal hohen und niedrigen Konzentrationen von Sexualhormonen ausgesetzt waren, wie z. B. beim sogenannten Androgenitalem Syndrom (AGS). Die Ergebnisse dieser Untersuchungen zeigten, dass Mädchen mit AGS bessere Noten in den räumlichen Tests erzielten als jene ohne AGS. Jedoch erzielten Männer mit AGS genauso gute oder schlechte Noten wie nicht betroffene Männer. Dies könnte nach Ansicht der Autoren daran liegen, dass Testosteronspiegel männlichen Föten mit AGS nicht zwangsläufig erhöht ist und andererseits ein abnorm hoher Testosteronspiegel zu schlechten Leistungen bei räumliche Aufgaben führen kann (Hirnstein & Hausmann, 2010).

Hinsichtlich aktivierender Effekte wird der Hormonspiegel gemessen und mit den kognitiven Tests korreliert. Einige Studien haben gezeigt, dass Frauen mit hohem Testosteronspiegel bessere Ergebnisse in räumlichen Tests erreichen als Frauen mit niedrigem Testosteronspiegel. Im Gegensatz dazu schneiden Männer mit niedrigem Testosteronspiegel bei räumlichen Tests besser ab als Männer mit hohem Testosteronspiegel. Die meisten der Studien, die den Einfluss von Sexualhormonen auf kognitive Geschlechtsdifferenzen untersuchten, haben festgestellt, dass Sexualhormone einen Einfluss auf die kognitiven Geschlechtsunterschiede haben, allerdings ist das Ausmaß des Einflusses bis jetzt nicht klar (vgl. Hirnstein & Hausmann, 2010, S. 76-78).

4.1.1.2 Psychologische Faktoren

Die Studienergebnisse zu psychologischen Geschlechtsunterschieden zeigten, dass die motivationalen und emotionalen Persönlichkeitsmerkmale eine große Rolle hinsichtlich der Wahl der Fächer Mathematik und Naturwissenschaft und auch für herausragende Leistungen in diesen Bereichen spielen. Die Mädchen interessieren sich für die geistigen Wissenschaften mehr als für Mathematik und Naturwissenschaft und dies wirkt sich negativ auf die Leistung in Mathematik und Naturwissenschaft aus. In einer Metaanalyse von zehn Studien, die Attribuierung von Erfolg und Misserfolg in Mathematik untersuchte, wurde festgestellt, dass der Grund für den Erfolg in Mathematik bei den Jungen von diesen selbst, aber auch von Personen des Lernumfelds, auf ihre Begabung und bei den Mädchen auf ihre starke Anstrengung zurückgeführt wird (Beerman & Heller, 1990 zit. nach Beerman et al., 1992, S. 46). Umgekehrt wurde als Ursache für den

¹ Mit organisierenden Effekten bezeichnen Hirnstein und Hausmann (2010) Auswirkungen der Hormone auf die prä- und neonatale Entwicklung der neuronalen Struktur bzw. Organisation.

Misserfolg in Mathematik bei den Jungen mangelnde Anstrengung und bei den Mädchen mangelnde Begabung angegeben.

In einer Metaanalyse von acht Studien (Beerman & Heller, 1990 zit. nach Beerman et al., 1992, S. 49), in der das Selbstkonzept in Mathematik untersucht wurde, zeigte sich, dass es entweder keinen Geschlechtsunterschied oder Geschlechtsunterschiede zugunsten der Jungen gab. Das Merkmal der Angst vor Mathematik wurde auch untersucht. Zunächst kann festgehalten werden, dass die Angst vor Mathematik sich negativ auf die Leistung in Mathematik auswirkt. Darüber hinaus zeigten die Mädchen mehr Angst als die Jungen (Reyes, 1984 zit. nach Beerman et al., 1992, S. 52).

Die psychologischen, kognitiven Geschlechtsunterschiede im MRT (Mental Rotation Test) (vgl. Hirnstein & Hausmann, 2010, S. 78) scheinen daher zu kommen, dass Frauen eine andere Methode bei der Lösung der MRT-Aufgaben einsetzen. Sie analysieren die Aufgaben verbal anstatt die Figuren mental zu rotieren. Dazu kommt, dass Frauen weniger Selbstvertrauen bei den MRT-Aufgaben haben als die Männer. Außerdem gibt es Geschlechtsunterschiede in der Wahrnehmungsgeschwindigkeit, die sich auf unterschiedliche Strategien beziehen. Diese Effekte sind deutlich im Zahlen-Symbol-Test festzustellen und weniger in anderen Verfahren wie z.B. beim Identical Pictures Test, weil im Zahlen-Symbol-Test das weibliche Denken auf eine stärkere verbale Nutzung basieren könnte.

4.1.1.3 Soziale Faktoren

Die kognitiven Geschlechtsdifferenzen werden nicht nur durch biologische und psychologische Faktoren determiniert, sondern auch durch sozialen Faktoren. Die Ergebnisse der Sozialisationsforschung haben belegt, dass die Erfahrungen, die durch den Sozialisationsprozess erworben werden, auf die Ausbildungsinteressen von Jungen und Mädchen unterschiedlich Einfluss nehmen (Beerman et al., 1992). Die Sozialisationsforschung interessiert sich dafür, wie Geschlechtsrollen erworben und wie geschlechtstypische Verhaltensweisen angeeignet werden.

Rollenmodelleffekte haben einen großen Einfluss auf die Interessen und die Entscheidungen für eine Fachrichtung. Es gibt traditionell in der deutschen Gesellschaft weniger weibliche Rollenmodelle in den Fachbereichen Mathematik und Naturwissenschaften. Dieser Umstand führt zu der Frage, ob die Interessen und die Entscheidungen für diese Fächer oder Bereiche bei

Mädchen vom Geschlecht des Lehrers abhängen und welche Voraussetzungen das Modell erfüllen muss. Es besteht Übereinstimmung darüber, dass ein regulärer Kontakt von Modell und Beobachter erforderlich ist. Außerdem sollte das Modell die Lernenden aktiv unterstützen und ermutigen, um gute Ergebnisse zu erzielen (Beerman et al., 1992). In einer Untersuchung von Casserly und Rock (1979, zit. nach Beerman et al., 1992, S. 55) wurde festgestellt, dass das Geschlecht des Lehrers und bestimmte Arten von Erfahrungen eine wichtige Rolle für das Interesse von Mädchen an Mathematik spielen. Darüber hinaus sind Lantz und Smith (1981, zit. nach Beerman et al., 1992, S. 55) zu dem Ergebnis gekommen, dass die Bezugspersonen und der subjektive Wert von Mathematik die besten Prädiktorvariablen hinsichtlich einer Entscheidung für Mathematik sind.

Somit stehen die Rollenmodelleffekte in Zusammenhang mit der Einstellung und der Erwartung von Eltern bezüglich Mathematik, die einen Einfluss auf den Sozialisationsprozess des Kindes ausüben. Die Forschungsergebnisse zu diesem Thema sind nicht einheitlich. Aber die meisten der Befunde zeigen, dass die Eltern die mathematische Begabung ihrer Töchter niedriger bewerten als jene ihrer Söhne. Auf diese Weise übt die Elternerwartung einen starken Einfluss auf die Leistung ihrer Kinder aus. Dies gilt auch für Lehrkräfte, welche die Jungen für begabter als die Mädchen halten. Nach ihrer Ansicht sind die Mädchen fleißig und ordentlich und die Jungen sind wissenschaftlich und ideenreich. Da diese Zuschreibungen bzw. Attributionen in der Lehrer-Schüler-Aktion zum Ausdruck kommen, haben sie letztlich sicherlich einen bedeutenden Einfluss auf die mathematische Leistung (vgl. Hoffmann, 1988 a/b; Malcom, 1988 zit. nach Beerman et al., 1992, S. 61).

Ein weiterer wesentlicher sozialer Einflussfaktor ist der sog. „Stereotype Threat“ (vgl. Hirnstein & Hausmann, 2010, S. 79). Der „Stereotype Threat“ besagt, dass eine Gruppe oder ein Individuum aufgrund eines negativen Vorurteils über die eigene Leistungsfähigkeit tatsächlich schlechte Leistung erbringt. Dieser Ausdruck wurde in Zusammenhang mit Unterschieden bei Intelligenztests geprägt. Die Effekte, die z. B. in einem mathematischen Test auftreten, können verschwinden, wenn der Test als allgemeiner Problemlösungstest und nicht als Mathematiktest dargestellt wird. Darüber hinaus können negative Stereotype über eine andere Gruppe bzw. positive Stereotype über die eigene Gruppe auch die eigene kognitive Leistung verbessern. So erreichen z.B. Frauen bessere Ergebnisse in einem Mathematiktest oder in einem Rotationstest, wenn sie mit einem positiven Stereotyp konfrontiert werden. Im Gegensatz dazu bekommen Männer schlechtere Noten in den gleichen Tests, wenn sie mit einem negativen Stereotyp konfrontiert werden.

4.1.1.4 Biopsychosoziale Faktoren

Der biopsychosoziale Ansatz (Beerman et al., 1992) stellt ein integratives Modell dar, das auf den empirischen Forschungsergebnisse und den überprüfenden Interventionsstudien basiert und das darauf hinweist, dass sich die Leistung in einem Fach aus einer Kombination verschiedener Einflussfaktoren ergibt. In dem Modell von Kulm (1980, zit. nach Beerman et al., 1992, S. 70) steht die Einstellung mit dem Verhalten in Zusammenhang und jeder Einflussfaktor kann sich positiv oder negativ auswirken, erst die Summe der Faktoren entscheidet den Einfluss auf das konkrete Verhalten. In diesem Modell wird die Bedeutung externer Faktoren wie z. B. die Lernsituation und vermittelnde Faktoren berücksichtigt. Fennema und Peterson (1985, zit. nach Beerman et al., 1992, S. 72) legen in ihrem Modell großen Wert auf bestimmte Verhaltensweisen wie die Fähigkeit, komplexe Aufgaben erfolgreich zu lösen, Ausdauer zu haben und unabhängig zu arbeiten. Diese Verhaltensweisen spielen eine wesentliche Rolle für den Erfolg in Mathematik und werden nach Fennema und Peterson (1985) „das autonome Lernverhalten“ genannt, auf das interne Faktoren wie die Selbstsicherheit im Erlernen von Mathematik, Ursachenzuschreibung für den Erfolg und Misserfolg in Mathematik, sowie externe Faktoren wie Klassenklima und vor allem die Interaktion zwischen Schüler und Lehrer, Einfluss ausüben.

Die Studien zur Geschlechterforschung (vgl. Hirnstein & Hausmann, 2010) stimmen darin überein, dass kognitive Geschlechtsunterschiede aus einer Interaktion von biologischen, psychologischen und sozialen Faktoren entstehen. Aber leider gibt es kaum wissenschaftliche Arbeiten, welche die biologischen, psychologischen und sozialen Faktoren integrativ untersuchten und die vorliegenden Ergebnisse sind unterschiedlich bis widersprüchlich. In einer Untersuchung (Hausmann, et al., 2009, zit. nach Hirnstein & Hausmann, 2010, S. 80) wurde der Einfluss von Geschlechtsstereotypen und Sexualhormonen auf kognitive Geschlechtsunterschiede erforscht.

In dieser Studie wurde die Stichprobe (100 Probanden) in zwei Gruppen eingeteilt. Das Geschlechtsstereotyp wurde durch einen Fragebogen nur bei einer Gruppe aktiviert, die zweite Gruppe (Kontrollgruppe) erhielt eine geschlechtsneutrale Version des Fragebogens. Danach mussten alle Probanden eine kognitive Testbatterie, die aus räumlichen (wie z.B. mentalen Rotationen) und verbalen Aufgaben bzw. Tests zur Wortflüssigkeit bestand, bearbeiten. Anschließend wurde der Testosteronspiegel der Versuchspersonen bestimmt. Die Ergebnisse

zeigten, dass die Frauen in verbalen Aufgaben und die Männer in MRT Test bessere Leistung erbrachten.

Diese Effekte im MRT-Test zeigten sich vor allem in der Gruppe, in der das Stereotyp vorher aktiviert wurde. Zusätzlich zeigte sich in der Gruppe der Männer bei der die Stereotyp aktiviert wurde, ein doppelt so hoher Testosteronwert wie in der Kontrollgruppe. Dieser erhöhte Testosteronausschuss könnte einen Einfluss auf die mentale Rotation haben. In einer späteren Studie (vgl. Hirnstein & Hausmann, 2010, S. 80) wurde auch weiter festgestellt, dass Männer im MRT und Frauen im verbalen Test besser abschnitten, aber der Effekt des aktivierten Stereotyps zeigte sich nur für den Test zur Wortflüssigkeit und nicht für den MRT.

Wie oben gezeigt wurde, unterscheiden sich Jungen und Mädchen in bestimmten kognitiven Fähigkeiten. Die Frage, ob die Geschlechtsdifferenzen auf einzelne Faktoren wie biologische (Anlage), soziale (Umwelt) oder auf mehrere Faktoren gleichzeitig zurückzuführen sind, ist nicht so einfach zu beantworten und führt zur „Anlage oder Umwelt“-Diskussion. Weder die angeborenen noch die sozialen Faktoren können allein auf die Geschlechtsunterschiede Einfluss nehmen. Diese Unterschiede ergeben sich aus einem Komplex oder einer Interaktion verschiedener Faktoren wie biologische (Geschlechtshormonelle Effekte), soziale (Geschlechtsrolle, Geschlechtsstereotypen) und psychologische (Motivation, Interesse).

4.1.2 Kulturelle Unterschiede in der Schulleistung

Die Ergebnisse hinsichtlich kognitiver Fähigkeiten zeigten, dass Schüler von einigen ethnischen Gruppen unterdurchschnittliche Leistung erbrachten. In den letzten Jahrzehnten nimmt das Niveau der Schulleistung jedoch zu. Während im Jahr 2000 in den USA 94% der Weißen, 87% Afroamerikaner und 63% der lateinamerikanischen Immigranten, die zwischen 25 und 29 Jahren alt waren, eine höhere Schule beendeten, erreichten in Deutschland 97% der Haupt- und Realschüler und 74,4% der Gymnasiasten einen Abschluss, aber nur 13,8% der Gymnasiasten schlossen ihr Studium ab. Diese Unterschiede führen die Forscher nicht auf die kognitiven Fähigkeiten zurück, sondern auf den sozioökonomischen Status. Allerdings verringern sich diese Differenzen, wenn man den sozioökonomischen Status zwischen ethnischen Gruppen kontrolliert (Byrnes, 2003; Uline & Johnson, 2005; Gleitman, Fridlund & Reisberg, 1999, zit. nach Woolfolk, 2008, S. 212-213).

In Deutschland wird das Problem der schlechten Leistung von Migrantenkindern im Schulsystem mit ethnischen Diskriminierungen durch Lehrer und Lehrerinnen in Verbindung gebracht (vgl. Kristen, 2006). Die Diskriminierung zeige sich darin, dass Kinder aus unterschiedlichen ethnischen Gruppen, die gleiche Ergebnisse erreichten, aufgrund ihrer Herkunft von den Lehrerinnen und Lehrern unterschiedlich bewertet werden. Die Ursachen der ethnischen Diskriminierung können auf individuellen Präferenzen von Lehrern und Lehrerinnen für oder gegen eine Gruppe oder auf Informationsdefizite zurückgeführt werden. d. h. wenn die Lehrerinnen und Lehrer keine oder unzureichende Informationen über die individuelle Leistung von Mitgliedern der Gruppe haben, wird die individuelle Leistung aufgrund der Gruppenzugehörigkeit bewertet. Deshalb wird die individuelle Leistung unangemessen beurteilt.

In einer eigenen Studie wurde anhand von Daten einer Teilstudie des DFG-Projekts „Bildungsentscheidungen in Migrantenfamilien“, die im Rahmen einer Studie zu ersten Übergangentscheidung von türkischen Familien im Vergleich zur deutschen Bevölkerung erhoben worden waren, untersucht, ob ethnische Unterschiede in der Notenvergabe auch nach Berücksichtigung der Testergebnisse in Fächern wie Deutsch und Mathematik sowie in Schullaufbahnempfehlungen festzustellen sind (vgl. Kristen, 2006).

Diese Untersuchung wurde gegen Ende 2000/2001 in Mannheim durchgeführt. Die Stichprobe wurde aus Schülerinnen und Schülern solcher vierten Klassenstufen in sechs Mannheimer Grundschulen gebildet, wo es eine ausreichend große Anzahl von türkischen und italienischen Kindern gab. Dabei wurden im Rahmen dieser Studie unterschiedliche Fähigkeiten der Schülerinnen und Schüler anhand von Leistungstests geprüft. Eingesetzt wurden dabei folgende Tests:

- Sprachfreie Tests wie der Coloured Progressive Matrices (CPM),
- Reproduktionen des Heidelberger Intelligenztests (HIT 3-4, Teil 3) und der
- kognitive Fähigkeitstests (KFT 4R) mit nur zwei Teilen, dem verbalen Teil 1 und quantitativen Teil 3.

Darüber hinaus sollten die Lehrkräfte, die Informationen über die Nationalitäten der Schüler und Schülerinnen erhielten bzw. hatten, in Klassenlisten die Noten der Schüler und Schülerinnen in den Fächern Deutsch und Mathematik sowie die Schullaufbahnempfehlung eintragen. Danach wurden die Leistungstestwerte mit den Einschätzungen der Lehrer und Lehrerinnen verglichen. Die deskriptiven Ergebnisse spiegeln die bekannten Muster ethnischer Unterschiede im Schulerfolg wider. In den Fächern Deutsch und Mathematik und in der Schullaufbahnempfehlung

schneiden deutsche Kinder besser ab. Um die Frage zu beantworten, ob nach Einbezug der Testwerte Nachteile für bestimmte Gruppen sichtbar werden, die auf ethnische Diskriminierung hinweisen, wurden multivariate Analysen durchgeführt. Hierbei zeigte sich, dass es keinen Hinweis darauf gibt, dass Lehrerinnen und Lehrer Empfehlungen in Abhängigkeit der ethnischen Herkunft aussprechen. Türkische und italienische Kinder erzielen nach Einbezug der Leistungswerte etwas schlechtere Noten als ihre deutschen Mitschülerinnen und Mitschüler. Ein Grund hierfür könnte in der Auswahl der Skalen der eingesetzten Testinstrumente liegen, die neben dem Wortschatz keine weiteren Deutschkompetenzen maßen (vgl. Kristen, 2006, S. 79-95).

Ethnische Leistungsunterschiede finden sich nicht nur in Deutschland, sondern etwa auch in Großbritannien. Sammons (1999, zit. nach Tikly, 2006, S. 97) kam in seiner Untersuchung von Vorschulkindern zu dem Ergebnis, dass es Leistungsunterschiede zwischen den Kindern, die aus unterschiedlichen ethnischen Gruppen stammen, in ihren intellektuellen Fähigkeiten bzw. ihren verbalen und nonverbalen Kompetenzen gibt. Während die britischen Kinder gute Ergebnisse erzielten, erbrachten pakistanische Kinder die niedrigsten Leistungen, gefolgt von afrikanischen Kindern. Außerdem verringerten sich diese Differenzen, wenn einige Faktoren wie z. B. die Bildung der Eltern und ihre Berufe berücksichtigt wurden. Zusätzlich wurde gezeigt, dass die Differenzen zwischen den Gruppen im Bereich der nonverbalen Leistung statistisch nicht signifikant sind, wenn diese Faktoren beachtet wurden.

Darüber hinaus wurde darauf hingewiesen, dass die ethnischen Unterschiede nicht nur in den Vorschulen existieren, sondern auch in den Schulen. Die meisten Untersuchungen, die in den Schulen durchgeführt wurden, stimmen darin überein, dass die indischen und chinesischen Kinder überdurchschnittliche Leistungen bei den Prüfungen des General Certificate of Secondary Education (GCSE) für das Jahr 2003 im Vergleich zu schwarzen Kindern und der Gruppe von Kindern, die aus Bangladesch kamen, erbrachten.

Das britische Bildungsministerium argumentiert bezüglich der ethnischen Differenzen zwischen den verschiedenen Gruppen von Migrantenkindern damit, dass die meisten der Migrantfamilien in benachteiligen Regionen wohnten. Sie leben normalerweise in großen Städten wie z.B. London, wo ein Großteil der Bevölkerung ausländisch ist. Zusätzlich zeigen einige Untersuchungen, dass die soziale Schicht einen starken Einfluss auf den Schulerfolg von Migrantenkindern ausübt und dieser Einfluss für die verschiedenen Gruppen unterschiedlich stark ausgeprägt ist (Troyna 1984; DES 1985; Drew/Gray 1990; zit. nach Tikly 2006, S. 99). Außerdem wurde nachgewiesen, dass die Schüler und Schülerinnen, die aus Familien mit niedrigerem

sozioökonomischen Status kommen, schlechtere Ergebnisse erzielen als die aus Familien mit höherem sozioökonomischen Status. Diese Differenz ist auch bei weißen Kindern auffällig (DfES, 2003, zit. nach Tikly, 2006, S. 99).

Wie oben anhand der vorgestellten Studien gezeigt wurde, kann man sagen, dass Migrantenkinder unterdurchschnittliche Leistungen oder schlechtere Ergebnisse aufweisen und diese Aussage nicht nur in Deutschland, sondern auch in anderen Ländern wie z. B. Großbritannien gilt. Die Frage, die sich hier stellt ist, welche Gründe dafür zu finden sind oder wie diese Unterschiede erklärt werden können. In der Literatur gibt es Erklärungen und Argumentationen, welche die Nachteile von Migrantenkinder im deutschen Schulsystem durch empirische Studien belegen (vgl. Diefenbach, 2007).

4.1.2.1 Die Erklärung durch kulturelle Defizite

Die Grundthese der Erklärung durch kulturelle Defizite lautet, dass Migrantenkinder „aufgrund ihres kulturellen Erbes Defizite hinsichtlich dessen aufweisen, was als „'Normalausstattung' an Verhaltensweisen, Kenntnissen und Fähigkeiten [vorausgesetzt wird]...“ die ein Kind oder ein Jugendlicher eines bestimmten Entwicklungsstandes in die Institutionen der Bildung und Erziehung mitbringe“ (Gogolin 2002: 264, zit. nach Diefenbach, 2007, S. 91). In dieser Erklärung werden zwei Varianten unterschieden. Die erste Variante gibt an, dass die Nachteile von Migrantenkinder/Schülern auf die Herkunfts-Lernkultur als ein Teil von der Herkunfts kultur zurückzuführen ist. Die zweite Variante erläutert diese Unterschiede durch eine defizitäre schichtspezifische Kultur oder Sozialisation.

Variante 1: Defizitäre Herkunfts-Lernkultur. Dieter Cleassens (1962) hat die Idee der Enkulturation in die deutsche Migrantenforschung eingeführt. Er geht davon aus, dass ein Kind mit Enkulturation in seiner Kernfamilie eine Basispersönlichkeit bildet. Der Autor behauptet: „das Erleben bestimmten Verhaltens der Eltern in den' Schlüsselsituationen' formt also letztlich die 'Basic Personality' und präjudiziert damit weitgehend späteres eigenes Verhaltens“ (Cleassens 1962, zit. nach Diefenbach, 2007, S. 92). Die Basispersönlichkeit nach Cleassens ist kaum veränderbar, weil die Grundlage der Basispersönlichkeit frühen Einflüssen zuzuschreiben ist. Wenn ein Kind aus einer Migrantenfamilie, aus einer bestimmten Kultur kommt und sich trotzdem anders verhält oder wegen der Migration eine andere Kultur repräsentiert, ergibt sich die Frage: welche Effekte hat die Basispersönlichkeit?

Wenn zwischen der Basispersönlichkeit des Kindes und der umgebenden Kultur eine Kluft besteht, gerät das Kind wahrscheinlich in Schwierigkeiten mit seiner Umgebung. Die Basispersönlichkeit des Kindes ist in einem neuen kulturell anders geprägten Umfeld defizitär (Diefenbach, 2007, S. 93).

Diese defizitäre Basispersönlichkeit ist nur eine von den kulturbedingten Defiziten. Es gibt natürlich weitere Faktoren, die einen Einfluss auf den Schulerfolg von Migrantenkindern ausüben, wie z. B. mangelnde Anerkennung von Lernen und Schulen. In einer empirischen Studie von Leenen, Grosch und Kreidt (1990, zit.nach Diefenbach, 2007, S. 93) haben die Forscher die Hypothese aufgestellt, „dass die türkischen Migranten eine traditionelle Haltung dem Lernen der Schule gegenüber haben, die durch Auswendiglernen des präsentierten Stoffs und die absolute Autorität der Lehrer in allen Fragen des schulischen Betragens und Erfolgs der Kinder gekennzeichnet sei“ (Diefenbach, 2007, S. 93). In diesem Fall leiden die türkischen Migrantenkinder unter einem Generationskonflikt, weil die Kinder entweder die negative Meinung der Eltern in Bezug auf das deutsche Schulsystem akzeptieren oder sie sich selbst an die neue Kultur und die Modernisierung anpassen müssen. Leenen et al., (1990, zit. nach Diefenbach, 2007, S. 93-95) führen den Misserfolg von Migrantenkindern/ Schülern im Bildungssystem auf das kulturelle Erbe zurück und um dieses Problem lösen zu können, müssen die Kinder über ein gewisses Ausmaß an Modernisierung und Individualisierung verfügen und vor allem die kulturellen Defizite überwinden.

Um genauer zu erfahren, welche Bedeutung die Bildung, Lernen und die Schule in den Migrantenfamilien hat, haben Leenen et al., (1991) eine erfolgreiche Gruppe von türkischen Jugendlichen befragt. Die Anzahl der Stichprobe waren 25 Studenten aus unterschiedlichen Hochschulstudiengängen und eine Kontrastgruppe mit 15 Personen, welche die Hauptschule abgeschlossen hatten. Anhand der durchgeführten Interviews stellte sich heraus, dass die Befragten für ihren Bildungserfolg einen kulturellen Konflikt mit ihren Eltern in Kauf nehmen mussten. Dieser Befund bestätigt die Behauptung der Autoren. Trotzdem wird nicht angenommen, dass die traditionelle türkische Lern-Kultur für den Misserfolg der Jugendlichen von Migrantenfamilien verantwortlich ist, weil der Informationsgehalt der Studie sehr niedrig ist. In einer Untersuchung von Nauck (1985, zit.nach Diefenbach, 2007, S. 93) wurde kaum ein autoritärer Erziehungsstil in türkischen Familien gefunden und viele andere Studien haben gezeigt, dass die türkischen Familien viel Respekt für die Werte von Leistung und Lernen hatten, sogar mehr als die deutschen Familien.

Der Erklärungsansatz der defizitären Herkunfts- Lernkultur stieß auf Kritik (vgl. Diefenbach, 2007, S. 95-100), weil nicht nur die türkischen Kinder, sondern auch die italienischen Kinder schlechtere Ergebnisse im Vergleich zu allen Nationalitäten erreichten, obwohl die italienische Kultur der deutschen Kultur ähnlicher ist als die türkische Kultur. Darüber hinaus wurde auch festgestellt, dass die Migrantenkinder, die in Deutschland geboren waren oder die vor dem siebten Lebensjahr nach Deutschland gekommen waren (sog. Bildungsinländer) im Allgemeinen mehr Sekundarschulabschlüsse aufwiesen als andere Migrantenkinder. Zusätzlich bestehen signifikante Unterschiede zwischen verschiedenen Nationalitäten, wenn nur der Faktor Bildungsinländer berücksichtigt wird (Haug, 2002, zit. nach Diefenbach, 2007, S. 96). Das stimmt teilweise mit dieser Erklärung überein. Deswegen wurde anhand der Daten des Integrationssurveys des Bundesinstitutes für Bevölkerungsforschung (BiB) geprüft, inwieweit der Bildungserfolg von Jugendlichen aus Migrantenfamilien unter Berücksichtigung der Bildungsinländer durch defizitäre Herkunftskultur erklärt werden kann. Diese These konnte allerdings nicht durch die Daten gestützt werden. Insgesamt lässt sich der Bildungserfolg der Migrantenkinder durch den Erklärungsansatz der defizitären Herkunfts- Lernkultur weder theoretisch noch empirisch nicht bestätigen.

Variante 2: Unterschichtskultur als defizitäres Sozialisationsumfeld. Laut dem Ansatz der Unterschichtskultur als defizitäres Sozialisationsumfeld werden die Misserfolge von Migrantenkindern/Schülern im deutschen Schulsystem durch kulturelle Defizite mit dem Sozialcharakter angegeben. Nach Riesman, Denney und Glazer (1958) ist der soziale Charakter „der Teil des 'Charakters', der signifikanten sozialen Gruppen gemeinsam ist und der nach der Definition der meisten zeitgenössischen Sozialwissenschaftler das Produkt der Erfahrungen dieser Gruppen darstellt. Der so verstandene Begriff des sozialen Charakters erlaubt uns, von dem Charakter von Klassen, Gruppen, Völkern und Nationen [!] zu sprechen“ (zit. nach Diefenbach, 2007, S. 101).

Der soziale Charakter wurde kritisiert, vor allem die Annahme, dass die Sozialisationsbedingungen in den jeweiligen verschiedenen sozialen Schichten ähnlich sind. Die Sozialisationsbedingungen wie z.B. das mangelnde Engagement der Eltern hinsichtlich der Schulleistung ihres Kindes oder die Wahrung der Autorität der Eltern hinsichtlich des Erziehungsstils haben einen Einfluss auf den Schulerfolg von Migrantenkindern. Dazu kommt, dass die Gleichsetzung von Migrantenfamilien mit Unterschichtsfamilien zweifelhaft ist (vgl. Diefenbach, 2007, S. 101).

Die Variante Unterschichtskultur als defizitäres Sozialisationsumfeld begründet die Nachteile von Migrantenkindern/ Schülern auf den Schulerfolg durch defizitäre Ressourcenausstattung der Unterschichtsfamilien wie z. B. in Bezug auf Wohnverhältnisse oder ökonomische und zeitliche Ressourcen etwa durch ein größeres Ausmaß an Müttererwerbstätigkeit aufgrund der fehlenden Finanzierung (vgl. Diefenbach, 2007, S. 101-10).

4.1.2.2 Die humankapitaltheoretische Erklärung

Diese Erklärung geht davon aus, dass es den Migrantenkindern stärker als den deutschen Kindern an Humankapital fehlt. Dieses ist sehr wichtig für den Erfolg im deutschen Schulsystem. Es zeigt sich in der Bildungsökonomie als „alle Investitionen, die in einem Menschen im Verlauf seiner Erziehung und Ausbildung gemacht werden und die ihm monetäre und nicht monetäre Erträge bringen“ (Diefenbach, 2007, S. 103).

Die familiale Sozialisation spielt eine große Rolle bei der Vermehrung von Humankapital, weil sich die Eltern darum bemühen, dass sie ihren Kindern Wissensbestände, Kenntnisse, Gewohnheiten usw. vermitteln, die den Erfolg in den Bildungsinstitutionen beeinflussen. Aus diesem Grund werden die Bildungsabschlüsse der Eltern und ihr Einkommen als Indikatoren für Humankapital bezeichnet (Krüsselberg, Auge & Hilzenbecher 1986:115, zit. nach Diefenbach, 2007, S. 103). Im Sinne der Humankapitaltheorie erbringen die Migrantenkinder schlechtere Leistungen, weil ihre Eltern keine Bildungsabschlüsse, eine geringere Bildung und ein geringeres Einkommen haben. Außerdem haben die Migrantenfamilien mehr Kinder als die deutschen Familien und das wirkt sich negativ auf die Akkumulation von Humankapital und dann auf den Schulerfolg bei den Migrantenkindern aus, d.h. Migrantenkinder haben weniger familiäre Ressourcen als deutsche Kinder.

Die Humankapitaltheorie wurde anhand empirischer Studien nur teilweise bestätigt. In den USA wurde gezeigt, dass die Bildung der Kinder mit der Bildung ihrer Eltern im positiven Zusammenhang steht. Darüber hinaus erreichen die Kinder aus Familien mit höherem Gehalt schneller einen Bildungsabschluss als die Kinder aus Familien mit niedrigem Einkommen (Becker, 1993a; 1993b, zit. nach. Diefenbach, 2007, S. 104). In Deutschland wurde die Humankapitaltheorie von Nauck, Diefenbach und Petri (1998, zit. nach Diefenbach, 2007, S. 104) anhand der Daten des Sozioökonomischen Panels (SOEP) untersucht. Die Ergebnisse zeigten, dass die deutschen

Kinder bessere Chancen haben, einen Schulabschluss zu erreichen als Migrantenkinder. Außerdem steht der Bildungserfolg von Jugendlichen aus Migrantenfamilien in einem signifikanten positiven, aber sehr geringen Zusammenhang mit dem ökonomischen und kulturellen Kapital der Herkunftsfamilie. Dies gilt aber nicht für den Bildungserfolg von deutschen Jugendlichen. Zusätzlich übt die Anzahl der Kinder im Haushalt bei Migrantenfamilien einen stärkeren Einfluss aus als von der Bildung der Eltern und von ihrem Gehalt auf die Wahrscheinlichkeit eines Kindes, einen höheren Schulabschluss als den Hauptschulabschluss zu erreichen. Dieser Einfluss konnte bei deutschen Familien nicht bestätigt werden.

Nach dieser Darstellung vom Problem des Misserfolgs der Migrantenkinder im deutschen Schulsystem und die Erklärung für dieses Phänomen ergibt sich, dass die Erklärung der kulturellen Defizite teilweise durch die Basispersönlichkeit und dem Generationskonflikt zutreffend ist, wenn die Migranteneltern eine negative Meinung gegen das deutsche Schulsystem haben. Jedoch sind die empirischen Befunde unterschiedlich und stützen die Annahme der kulturellen Defizite nicht. In Bezug auf die Humankapitaltheoretische Erklärung wurde empirisch teilweise festgestellt, dass die Bildung der Eltern und die Anzahl der Geschwister als Indikatoren für das Humankapital gelten.

4.2. Intelligenz als Prädiktor für Erfolg in Schule und Beruf

4.2.1 Berufserfolg

Die psychologische Forschung kommt zu dem Ergebnis, dass der Ausbildungserfolg durch Intelligenz gut vorhersagbar ist. Außerdem wurde festgestellt, dass Intelligenz nicht nur als der beste Einzelprädiktor von Ausbildungserfolg ist, sondern auch 25 Prozent einen beachtlichen Anteil der Varianz des Ausbildungserfolgs aufklärt. Dieser Befund wurde durch viele Untersuchungen belegt, die in unterschiedlichen Metaanalysen zusammengefasst sind (Schmidt & Hunter, 1998, zit. nach Schmidt-Atzert, Deter & Jaeckel. 2004, S. 147). Dieses Ergebnis zeigt den Zusammenhang zwischen der allgemeinen Intelligenz und dem Ausbildungserfolg. Intelligenz wird definiert als die Fähigkeit abstrakt zu denken und Probleme zu lösen, aber in der psychologischen Forschung wird sie als die allen kognitiven Leistungstests gemeinsame Varianz definiert bzw. operationalisiert. Jeder kognitiver Leistungstest erfasst zu einem Teil die allgemeine Intelligenz. Zusätzlich können auch eine oder mehrere spezifische Fähigkeiten

gemessen werden, wie z.B. verbale und quantitative Fähigkeiten (vgl. Schmidt-Atzert et al., 2004, S. 148; siehe auch Kapitel 2). Viele Studien beschäftigen sich mit der Beantwortung der Frage, ob die allgemeine Intelligenz alleine zur Prognose von Berufserfolg geeignet ist oder ob spezifische kognitive Fähigkeiten zusätzlich dazu beitragen.

Im Jahr 1991 haben die Wissenschaftler Ree und Earle (zit. nach Schmidt-Atzert et al., 2004, S. 148) versucht, diese Frage zu beantworten. Sie untersuchten über 78.000 Mitglieder der US amerikanischen Luftwaffe. Die Teilnehmer hatten eine militärische Grundausbildung und eine Ausbildung in 82 unterschiedlichen Berufen durchlaufen. Die Ausbildung enthielt verschiedene Aufgaben, wie z. B. den Umgang mit Polizeihunden oder eine Tätigkeit wie Feuerwehrmann. Bei dieser Studie wurde die Armed Services Vocational Aptitude Battery (ASVAB) als Prädiktor für den Berufserfolg eingesetzt. Diese Testbatterie besteht aus 10 Subtests, wie z. B. Tests für rechnerisches Denken, Kenntnis von Wörtern, mechanisches Verständnis usw. Die allgemeine Intelligenz und die spezifischen Fähigkeiten wurden durch Faktorenanalyse bestimmt. Die Ergebnisse zeigten, dass die spezifischen Fähigkeiten mehr Varianz aufklärten als die generelle Intelligenz. Allerdings wird der Unterschied auch auf die Größe der Stichprobe zurückgeführt und diese Differenz hat eine geringere praktische Bedeutung. In weiteren Untersuchungen haben Ree und Earles (1996, zit. nach Schmidt-Atzert et al., 2004, S. 148) allerdings festgestellt, dass die allgemeine Intelligenz der beste Einzelprädiktor ist und die spezifischen Fähigkeiten kaum oder keine zusätzliche Varianz aufklären.

In einer Untersuchung von Schmidt-Atzert, Deter und Jaeckel (2004) wurde überprüft, ob der Befund von Ree und Earles auf das deutsche Ausbildungssystem übertragen werden kann. In dieser Studie wurden 646 Auszubildende aus vier technischen Berufen und zwei Büroberufen in der chemischen Industrie untersucht. Die Forscher setzten zur Messung der Intelligenz ausgewählte Subtests des Wilde-Intelligenz-Tests von Jäger und Athoff (1983), der auf der Grundlage des Thurstone-Modells (1938) konzipiert ist (siehe Kapitel 2) ein, sowie weitere Tests. Die Auswahl der weiteren Tests orientierte sich daran, Fähigkeiten abzudecken, die im Betrieb wichtig für den jeweiligen Beruf sind. Für die Mechanikerberufe wurden Testverfahren ausgewählt, die z.B. praktisch-technisches Verständnis abfragten. Für die Bürokommunikation wurde der Bürotest von Marschner (1967) eingesetzt. Allen Bewerbern wurde der Test d2 von Brickenkamp, 1962 vorgelegt. Die Tests wurden vor Beginn der Ausbildung durchgeführt. Als Kriterien für den Ausbildungserfolg wurden die schriftliche und praktische Abschlussprüfung sowie für einige Auszubildende die betriebliche Beurteilung herangezogen. Die allgemeine Intelligenz erwies sich außer bei dem Beruf

Bürokommunikation als der beste Prädiktor zur Prognose von theoretischen Kenntnissen. Etwa 4 Jahre nach der Ausbildung zeigte die allgemeine Intelligenz eine durchschnittliche Validität von $r = 0,69$.

In Bezug auf die spezifischen kognitiven Fähigkeiten nach Thurstone leistete nur eine spezifische Fähigkeit (räumliches Denken) bei einem Beruf (Mechanikerberufe) einen zusätzlichen Beitrag bei den theoretischen Kenntnissen. Nach Schmidt-Atzert et al., (2004, S. 147) eignet sich die Kombination von allgemeiner Intelligenz und berufsspezifischem Vorwissen am besten zur Vorhersage von Ausbildungserfolg.

In einer Metaanalyse von Hülsheger, Maier, Stumpp und Muck (2006) wurden die Validitäten unterschiedlicher Intelligenztests wie der IST-70 (Amthauer, 1973), der WIT (Jäger & Althoff, 1994), das PSB (Horn, 1969), das LPS (Horn, 1983), SPM (Kratzmeier, 1979) und CFT 3 (Weiß, 1971) quantitativ zur Vorhersage von Ausbildungserfolg in Deutschland integriert und die verschiedenen Intelligenztests aufgrund ihrer mittleren korrigierten Validität verglichen. Außerdem wurde untersucht, ob die Validitäten der Intelligenztests in Abhängigkeit von Ausbildungserfolgskriterien wie z. B. Noten oder Leistungsbeurteilungen variieren. Die Analysen beruhten auf 90 unabhängigen Stichproben. Obwohl die Ergebnisse der Analysen Validitätsunterschiede zwischen einigen der eingesetzten Tests wie z. B. zwischen dem WIT, dessen operationale Validität .54 aufgewiesen hat, und dem IST-70, dessen operationale Validität nur .48 betrug, ergaben, haben sich trotzdem alle untersuchten Tests als valide Prädiktoren von Ausbildungserfolg erwiesen. Außerdem wurde festgestellt, dass die verschiedenen hohen Validitäten der Intelligenztests je nach Ausbildungserfolgskriterium (Noten, Beurteilungen) auf verschiedene Reliabilitäten dieser Kriterien oder auf die Unterschiede in der Messgenauigkeit zurückgeführt werden konnten. Hier weisen Noten eine höhere Reliabilität als Beurteilungen auf (Hülsheger et al., 2006, S. 145-162).

4.2.2 Intelligenz und schulische Leistungen

Intelligenz gilt im schulischen Feld als ein wichtiger Prädiktor von schulischen Leistungen (Perleth, 2008). Die Forschungsergebnisse bezüglich der Beziehung zwischen allgemeiner Intelligenz und Schulleistung zeigten, dass starke Zusammenhänge zwischen allgemeiner Intelligenz und Schulleistung bestehen. Die Korrelationskoeffizienten zwischen allgemeiner Intelligenz und Schulleistung liegen zwischen $r = .40$ bis $r = .70$ (Jencks, 1979; Mackintosh, 1998, zit. nach Maltby, Day & Macaskill, 2011, S. 724). In der Münchener Längsschnittstudie zur Hochbegabungsentwicklung (1985-1988, zit. nach Heller, 2008, S. 184-185) wurde

nachgewiesen, dass die allgemeine Intelligenz (erfasst durch die KFT Gesamtleistung) ein sinnvoller Prädiktor für die schulische Leistung darstellt. Darüber hinaus zeigten die KFT Dimensionen unterschiedliche hohe Zusammenhänge mit den Schulnoten in Deutsch, Englisch und Mathematik. Außerdem wurden nicht kognitive Persönlichkeitsmerkmale wie z. B Motivation, Interesse usw. als vermittelnde Faktoren zwischen Intelligenz und Leistung gekennzeichnet.

Im außerschulischen Bereich und im beruflichen Bereich spielt Intelligenz eine geringe Rolle für die Prognose von Leistung. Nach Ericsson und Charness (1994, zit. nach Perleth, 2000, S. 46) gibt es keinen Einfluss von Begabung auf außergewöhnliche Leistungen. Zusätzlich sind Grundkomponenten kognitiver Leistung wie Wahrnehmungsgeschwindigkeit, Reaktionszeiten und auch die körperlichen Merkmale wie Herz, Lunge und die Stärke der Knochen trainierbar. Ericsson argumentiert gegen die Bedeutung von Begabungseffekten für hervorragende Leistungen; seiner Auffassung nach können also Begabungsmerkmale außergewöhnliche Leistung nicht vorhersagen und Begabungstests messen einfach tatsächlich erworbenes Wissen. Dazu käme, dass die Wunderkinder, die früh außergewöhnliche Leistungen in Musik oder in Schach zeigten, dieselben Lernschritte wie andere Kinder durchlaufen, die jeweiligen Stufen aber schneller erreichen.

Im Gegensatz dazu wurde nachgewiesen, dass jedes Leistungsniveau von der Förderung der Umwelt wie z. B: Eltern und Schule abhängt. Nach Ansicht von Ericsson und Mitautoren (Ericsson & Smith, 1991; Ericsson, Krampe & Tesch-Römer, 1993; Ericsson, Krampe & Heizmann, 1993; Ericsson & Charness, 1994; Ericsson, 1996, zit. nach Perleth, 2000, S. 46) spielt die Erfahrungen für die Entwicklung von Expertise eine weitaus größere Rolle als die Intelligenz, aber nicht jede Erfahrung führt zu Expertise, sondern nur die Erfahrung, die in Lernprozessen durch aktives und zielgerichtetes Lernen erworben wurde.

4.3 Zusammenfassung

In diesem Kapitel wurde gezeigt, wie hoch Intelligenz mit einigen Variablen wie z.B. Geschlecht, Kultur, Berufsausbildung und schulischen Leistungen korreliert. In Bezug auf den Zusammenhang zwischen Geschlecht und allgemeiner Intelligenz wurde festgestellt, dass Jungen in Mathematik und Naturwissenschaft besser abschneiden als Mädchen. Diese wiederum erzielten bessere Ergebnisse im Bereich Sprache und in geistig-kulturellen

Gebieten. Diese Geschlechtsunterschiede basieren auf biologischen (geschlechtshormonelle Effekte), psychologischen (Motivation, Interesse) und sozialen (Geschlechtsrolle, Geschlechtsstereotypen) Faktoren.

Des Weiteren wurde in diesem Kapitel das allgemeine Problem des Misserfolgs von Migrantenkindern im Schulsystem (besonders hier im deutschen Schulsystem) erläutert und wie man diese ethnischen Effekte erklären kann. Die angewendeten kognitiven Intelligenztests in einigen Untersuchungen und Zeugnisse (Schulnoten) zeigen, dass die Migrantenkinder schlechtere Leistungen als die deutschen Kinder erbringen. Diese ethnischen Unterschiede wurden anhand von zwei Erklärungsmodellen, der Theorie der kulturellen Defizite und der Humankapitaltheorie erläutert. Die Erklärungen durch kulturelle Defizite wurden in defizitäre Herkunfts-Lernkultur und in Unterschichtskultur als defizitäres Sozialisationsumfeld eingeteilt. Die Erklärung durch Humankapital nimmt an, dass die Migrantenkinder einen größeren Mangel an Humankapital erfahren als die deutschen Kinder. Nach dieser Erklärung wird die schlechte Leistung von Migrantenkindern auf den niedrigen Schulabschluss und das geringe Einkommen ihrer Eltern zurückgeführt.

Die Beziehung zwischen Intelligenz und Leistung wurde in zwei Bereichen untersucht, dem der beruflichen Erfolge und dem der schulischen Leistungen. Die amerikanischen und europäischen Studien, die den Zusammenhang zwischen Intelligenz und Berufsausbildung untersuchten, haben festgestellt, dass die allgemeine Fähigkeit den besten Prädikator für die Vorhersage von Berufserfolg darstellt und die spezifischen Fähigkeiten kaum oder keine zusätzliche Varianz erklären.

Im beruflichen Bereich spielt allerdings nach Expertiseforschern wie z. B. Ericsson und Charness (1994) Begabung keine oder nur eine geringe Rolle für herausragende Leistungen von Experten, sondern vielmehr die Erfahrungen, die im Lernprozess erworben wurden. Wie Perleth (1997) ausführt, führt die Untersuchung von Experten auf hohem Niveau jedoch möglicherweise zu einer Unterschätzung der Rolle der Intelligenz beim Erwerb von Expertise (siehe auch Rost, 2009).

Im schulischen Bereich wurde einen starken Zusammenhang zwischen Intelligenz und Schulleistung bestätigt. Deswegen wurde Intelligenz als ein wichtiger und sinnvoller Prädiktor von der schulischen Leistung dargestellt.

5. Empirischer Teil: Untersuchungen zum KFT 4-12+ R

5.1 Fragestellungen und Hypothesen

Im theoretischen Teil wurde aufgezeigt, dass Intelligenz nicht direkt gemessen werden kann, sondern nur mittels besonderer Messinstrumente, nämlich Intelligenztests. Dazu kommt, dass einige Wissenschaftler wie Boring (1923) Intelligenz als „das, was der IQ test mißt“ (zit. nach Holling & Kännig, 1999, S. 24) definiert haben. Das zeigt die Wichtigkeit von Intelligenztests bei der Intelligenzmessung. Gegenstand dieser Studie ist die Qualität eines der vorgestellten Intelligenztests (KFT 4-12 R). Die Bewertung der Qualität von Skalen orientiert sich stets an Itemanalysen (Schwierigkeitsindex und Trennschärfekoeffizient) und den drei primären Testgütekriterien (Objektivität, Reliabilität und Validität). Da Objektivität statistisch nicht berechnet werden kann, werden hier jene zwei Gütekriterien (Reliabilität und Validität) des KFT 4-12+R überprüft, die statistisch ermittelbar sind.

Die Beschäftigung mit Itemanalyse oder Itemkriterien „verfolgt einen zweifachen Zweck:

- a) Revision des Tests im Hinblick auf eine höhere Reliabilität und Validität durch Verbesserung oder Elimination nicht genügend geeigneter Aufgaben.
- b) Überprüfung der Testpunktverteilung hinsichtlich einer höheren Reliabilität und einer besseren Normierbarkeit“ (Lienert & Raatz, 1998, S. 57).

Die Itemanalyse erfolgt durch die Berechnung des Itemschwierigkeitsindex und des Trennschärfekoeffizienten. Der Itemschwierigkeitsindex „ist gleich dem prozentualen Anteil P der auf diese Aufgaben entfallenden richtigen Antworten in einer Analysenstichprobe von der Größe N; der Schwierigkeitsindex liegt also bei schwierigen Aufgaben niedrig, bei leichten hoch“ (Lienert & Raatz, 1998, S. 73). Der Trennschärfekoeffizient wird durch die „Korrelation zwischen dem Aufgabenpunktwert- meist 0 oder 1 – und dem Rohwert eines jeden von N Pbn der Analysen-stichprobe“ definiert (Lienert & Raatz, 1998, S. 78)

Aufgrund der Bedeutung von Itemmerkmalen als Voraussetzung für hohe Reliabilität und Validität und der Möglichkeit, durch Eliminierung oder Verbesserung weniger geeigneter Items Reliabilität und Validität zu verbessern, wurde die erste Untersuchungsfrage formuliert:

Frage 1: Gibt es unterschiedliche Verteilungen der Itemschwierigkeiten und der Itemtrennschärfen innerhalb der Subtests, Dimensionen und der Gesamtleistung des KFT 4-12+R in den Analysestichproben der beiden Klassenstufen 5 und 7?

Das erste Gütekriterium, das nach Prüfung der Objektivität eines Tests analysiert werden muss, ist die Reliabilität oder Zuverlässigkeit, die zeigt, wie genau ein Test ein bestimmtes Merkmal misst. Der Grad der Genauigkeit wird durch den Reliabilitätskoeffizienten geschätzt, der nach unterschiedlichen Methoden (Testhalbierungsmethode, Innere Konsistenz und Retest-Reliabilität) berechnet wird. In der hier vorliegenden Untersuchung kann folgende Fragestellung bearbeitet werden:

Frage 2: Wie hoch sind die Reliabilitätskoeffizienten der einzelnen KFT4-12 R-Untertests, Testteile (KFT-V, KFT-Q, KFT-N) und der Testgesamtleistung (KFT-GL) in der für diese Analysen zusammengestellten Reliabilitätsstichprobe (siehe weiter unten)?

Das nächste zu prüfende Kriterium ist die Validität oder Gültigkeit, die angibt, inwieweit ein Test das zu messende Merkmal tatsächlich misst. Validität wird nicht nur durch verschiedene Koeffizienten (Kriteriumsvalidität durch Korrelation, Konstruktvalidität durch Faktorenanalyse) ermittelt, sondern auch durch logische Analyse (Inhaltsvalidität). In der hier vorgelegten Arbeit wird die Validität aber nur durch berechnete Koeffizienten eingeschätzt. Folgende Fragestellungen wurden dabei verfolgt:

Frage 3: Welche Faktorenstruktur ergibt sich für den KFT 4-12 R (Aspekte der faktoriellen Validität)?

Frage 4: Besteht ein Zusammenhang zwischen den Ergebnissen der Schüler/innen in den KFT 4-12 R Untertests, Dimensionen (V, Q, N), der Gesamtleistung (GL) und den Schulnoten, die als Validierungskriterium für den KFT 4-12 R verwendet werden (Aspekt der Kriteriumsvalidität)?

Das Ziel der hier präsentierten Studie ist nicht nur die Überprüfung der Gütekriterien des KFT 4-12+R, sondern auch die Untersuchung von Einflüssen verschiedener Faktoren auf die allgemeine Intelligenz. Viele Studien haben beispielsweise Geschlechtsunterschiede untersucht, aber die Ergebnisse stimmen nicht durchgehend überein. Darum werden auch in der vorliegenden Untersuchung Geschlechtsunterschiede überprüft und dann anhand bisheriger Ergebnisse diskutiert.

Darüber hinaus wird der Einfluss von kulturellen Milieus anhand der Muttersprache der Schülerinnen und Schüler untersucht. Es werden Schüler, die teilweise aus unterschiedlichen Ländern nach Deutschland gekommen sind und die zu Hause überwiegend eine andere Sprache als Deutsch sprechen, verglichen, um herauszufinden, ob sich für einzelnen Gruppen bzw. Ethnien Unterschiede in den kognitiven Fähigkeiten finden. Daraus ergeben sich insgesamt folgende Fragestellungen:

Frage 5: Unterscheiden sich Mädchen und Jungen in den KFT 4-12+R-Dimensionen (KFT-V, KFT-Q, KFT-N) und der Gesamtleistung (KFT-GL) in der Gesamtstichprobe im Zeitraum von 1998 bis 2010 in Klasse 5 bzw. bis 2011 in Klasse 7?

Frage 6: Unterscheiden sich Schüler, die zu Hause überwiegend Deutsch sprechen (meiste Schüler deutscher Herkunft), von denen, die zu Hause überwiegend eine andere Sprache spechen (meist Migrantenkinder), in den KFT4-12+R-Dimensionen (KFT-V, KFT-Q, KFT-N) und der Gesamtleistung (KFT-GL) in den fünften und siebten Klassen der Kohorten 1998 bis 2010 in Klasse 5 bzw. bis 2011 in Klasse 7?

Die Ludwig-Maximilians-Universität München (bis Frühjahr 1999) bzw. die Universität Rostock (ab Sommer 1999) arbeitet mit der Willy-Brandt-Gesamtschule an einem Projekt über die Nützlichkeit der kognitiven Fähigkeitsmessung, erfasst mit dem KFT 4-12+R (deutsche Version), für die spezifischen Fragestellungen der Schulpsychologen an dieser Schule. Ab 1977 wurden jeweils alle Schüler/innen der fünften, sechsten und siebten Klassen getestet. Die vorliegende Studie analysiert auch Veränderungen (Kohorteneffekte) hinsichtlich der intellektuellen Fähigkeiten, gemessen mit dem KFT 4-12+R, in den fünften und siebten Klassen der letzten Jahre, genauer im Zeitraum von 1998 bis 2010 (Klassenstufe 5) bzw. bis 2011 (Klassenstufe 7). Weiter wurden die einzelnen Schüler zu zwei Zeitpunkten, d.h. sowohl in Klassenstufe 5 wie auch in Klassenstufe 7, untersucht und die Ergebnisse können individuell zugeordnet werden. Das ermöglicht es, eine Analyse der Entwicklung der kognitiven Fähigkeiten von Personen vorzunehmen, was die Ergebnisse aussagekräftiger macht. Deswegen wurden hier die folgenden Untersuchungsfragen gestellt:

Frage 7: Wie verändern sich die Ergebnisse in den KFT4-12+R-Dimensionen (KFT-V, KFT-Q, KFT-N) und in der Gesamtleistung (KFT-GL) in den fünften und siebten Klassen in den letzten Jahren (1998 bis 2011).

Frage 8: Zeigen sich insgesamt oder für einzelne Schülergruppen Entwicklungen in den intellektuellen Fähigkeitsbereichen der Schüler anhand der KFT4-12+R-Dimensionen (KFT-V, KFT-Q, KFT-N) von der Klasse 5 zur Klasse 7?

Aus diesen Fragestellungen wurden die folgenden Hypothesen für die statistische Datenanalyse abgeleitet:

Hypothese 1: Die Itemschwierigkeiten und Itemtrennschärfen der Subtests, Dimensionen und der Gesamtleistung des KFT 4-12+R sowie ihre Verteilungen können in den Analysestichproben der beiden Klassenstufen 5 und 7 als mindestens ausreichend gelten.

Hypothese 2: Die Reliabilitätskoeffizienten der einzelnen KFT4-12 +R Untertests, Testteile (KFT-V, KFT-Q, KFT-N) und der Gesamtleistung (KFT-GL) in der Reliabilitätsstichprobe können nach den üblichen Kriterien als ausreichend hoch gelten.

Hypothese 3: Die faktorielle Struktur der KFT4-12 R Subtests zeigt die drei KFT-typischen Faktoren, nämlich die KFT-Dimensionen (KFT-V, KFT-Q, KFT-N).

Hypothese 4: Die Ergebnisse der Schüler/innen in den KFT 4-12+ R Untertests, Dimensionen (KFT-V, KFT-Q, KFT-N) und der Gesamtleistung korrelieren signifikant und ausreichend hoch mit den Schulnoten.

Hypothese 5: Es gibt einen signifikanten Unterschied in den KFT 4-12+R Dimensionen (KFT-V, KFT-Q, KFT-N) und der Gesamtleistung (KFT-GL) zwischen Jungen und Mädchen in der Gesamtstichprobe bzw. in Jahrgangsstufen.

Hypothese 6: Es gibt einen signifikanten Unterschied in den KFT4-12+R Untertests, Dimensionen (KFT-V, KFT-Q, KFT-N) und der Gesamtleistung (KFT-GL) zwischen den Schülern/innen mit der Muttersprache Deutsch und den Schülern/innen, die zu Hause andere Sprachen als Deutsch sprechen.

Hypothese 7: Die intellektuellen Fähigkeiten in Bezug auf die KFT 4-12 +R Dimensionen (KFT-V, KFT-Q, KFT-N) und die Gesamtleistung (KFT-GL) der Schüler/innen der beiden Klassenstufen 5 und 7 unterscheiden sich in den einzelnen Jahren 1998 bis zum Jahr 2010 signifikant (Kohorteneffekte).

Hypothese 8: Die kognitiven Fähigkeiten hinsichtlich der KFT4-12+ R Dimensionen (KFT-V, KFT-Q, KFT-N) der Schüler/innen verbessern sich von Klasse 5 zu Klasse 7 (Entwicklungs-effekte).

5.2 Methode

5.2.1 Beschreibung der Stichprobe

Insgesamt wurden in den Jahren 1998 bis 2011 4320 Datensätzen von Schülern/innen der Willy-Brandt-Gesamtschule in München erhoben. Die Tests wurden von Schulpsychologen/innen, Sozialpädagogen/innen und bestimmten, trainierten Klassenlehrern/ innen durchgeführt.

Zur Schule und zum Projekt informierten erstmals Perleth, Schauer, Hofmann, Volk und Wernberger (1989, S. 261-265): Die Willy-Brandt-Gesamtschule und die Universität München (seit 1999: Universität Rostock) kooperierten zusammen an einem Projekt angewandter Forschung über die Nützlichkeit des kognitiven Fähigkeits-Test KFT (ab 1997 wurde die aktuelle Version KFT 4-12+R verwendet; Heller & Perleth, 2000), der deutschen Version des Cognitive Abilities Tests (CAT, Thorndike & Hagen, 1971), für die spezifischen Anforderungen der Arbeit der Schulpsychologen an dieser besonderen Schule. Seit 1977 wurde der KFT 4-13 (R) jedes Jahr in den fünften, sechsten und siebten Klassen der Willy-Brandt-Gesamtschule durchgeführt. Die Ergebnisse der Durchführung wurden und werden verwendet, um die Arbeit der Schulberater beispielsweise bei Empfehlungen des Kursniveaus oder der Wahl von Fremdsprachen zu optimieren.

In Deutschland gibt es kein einheitliches Schulsystem, sondern jedes Bundesland hat ein eigenes Bildungsministerium und eigene Schulgesetze. In Bayern und damit in München müssen alle Schüler zuerst für vier Jahre die Grundschule besuchen. Das normale bayrische Schulsystem auf der sekundären Stufe besteht aus drei Schularten mit unterschiedlichen Leistungsniveaus, der Hauptschule, Realschule und dem Gymnasium. Die Hauptschule ist das unterste Leistungsniveau. An der Realschule können die Schüler/innen einen mittleren Schulabschluss erhalten. Das Gymnasium ist die akademischste Schule und führt zum Abitur, das ist ein Schulabschluss, der für das Studium an einer Universität nötig ist. Hinsichtlich des bayerischen Schulsystems ist zu erwähnen, dass die Entscheidung über die Schullaufbahn und mögliche spätere Berufe sehr früh getroffen wird. Außerdem ist der Wechsel zwischen den Schularten zwar nicht unmöglich aber doch sehr schwierig. Darüber hinaus besuchten im Untersuchungszeitraum die Schüler/innen die Schulen überwiegend nur am Vormittag, es gibt keinen Unterricht am Nachmittag. Klassenwiederholungen kommen relativ häufig vor. Außerdem verlässt eine große Zahl der Schüler die Schulen ohne qualifiziertem Abschluss.

Deswegen wurden in den 70er Jahren in verschiedenen deutschen Bundesländern Gesamtschulen konzipiert, damit alle Schüler eine Ausbildung entsprechend ihrer Fähigkeiten bekommen, die soziale Integration gefördert wird, alle Schüler eine individuelle Beratung erhalten und um die individuellen Vorlieben und Interessen zu berücksichtigen. In Gesamtschulen findet keine Aufteilung der Schüler/innen in Schulen unterschiedlicher Anforderungen statt, sondern die Schüler werden in einzelnen Fächern in unterschiedliche Kursniveaus eingeteilt. Diese Schulen wurden in Deutschland in den 1970er Jahren eingeführt, damit auch Schüler/innen aus unteren sozioökonomischen Klassen, die ansonsten nur reduzierte Chancen auf eine gute Ausbildung hatten, eine solche erhalten konnten.

Die Willy-Brandt-Gesamtschule München wurde im September 1970 mit 200 Schülern in Klasse 5 eröffnet. Seit dieser Zeit werden 150 bis 200 Schüler jährlich aufgenommen. Das Hauptproblem dieser Schule ist, dass sie heute die einzige Gesamtschule in Bayern darstellt. Deswegen müssen die ausgestellten Zertifikate von dieser Schule mit den Zertifikaten der übrigen Schulen vergleichbar sein. Außerdem müssen die Schüler, welche die Gesamtschule verlassen, eine bestimmte Zahl von Kursen auf höherem Niveau besuchen, um Zertifikate (Zeugnisse) über eine höhere Qualifizierung zu erhalten, so dass sie etwa die Oberstufe eines Gymnasiums besuchen können.

In der Willy-Brandt-Gesamtschule gibt es Beratungslehrer und Psychologen, die zusammenarbeiten und Beratung und Förderprogramme im Bereich Lern von - und Leistungsproblemen, emotionalen Störungen, bei Konflikte zwischen den Lehrern und den Schülern usw. anbieten. Darüber hinaus sind sie für die jährliche Anwendung des Intelligenztests in den fünften, sechsten (nur bis zum Schuljahr 2003/2004) und siebten Klassen verantwortlich.

Die Willy-Brandt-Gesamtschule München umfasst die drei Schulformen Hauptschule, Realschule und Gymnasium. Nach der Grundschule können die Schüler die Willy-Brandt-Gesamtschule bis zur 10. Klasse besuchen. Ab der 6. Klasse besuchen die Schüler je nach individueller Leistung Kurse in den Fächern Englisch und Mathematik. Ab der 7. Klasse besuchen die Schüler verschiedene auf ihre Leistung abgestimmte Kurse in Englisch, Deutsch und Mathematik. Ab der 8. Klasse werden die Schüler bezüglich der leistungsbezogenen Wahl der Schullaufbahn in Kerngruppen zusammengefasst. Dann werden die Entscheidungen getroffen, ob die Schüler entweder in Kursen unterrichtet werden, die insgesamt dem Haupt- oder Realschulniveau oder dem Niveau eines Gymnasiums entsprechen.

Die Willy-Brandt-Gesamtschule beteiligt sich am UNESCO-Schulprojekt und ist seit 2009 „Anerkannte UNESCO-Projektschule“. In der Schule unterrichten und arbeiten ungefähr 85 Lehrer und 15 Sozialpädagogen. Fast 1000 Schüler besuchen die Willy-Brandt-Gesamtschule. Den Schülern werden zusätzlich zum normalen Unterricht auch besondere Kurse oder Schul-

angebote für Sport, Kunst usw. geboten. Es gibt auch einen Wikipedia-Eintrag zu der Schule:
http://de.wikipedia.org/wiki/Willy-Brandt-Gesamtschule_M%C3%BCnchen

An den hier berichteten Untersuchungen nahmen nicht alle Schüler aus allen Klassen teil, sondern nur die Schüler der Klassen 5 (N5= 2091) und 7 (N7= 2229), im Alter von etwa 10 und 12 Jahren. Die folgende Tabelle 5.1.1 gibt den Umfang der Stichprobe pro Jahrgang in den Klassenstufen 5 und 7 an:

Tabelle 5.2.1.1: Verteilung der Stichprobe pro Jahrgang vom Schuljahr 1998/1999 bis 2010/2011 bzw. 2011/2012 in den beiden Klassenstufen 5 und 7:

Jahrgang	Klasse 5			Klasse 7		
	Gesamt	Jungen	Mädchen	Gesamt	Jungen	Mädchen
1998/1999	169	92 (54,4%)	77 (45,6%)	162	75 (46,3%)	87 (53,7%)
1999/2000	166	81 (48,8%)	85 (51,2%)	168	89 (53%)	79 (47%)
2000/2001	165	81 (49,1%)	84 (50,9%)	160	84 (52,5%)	76 (47,5%)
2001/2002	163	84 (51,5%)	79 (48,5%)	165	85 (51,5%)	80 (48,5%)
2002/2003	167	86 (51,5%)	81 (48,5%)	166	83 (50%)	83 (50%)
2003/2004	161	86 (53,4%)	75 (46,6%)	161	82 (50,9%)	79 (49,1%)
2004/2005	159	90 (56,6%)	69 (43,4%)	168	86 (51,2%)	82 (48,8%)
2005/2006	154	92 (59,7%)	62 (40,3%)	150	76 (50,7%)	74 (49,3%)
2006/2007	159	88 (55,3%)	71 (44,7%)	155	81 (52,3%)	74 (47,7%)
2007/2008	153	94 (61,4%)	59 (38,6%)	160	93 (58,1%)	67 (41,9%)
2008/2009	160	79 (49,4%)	81 (50,6%)	159	83 (52,2%)	76 (47,8%)
2009/2010	158	81 (51,3%)	77 (48,7%)	154	93 (60,4%)	61 (39,6%)
2010/2011	157	79 (50,3%)	78 (49,7%)	143	75 (52,4%)	68 (47,6%)
2011/2012				158	80 (50,6%)	78 (49,4%)
Summe	2091	1113 (53,23%)	978 (46,77%)	2229	1165 (52,27%)	1064 (47,73%)

Diese Tabelle zeigt die Gesamtzahl der Schüler/innen, die sich an dieser Untersuchung bzw. den verschiedenen Testerehebungen beteiligt haben. Aufgrund fehlender Daten konnten aber nicht die Ergebnisse aller Schüler/innen in die statistische Auswertung miteinbezogen werden. In der vorliegenden Untersuchung werden daher die folgenden Sub-Stichproben unterschieden:

Reliabilitätsstichprobe. Um die Reliabilität des KFT 4-12 +R Tests zu ermitteln, wurde eine Stichprobe von den Klassenstufen 5 (N5 = 2090) und 7 (N7 = 2225) einbezogen. Diese verteilt

sich wie folgt auf die zwei Testformen: N5A=955 (Klassenstufe 5, Testform A), N5B = 1135 (Klassenstufe 5, Testform B), N7A = 1112 (Klassenstufe 7, Testform A) und N7B = 1113 (Klassenstufe 7, Testform B).

Validitätsstichprobe. Es wird in dieser Untersuchung die faktorielle Validität des KFT 4-12+ R an einer Stichprobe der Klassenstufen 5 (N5 = 2090) und 7 (N7 = 2225), die Kriteriums- bzw. Übereinstimmungsvalidität an einer Stichprobe der Klassenstufe 7 (N7 = 154) und die prognostische Validität ebenfalls an einer Stichprobe von Klassenstufe 7 (N7 = 133) bestimmt.

Tabelle 5.2.1.2: Verteilung der Teil-Stichproben der einzelnen Sprachgruppen in den Klassenstufen 5 und 7

	Klasse 5			Klasse 7			von Klasse 5 zur Klasse 7		
	Jungen	Mädchen	Gesamt	Jungen	Mädchen	Gesamt	Jungen	Mädchen	Gesamt
Deutsch	608 (29,1%)	506 (24,2%)	1114 (53,4%)	692 (31,3%)	556 (25,1%)	1248 (56,4%)	426 (32,3%)	321 (24,4%)	747 (56,7%)
Türkisch	222 (10,6 %)	216 (10,3%)	438 (20,9%)	205 (9,3%)	249 (11,3%)	454 (20,5%)	143 (10,9%)	155 (11,8%)	298 (22,6%)
Osteuropa	137 (6,6%)	123 (5,9%)	260 (12,5%)	131 (5,9%)	131 (5,9%)	262 (11,8%)	72 (5,5%)	75 (5,7%)	147 (11,2%)
Südeuropa	38 (1,8%)	44 (2,1%)	82 (3,9%)	36 (1,6%)	38 (1,7%)	74 (3,3%)	14 (1,1%)	18 (1,4%)	32 (2,4%)
Westeuropa/ USA	12 (0,6%)	16 (0,8%)	28 (1,3%)	7 (0,3%)	11 (0,5%)	18 (0,8%)	3 (0,2%)	4 (0,3%)	7 (0,5%)
Asien	50 (2,4%)	37 (1,8%)	87 (4,2%)	45 (2%)	39 (1,8%)	84 (3,8%)	27 (2,1%)	19 (1,4%)	46 (3,5%)
Afrika/ Arabien	43 (2,1%)	36 (1,7%)	79 (3,8%)	41 (1,9%)	32 (1,4%)	73 (3,3%)	22 (1,7%)	18 (1,4%)	40 (3%)
Summe	1110 (53,2%)	978 (46,8%)	2088 (100%)	1157 (52,3%)	1056 (47,7%)	2213 (100%)	707 (53,7%)	610 (46,3%)	1317 (100%)

Legende: Prozentwerte auf den Stichprobenumfang der Klassenstufen 5 und 7 sowie 5+7 bezogen.

Sprachstichproben. Zur Berechnung von Unterschieden in den Leistungen zwischen Kindern mit deutscher und solchen Kindern, deren Muttersprache kein Deutsch ist oder die mit ihren Eltern zu Hause kaum oder kein Deutsch sprechen, wurden neben „Deutsch“ folgende sechs Teil-Stichproben von Sprachgruppen von Schülern/innen der Klassenstufen 5 und 7 gebildet: Türkisch, Osteuropa (die meisten Schüler dieser Gruppe sprechen als Muttersprache oder Verkehrssprache zu Hause Bosnisch, Serbisch, Ukrainisch, Russisch oder Polnisch), Südeuropa (die meisten Schüler/innen sprechen zu Hause überwiegend Spanisch, Italienisch und Griechisch), Westeuropa/USA (die kleinste Gruppe und umfasst Schüler/innen, die zu Hause bzw. als Muttersprache Französisch oder Englisch sprechen), Asien (die meisten Schüler sprechen Afgha-

nisch, Pakistanisch, Vietnamesisch und Persisch), und Afrika/Araber (die meisten sprechen als Muttersprache Arabisch, wobei die häufigsten Dialekte Tunesisch, Ägyptisch und Libanesisch sind). Für eine Gesamtübersicht hinsichtlich der Verteilung der Gesamtstichprobe auf die Teilstichproben bzgl. der Muttersprache bzw. der zu Hause gesprochenen Sprache siehe Tabelle 5.2.1.2.

5.2.2 Darstellung des KFT 4-12+ R

Der kognitive Fähigkeits-Test für 4. bis 12. Klassen (KFT 4-12+ R; vgl. Heller & Perleth, 2000) ist ein differentieller Intelligenztest zur Erfassung der intellektuellen Fähigkeiten von Schülern der 4. bis 12./13. Klassen. Er ist die übersetzte und auf deutsche Verhältnisse übertragene Form der Cognitive Abilities Tests von Thorndike und Hagen. Der KFT ist als Gruppen- und Einzeltest durchführbar und als Power-Speed-Test zu betrachten. Er ist in zwei Formen A und B vorhanden und setzt sich aus einem verbalen (KFT-V), einem quantitativen (KFT-Q) und einem

Tabelle 5.2.2.1: Testaufbau und Zeitbedarf des KFT 4-12+ R

	Aufgabenanzahl (Insgesamt)	Zeitbedarf in Min:Sek		
		Instruktion	Bearbeitungszeit	Summe
Einführung		4:00	4:00
Verbaler Teil				
V1 Wortschatz	25	3:00	7:00	10:00
V2 Wortklassifikationen	25	1:30	9:00	10:30
V3 Wortanalogien	20	2:30	7:00	9:30
V-Teil Gesamt (KFT-V)	70	7:00	23:00	30:00
Quantitativer Teil				
Q1 Mengenvergleiche	25	3:30	10:00	13:30
Q2 Zahlenreihen	20	2:00	9:00	11:00
Q3 Gleichungenbilden	15	5:30	13:00	18:30
Q-Teil Gesamt (KFT-Q)	60	11:00	32:00	43:00
Nonverbaler Teil				
N1 Figurenklassifikation	25	3:00	9:00	12:00
N2 Figurenanalogien	25	2:30	8:00	10:30
N3 Faltaufgaben	15	3:00	8:00	11:00
N-Teil Gesamt (KFT-N)	65	8:30	25:00	33:30
KFT Gesamt (KFT-GL)	195	30:30	80:00	130:00 bzw. 3 Schulstunden

nonverbalen Teil (KFT-N) zusammen. Jeder Teil besteht aus drei Subtests mit 12 bis 25 Items pro Klassenstufe. Es gibt zeitliche Beschränkung für die Bearbeitung jedes Untertests wie in der folgenden Tabelle 5.2.2.1 angeführt:

Verbaler Teil (V-Teil). Dieser Teil besteht aus den folgenden Subtests:

Wortschatz (V-Test 1 bzw. V1). Bei diesem Test werden die Teilnehmer gebeten, ein Wort aus fünf Wörtern auszuwählen, das die gleiche oder ähnliche Bedeutung (Synonym oder Oberbegriff) hat wie ein vorgegebenes Wort. Danach muss auf dem Antwortbogen die richtige Antwort bei der Nummer der Aufgabe in der Spalte zum **V-Test 1** markiert werden. Die Aufgaben des Wortschatztests werden durch das folgende Beispiel B1 verdeutlicht:

B1 Rose **A** Musik **B** Blume **C** Speise **D** Aussicht **E** Lasso

Wortklassifikationen (V-Test 2 bzw. V2). Bei diesem Test werden drei Wörter vorgegeben. Sie haben einen gemeinsamen Oberbegriff. Die Testteilnehmer/innen müssen ein Wort aus fünf weiteren vorgegebenen Wörtern auswählen, das zu dem gleichen Oberbegriff gehört. Dann muss auf dem Antwortbogen die richtige Antwort bei der Nummer der Aufgabe in der Spalte **V-Test 2** markiert werden. Die Aufgaben dieses Tests werden durch das folgende Beispiel B1 verdeutlicht:

B1 Maus Wolf Bär

A Rose **B** Löwe **C** laufen **D** hungrig **E** braun

Wortanalogien (V-Test 3 zw. V3). Bei diesem Test werden ein Paar von Wörtern, die miteinander in einem Zusammenhang stehen, und ein drittes Wort vorgegeben. Die Teilnehmer müssen ein Wort aus den fünf folgenden Wörtern auswählen, das mit dem dritten Wort in derselben Beziehung steht wie die ersten beiden Wörter zueinander. Dann muss auf dem Antwortbogen die richtige Antwort bei der Nummer der Aufgabe in der Spalte **V-Test 3** markiert werden. Die Aufgaben dieses Tests werden durch das folgende Beispiel verdeutlicht:

B1 groß : riesig → klein :

A Junge **B** erwachsen **C** winzig **D** wenig **E** mehr

Quantitativer Teil (Q-Teil). Dieser Teil umfasst drei Subtests:

Mengenvergleiche (Q-Test 1 bzw. Q1). Bei diesem Test werden zwei Felder vorgegeben, in denen Mengen, Zahlen, Flächen, Winkel usw. enthalten sind. Hier muss die Menge im Feld 1

mit der Menge im Feld 2 verglichen werden. Danach muss auf dem Antwortbogen die richtige Antwort in der Spalte Q-Test 1 markiert werden: A (wenn die Menge im **linken Feld 1 größer** als die in Feld 2 ist), B (wenn die beiden Mengen **gleich groß** sind) oder C (wenn die Menge im **rechten Feld 2 größer** ist). Die Aufgaben dieses Tests werden durch das folgende Beispiel verdeutlicht:

	Feld 1	Feld 2
B1		
	A (Wert im linken Feld größer) B (Werte gleich groß) C (Wert im rechten Feld größer)	

Zahlenreihen (Q-Test 2 bzw. Q2). Bei diesem Test wird eine Reihe von fünf Zahlen und Buchstaben dargestellt, die nach einer Regel angeordnet sind. Zunächst sollen die Teilnehmer die Regel herausfinden. Dann soll die entsprechende Zahl aus den Antwortalternativen rechts ausgewählt werden. Die richtige Antwort (der richtige Buchstabe) muss auf dem Antwortbogen bei der Nummer der Aufgabe in der Spalte **Q-Test-2** markiert werden. Die Aufgaben dieses Tests werden durch das folgende Beispiel verdeutlicht:

B1	10 9 8 7 6	→	A 3 B 4 C 5 D 6 E 7
----	------------------------	---	---------------------------------

Gleichungenbilden (Q-Test 3 bzw. Q3). Dieser Test besteht aus Zahlen und Operationszeichen, mit denen die Zahlen verbunden werden sollen. Bei diesem Test sollen alle Zahlen und Operationszeichen nur einmal verwendet werden. Die richtige Antwort muss aus fünf Auswahlantworten ausgewählt und auf dem Antwortbogen bei der Nummer der Aufgabe in der Spalte **Q-Test 3** markiert werden. Die Aufgaben dieses Tests werden durch das folgende Beispiel verdeutlicht:

B1	2 3	-	→	A 1 B 2 C 3 D 4 E 5
----	--------	---	---	---------------------------------

Nonverbaler Teil (N-Teil). Dieser Teil setzt sich aus den folgenden Subtests zusammen:

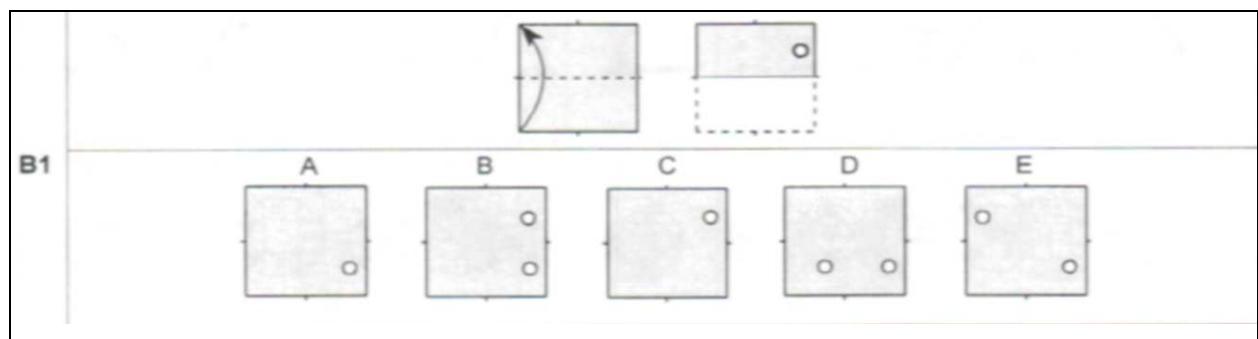
Figurenklassifikation (N-Test 1 bzw. N1). Bei diesem Test wird eine Reihe von Figuren dargestellt, die nach Klassen (Form, Lage, usw.) geordnet sind. Die Teilnehmer sollen eine der fünf folgenden Figuren auswählen, die zu der Klasse der vorgegebenen Figuren gehört. Dann muss die richtige Antwort auf dem Antwortbogen bei der Nummer der Aufgabe in der Spalte **N-Test 1** markiert werden. Die Aufgaben dieses Tests werden durch das folgende Beispiel verdeutlicht:

B1		A B C D E
----	--	---

Figurenanalogien N-Test 2 (N-Test 2 bzw. N2). Bei diesem Test werden ein Paar von Figuren, die in einer Beziehung zueinander stehen, und noch eine dritte Figur vorgegeben. Hier muss eine Figur aus den Antwortalternativen ausgewählt werden, die mit der dritten Figur in einer Beziehung steht wie die ersten beiden Figuren zueinander. Danach muss die richtige Antwort auf dem Antwortbogen bei der Nummer der Aufgabe in der Spalte **N-Test 2** markiert werden. Die Aufgaben dieses Tests werden durch das folgende Beispiel verdeutlicht:

B1		:	A B C D E
----	--	---	---

Faltaufgaben N-Test 3 (N-Test 3 bzw. N3). Bei diesem Test wird ein graues Papier gefaltet. Dann werden Löcher in das gefaltete Papier gestanzt. Anschließend muss entschieden werden wie das Papier nach dem Auffalten aussieht. Danach muss die richtige Antwort auf dem Antwortbogen bei der Nummer der Aufgaben in der Spalte **N-Test 3** markiert werden. Die Aufgaben dieses Tests werden durch das folgende Beispiel verdeutlicht:



5.2.3 Zensuren

In der vorliegenden Studie konnten die Zensuren einer Teilstichprobe der Schüler/innen aus Klassestufe 7 ($N= 154$) erhoben werden. Die Zensuren wurden als Kriterium herangezogen, um die **Kriteriums-/Übereinstimmungsvalidität** des KFT 4-12+R zu berechnen, bei der die Schulnoten mit den Ergebnissen der Skalen des KFT 4-12+R (erhoben zum selben Zeitpunkt) korreliert wurden. Zur Berechnung der **prognostischen Validität** wurden die Ergebnisse der Skalen des KFT 4-12+R in Klasse 5 herangezogen. Diese wurden mit den Schulzensuren in

Klasse 7 in Beziehung gesetzt und die Korrelation zwischen den beiden Ergebnissen berechnet. Die Zensuren folgender Fächer wurden dabei im Zuge der Analysen herangezogen: Deutsch, Englisch und Mathematik. Diese Zensuren stammen aus unterschiedlichen Kursniveaus. Für jedes Fach gibt es drei Niveaus: 1 (höchstes Niveau), 2 (mittleres Niveau) und 3 (niedrigstes Niveau). Darüber hinaus liegt für jede/n Schüler/in der „Zensurenstichprobe“ eine Schullaufbahnempfehlung aus der 7. Klasse vor. Diese Information konnten den Unterlagen der Schule entnommen werden.

5.2.4 Statistische Verfahren

Nach der Durchführung der KFT 4-12+R Tests wurden die Daten mit einem Scanner eingelesen und mit dem Programm SPSS analysiert. Außerdem wurden die eingelesenen Daten durch studentische Hilfskräfte kontrolliert. Die folgenden statistischen Methoden wurden verwendet, um die Fragestellungen dieser Untersuchung zu beantworten (vgl. zu den Verfahren Bühl, 2010):

- **Fragestellung 1.** Die Itemanalysen wurden anhand des „*Schwierigkeitsindex*“ und des „*Trennschärfeindex*“ durchgeführt.
- **Fragestellung 2.** Die Reliabilität des KFT 4-12+R (Split-Half-Reliabilität, Innere Konsistenz und Retest-Reliabilität) wurde durch die „*Korrelationsmethode*“ sowie einschlägige Koeffizienten, die auf den Interkorrelationen der Items und Skalen basieren, bestimmt (SPSS-Prozeduren „RELIABILITY“).
- **Fragestellung 3 und 4.** Die faktorielle Validität des KFT 4-12+R wurde mittels „*Faktoren-analysen*“ analysiert und die Kriteriumsvalidität des KFT 4-12+R mittels „*Korrelations-methoden*“ bestimmt (Übereinstimmungsvalidität und prognostische Validität; SPSS-Prozedur „FACTOR“).
- **Fragestellung 5.** Die Berechnungen von Geschlechtsunterschieden im Zeitraum von 1998 bis 2010 in Klasse 5 und bis 2011 in Klasse 7 erfolgten mittels *Varianzanalyse* (SPSS-Prozedur „ANOVA“).
- **Fragestellung 6.** Die ethnischen Unterschiede zwischen deutschen Schülern/innen und aus-ländischen Schülern/innen wurden mittels *Varianzanalyse* berechnet (SPSS-Prozedur „ANOVA“).
- **Fragestellung 7.** Die Berechnungen der Veränderungen der kognitiven Fähigkeiten in den Klassenstufen 5 und 7 wurden durch *Varianzanalysen* ermittelt (SPSS-Prozeduren „ANOVA“, „ONEWAY“ mit den Subroutinen „POST-HOC SCHEFFE-TEST“).

- **Fragestellung 8.** Die Veränderungen der intellektuellen Fähigkeiten von Klasse 5 zu Klasse 7 wurden zuerst durch die Berechnung der Differenzen zwischen den Ergebnissen der KFT 4- 12+R Dimensionen und der Gesamtleistung zum ersten (Klasse 5) und zweiten Messzeitpunkt (Klassen 7) bestimmt, dann wurde die einfache **Varianzanalyse für Differenzwerte** (SPSS-Prozedur „ONEWAY“) anstatt einer „**Varianzanalyse mit Messwiederholung**“ durchgeführt.

5.3 Ergebnisse zur Testkonstruktion des KFT 4-12+ R

5.3.1 Itemanalysen

Es wird hier die Frage beantwortet, ob der KFT 4-12+R über eine ausreichende Differenzierungsfähigkeit verfügt, indem die Itemschwierigkeiten (wie schwer oder leicht ein Item von allen Versuchspersonen gelöst werden konnte) sowie die Itemtrennschärfen (wie gut ein Item inhaltlich die Skala widerspiegelt) herangezogen wurden. Items die entweder von allen Probanden oder von keinem Probanden gelöst werden, sowie jene Items, deren Korrelationskoeffizienten mit der Skala gering ausfallen sind hinsichtlich ihrer Differenzierungsfähigkeit ungeeignet, d.h. Items mit Schwierigkeitsindex $p < .20$ oder $p > .80$ und mit Trennschärfeindex $r_{it} < .30$ sollten entfernt werden. Um diese Frage zu beantworten wurden insgesamt 4315 Datensätzen der Klassenstufen 5 und 7 herangezogen, welche sich wie folgt auf die beiden Testformen verteilen: N5A = 955, N5B = 1135, N7A = 1112 und N7B = 1113.

Die Ergebnisse zeigen, dass die Schwierigkeiten einzelner Items der Skalen des KFT-4-12+R in den beiden Klassenstufen und Testformen (siehe Tabellen 5.3.1.3 bis 5.3.1.28) nicht ausreichend hoch sind. Im Allgemeinen weisen die Items der Skalen des KFT 4-12+R in den beiden Testformen und beiden Klassen 5 und 7 jedoch mittlere Schwierigkeiten auf (Tabelle 5.3.1.1). Die gilt nicht für die Trennschärfe, die in den Untertests KFT-Q1, KFT-Q2, Testteilen KFT-V, KFT-Q, und der KFT-Gesamtleistung für beide Testformen in den Klassenstufen 5 und 7 niedrig ist. Darüber hinaus fallen die Trennschärfen der Subtests KFT-V1, KFT-V3 in Klasse 5 Testform B sowie KFT-V2 in Klasse 7 in den beiden Testformen gering aus (siehe Tabellen 5.3.1.2 bis 5.3.1.28).

Abbildung 5.3.1.1 zeigt, dass sich für die verbalen Untertests KFT-V1, KFT-V3, und die Dimension KFT-V, die quantitativen Subtests KFT-Q1, KFT-Q3, und die Dimension KFT-Q in Klasse 5 für beide Testformen sowie KFT-N1, KFT-N, KFT-V2 Form B und die KFT-GL für

beide Formen der Klasse 5 mittlere Schwierigkeitsgrade bzw. normale Schwierigkeitsverteilungen ergeben. Das gilt für den V-Teil von Klasse 7 in den beiden Testformen, die quantitativen Untertests (KFT-Q1, KFT-Q2 Form B, und KFT-Q3 für beide Formen) und die Dimension KFT-Q sowie die KFT-Gesamtleistung für beide Formen, aber das gilt nicht für den N-Teil (Abbildung 5.3.1.2). Dazu kommt, dass sich schiefe Verteilungen in den Untertests und Testteilen in beiden Klassen 5 und 7 auf die Differenzierungsfähigkeit des KFT 4-12+R auswirken. Darüber hinaus fällt in den Abbildungen 5.3.1.1 und 5.3.1.2 auf, dass einige Skalen des KFT 4-12+R für beide Testformen in Klassenstufen 5 (KFT-V2, KFT-Q2 und KFT-N1) sowie 7 (Subtests des KFT-N-Teils und KFT-Q2) sehr ungleiche Schwierigkeitsgrade haben.

Mittelwerte, Standardabweichungen, Schiefe, Exzess und andere Merkmale der Verteilungen der KFT-Skalen in den beiden Testformen für die beiden Klassenstufen 5 und 7 sind in den Tabellen 5.3.1.29 bis 5.3.1.32 zusammengestellt. Diese Tabellen geben an, dass die Schüler/innen in beiden Klassenstufen (5 und 7) und in den beiden Testformen bessere Leistung im nonverbalen Teil als im verbalen und quantitativen Teil erbracht haben.

Tabelle 5.3.1.1: Durchschnittliche Itemschwierigkeiten der Skalen des KFT 4-12+R für beide Testformen und die Klassenstufen 5 und 7

	Klasse 5 (p)		Klasse 7 (p)	
	Form A	Form B	Form A	Form B
KFT-V1	,51	,50	,50	,52
KFT-V2	,64	,51	,64	,57
KFT-V3	,46	,45	,49	,56
KFT-Q1	,45	,46	,42	,50
KFT-Q2	,79	,55	,79	,63
KFT-Q3	,47	,43	,46	,41
KFT-N1	,77	,47	,79	,62
KFT-N2	,61	,63	,67	,73
KFT-N3	,68	,65	,66	,71
KFT-V	,55	,49	,55	,55
KFT-Q	,57	,49	,55	,52
KFT-N	,69	,58	,72	,68
KFT-GL	,60	,52	,61	,58

Legende: V1: Wortschatz, V2: Wortklassifikation, V3: Wortanalogien, Q1: Mengenvergleich, Q2: Zahlenreihen, Q3: Gleichungenbilden, N1: Figurenklassifikation, N2: Figuren-analogien, N3: Faltaufgaben.

Tabelle 5.3.1.2: Durchschnittliche Itemtrennschärfen der Skalen des KFT 4-12+ für beide Testformen und die Klassenstufen 5 und 7

	Klasse 5 (r)		Klasse 7 (r)	
	Form A	Form B	Form A	Form B
KFT-V1	,30	,27	,27	,24
KFT-V2	,34	,32	,22	,18
KFT-V3	,30	,29	,27	,26
KFT-Q1	,10	,17	,18	,18
KFT-Q2	,39	,37	,33	,34
KFT-Q3	,25	,24	,28	,22
KFT-N1	,40	,38	,47	,41
KFT-N2	,54	,53	,42	,45
KFT-N3	,37	,39	,41	,37
KFT-V	,28	,27	,26	,23
KFT-Q	,20	,24	,23	,23
KFT-N	,38	,38	,40	,37
KFT-GL	,24	,24	,25	,24

Tabelle 5.3.1.3: Schwierigkeits- und Trennschärfeindizes für den Untertest V1 des KFT 4-12+R in Klasse 5 für beide Testformen A und B

Items	Testform A			Testform B		
	Schwierigkeitsindex (p)	Trennschärfe-index (r)	N	Schwierigkeitsindex (p)	Trennschärfe-index(r)	N
v1_1	,90	,165	954	,71	,334	1134
v1_2	,70	,330	954	,58	,336	1134
v1_3	,75	,421	954	,77	,183	1134
v1_4	,60	,335	954	,68	,113	1134
v1_5	,93	,240	954	,75	,125	1134
v1_6	,75	,125	954	,71	,194	1134
v1_7	,68	,448	954	,49	,118	1134
v1_8	,39	,336	954	,47	,232	1134
v1_9	,74	-,024	954	,66	,342	1134
v1_10	,55	,295	954	,68	,392	1134
v1_11	,71	,444	954	,71	,339	1134
v1_12	,69	,388	954	,60	,284	1134
v1_13	,45	,134	954	,44	,225	1134
v1_14	,39	,284	954	,54	,403	1134
v1_15	,06	,125	954	,49	,175	1134
v1_16	,51	,514	954	,30	,330	1134
v1_17	,50	,382	954	,62	,382	1134
v1_18	,50	,355	954	,55	,346	1134
v1_19	,41	,358	954	,24	,233	1134
v1_20	,62	,397	954	,10	,251	1134
v1_21	,29	,317	954	,36	,376	1134
v1_22	,25	,393	954	,50	,324	1134
v1_23	,21	,367	954	,17	,201	1134
v1_24	,13	,232	954	,15	,263	1134
v1_25	,14	,092	954	,36	,365	1134

Tabelle 5.3.1.4: Schwierigkeits- und Trennschärfeindizes für den Untertest V2 des KFT 4-12+R in Klasse 5 für beide Testformen A und B

Items	Testform A			Testform B		
	Schwierigkeitsindex (p)	Trennschärfe-index (r)	N	Schwierigkeitsindex (p)	Trennschärfe-index (r)	N
v2_1	,89	,273	955	,86	,305	1134
v2_2	,64	,464	955	,45	,506	1134
v2_3	,73	,279	955	,54	,385	1134
v2_4	,70	,261	955	,59	,472	1134
v2_5	,77	,575	955	,74	,228	1134
v2_6	,65	,373	955	,72	,447	1134
v2_7	,77	,536	955	,54	,332	1134
v2_8	,46	,088	955	,62	,465	1134
v2_9	,71	,515	955	,39	,312	1134
v2_10	,59	,328	955	,46	,374	1134
v2_11	,72	,448	955	,61	,266	1134
v2_12	,50	,510	955	,55	,189	1134
v2_13	,85	,243	955	,55	,496	1134
v2_14	,64	,300	955	,55	,366	1134
v2_15	,57	,156	955	,54	,390	1134
v2_16	,50	,141	955	,59	,358	1134
v2_17	,79	,421	955	,35	-,040	1134
v2_18	,58	,420	955	,28	,130	1134
v2_19	,40	,277	955	,29	,097	1134
v2_20	,28	,184	955	,33	,321	1134
v2_21	,68	,497	955	,20	,177	1134
v2_22	,81	,318	955	,53	,243	1134
v2_23	,63	,397	955	,28	,296	1134
v2_24	,54	,227	955	,59	,457	1134
v2_25	,72	,302	955	,58	,335	1134

Tabelle 5.3.1.5: Schwierigkeits- und Trennschärfeindizes für den Untertest V3 des KFT 4-12+R in Klasse 5 für beide Testformen A und B

Items	Testform A			Testform B		
	Schwierigkeitsindex (p)	Trennschärfe-index (r)	N	Schwierigkeitsindex (p)	Trennschärfe-index (r)	N
v3_1	,79	,246	954	,75	,240	1134
v3_2	,73	,401	954	,24	,305	1134
v3_3	,61	,580	954	,38	,340	1134
v3_4	,51	,390	954	,59	,475	1134
v3_5	,59	,358	954	,57	,481	1134
v3_6	,53	,447	954	,35	,387	1134
v3_7	,31	,225	954	,68	,195	1134
v3_8	,64	,416	954	,52	,520	1134
v3_9	,35	,328	954	,36	,104	1134
v3_10	,70	,203	954	,73	,264	1134

Items	Testform A			Testform B		
	Schwierigkeitsindex (p)	Trennschärfe-index (r)	N	Schwierigkeitsindex (p)	Trennschärfe-index (r)	N
v3_11	,34	,254	954	,14	,140	1134
v3_12	,27	,366	954	,53	,142	1134
v3_13	,56	,279	954	,49	,245	1134
v3_14	,42	,269	954	,47	,480	1134
v3_15	,15	,278	954	,40	,303	1134
v3_16	,40	,267	954	,47	,185	1134
v3_17	,57	,022	954	,13	,193	1134
v3_18	,15	,264	954	,51	,473	1134
v3_19	,48	,146	954	,41	,205	1134
v3_20	,20	,238	954	,22	,233	1134

Tabelle 5.3.1.6: Schwierigkeits- und Trennschärfeindizes für den Untertest Q1 des KFT 4-12+R in Klasse 5 für beide Testformen A und B

Items	Testform A			Testform B		
	Schwierigkeitsindex (p)	Trennschärfe-index (r)	N	Schwierigkeitsindex (p)	Trennschärfe-index (r)	N
q1_1	,76	,116	955	,86	,100	1133
q1_2	,96	,092	955	,95	,108	1133
q1_3	,94	,123	955	,77	,132	1133
q1_4	,83	,048	955	,80	,174	1133
q1_5	,85	,136	955	,85	,213	1133
q1_6	,78	,136	955	,86	,089	1133
q1_7	,69	,176	955	,16	,125	1133
q1_8	,62	,148	955	,41	,117	1133
q1_9	,54	,145	955	,78	,145	1133
q1_10	,62	,184	955	,62	,240	1133
q1_11	,56	,084	955	,48	,211	1133
q1_12	,27	,054	955	,56	,203	1133
q1_13	,09	,049	955	,42	,140	1133
q1_14	,35	,171	955	,38	,198	1133
q1_15	,08	-,021	955	,14	,170	1133
q1_16	,57	,166	955	,49	,249	1133
q1_17	,46	,097	955	,16	,256	1133
q1_18	,10	,058	955	,52	,244	1133
q1_19	,26	,103	955	,16	,121	1133
q1_20	,23	,056	955	,22	,232	1133
q1_21	,24	,082	955	,17	,134	1133
q1_22	,07	,016	955	,22	,228	1133
q1_23	,19	,166	955	,31	,252	1133
q1_24	,19	,125	955	,19	,125	1133
q1_25	,03	,073	955	,16	,132	1133

Tabelle 5.3.1.7: Schwierigkeits- und Trennschärfeindizes für den Untertest Q2 des KFT 4-12+R in Klasse 5 für beide Testformen A und B

Items	Testform A			Testform B		
	Schwierigkeitsindex (p)	Trennschärfe-index (r)	N	Schwierigkeitsindex (p)	Trennschärfe-index (r)	N
q2_1	,80	,291	953	,90	,183	1130
q2_2	,94	,290	953	,79	,240	1130
q2_3	,92	,436	953	,61	,426	1130
q2_4	,86	,377	953	,62	,380	1130
q2_5	,86	,314	953	,69	,419	1130
q2_6	,89	,401	953	,54	,326	1130
q2_7	,88	,403	953	,79	,386	1130
q2_8	,83	,465	953	,52	,497	1130
q2_9	,94	,430	953	,60	,366	1130
q2_10	,94	,388	953	,34	,451	1130
q2_11	,86	,397	953	,68	,419	1130
q2_12	,87	,378	953	,59	,383	1130
q2_13	,82	,447	953	,77	,385	1130
q2_14	,53	,374	953	,46	,438	1130
q2_15	,58	,367	953	,53	,354	1130
q2_16	,59	,399	953	,45	,486	1130
q2_17	,68	,354	953	,37	,257	1130
q2_18	,74	,456	953	,27	,354	1130
q2_19	,60	,381	953	,32	,361	1130
q2_20	,71	,378	953	,24	,296	1130

Tabelle 5.3.1.8: Schwierigkeits- und Trennschärfeindizes für den Untertest Q3 des KFT 4-12+R in Klasse 5 für beide Testformen A und B

Items	Testform A			Testform B		
	Schwierigkeitsindex (p)	Trennschärfe-index (r)	N	Schwierigkeitsindex (p)	Trennschärfe-index (r)	N
q3_1	,85	,163	952	,87	,234	1135
q3_2	,80	,299	952	,79	,285	1135
q3_3	,75	,312	952	,71	,273	1135
q3_4	,66	,314	952	,57	,247	1135
q3_5	,62	,334	952	,44	,347	1135
q3_6	,56	,334	952	,66	,243	1135
q3_7	,57	,436	952	,52	,207	1135
q3_8	,45	,389	952	,38	,345	1135
q3_9	,33	,331	952	,31	,340	1135
q3_10	,42	,360	952	,17	,162	1135
q3_11	,24	,242	952	,35	,341	1135
q3_12	,43	,128	952	,16	,156	1135
q3_13	,27	,261	952	,12	,122	1135
q3_14	,06	-,039	952	,11	,075	1135
q3_15	,12	-,095	952	,25	,250	1135

Tabelle 5.3.1.9: Schwierigkeits- und Trennschärfeindizes für den Untertest N1 des KFT 4-12+R in Klasse 5 für beide Testformen A und B

Items	Testform A			Testform B		
	Schwierigkeits-index (p)	Trennschärfe-index (r)	N	Schwierigkeits-index (p)	Trennschärfe-index (r)	N
n1_1	,80	,254	951	,90	,178	1135
n1_2	,93	,254	951	,77	,418	1135
n1_3	,95	,177	951	,60	,087	1135
n1_4	,91	,242	951	,62	,120	1135
n1_5	,93	,332	951	,60	,223	1135
n1_6	,89	,591	951	,38	,332	1135
n1_7	,90	,213	951	,66	,438	1135
n1_8	,86	,376	951	,60	,455	1135
n1_9	,85	,609	951	,61	,232	1135
n1_10	,91	,193	951	,38	,298	1135
n1_11	,67	,340	951	,54	,551	1135
n1_12	,73	,428	951	,46	,324	1135
n1_13	,80	,306	951	,44	,594	1135
n1_14	,70	,437	951	,40	,587	1135
n1_15	,77	,582	951	,40	,130	1135
n1_16	,76	,559	951	,39	,596	1135
n1_17	,78	,599	951	,43	,574	1135
n1_18	,72	,475	951	,38	,545	1135
n1_19	,67	,379	951	,35	,335	1135
n1_20	,26	,146	951	,28	,454	1135
n1_21	,70	,628	951	,35	,424	1135
n1_22	,78	,408	951	,36	,506	1135
n1_23	,77	,523	951	,45	,372	1135
n1_24	,60	,438	951	,29	,555	1135
n1_25	,67	,417	951	,25	,194	1135

Tabelle 5.3.1.10: Schwierigkeits- und Trennschärfeindizes für den Untertest N2 des KFT 4-12+R in Klasse 5 für beide Testformen A und B

Items	Testform A			Testform B		
	Schwierigkeits-index (p)	Trennschärfe-index (r)	N	Schwierigkeits-index (p)	Trennschärfe-index (r)	N
n2_1	,66	,643	953	,72	,531	1135
n2_2	,74	,664	953	,75	,428	1135
n2_3	,74	,650	953	,77	,393	1135
n2_4	,71	,674	953	,75	,563	1135
n2_5	,77	,444	953	,79	,168	1135
n2_6	,79	,572	953	,81	,330	1135
n2_7	,60	,575	953	,83	,409	1135
n2_8	,79	,573	953	,58	,538	1135
n2_9	,71	,635	953	,71	,595	1135
n2_10	,59	,507	953	,62	,523	1135
n2_11	,81	,337	953	,67	,662	1135
n2_12	,58	,689	953	,68	,659	1135

Items	Testform A			Testform B		
	Schwierigkeits-index (p)	Trennschärfe-index (r)	N	Schwierigkeits-index (p)	Trennschärfe-index (r)	N
n2_13	,71	,727	953	,56	,581	1135
n2_14	,47	,346	953	,74	,558	1135
n2_15	,46	,459	953	,51	,605	1135
n2_16	,49	,522	953	,58	,558	1135
n2_17	,53	,448	953	,24	,305	1135
n2_18	,66	,614	953	,54	,555	1135
n2_19	,59	,615	953	,55	,576	1135
n2_20	,32	,424	953	,62	,719	1135
n2_21	,50	,530	953	,63	,643	1135
n2_22	,49	,476	953	,53	,661	1135
n2_23	,44	,467	953	,52	,592	1135
n2_24	,62	,521	953	,58	,590	1135
n2_25	,44	,472	953	,56	,565	1135

Tabelle 5.3.1.11: Schwierigkeits- und Trennschärfeindizes für den Untertest N3 des KFT 4-12+R in Klasse 5 für beide Testformen A und B

Items	Testform A			Testform B		
	Schwierigkeits-index (p)	Trennschärfe-index (r)	N	Schwierigkeits-index (p)t	Trennschärfe-index (r)	N
n3_1	,79	,376	953	,68	,355	1134
n3_2	,62	,312	953	,73	,404	1134
n3_3	,77	,389	953	,57	,309	1134
n3_4	,80	,324	953	,82	,344	1134
n3_5	,84	,290	953	,74	,497	1134
n3_6	,75	,472	953	,77	,343	1134
n3_7	,40	,343	953	,77	,518	1134
n3_8	,77	,513	953	,35	,301	1134
n3_9	,76	,528	953	,75	,527	1134
n3_10	,85	,334	953	,61	,364	1134
n3_11	,65	,354	953	,58	,496	1134
n3_12	,49	,383	953	,80	,423	1134
n3_13	,68	,402	953	,53	,403	1134
n3_14	,40	,238	953	,68	,446	1134
n3_15	,64	,345	953	,34	,203	1134

Tabelle 5.3.1.12: Schwierigkeits- und Trennschärfeindizes für die Dimension V des KFT 4-12+R in Klasse 5 für beide Testformen A und B

Items	Testform A			Testform B		
	Schwierigkeitsindex (p)	Trennschärfeindex (r)	N	Schwierigkeitsindex (p)	Trennschärfeindex (r)	N
v1_1	,90	,184	954	,71	,345	1134
v1_2	,70	,334	954	,58	,326	1134
v1_3	,75	,390	954	,77	,135	1134
v1_4	,60	,309	954	,68	,085	1134
v1_5	,93	,207	954	,75	,105	1134
v1_6	,75	,123	954	,71	,122	1134
v1_7	,68	,398	954	,49	,067	1134
v1_8	,39	,301	954	,47	,240	1134
v1_9	,74	-,032	954	,66	,313	1134
v1_10	,55	,262	954	,68	,412	1134
v1_11	,71	,460	954	,71	,338	1134
v1_12	,69	,354	954	,60	,275	1134
v1_13	,45	,114	954	,44	,246	1134
v1_14	,39	,288	954	,54	,397	1134
v1_15	,06	,085	954	,49	,150	1134
v1_16	,51	,477	954	,30	,326	1134
v1_17	,50	,367	954	,62	,332	1134
v1_18	,50	,323	954	,55	,317	1134
v1_19	,41	,320	954	,24	,236	1134
v1_20	,62	,339	954	,10	,256	1134
v1_21	,29	,257	954	,36	,332	1134
v1_22	,25	,326	954	,50	,279	1134
v1_23	,21	,358	954	,17	,175	1134
v1_24	,13	,169	954	,15	,219	1134
v1_25	,14	,089	954	,36	,346	1134
v2_1	,89	,170	954	,86	,223	1134
v2_2	,64	,397	954	,45	,483	1134
v2_3	,73	,184	954	,54	,309	1134
v2_4	,70	,210	954	,59	,334	1134
v2_5	,77	,404	954	,74	,187	1134
v2_6	,65	,400	954	,72	,339	1134
v2_7	,77	,443	954	,54	,362	1134
v2_8	,46	,162	954	,62	,349	1134
v2_9	,71	,383	954	,39	,378	1134
v2_10	,59	,288	954	,46	,397	1134
v2_11	,72	,431	954	,61	,245	1134
v2_12	,51	,509	954	,55	,263	1134
v2_13	,85	,190	954	,55	,422	1134
v2_14	,64	,311	954	,55	,283	1134
v2_15	,57	,246	954	,54	,348	1134
v2_16	,50	,100	954	,59	,391	1134
v2_17	,80	,335	954	,35	,007	1134
v2_18	,58	,429	954	,28	,118	1134
v2_19	,40	,309	954	,29	,084	1134
v2_20	,29	,169	954	,33	,335	1134
v2_21	,68	,383	954	,20	,174	1134

Items	Testform A			Testform B		
	Schwierigkeitsindex (p)	Trennschärfe-index (r)	N	Schwierigkeitsindex (p)	Trennschärfe-index (r)	N
v2_22	,81	,308	954	,53	,261	1134
v2_23	,63	,373	954	,28	,302	1134
v2_24	,54	,157	954	,59	,348	1134
v2_25	,72	,300	954	,58	,328	1134
v3_1	,79	,312	954	,75	,218	1134
v3_2	,73	,252	954	,24	,233	1134
v3_3	,61	,456	954	,38	,264	1134
v3_4	,51	,305	954	,59	,354	1134
v3_5	,59	,298	954	,57	,383	1134
v3_6	,53	,375	954	,35	,375	1134
v3_7	,31	,174	954	,68	,225	1134
v3_8	,64	,345	954	,52	,403	1134
v3_9	,35	,283	954	,36	,166	1134
v3_10	,70	,153	954	,73	,261	1134
v3_11	,34	,241	954	,14	,149	1134
v3_12	,27	,454	954	,53	,130	1134
v3_13	,56	,244	954	,49	,217	1134
v3_14	,42	,294	954	,47	,404	1134
v3_15	,15	,320	954	,40	,320	1134
v3_16	,40	,262	954	,47	,220	1134
v3_17	,57	,102	954	,13	,237	1134
v3_18	,15	,213	954	,51	,354	1134
v3_19	,48	,208	954	,41	,211	1134
v3_20	,20	,210	954	,22	,195	1134

Tabelle 5.3.1.13: Schwierigkeits- und Trennschärfeindizes für die Dimension Q des KFT 4-12+R in Klasse 5 für beide Testformen A und B

Items	Testform A			Testform B		
	Schwierigkeitsindex (p)	Trennschärfe-index (r)	N	Schwierigkeitsindex (p)	Trennschärfe-index (r)	N
q1_1	,76	,120	951	,86	,090	1127
q1_2	,96	,098	951	,95	,100	1127
q1_3	,94	,115	951	,77	,197	1127
q1_4	,83	,067	951	,80	,195	1127
q1_5	,85	,196	951	,86	,187	1127
q1_6	,78	,175	951	,86	,081	1127
q1_7	,70	,237	951	,16	,202	1127
q1_8	,62	,129	951	,42	,116	1127
q1_9	,54	,236	951	,78	,170	1127
q1_10	,62	,196	951	,63	,179	1127
q1_11	,56	,186	951	,49	,255	1127
q1_12	,27	,059	951	,56	,209	1127
q1_13	,09	,-011	951	,42	,108	1127
q1_14	,35	,190	951	,38	,172	1127
q1_15	,08	,-010	951	,14	,161	1127
q1_16	,57	,171	951	,49	,210	1127

Items	Testform A			Testform B		
	Schwierigkeitsindex (p)	Trennschärfe-index (r)	N	Schwierigkeitsindex (p)	Trennschärfe-index (r)	N
q1_17	,46	,031	951	,16	,205	1127
q1_18	,10	,035	951	,52	,250	1127
q1_19	,26	,093	951	,16	,140	1127
q1_20	,23	-,003	951	,22	,172	1127
q1_21	,24	,074	951	,17	,146	1127
q1_22	,07	-,003	951	,22	,211	1127
q1_23	,19	,101	951	,31	,210	1127
q1_24	,19	,056	951	,19	,129	1127
q1_25	,03	-,037	951	,16	,117	1127
q2_1	,80	,237	951	,90	,186	1127
q2_2	,94	,215	951	,79	,247	1127
q2_3	,92	,324	951	,61	,371	1127
q2_4	,86	,284	951	,62	,357	1127
q2_5	,86	,244	951	,69	,333	1127
q2_6	,89	,319	951	,54	,237	1127
q2_7	,88	,330	951	,79	,300	1127
q2_8	,83	,341	951	,51	,439	1127
q2_9	,94	,306	951	,60	,272	1127
q2_10	,94	,303	951	,34	,415	1127
q2_11	,86	,307	951	,68	,374	1127
q2_12	,87	,305	951	,59	,302	1127
q2_13	,82	,319	951	,78	,310	1127
q2_14	,53	,433	951	,46	,429	1127
q2_15	,58	,372	951	,53	,342	1127
q2_16	,59	,395	951	,45	,455	1127
q2_17	,68	,343	951	,37	,248	1127
q2_18	,74	,441	951	,27	,333	1127
q2_19	,60	,337	951	,32	,402	1127
q2_20	,71	,357	951	,24	,370	1127
q3_1	,85	,167	951	,87	,239	1127
q3_2	,80	,253	951	,79	,249	1127
q3_3	,75	,292	951	,71	,201	1127
q3_4	,66	,244	951	,57	,292	1127
q3_5	,62	,316	951	,44	,264	1127
q3_6	,56	,285	951	,66	,225	1127
q3_7	,57	,353	951	,52	,224	1127
q3_8	,45	,333	951	,38	,304	1127
q3_9	,33	,283	951	,31	,291	1127
q3_10	,42	,240	951	,17	,172	1127
q3_11	,25	,158	951	,35	,296	1127
q3_12	,43	,187	951	,16	,114	1127
q3_13	,27	,209	951	,12	,158	1127
q3_14	,06	-,063	951	,11	,088	1127
q3_15	,12	-,035	951	,25	,285	1127

Tabelle 5.3.1.14: Schwierigkeits- und Trennschärfeindizes für die Dimension N des KFT 4-12+R in Klasse 5 für beide Testformen A und B

Items	Testform A			Testform B		
	Schwierigkeitsindex (p)	Trennschärfe-index (r)	N	Schwierigkeitsindex (p)	Trennschärfe-index (r)	N
n1_1	,80	,184	951	,90	,214	1134
n1_2	,93	,207	951	,77	,345	1134
n1_3	,95	,148	951	,60	,147	1134
n1_4	,91	,181	951	,62	,141	1134
n1_5	,93	,269	951	,60	,222	1134
n1_6	,89	,421	951	,38	,309	1134
n1_7	,90	,205	951	,66	,371	1134
n1_8	,86	,281	951	,60	,396	1134
n1_9	,85	,474	951	,61	,179	1134
n1_10	,91	,243	951	,38	,299	1134
n1_11	,67	,299	951	,54	,426	1134
n1_12	,73	,372	951	,46	,256	1134
n1_13	,80	,222	951	,44	,496	1134
n1_14	,70	,406	951	,40	,479	1134
n1_15	,77	,466	951	,40	,131	1134
n1_16	,76	,463	951	,39	,470	1134
n1_17	,78	,462	951	,43	,500	1134
n1_18	,72	,385	951	,38	,421	1134
n1_19	,67	,332	951	,35	,277	1134
n1_20	,26	,221	951	,28	,378	1134
n1_21	,70	,522	951	,35	,366	1134
n1_22	,78	,321	951	,36	,392	1134
n1_23	,77	,412	951	,45	,308	1134
n1_24	,60	,372	951	,29	,454	1134
n1_25	,67	,338	951	,25	,206	1134
n2_1	,67	,540	951	,72	,472	1134
n2_2	,74	,551	951	,75	,415	1134
n2_3	,74	,575	951	,78	,325	1134
n2_4	,71	,598	951	,75	,497	1134
n2_5	,77	,414	951	,80	,180	1134
n2_6	,79	,503	951	,81	,308	1134
n2_7	,60	,529	951	,83	,402	1134
n2_8	,79	,507	951	,58	,526	1134
n2_9	,72	,551	951	,71	,538	1134
n2_10	,59	,462	951	,62	,494	1134
n2_11	,81	,304	951	,67	,573	1134
n2_12	,59	,636	951	,68	,570	1134
n2_13	,71	,638	951	,56	,523	1134
n2_14	,47	,298	951	,74	,484	1134
n2_15	,46	,430	951	,51	,574	1134
n2_16	,49	,478	951	,58	,521	1134
n2_17	,53	,434	951	,24	,287	1134
n2_18	,66	,552	951	,54	,522	1134
n2_19	,59	,542	951	,55	,541	1134
n2_20	,32	,404	951	,62	,618	1134
n2_21	,50	,494	951	,63	,566	1134

Items	Testform A			Testform B		
	Schwierigkeitsindex (p)	Trennschärfe-index (r)	N	Schwierigkeitsindex (p)	Trennschärfe-index (r)	N
n2_22	,49	,474	951	,53	,613	1134
n2_23	,45	,424	951	,52	,511	1134
n2_24	,62	,450	951	,58	,521	1134
n2_25	,44	,433	951	,56	,481	1134
n3_1	,79	,260	951	,68	,300	1134
n3_2	,62	,201	951	,73	,306	1134
n3_3	,77	,299	951	,57	,246	1134
n3_4	,80	,231	951	,82	,296	1134
n3_5	,84	,216	951	,74	,349	1134
n3_6	,75	,341	951	,77	,199	1134
n3_7	,40	,260	951	,77	,377	1134
n3_8	,77	,352	951	,35	,276	1134
n3_9	,76	,391	951	,75	,402	1134
n3_10	,85	,309	951	,61	,236	1134
n3_11	,66	,243	951	,58	,411	1134
n3_12	,49	,285	951	,80	,310	1134
n3_13	,68	,308	951	,53	,305	1134
n3_14	,40	,170	951	,68	,316	1134
n3_15	,64	,253	951	,34	,218	1134

Tabelle 5.3.1.15: Schwierigkeits- und Trennschärfeindizes für die Gesamtleistung GL des KFT 4-12+R in Klasse 5 für beide Testformen A und B

Items	Testform A			Testform B		
	Schwierigkeitsindex (p)	Trennschärfe-index (r)	N	Schwierigkeitsindex (p)	Trennschärfe-index (r)	N
v1_1	,90	,133	946	,71	,265	1124
v1_2	,70	,255	946	,58	,249	1124
v1_3	,75	,321	946	,77	,117	1124
v1_4	,60	,204	946	,68	,103	1124
v1_5	,93	,189	946	,75	,090	1124
v1_6	,75	,133	946	,71	,116	1124
v1_7	,67	,340	946	,49	,077	1124
v1_8	,39	,216	946	,47	,171	1124
v1_9	,74	,016	946	,66	,281	1124
v1_10	,55	,234	946	,68	,304	1124
v1_11	,71	,330	946	,71	,228	1124
v1_12	,69	,267	946	,60	,263	1124
v1_13	,45	,107	946	,44	,240	1124
v1_14	,39	,215	946	,54	,245	1124
v1_15	,06	,094	946	,49	,130	1124
v1_16	,51	,363	946	,30	,232	1124
v1_17	,50	,255	946	,62	,262	1124
v1_18	,50	,251	946	,55	,272	1124
v1_19	,41	,255	946	,24	,182	1124
v1_20	,62	,268	946	,10	,194	1124
v1_21	,29	,209	946	,36	,227	1124

Items	Testform A			Testform B		
	Schwierigkeitsindex (p)	Trennschärfe-index (r)	N	Schwierigkeitsindex (p)	Trennschärfe-index (r)	N
v1_22	,25	,287	946	,50	,196	1124
v1_23	,21	,275	946	,17	,139	1124
v1_24	,13	,111	946	,15	,187	1124
v1_25	,14	,092	946	,36	,291	1124
v2_1	,89	,165	946	,86	,194	1124
v2_2	,64	,332	946	,45	,340	1124
v2_3	,73	,177	946	,54	,217	1124
v2_4	,70	,172	946	,59	,234	1124
v2_5	,77	,335	946	,74	,159	1124
v2_6	,65	,311	946	,72	,279	1124
v2_7	,77	,363	946	,54	,254	1124
v2_8	,46	,126	946	,62	,282	1124
v2_9	,72	,323	946	,39	,280	1124
v2_10	,59	,246	946	,46	,287	1124
v2_11	,72	,321	946	,61	,211	1124
v2_12	,51	,436	946	,55	,169	1124
v2_13	,85	,156	946	,55	,311	1124
v2_14	,64	,230	946	,55	,233	1124
v2_15	,57	,171	946	,54	,271	1124
v2_16	,50	,069	946	,59	,285	1124
v2_17	,80	,243	946	,35	,010	1124
v2_18	,58	,348	946	,28	,102	1124
v2_19	,40	,205	946	,28	,094	1124
v2_20	,29	,166	946	,33	,274	1124
v2_21	,68	,305	946	,20	,121	1124
v2_22	,81	,246	946	,53	,156	1124
v2_23	,63	,229	946	,29	,194	1124
v2_24	,54	,093	946	,59	,292	1124
v2_25	,72	,225	946	,58	,236	1124
v3_1	,79	,264	946	,75	,240	1124
v3_2	,74	,255	946	,24	,221	1124
v3_3	,61	,410	946	,38	,251	1124
v3_4	,51	,292	946	,59	,363	1124
v3_5	,59	,290	946	,58	,368	1124
v3_6	,53	,355	946	,35	,316	1124
v3_7	,31	,159	946	,68	,201	1124
v3_8	,64	,334	946	,52	,387	1124
v3_9	,36	,253	946	,36	,112	1124
v3_10	,70	,177	946	,74	,232	1124
v3_11	,34	,244	946	,14	,126	1124
v3_12	,27	,355	946	,54	,109	1124
v3_13	,56	,251	946	,49	,215	1124
v3_14	,42	,265	946	,47	,368	1124
v3_15	,15	,267	946	,40	,258	1124
v3_16	,41	,212	946	,47	,185	1124
v3_17	,57	,115	946	,14	,188	1124
v3_18	,15	,204	946	,51	,372	1124
v3_19	,48	,199	946	,41	,211	1124
v3_20	,20	,174	946	,22	,188	1124

Items	Testform A			Testform B		
	Schwierigkeitsindex (p)	Trennschärfe-index (r)	N	Schwierigkeitsindex (p)	Trennschärfe-index (r)	N
q1_1	,76	,129	946	,86	,086	1124
q1_2	,96	,127	946	,95	,107	1124
q1_3	,94	,086	946	,77	,240	1124
q1_4	,83	,078	946	,80	,158	1124
q1_5	,85	,192	946	,86	,161	1124
q1_6	,78	,102	946	,86	,049	1124
q1_7	,70	,245	946	,16	,193	1124
q1_8	,62	,118	946	,41	,100	1124
q1_9	,54	,211	946	,78	,146	1124
q1_10	,62	,154	946	,63	,119	1124
q1_11	,56	,180	946	,48	,253	1124
q1_12	,27	,006	946	,56	,128	1124
q1_13	,09	,062	946	,42	,076	1124
q1_14	,35	,166	946	,38	,029	1124
q1_15	,08	-,001	946	,14	,080	1124
q1_16	,58	,107	946	,49	,084	1124
q1_17	,46	-,044	946	,15	,136	1124
q1_18	,10	,036	946	,52	,199	1124
q1_19	,26	,025	946	,16	,095	1124
q1_20	,23	-,030	946	,22	,047	1124
q1_21	,24	,035	946	,17	,136	1124
q1_22	,07	-,040	946	,22	,113	1124
q1_23	,19	,115	946	,31	,054	1124
q1_24	,19	,068	946	,19	,087	1124
q1_25	,03	-,093	946	,16	-,006	1124
q2_1	,80	,139	946	,90	,190	1124
q2_2	,94	,162	946	,79	,256	1124
q2_3	,92	,228	946	,61	,272	1124
q2_4	,86	,297	946	,62	,329	1124
q2_5	,87	,176	946	,69	,304	1124
q2_6	,89	,248	946	,54	,233	1124
q2_7	,88	,280	946	,79	,284	1124
q2_8	,83	,256	946	,51	,301	1124
q2_9	,94	,197	946	,60	,264	1124
q2_10	,95	,178	946	,34	,304	1124
q2_11	,86	,228	946	,68	,311	1124
q2_12	,87	,249	946	,59	,270	1124
q2_13	,83	,207	946	,78	,225	1124
q2_14	,53	,413	946	,46	,352	1124
q2_15	,58	,348	946	,52	,288	1124
q2_16	,59	,363	946	,45	,293	1124
q2_17	,68	,323	946	,37	,167	1124
q2_18	,74	,349	946	,27	,222	1124
q2_19	,60	,328	946	,32	,311	1124
q2_20	,71	,311	946	,24	,257	1124
q3_1	,86	,149	946	,87	,157	1124
q3_2	,80	,236	946	,79	,223	1124
q3_3	,75	,226	946	,72	,187	1124
q3_4	,66	,205	946	,57	,235	1124

Items	Testform A			Testform B		
	Schwierigkeitsindex (p)	Trennschärfe-index (r)	N	Schwierigkeitsindex (p)	Trennschärfe-index (r)	N
q3_5	,62	,230	946	,44	,181	1124
q3_6	,56	,191	946	,66	,178	1124
q3_7	,57	,289	946	,52	,213	1124
q3_8	,45	,261	946	,38	,214	1124
q3_9	,33	,214	946	,31	,209	1124
q3_10	,42	,159	946	,17	,129	1124
q3_11	,25	,107	946	,35	,201	1124
q3_12	,44	,104	946	,16	,086	1124
q3_13	,27	,116	946	,12	,050	1124
q3_14	,06	,-055	946	,11	,-003	1124
q3_15	,12	,-070	946	,25	,144	1124
n1_1	,81	,163	946	,90	,182	1124
n1_2	,93	,178	946	,77	,268	1124
n1_3	,95	,146	946	,60	,137	1124
n1_4	,91	,143	946	,62	,128	1124
n1_5	,93	,239	946	,61	,201	1124
n1_6	,89	,316	946	,38	,286	1124
n1_7	,90	,206	946	,66	,290	1124
n1_8	,87	,229	946	,60	,329	1124
n1_9	,85	,373	946	,61	,166	1124
n1_10	,91	,233	946	,38	,281	1124
n1_11	,67	,260	946	,54	,308	1124
n1_12	,74	,311	946	,47	,198	1124
n1_13	,80	,197	946	,45	,410	1124
n1_14	,70	,324	946	,40	,369	1124
n1_15	,77	,358	946	,41	,122	1124
n1_16	,76	,394	946	,39	,375	1124
n1_17	,78	,387	946	,43	,399	1124
n1_18	,73	,300	946	,39	,357	1124
n1_19	,67	,293	946	,35	,217	1124
n1_20	,26	,191	946	,28	,325	1124
n1_21	,70	,433	946	,36	,307	1124
n1_22	,79	,265	946	,36	,323	1124
n1_23	,77	,314	946	,46	,268	1124
n1_24	,61	,309	946	,29	,377	1124
n1_25	,67	,297	946	,25	,193	1124
n2_1	,67	,409	946	,72	,378	1124
n2_2	,74	,429	946	,75	,363	1124
n2_3	,74	,460	946	,78	,288	1124
n2_4	,71	,468	946	,75	,399	1124
n2_5	,77	,349	946	,79	,171	1124
n2_6	,79	,430	946	,81	,271	1124
n2_7	,60	,448	946	,83	,345	1124
n2_8	,79	,412	946	,59	,467	1124
n2_9	,72	,458	946	,71	,437	1124
n2_10	,59	,419	946	,62	,412	1124
n2_11	,81	,262	946	,68	,482	1124
n2_12	,59	,548	946	,68	,477	1124
n2_13	,71	,516	946	,56	,445	1124

Items	Testform A			Testform B		
	Schwierigkeitsindex (p)	Trennschärfe-index (r)	N	Schwierigkeitsindex (p)	Trennschärfe-index (r)	N
n2_14	,47	<u>,267</u>	946	,75	,410	1124
n2_15	,47	,387	946	,51	,495	1124
n2_16	,49	,406	946	,58	,419	1124
n2_17	,53	,369	946	,24	<u>,252</u>	1124
n2_18	,66	,466	946	,54	,446	1124
n2_19	,59	,468	946	,55	,467	1124
n2_20	,33	,379	946	,62	,522	1124
n2_21	,50	,439	946	,63	,473	1124
n2_22	,50	,434	946	,53	,507	1124
n2_23	,45	,386	946	,52	,415	1124
n2_24	,62	,390	946	,58	,451	1124
n2_25	,44	,394	946	,56	,391	1124
n3_1	,79	<u>,263</u>	946	,68	,292	1124
n3_2	,62	<u>,229</u>	946	,73	,305	1124
n3_3	,77	<u>,293</u>	946	,57	<u>,235</u>	1124
n3_4	<u>,81</u>	<u>,219</u>	946	<u>,82</u>	<u>,275</u>	1124
n3_5	<u>,84</u>	<u>,246</u>	946	,74	,314	1124
n3_6	,75	,335	946	,77	<u>,213</u>	1124
n3_7	,40	<u>,282</u>	946	,77	,362	1124
n3_8	,77	,308	946	,35	,314	1124
n3_9	,76	,348	946	,75	,407	1124
n3_10	<u>,85</u>	<u>,272</u>	946	,61	<u>,212</u>	1124
n3_11	,66	<u>,265</u>	946	,59	,389	1124
n3_12	,49	<u>,260</u>	946	,80	,304	1124
n3_13	,68	,298	946	,53	,305	1124
n3_14	,40	<u>,213</u>	946	,68	,328	1124
n3_15	,64	<u>,246</u>	946	,34	<u>,237</u>	1124

Tabelle 5.3.1.16: Schwierigkeits- und Trennschärfeindizes für den Untertest V1 des KFT 4-12+R in Klasse 7 für beide Testformen A und B

Items	Testform A			Testform B		
	Schwierigkeitsindex (p)	Trennschärfe-index (r)	N	Schwierigkeitsindex (p)	Trennschärfe-index (r)	N
v1_11	,88	,316	1103	,86	,268	1110
v1_12	,83	,290	1103	,83	,167	1110
v1_13	,63	,136	1103	,58	,252	1110
v1_14	,60	,320	1103	,81	,339	1110
v1_15	,09	,124	1103	,61	,154	1110
v1_16	,72	,430	1103	,54	,255	1110
v1_17	,64	,333	1103	,89	,253	1110
v1_18	,66	,358	1103	,76	,229	1110
v1_19	,60	,298	1103	,38	,254	1110
v1_20	,83	,230	1103	,48	,417	1110
v1_21	,48	,283	1103	,78	,229	1110
v1_22	,64	,391	1103	,84	,173	1110
v1_23	,41	,313	1103	,32	,348	1110
v1_24	,29	,306	1103	,63	,358	1110
v1_25	,22	,165	1103	,68	,353	1110
v1_26	,86	,248	1103	,16	,198	1110
v1_27	,24	,329	1103	,42	,281	1110
v1_28	,25	,124	1103	,18	,104	1110
v1_29	,35	,348	1103	,27	,320	1110
v1_30	,16	,167	1103	,17	,100	1110
v1_31	,34	,190	1103	,25	,178	1110
v1_32	,10	,001	1103	,39	,332	1110
v1_33	,63	,391	1103	,30	,165	1110
v1_34	,37	,241	1103	,59	,174	1110
v1_35	,72	,336	1103	,38	,213	1110

Tabelle 5.3.1.17: Schwierigkeits- und Trennschärfeindizes für den Untertest V2 des KFT 4-12+R in Klasse 7 für beide Testformen A und B

Items	Testform A			Testform B		
	Schwierigkeitsindex (p)	Trennschärfe-index (r)	N	Schwierigkeitsindex (p)	Trennschärfe-index (r)	N
v2_11	,82	,326	1103	,78	,110	1111
v2_12	,74	,378	1103	,58	,267	1111
v2_13	,92	,153	1103	,78	,309	1111
v2_14	,78	,281	1103	,69	,221	1111
v2_15	,72	,192	1103	,76	,367	1111
v2_16	,65	,152	1103	,83	,316	1111
v2_17	,93	,259	1103	,51	,119	1111
v2_18	,85	,316	1103	,37	,168	1111
v2_19	,63	,325	1103	,28	,034	1111
v2_20	,48	,184	1103	,55	,373	1111
v2_21	,79	,105	1103	,30	,229	1111
v2_22	,95	,194	1103	,79	,109	1111

Items	Testform A			Testform B		
	Schwierigkeitsindex (p)	Trennschärfe-index (r)	N	Schwierigkeitsindex (p)	Trennschärfe-index (r)	N
v2_23	,82	,263	1103	,50	,187	1111
v2_24	,69	,112	1103	,84	,290	1111
v2_25	,85	,182	1103	,81	,294	1111
v2_26	,63	,237	1103	,38	,112	1111
v2_27	,43	,339	1103	,42	,223	1111
v2_28	,83	,260	1103	,67	,113	1111
v2_29	,61	,142	1103	,13	-,063	1111
v2_30	,25	,176	1103	,48	,195	1111
v2_31	,50	,195	1103	,80	,154	1111
v2_32	,36	,318	1103	,73	,280	1111
v2_33	,41	,305	1103	,67	,113	1111
v2_34	,07	,121	1103	,44	-,101	1111
v2_35	,34	,009	1103	,22	,148	1111

Tabelle 5.3.1.18: Schwierigkeits- und Trennschärfeindizes für den Untertest V3 des KFT 4-12+R in Klasse 7 für beide Testformen A und B

Items	Testform A			Testform B		
	Schwierigkeitsindex (p)	Trennschärfe-index (r)	N	Schwierigkeitsindex (p)	Trennschärfe-index (r)	N
v3_9	,52	,394	1103	,55	,066	1110
v3_10	,72	,228	1103	,86	,245	1110
v3_11	,51	,410	1103	,87	,255	1110
v3_12	,41	,480	1103	,68	,147	1110
v3_13	,80	,281	1103	,71	,301	1110
v3_14	,62	,360	1103	,65	,439	1110
v3_15	,52	,463	1103	,71	,221	1110
v3_16	,65	,330	1103	,74	,180	1110
v3_17	,56	,099	1103	,45	,318	1110
v3_18	,42	,332	1103	,71	,404	1110
v3_19	,72	,170	1103	,64	,314	1110
v3_20	,37	,433	1103	,42	,277	1110
v3_21	,78	,159	1103	,38	,427	1110
v3_22	,18	,057	1103	,55	,199	1110
v3_23	,28	,214	1103	,32	,276	1110
v3_24	,48	,258	1103	,39	,287	1110
v3_25	,20	,193	1103	,65	,322	1110
v3_26	,48	,223	1103	,58	,321	1110
v3_27	,31	,187	1103	,24	,260	1110
v3_28	,23	,124	1103	,09	,036	1110

Tabelle 5.3.1.19: Schwierigkeits- und Trennschärfeindizes für den Untertest Q1 des KFT 4-12+R in Klasse 7 für beide Testformen A und B

Items	Testform A			Testform B		
	Schwierigkeitsindex (p)	Trennschärfe-index (r)	N	Schwierigkeitsindex (p)	Trennschärfe-index (r)	N
q1_11	,55	<u>-,031</u>	1104	<u>,87</u>	<u>,189</u>	1109
q1_12	,47	<u>,228</u>	1104	,78	<u>,061</u>	1109
q1_13	,59	<u>,250</u>	1104	,50	<u>,147</u>	1109
q1_14	,68	<u>,162</u>	1104	,66	<u>,114</u>	1109
q1_15	,47	,297	1104	,54	<u>,212</u>	1109
q1_16	<u>,83</u>	<u>,122</u>	1104	,50	<u>,055</u>	1109
q1_17	,49	<u>,019</u>	1104	,52	<u>,198</u>	1109
q1_18	,42	<u>,236</u>	1104	<u>,81</u>	<u>,141</u>	1109
q1_19	<u>,17</u>	<u>,032</u>	1104	,34	<u>,221</u>	1109
q1_20	,25	<u>-,046</u>	1104	,48	<u>,242</u>	1109
q1_21	,40	<u>,182</u>	1104	,66	<u>,235</u>	1109
q1_22	<u>,17</u>	<u>,215</u>	1104	,45	<u>,216</u>	1109
q1_23	,56	,374	1104	,48	<u>,184</u>	1109
q1_24	,50	,309	1104	,57	<u>,252</u>	1109
q1_25	,25	<u>,285</u>	1104	,40	<u>,146</u>	1109
q1_26	,43	,305	1104	,54	<u>,272</u>	1109
q1_27	,37	<u>,082</u>	1104	,25	<u>,184</u>	1109
q1_28	,57	<u>,187</u>	1104	,26	<u>,212</u>	1109
q1_29	,30	,376	1104	,46	<u>,211</u>	1109
q1_30	,27	<u>,199</u>	1104	,52	<u>,130</u>	1109
q1_31	,52	<u>,239</u>	1104	,44	<u>,250</u>	1109
q1_32	,21	<u>,187</u>	1104	,33	<u>,238</u>	1109
q1_33	,25	<u>,180</u>	1104	,62	<u>,163</u>	1109
q1_34	,21	<u>,006</u>	1104	<u>,18</u>	<u>,064</u>	1109
q1_35	,48	<u>,059</u>	1104	,33	<u>,180</u>	1109

Tabelle 5.3.1.20: Schwierigkeits- und Trennschärfeindizes für den Untertest Q2 des KFT 4-12+R in Klasse 7 für beide Testformen A und B

Items	Testform A			Testform B		
	Schwierigkeitsindex (p)	Trennschärfe-index (r)	N	Schwierigkeitsindex (p)	Trennschärfe-Index (r)	N
q2_9	,95	,248	1104	,75	,323	1111
q2_10	,96	,203	1104	,77	,311	1111
q2_11	,90	,203	1104	,86	,305	1111
q2_12	,94	,323	1104	,83	,241	1111
q2_13	,94	,266	1104	,92	,166	1111
q2_14	,76	,427	1104	,78	,351	1111
q2_15	,75	,335	1104	,77	,309	1111
q2_16	,78	,438	1104	,81	,302	1111
q2_17	,78	,298	1104	,59	,280	1111
q2_18	,86	,343	1104	,61	,333	1111
q2_19	,75	,247	1104	,68	,445	1111
q2_20	,85	,260	1104	,57	,469	1111
q2_21	,77	,465	1104	,51	,348	1111
q2_22	,84	,341	1104	,60	,371	1111
q2_23	,76	,436	1104	,40	,333	1111
q2_24	,63	,332	1104	,45	,392	1111
q2_25	,66	,420	1104	,55	,461	1111
q2_26	,71	,349	1104	,55	,436	1111
q2_27	,55	,304	1104	,38	,349	1111
q2_28	,66	,456	1104	,33	,383	1111

Tabelle 5.3.1.21: Schwierigkeits- und Trennschärfeindizes für den Untertest Q3 des KFT 4-12+R in Klasse 7 für beide Testformen A und B

Items	Testform A			Testform B		
	Schwierigkeitsindex (p)	Trennschärfe-index (r)	N	Schwierigkeitsindex (p)	Trennschärfe-index (r)	N
q3_7	,81	,264	1102	,70	,145	1110
q3_8	,65	,317	1102	,56	,273	1110
q3_9	,63	,307	1102	,65	,331	1110
q3_10	,62	,355	1102	,39	,250	1110
q3_11	,43	,303	1102	,63	,277	1110
q3_12	,70	,241	1102	,37	,173	1110
q3_13	,49	,335	1102	,45	,269	1110
q3_14	,44	,248	1102	,20	,110	1110
q3_15	,25	,139	1102	,59	,354	1110
q3_16	,34	,338	1102	,36	,279	1110
q3_17	,40	,259	1102	,19	,001	1110
q3_18	,49	,321	1102	,26	,251	1110
q3_19	,22	,213	1102	,37	,262	1110
q3_20	,22	,238	1102	,27	,256	1110
q3_21	,26	,289	1102	,19	,084	1110

Tabelle 5.3.1.22: Schwierigkeits- und Trennschärfeindizes für den Untertest N1 des KFT 4-12+R in Klasse 7 für beide Testformen A und B

Items	Testform A			Testform B		
	Schwierigkeitsindex (p)	Trennschärfe-Index (r)	N	Schwierigkeitsindex (p)	Trennschärfe-Index (r)	N
n1_11	,78	,473	1112	,81	,386	1111
n1_12	,70	,266	1112	,65	,313	1111
n1_13	,81	,419	1112	,78	,541	1111
n1_14	,86	,377	1112	,69	,555	1111
n1_15	,86	,654	1112	,51	,152	1111
n1_16	,85	,610	1112	,68	,575	1111
n1_17	,90	,406	1112	,75	,504	1111
n1_18	,87	,317	1112	,64	,479	1111
n1_19	,81	,471	1112	,42	,178	1111
n1_20	,57	,366	1112	,58	,472	1111
n1_21	,83	,603	1112	,65	,387	1111
n1_22	,84	,498	1112	,65	,530	1111
n1_23	,88	,535	1112	,79	,489	1111
n1_24	,73	,474	1112	,67	,517	1111
n1_25	,82	,514	1112	,54	,389	1111
n1_26	,73	,531	1112	,66	,568	1111
n1_27	,88	,624	1112	,67	,525	1111
n1_28	,75	,415	1112	,61	,425	1111
n1_29	,76	,524	1112	,83	,628	1111
n1_30	,84	,493	1112	,64	,391	1111
n1_31	,87	,620	1112	,45	,359	1111
n1_32	,66	,372	1112	,40	,177	1111
n1_33	,85	,635	1112	,59	,321	1111
n1_34	,67	,371	1112	,33	,164	1111
n1_35	,66	,187	1112	,41	,336	1111

Tabelle 5.3.1.23: Schwierigkeits- und Trennschärfeindizes für den Untertest N2 des KFT 4-12+R in Klasse 7 für beide Testformen A und B

Items	Testform A			Testform B		
	Schwierigkeitsindex (p)	Trennschärfe-Index (r)	N	Schwierigkeitsindex (p)	Trennschärfe-Index (r)	N
n2_11	,89	,336	1111	,76	,402	1113
n2_12	,78	,459	1111	,90	,555	1113
n2_13	,85	,576	1111	,71	,468	1113
n2_14	,66	,183	1111	,86	,344	1113
n2_15	,60	,476	1111	,76	,553	1113
n2_16	,62	,549	1111	,85	,422	1113
n2_17	,72	,458	1111	,39	,204	1113
n2_18	,79	,465	1111	,78	,467	1113
n2_19	,66	,535	1111	,82	,476	1113
n2_20	,49	,452	1111	,81	,573	1113
n2_21	,72	,499	1111	,90	,541	1113
n2_22	,69	,492	1111	,83	,608	1113

Items	Testform A			Testform B		
	Schwierigkeitsindex (p)	Trennschärfe-index (r)	N	Schwierigkeitsindex (p)	Trennschärfe-index (r)	N
n2_23	,68	,405	1111	,78	,525	1113
n2_24	,83	,487	1111	,82	,496	1113
n2_25	,69	,375	1111	,76	,539	1113
n2_26	,63	,416	1111	,73	,565	1113
n2_27	,65	,522	1111	,74	,513	1113
n2_28	,63	,453	1111	,75	,489	1113
n2_29	,79	,434	1111	,62	,397	1113
n2_30	,74	,380	1111	,80	,485	1113
n2_31	,56	,278	1111	,50	,367	1113
n2_32	,58	,374	1111	,55	,277	1113
n2_33	,57	,367	1111	,57	,433	1113
n2_34	,51	,220	1111	,74	,219	1113
n2_35	,56	,430	1111	,42	,287	1113

Tabelle 5.3.1.24: Schwierigkeits- und Trennschärfeindizes für den Untertest N3 des KFT 4-12+R in Klasse 7 für beide Testformen A und B

Items	Testform A			Testform B		
	Schwierigkeitsindex (p)	Trennschärfe-index (r)	N	Schwierigkeitsindex (p)	Trennschärfe-index	N
n3_7	,44	,398	1110	,87	,379	1113
n3_8	,85	,510	1110	,56	,353	1113
n3_9	,84	,449	1110	,87	,489	1113
n3_10	,88	,406	1110	,82	,377	1113
n3_11	,79	,427	1110	,76	,426	1113
n3_12	,63	,440	1110	,90	,316	1113
n3_13	,75	,413	1110	,76	,422	1113
n3_14	,54	,371	1110	,82	,393	1113
n3_15	,72	,436	1110	,57	,312	1113
n3_16	,52	,252	1110	,75	,442	1113
n3_17	,67	,514	1110	,71	,440	1113
n3_18	,64	,471	1110	,52	,205	1113
n3_19	,48	,277	1110	,69	,376	1113
n3_20	,46	,360	1110	,47	,337	1113
n3_21	,69	,410	1110	,52	,366	1113

Tabelle 5.3.1.25: Schwierigkeits- und Trennschärfeindizes für die Dimension V des KFT 4-12+R in Klasse 7 für beide Testformen A und B

Items	Testform A			Testform B		
	Schwierigkeitsindex (p)	Trennschärfeindex (r)	N	Schwierigkeitsindex (p)	Trennschärfeindex (r)	N
v1_11	<u>.88</u>	,315	1103	<u>.86</u>	<u>.275</u>	1109
v1_12	<u>.83</u>	<u>.281</u>	1103	<u>.83</u>	<u>.173</u>	1109
v1_13	,63	<u>.130</u>	1103	,58	<u>.276</u>	1109
v1_14	,60	,360	1103	<u>.81</u>	,330	1109
v1_15	<u>.09</u>	<u>.082</u>	1103	,61	<u>.193</u>	1109
v1_16	,72	,477	1103	,54	,314	1109
v1_17	,64	,371	1103	<u>.89</u>	<u>.268</u>	1109
v1_18	,66	,367	1103	,76	<u>.261</u>	1109
v1_19	,60	<u>.292</u>	1103	,38	<u>.240</u>	1109
v1_20	<u>.83</u>	<u>.220</u>	1103	,48	,390	1109
v1_21	,48	,297	1103	,78	<u>.236</u>	1109
v1_22	,64	,441	1103	<u>.84</u>	<u>.181</u>	1109
v1_23	,41	,317	1103	,32	,305	1109
v1_24	,29	,313	1103	,63	,363	1109
v1_25	,22	<u>.166</u>	1103	,68	,369	1109
v1_26	<u>.86</u>	<u>.246</u>	1103	<u>.16</u>	<u>.149</u>	1109
v1_27	,24	,341	1103	,42	,305	1109
v1_28	,25	<u>.134</u>	1103	<u>.18</u>	<u>.070</u>	1109
v1_29	,35	,348	1103	,27	,306	1109
v1_30	<u>.16</u>	<u>.165</u>	1103	<u>.17</u>	<u>.099</u>	1109
v1_31	,34	<u>.172</u>	1103	,25	<u>.155</u>	1109
v1_32	<u>.10</u>	<u>-.006</u>	1103	,39	,311	1109
v1_33	,63	,391	1103	,30	<u>.156</u>	1109
v1_34	,37	<u>.234</u>	1103	,59	<u>.146</u>	1109
v1_35	,72	,332	1103	,38	<u>.171</u>	1109
v2_11	<u>.82</u>	,316	1103	,78	<u>.135</u>	1109
v2_12	,74	,415	1103	,58	,335	1109
v2_13	<u>.92</u>	<u>.116</u>	1103	,78	,308	1109
v2_14	,78	<u>.293</u>	1103	,69	<u>.204</u>	1109
v2_15	,72	<u>.237</u>	1103	,76	,362	1109
v2_16	,65	<u>.124</u>	1103	<u>.83</u>	,375	1109
v2_17	<u>.93</u>	<u>.240</u>	1103	,51	<u>.145</u>	1109
v2_18	<u>.85</u>	<u>.291</u>	1103	,37	<u>.155</u>	1109
v2_19	,63	,347	1103	,28	<u>.034</u>	1109
v2_20	,48	<u>.191</u>	1103	,55	,438	1109
v2_21	,79	<u>.069</u>	1103	,30	<u>.193</u>	1109
v2_22	<u>.95</u>	<u>.175</u>	1103	,79	<u>.075</u>	1109
v2_23	<u>.82</u>	<u>.257</u>	1103	,50	<u>.220</u>	1109
v2_24	,69	<u>.090</u>	1103	<u>.84</u>	<u>.208</u>	1109
v2_25	<u>.85</u>	<u>.213</u>	1103	<u>.82</u>	,342	1109
v2_26	,63	<u>.233</u>	1103	,38	<u>.149</u>	1109
v2_27	,43	,385	1103	,42	<u>.278</u>	1109
v2_28	<u>.83</u>	<u>.248</u>	1103	,67	<u>.089</u>	1109
v2_29	,61	<u>.179</u>	1103	<u>.13</u>	<u>-.111</u>	1109
v2_30	,25	<u>.261</u>	1103	,48	<u>.264</u>	1109
v2_31	,50	<u>.235</u>	1103	,80	<u>.143</u>	1109

Items	Testform A			Testform B		
	Schwierigkeitsindex (p)	Trennschärfe-index (r)	N	Schwierigkeitsindex (p)	Trennschärfe-index (r)	N
v2_32	,36	,325	1103	,73	,280	1109
v2_33	,41	,400	1103	,67	,117	1109
v2_34	,07	,180	1103	,44	-,108	1109
v2_35	,34	,013	1103	,22	,181	1109
v3_9	,52	,369	1103	,55	,107	1109
v3_10	,72	,176	1103	,86	,230	1109
v3_11	,51	,396	1103	,87	,232	1109
v3_12	,41	,533	1103	,68	,104	1109
v3_13	,80	,302	1103	,71	,288	1109
v3_14	,62	,357	1103	,65	,346	1109
v3_15	,52	,526	1103	,71	,294	1109
v3_16	,65	,346	1103	,74	,279	1109
v3_17	,56	,097	1103	,45	,380	1109
v3_18	,42	,365	1103	,71	,312	1109
v3_19	,72	,170	1103	,64	,241	1109
v3_20	,37	,424	1103	,42	,266	1109
v3_21	,78	,157	1103	,38	,414	1109
v3_22	,18	,119	1103	,55	,165	1109
v3_23	,28	,207	1103	,32	,246	1109
v3_24	,48	,268	1103	,39	,231	1109
v3_25	,20	,194	1103	,65	,276	1109
v3_26	,48	,241	1103	,58	,297	1109
v3_27	,31	,217	1103	,24	,302	1109
v3_28	,23	,142	1103	,09	,041	1109

Tabelle 5.3.1.26: Schwierigkeits- und Trennschärfeindizes für die Dimension Q des KFT 4-12+R in Klasse 7 für beide Testformen A und B

Items	Testform A			Testform B		
	Schwierigkeitsindex (p)	Trennschärfe-index (r)	N	Schwierigkeitsindex (p)	Trennschärfe-index (r)	N
q1_11	,55	,009	1102	,88	,248	1104
q1_12	,47	,196	1102	,78	,137	1104
q1_13	,59	,305	1102	,50	,112	1104
q1_14	,68	,161	1102	,66	,084	1104
q1_15	,47	,372	1102	,54	,272	1104
q1_16	,83	,170	1102	,51	,053	1104
q1_17	,49	,039	1102	,52	,218	1104
q1_18	,42	,223	1102	,81	,173	1104
q1_19	,17	,004	1102	,34	,218	1104
q1_20	,25	-,065	1102	,48	,191	1104
q1_21	,40	,195	1102	,66	,325	1104
q1_22	,17	,197	1102	,44	,190	1104
q1_23	,56	,385	1102	,48	,148	1104
q1_24	,50	,278	1102	,58	,279	1104
q1_25	,25	,269	1102	,40	,139	1104
q1_26	,44	,340	1102	,54	,269	1104

Items	Testform A			Testform B		
	Schwierigkeitsindex (p)	Trennschärfe-index (r)	N	Schwierigkeitsindex (p)	Trennschärfe-index (r)	N
q1_27	,37	<u>,103</u>	1102	,25	<u>,162</u>	1104
q1_28	,57	<u>,230</u>	1102	,26	<u>,171</u>	1104
q1_29	,30	,385	1102	,46	<u>,222</u>	1104
q1_30	,27	<u>,198</u>	1102	,53	<u>,112</u>	1104
q1_31	,52	<u>,227</u>	1102	,44	<u>,252</u>	1104
q1_32	,21	<u>,146</u>	1102	,33	<u>,194</u>	1104
q1_33	,25	<u>,154</u>	1102	,62	<u>,158</u>	1104
q1_34	,21	<u>-,036</u>	1102	<u>,18</u>	<u>,050</u>	1104
q1_35	,48	<u>,029</u>	1102	,33	<u>,218</u>	1104
q2_9	<u>,95</u>	<u>,195</u>	1102	,75	<u>,282</u>	1104
q2_10	<u>,96</u>	<u>,143</u>	1102	,77	,299	1104
q2_11	<u>,90</u>	<u>,121</u>	1102	<u>,86</u>	<u>,252</u>	1104
q2_12	<u>,94</u>	<u>,243</u>	1102	<u>,83</u>	<u>,209</u>	1104
q2_13	<u>,94</u>	<u>,188</u>	1102	<u>,92</u>	<u>,141</u>	1104
q2_14	,76	,411	1102	,78	,302	1104
q2_15	,75	,298	1102	,77	<u>,279</u>	1104
q2_16	,78	,368	1102	<u>,82</u>	<u>,217</u>	1104
q2_17	,78	<u>,248</u>	1102	,59	<u>,248</u>	1104
q2_18	<u>,86</u>	<u>,240</u>	1102	,61	,310	1104
q2_19	,75	<u>,178</u>	1102	,68	,396	1104
q2_20	<u>,85</u>	<u>,187</u>	1102	,57	,434	1104
q2_21	,77	,365	1102	,51	,311	1104
q2_22	<u>,84</u>	<u>,290</u>	1102	,60	,333	1104
q2_23	,76	,378	1102	,40	,327	1104
q2_24	,63	,305	1102	,45	,369	1104
q2_25	,67	,387	1102	,55	,407	1104
q2_26	,71	,313	1102	,55	,426	1104
q2_27	,55	,317	1102	,38	,359	1104
q2_28	,66	,436	1102	,33	,384	1104
q3_7	<u>,81</u>	<u>,261</u>	1102	,70	<u>,219</u>	1104
q3_8	,65	<u>,293</u>	1102	,56	,308	1104
q3_9	,63	<u>,266</u>	1102	,65	<u>,254</u>	1104
q3_10	,62	,323	1102	,39	<u>,219</u>	1104
q3_11	,43	<u>,230</u>	1102	,63	,316	1104
q3_12	,70	<u>,261</u>	1102	,38	<u>,099</u>	1104
q3_13	,49	,308	1102	,45	<u>,240</u>	1104
q3_14	,44	<u>,209</u>	1102	,20	<u>,158</u>	1104
q3_15	,25	<u>,161</u>	1102	,59	,334	1104
q3_16	,34	,305	1102	,37	<u>,290</u>	1104
q3_17	,40	<u>,276</u>	1102	<u>,18</u>	<u>,033</u>	1104
q3_18	,49	<u>,289</u>	1102	,26	<u>,209</u>	1104
q3_19	,22	<u>,197</u>	1102	,37	<u>,195</u>	1104
q3_20	,22	<u>,201</u>	1102	,27	<u>,205</u>	1104
q3_21	,26	<u>,227</u>	1102	<u>,19</u>	<u>,090</u>	1104

Tabelle 5.3.1.27: Schwierigkeits- und Trennschärfeindizes für die Dimension N des KFT 4-12+R in Klasse 7 für beide Testformen A und B

Items	Testform A			Testform B		
	Schwierigkeitsindex (p)	Trennschärfeindex (r)	N	Schwierigkeitsindex (p)	Trennschärfe-Index (r)	N
n1_11	,78	,432	1110	,81	,365	1111
n1_12	,70	,294	1110	,65	,253	1111
n1_13	,81	,383	1110	,78	,497	1111
n1_14	,86	,316	1110	,69	,471	1111
n1_15	,86	,534	1110	,51	,133	1111
n1_16	,85	,552	1110	,68	,485	1111
n1_17	,90	,325	1110	,75	,420	1111
n1_18	,87	,231	1110	,64	,404	1111
n1_19	,82	,459	1110	,42	,157	1111
n1_20	,57	,409	1110	,58	,454	1111
n1_21	,83	,491	1110	,65	,352	1111
n1_22	,84	,455	1110	,65	,453	1111
n1_23	,88	,439	1110	,79	,472	1111
n1_24	,73	,451	1110	,67	,502	1111
n1_25	,82	,445	1110	,54	,426	1111
n1_26	,73	,460	1110	,66	,514	1111
n1_27	,88	,517	1110	,67	,483	1111
n1_28	,75	,388	1110	,61	,348	1111
n1_29	,76	,495	1110	,83	,535	1111
n1_30	,84	,407	1110	,64	,361	1111
n1_31	,87	,525	1110	,45	,313	1111
n1_32	,66	,378	1110	,40	,209	1111
n1_33	,85	,550	1110	,59	,300	1111
n1_34	,67	,373	1110	,33	,147	1111
n1_35	,66	,207	1110	,41	,315	1111
n2_11	,89	,380	1110	,76	,362	1111
n2_12	,78	,382	1110	,90	,479	1111
n2_13	,85	,553	1110	,71	,425	1111
n2_14	,66	,143	1110	,86	,344	1111
n2_15	,60	,492	1110	,76	,520	1111
n2_16	,62	,526	1110	,85	,394	1111
n2_17	,72	,440	1110	,39	,186	1111
n2_18	,79	,473	1110	,78	,424	1111
n2_19	,66	,485	1110	,82	,428	1111
n2_20	,49	,400	1110	,81	,493	1111
n2_21	,72	,482	1110	,90	,487	1111
n2_22	,69	,464	1110	,83	,563	1111
n2_23	,68	,365	1110	,78	,436	1111
n2_24	,83	,459	1110	,82	,424	1111
n2_25	,69	,319	1110	,76	,433	1111
n2_26	,63	,388	1110	,73	,460	1111
n2_27	,65	,501	1110	,74	,463	1111
n2_28	,63	,431	1110	,75	,439	1111
n2_29	,79	,443	1110	,62	,369	1111
n2_30	,74	,315	1110	,80	,392	1111
n2_31	,56	,220	1110	,50	,358	1111

Items	Testform A			Testform B		
	Schwierigkeitsindex (p)	Trennschärfe-index (r)	N	Schwierigkeitsindex (p)	Trennschärfe-Index (r)	N
n2_32	,58	,338	1110	,55	,357	1111
n2_33	,57	,344	1110	,56	,439	1111
n2_34	,51	,216	1110	,74	,231	1111
n2_35	,56	,405	1110	,42	,302	1111
n3_7	,44	,363	1110	,87	,290	1111
n3_8	,85	,460	1110	,56	,303	1111
n3_9	,84	,423	1110	,87	,408	1111
n3_10	,88	,372	1110	,82	,260	1111
n3_11	,79	,362	1110	,76	,397	1111
n3_12	,63	,383	1110	,90	,251	1111
n3_13	,75	,421	1110	,76	,350	1111
n3_14	,54	,352	1110	,82	,331	1111
n3_15	,72	,389	1110	,57	,289	1111
n3_16	,52	,231	1110	,75	,378	1111
n3_17	,67	,466	1110	,71	,373	1111
n3_18	,64	,428	1110	,52	,181	1111
n3_19	,48	,223	1110	,69	,298	1111
n3_20	,46	,368	1110	,47	,312	1111
n3_21	,69	,354	1110	,52	,337	1111

Tabelle 5.3.1.28: Schwierigkeits- und Trennschärfeindizes für die Gesamtleistung GL des KFT 4-12+R in Klasse 7 für beide Testformen A und B

Items	Testform A			Testform B		
	Schwierigkeitsindex (p)	Trennschärfe-index (r)	N	Schwierigkeitsindex (p)	Trennschärfe-index (r)	N
v1_11	,88	,273	1095	,86	,224	1098
v1_12	,83	,225	1095	,83	,202	1098
v1_13	,63	,135	1095	,58	,238	1098
v1_14	,60	,293	1095	,81	,234	1098
v1_15	,09	,031	1095	,60	,142	1098
v1_16	,72	,397	1095	,54	,286	1098
v1_17	,64	,270	1095	,89	,263	1098
v1_18	,66	,321	1095	,76	,232	1098
v1_19	,60	,241	1095	,38	,195	1098
v1_20	,83	,193	1095	,48	,251	1098
v1_21	,48	,199	1095	,78	,227	1098
v1_22	,64	,329	1095	,84	,133	1098
v1_23	,41	,229	1095	,32	,276	1098
v1_24	,29	,232	1095	,63	,290	1098
v1_25	,22	,140	1095	,68	,301	1098
v1_26	,86	,211	1095	,16	,137	1098
v1_27	,25	,230	1095	,42	,270	1098
v1_28	,25	,112	1095	,18	,057	1098
v1_29	,35	,287	1095	,27	,240	1098
v1_30	,16	,115	1095	,17	,081	1098
v1_31	,34	,174	1095	,25	,122	1098

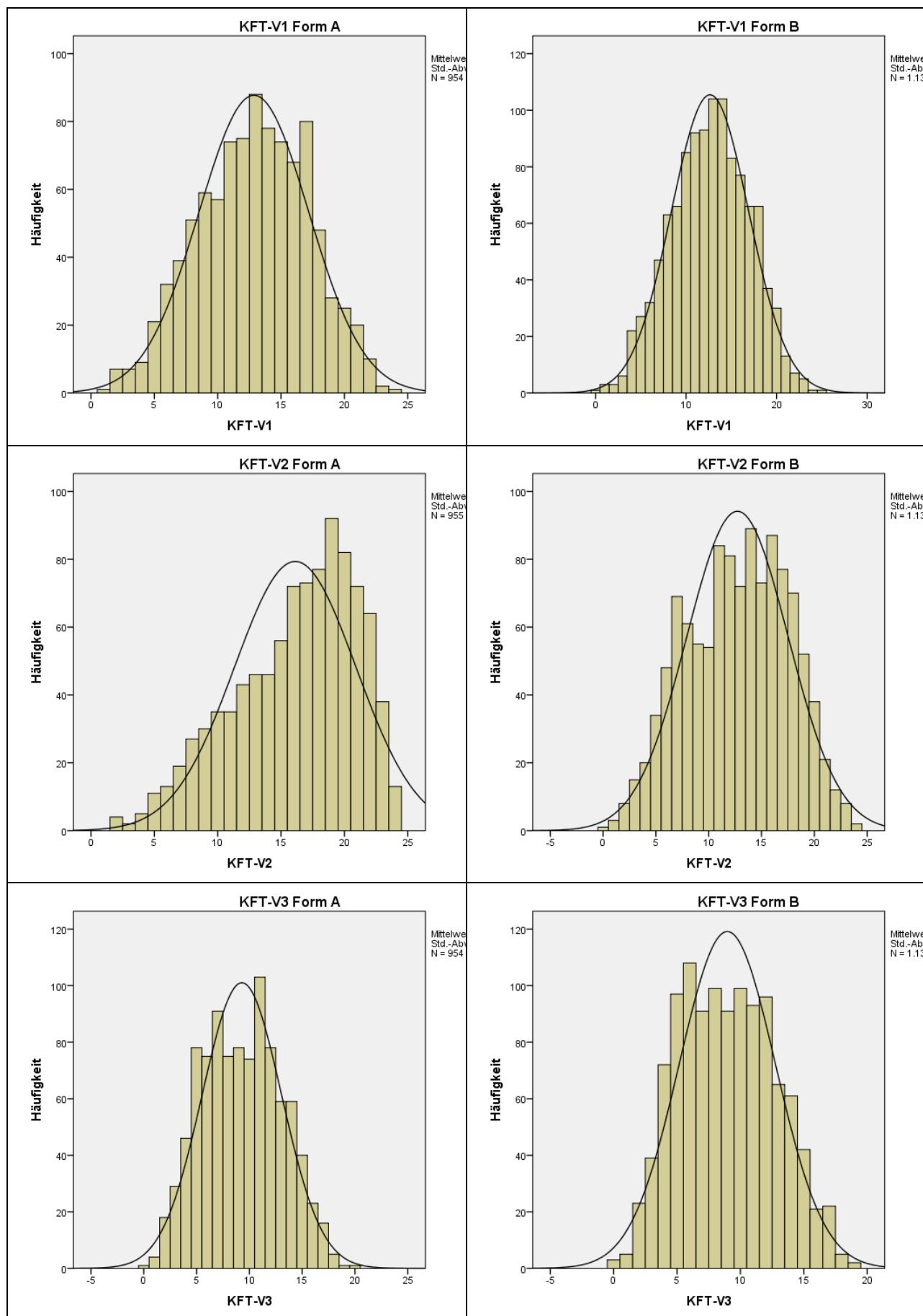
Items	Testform A			Testform B		
	Schwierigkeitsindex (p)	Trennschärfe-index (r)	N	Schwierigkeitsindex (p)	Trennschärfe-index (r)	N
v1_32	,10	-,002	1095	,38	,272	1098
v1_33	,63	,344	1095	,31	,109	1098
v1_34	,37	,191	1095	,60	,136	1098
v1_35	,72	,296	1095	,39	,160	1098
v2_11	,82	,253	1095	,78	,113	1098
v2_12	,74	,326	1095	,58	,269	1098
v2_13	,92	,127	1095	,78	,259	1098
v2_14	,78	,239	1095	,69	,191	1098
v2_15	,71	,194	1095	,76	,292	1098
v2_16	,65	,136	1095	,83	,289	1098
v2_17	,93	,181	1095	,51	,088	1098
v2_18	,85	,228	1095	,36	,107	1098
v2_19	,63	,280	1095	,28	-,002	1098
v2_20	,48	,208	1095	,55	,338	1098
v2_21	,79	,071	1095	,30	,163	1098
v2_22	,95	,143	1095	,79	,098	1098
v2_23	,82	,186	1095	,50	,169	1098
v2_24	,69	,069	1095	,84	,189	1098
v2_25	,85	,236	1095	,81	,256	1098
v2_26	,63	,220	1095	,38	,124	1098
v2_27	,42	,317	1095	,42	,245	1098
v2_28	,83	,215	1095	,67	,078	1098
v2_29	,61	,159	1095	,13	-,078	1098
v2_30	,25	,228	1095	,48	,222	1098
v2_31	,49	,198	1095	,80	,126	1098
v2_32	,36	,279	1095	,73	,271	1098
v2_33	,41	,307	1095	,68	,107	1098
v2_34	,07	,091	1095	,44	-,116	1098
v2_35	,34	-,014	1095	,22	,142	1098
v3_9	,52	,348	1095	,55	,076	1098
v3_10	,72	,193	1095	,86	,206	1098
v3_11	,51	,329	1095	,87	,259	1098
v3_12	,41	,434	1095	,69	,112	1098
v3_13	,80	,253	1095	,70	,319	1098
v3_14	,62	,347	1095	,65	,337	1098
v3_15	,52	,471	1095	,71	,229	1098
v3_16	,65	,296	1095	,74	,247	1098
v3_17	,56	,120	1095	,45	,327	1098
v3_18	,42	,341	1095	,71	,338	1098
v3_19	,72	,191	1095	,64	,262	1098
v3_20	,37	,373	1095	,42	,238	1098
v3_21	,78	,158	1095	,38	,384	1098
v3_22	,18	,090	1095	,54	,152	1098
v3_23	,28	,191	1095	,32	,206	1098
v3_24	,48	,202	1095	,38	,244	1098
v3_25	,20	,177	1095	,65	,335	1098
v3_26	,48	,247	1095	,58	,273	1098
v3_27	,31	,163	1095	,24	,245	1098
v3_28	,23	,122	1095	,09	,013	1098

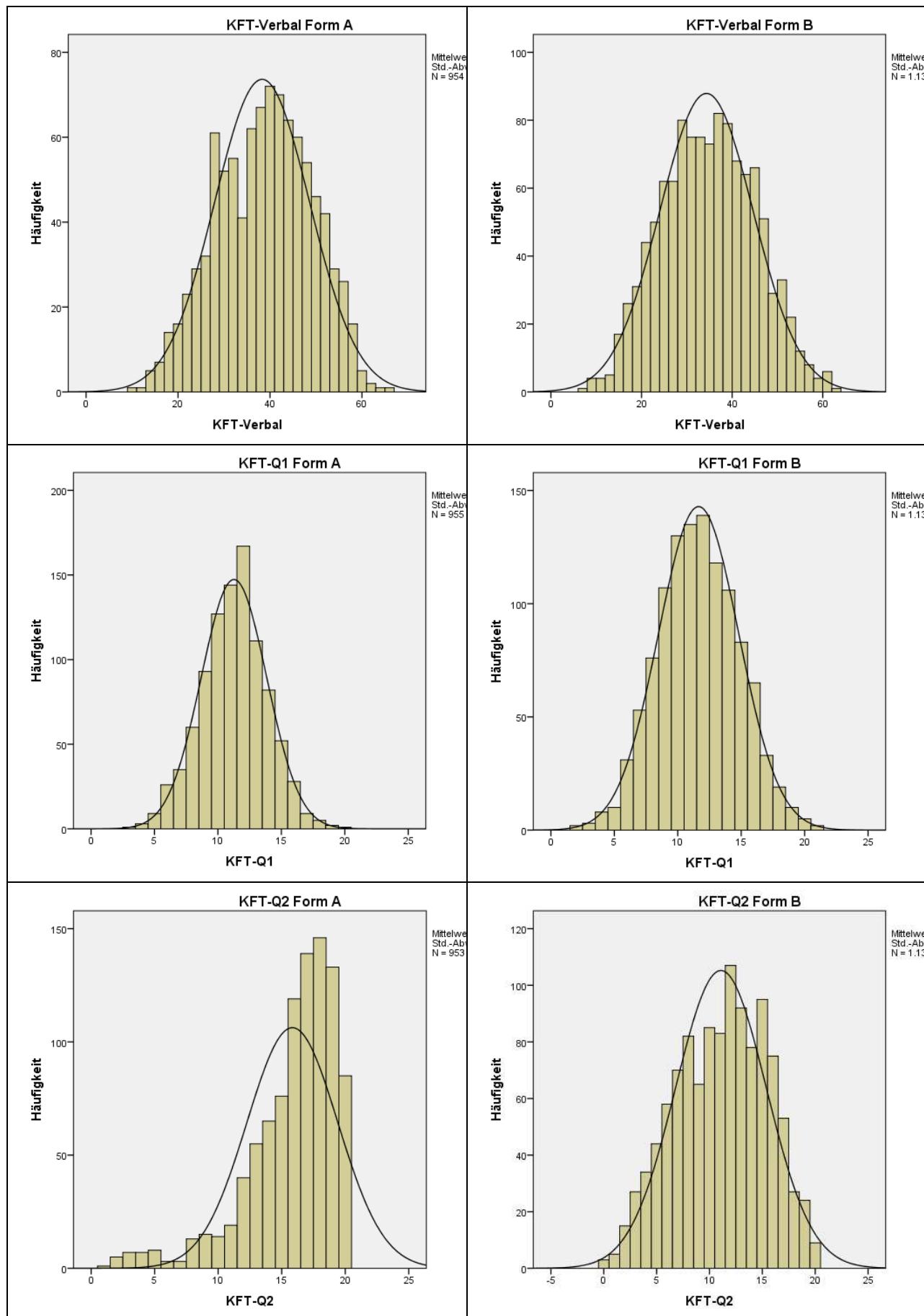
Items	Testform A			Testform B		
	Schwierigkeitsindex (p)	Trennschärfe-index (r)	N	Schwierigkeitsindex (p)	Trennschärfe-index (r)	N
q1_11	,56	<u>,019</u>	1095	<u>,88</u>	<u>,211</u>	1098
q1_12	,47	<u>,110</u>	1095	,78	<u>,141</u>	1098
q1_13	,59	,311	1095	,50	<u>,067</u>	1098
q1_14	,68	<u>,129</u>	1095	,66	<u>,091</u>	1098
q1_15	,47	,305	1095	,54	<u>,259</u>	1098
q1_16	<u>,83</u>	<u>,147</u>	1095	,50	<u>-,033</u>	1098
q1_17	,49	<u>-,009</u>	1095	,52	<u>,251</u>	1098
q1_18	,42	<u>,236</u>	1095	<u>,81</u>	<u>,171</u>	1098
q1_19	<u>,17</u>	<u>-,024</u>	1095	,34	<u>,198</u>	1098
q1_20	,25	<u>-,058</u>	1095	,48	<u>,133</u>	1098
q1_21	,40	<u>,170</u>	1095	,65	,328	1098
q1_22	<u>,17</u>	<u>,133</u>	1095	,45	<u>,175</u>	1098
q1_23	,56	,344	1095	,48	<u>,101</u>	1098
q1_24	,49	<u>,220</u>	1095	,58	<u>,242</u>	1098
q1_25	,25	<u>,245</u>	1095	,40	<u>,114</u>	1098
q1_26	,44	<u>,288</u>	1095	,54	<u>,277</u>	1098
q1_27	,37	<u>,071</u>	1095	,25	<u>,104</u>	1098
q1_28	,58	<u>,238</u>	1095	,26	<u>,120</u>	1098
q1_29	,30	,327	1095	,46	<u>,273</u>	1098
q1_30	,27	<u>,149</u>	1095	,53	<u>,070</u>	1098
q1_31	,52	<u>,181</u>	1095	,45	<u>,190</u>	1098
q1_32	,22	<u>,154</u>	1095	,33	<u>,128</u>	1098
q1_33	,26	<u>,147</u>	1095	,62	<u>,081</u>	1098
q1_34	,21	<u>-,069</u>	1095	<u>,18</u>	<u>,020</u>	1098
q1_35	,48	<u>,039</u>	1095	,33	<u>,191</u>	1098
q2_9	<u>,95</u>	<u>,180</u>	1095	,75	,337	1098
q2_10	<u>,96</u>	<u>,096</u>	1095	,77	,303	1098
q2_11	<u>,90</u>	<u>,110</u>	1095	<u>,86</u>	<u>,230</u>	1098
q2_12	<u>,94</u>	<u>,232</u>	1095	<u>,83</u>	<u>,268</u>	1098
q2_13	<u>,94</u>	<u>,140</u>	1095	<u>,92</u>	<u>,164</u>	1098
q2_14	,76	,380	1095	,78	<u>,281</u>	1098
q2_15	,75	<u>,275</u>	1095	,77	<u>,222</u>	1098
q2_16	,78	,353	1095	<u>,82</u>	<u>,220</u>	1098
q2_17	,78	<u>,276</u>	1095	,59	<u>,230</u>	1098
q2_18	<u>,86</u>	<u>,183</u>	1095	,61	<u>,257</u>	1098
q2_19	,76	<u>,217</u>	1095	,68	,350	1098
q2_20	<u>,85</u>	<u>,176</u>	1095	,57	,363	1098
q2_21	,77	,371	1095	,51	<u>,262</u>	1098
q2_22	<u>,84</u>	<u>,281</u>	1095	,60	<u>,290</u>	1098
q2_23	,76	,323	1095	,40	<u>,293</u>	1098
q2_24	,63	,295	1095	,45	,347	1098
q2_25	,66	,340	1095	,55	,341	1098
q2_26	,71	,322	1095	,55	,365	1098
q2_27	,55	<u>,266</u>	1095	,39	<u>,266</u>	1098
q2_28	,66	,405	1095	,33	<u>,278</u>	1098
q3_7	<u>,81</u>	<u>,203</u>	1095	,70	<u>,238</u>	1098
q3_8	,65	<u>,264</u>	1095	,56	<u>,251</u>	1098
q3_9	,63	<u>,225</u>	1095	,65	<u>,197</u>	1098
q3_10	,62	,301	1095	,39	<u>,184</u>	1098

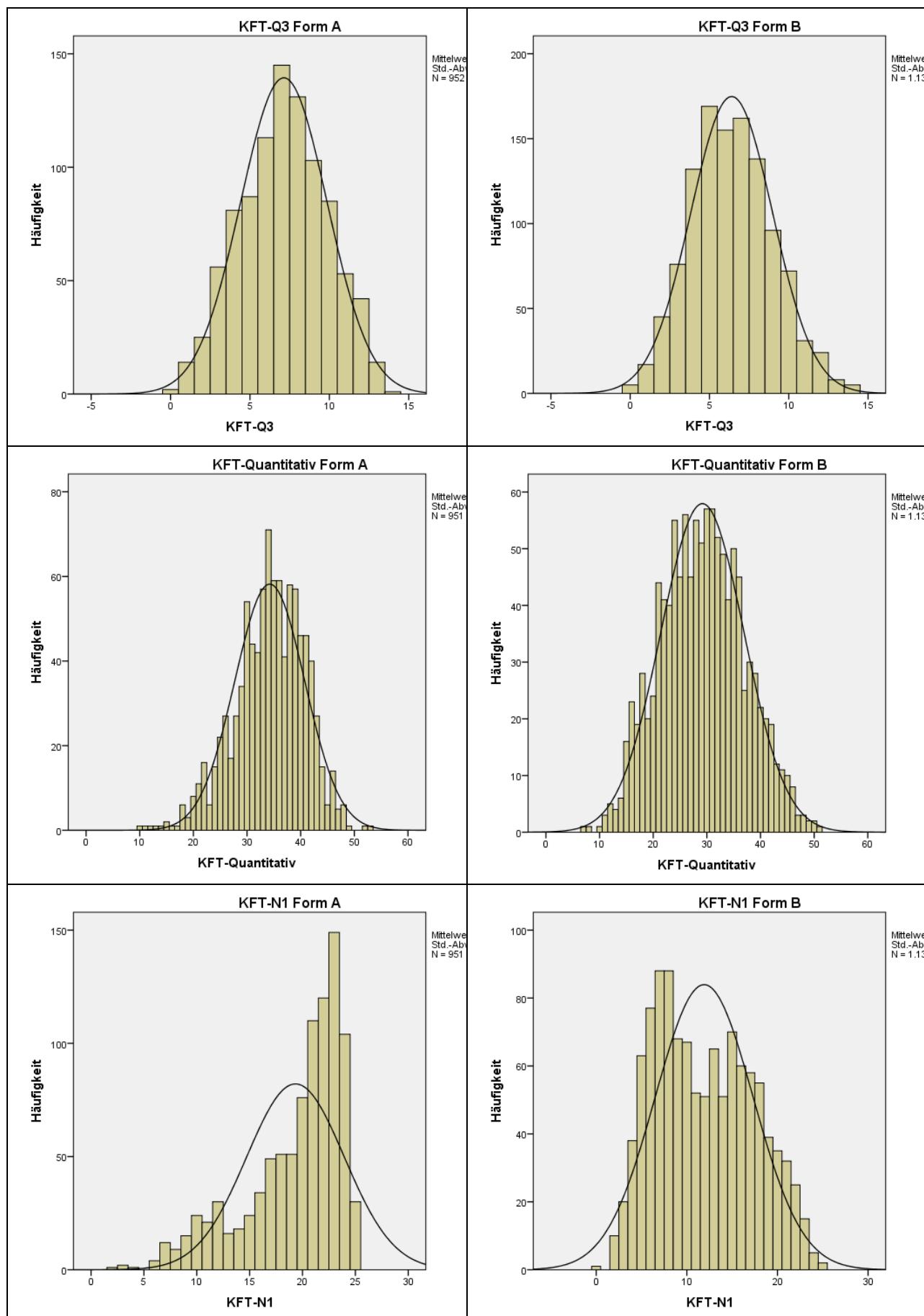
Items	Testform A			Testform B		
	Schwierigkeitsindex (p)	Trennschärfe-index (r)	N	Schwierigkeitsindex (p)	Trennschärfe-index (r)	N
q3_11	,43	<u>,158</u>	1095	,63	<u>,291</u>	1098
q3_12	,70	<u>,213</u>	1095	,37	<u>,083</u>	1098
q3_13	,49	<u>,237</u>	1095	,45	<u>,238</u>	1098
q3_14	,44	<u>,154</u>	1095	,20	<u>,160</u>	1098
q3_15	,25	<u>,137</u>	1095	,59	<u>,267</u>	1098
q3_16	,33	<u>,230</u>	1095	,37	<u>,209</u>	1098
q3_17	,40	<u>,199</u>	1095	<u>,18</u>	<u>-,001</u>	1098
q3_18	,49	<u>,181</u>	1095	,26	<u>,136</u>	1098
q3_19	,22	<u>,158</u>	1095	,37	<u>,137</u>	1098
q3_20	,22	<u>,151</u>	1095	,27	<u>,144</u>	1098
q3_21	,26	<u>,179</u>	1095	<u>,19</u>	<u>,080</u>	1098
n1_11	,78	,333	1095	<u>,81</u>	<u>,287</u>	1098
n1_12	,70	<u>,227</u>	1095	,65	<u>,189</u>	1098
n1_13	<u>,82</u>	<u>,278</u>	1095	,78	,414	1098
n1_14	<u>,86</u>	<u>,278</u>	1095	,69	,424	1098
n1_15	<u>,86</u>	,388	1095	,51	<u>,158</u>	1098
n1_16	<u>,85</u>	,404	1095	,68	,415	1098
n1_17	<u>,90</u>	<u>,288</u>	1095	,75	,377	1098
n1_18	<u>,87</u>	<u>,188</u>	1095	,64	,334	1098
n1_19	<u>,82</u>	,374	1095	,42	<u>,135</u>	1098
n1_20	,57	,346	1095	,58	,419	1098
n1_21	<u>,83</u>	,345	1095	,66	,348	1098
n1_22	<u>,84</u>	,326	1095	,65	,397	1098
n1_23	<u>,89</u>	,295	1095	,79	,408	1098
n1_24	,73	,354	1095	,67	,459	1098
n1_25	<u>,82</u>	,317	1095	,54	,414	1098
n1_26	,73	,383	1095	,66	,429	1098
n1_27	<u>,88</u>	,382	1095	,67	,437	1098
n1_28	,75	,369	1095	,61	,323	1098
n1_29	,76	,375	1095	<u>,83</u>	,462	1098
n1_30	<u>,84</u>	,366	1095	,64	,323	1098
n1_31	<u>,88</u>	,386	1095	,45	<u>,255</u>	1098
n1_32	,66	<u>,293</u>	1095	,40	<u>,214</u>	1098
n1_33	<u>,85</u>	,397	1095	,59	<u>,273</u>	1098
n1_34	,67	,299	1095	,33	<u>,144</u>	1098
n1_35	,66	<u>,216</u>	1095	,42	,328	1098
n2_11	<u>,89</u>	<u>,240</u>	1095	,76	,339	1098
n2_12	,78	,390	1095	<u>,90</u>	,386	1098
n2_13	<u>,85</u>	,420	1095	,71	,388	1098
n2_14	,66	<u>,181</u>	1095	<u>,86</u>	<u>,290</u>	1098
n2_15	,61	,455	1095	,76	,446	1098
n2_16	,62	,464	1095	<u>,85</u>	,337	1098
n2_17	,73	,381	1095	,39	<u>,147</u>	1098
n2_18	,79	,357	1095	,78	,361	1098
n2_19	,66	,369	1095	<u>,83</u>	,374	1098
n2_20	,49	,401	1095	<u>,81</u>	,406	1098
n2_21	,72	,396	1095	<u>,90</u>	,413	1098
n2_22	,69	,381	1095	<u>,83</u>	,465	1098
n2_23	,68	,365	1095	,78	,371	1098

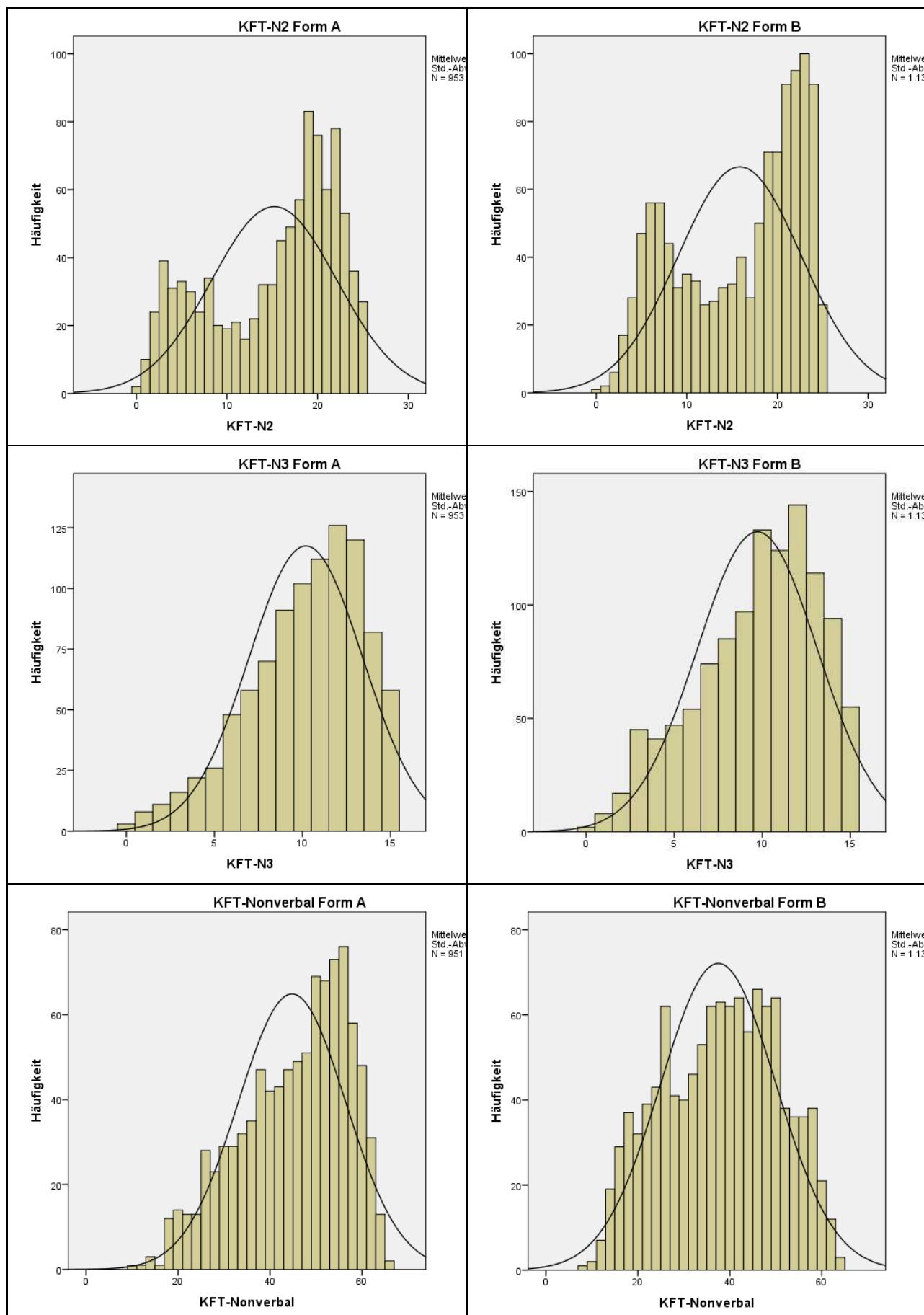
Items	Testform A			Testform B		
	Schwierigkeitsindex (p)	Trennschärfe-index (r)	N	Schwierigkeitsindex (p)	Trennschärfe-index (r)	N
n2_24	,83	,352	1095	,82	,365	1098
n2_25	,69	,327	1095	,76	,349	1098
n2_26	,64	,347	1095	,72	,411	1098
n2_27	,65	,418	1095	,74	,436	1098
n2_28	,63	,375	1095	,74	,383	1098
n2_29	,79	,367	1095	,62	,334	1098
n2_30	,74	,326	1095	,80	,345	1098
n2_31	,56	,266	1095	,50	,331	1098
n2_32	,58	,360	1095	,55	,306	1098
n2_33	,57	,307	1095	,56	,441	1098
n2_34	,51	,251	1095	,74	,182	1098
n2_35	,56	,363	1095	,42	,314	1098
n3_7	,44	,340	1095	,87	,281	1098
n3_8	,85	,354	1095	,56	,294	1098
n3_9	,84	,338	1095	,87	,380	1098
n3_10	,88	,264	1095	,82	,235	1098
n3_11	,79	,283	1095	,76	,380	1098
n3_12	,63	,348	1095	,90	,239	1098
n3_13	,76	,340	1095	,76	,334	1098
n3_14	,54	,344	1095	,82	,321	1098
n3_15	,72	,338	1095	,57	,311	1098
n3_16	,52	,177	1095	,75	,358	1098
n3_17	,67	,406	1095	,71	,378	1098
n3_18	,64	,403	1095	,52	,188	1098
n3_19	,47	,263	1095	,69	,276	1098
n3_20	,46	,371	1095	,47	,298	1098
n3_21	,69	,309	1095	,52	,315	1098

Die folgende Abbildung enthält die Verteilungsmerkmale der Skalen des KFT 4-12+R für Klassenstufe 5.









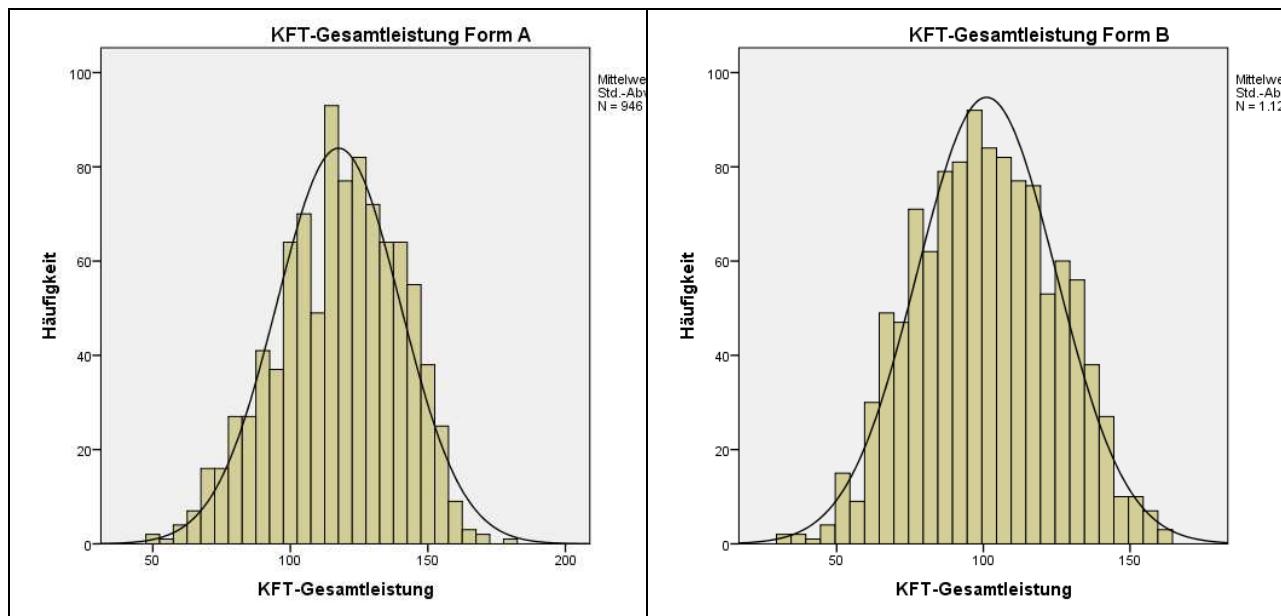


Abbildung 5.3.1.1: Verteilungsmerkmale der KFT-Subtests, Testdimensionen und der Gesamtleistung für Klasse 5 Form A und B

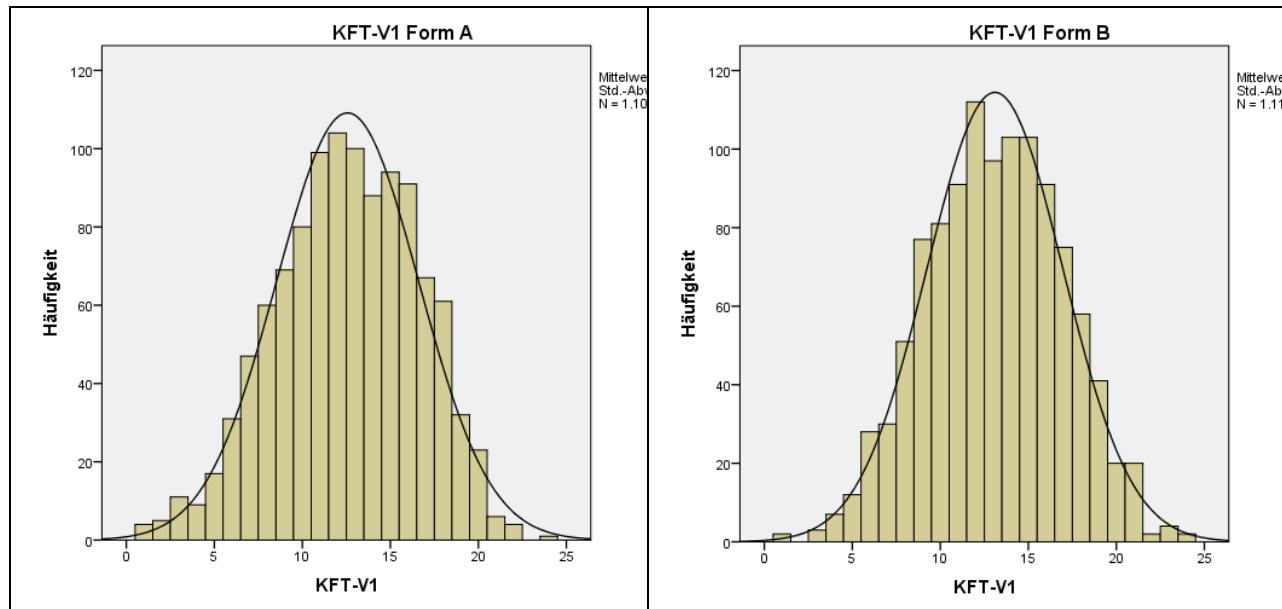
Tabelle 5.3.1.29: Verteilungsmerkmale der KFT-Subtests, Testdimensionen und der Gesamtleistung für Klasse 5 Form A

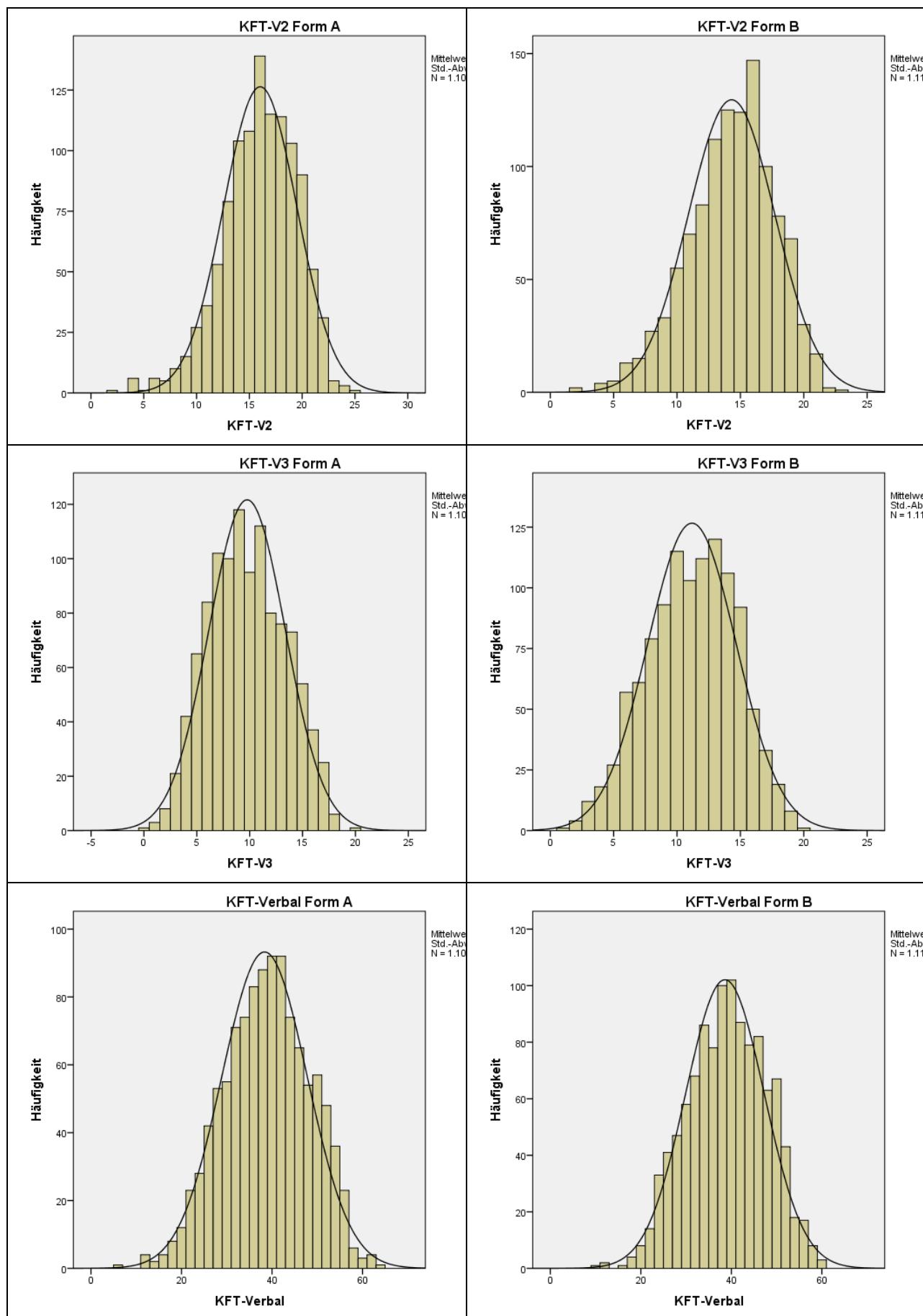
KFT	N5A	Mittel	StadAbw	Schiefe	Exzeß	Pro25	Pro50	Pro75
V1	954	12,86	4,337	-,132	-,485	10	13	16
V2	955	16,12	4,801	-,596	-,379	13	17	20
V3	954	9,29	3,766	,086	-,709	6	9	12
V	954	38,28	10,336	-,141	-,583	30	39	46
Q1	955	11,26	2,585	-,061	,080	10	11	13
Q2	953	15,84	3,577	-1,518	2,675	14	17	18
Q3	952	7,13	2,724	-,052	-,514	5	7	9
Q	951	34,23	6,519	-,456	,279	30	35	39
N1	951	19,33	4,622	-1,145	,588	17	21	23
N2	953	15,22	6,912	-,560	-,978	9	17	21
N3	953	10,19	3,235	-,686	,008	8	11	13
N	951	44,78	11,696	-,554	-,536	37	47	54
GL	946	117,51	22,473	-,290	-,384	102	119	135

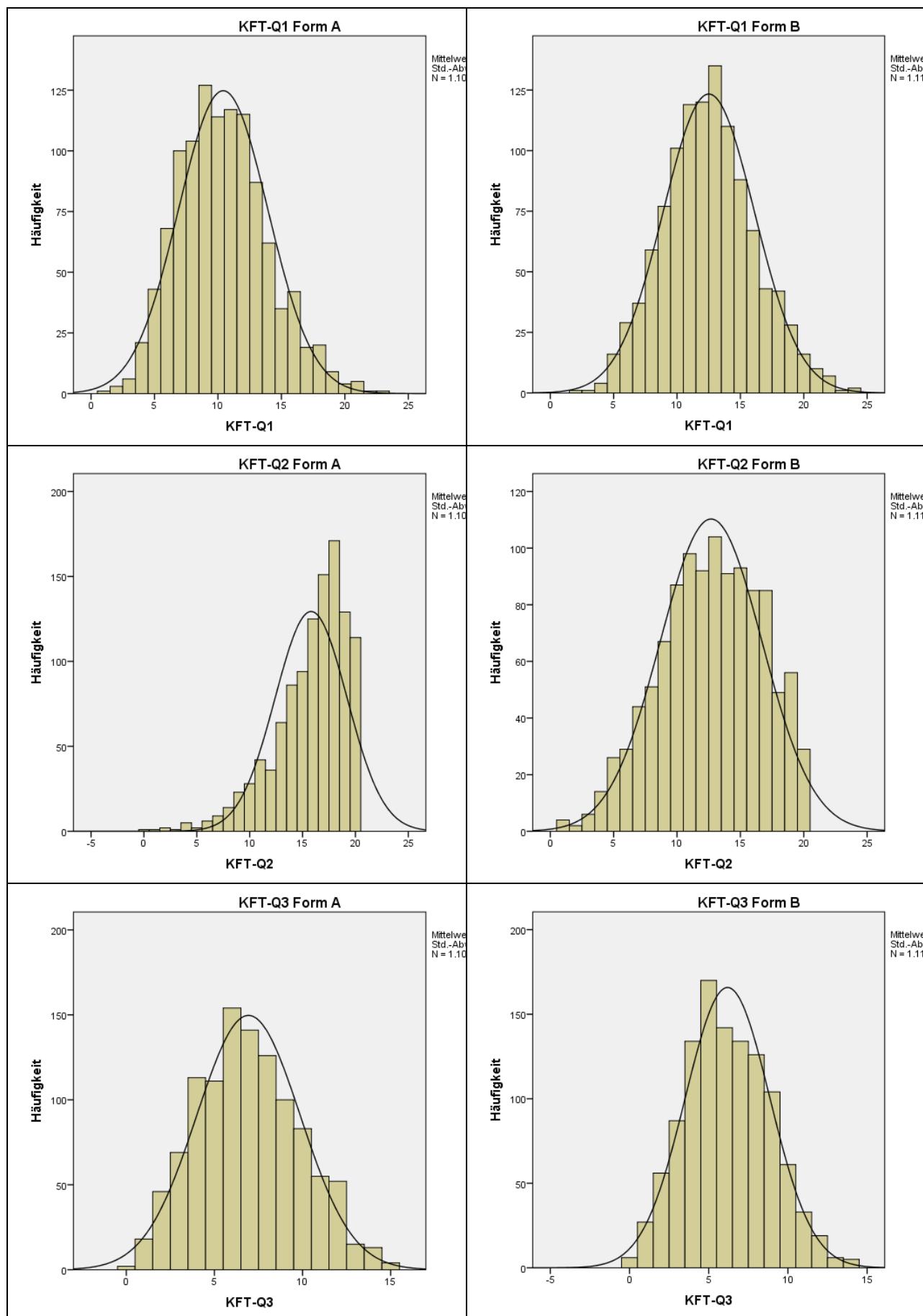
Tabelle 5.3.1.30: Verteilungsmerkmale der KFT-Subtests, Testdimensionen und der Gesamtleistung für Klasse 5 Form B

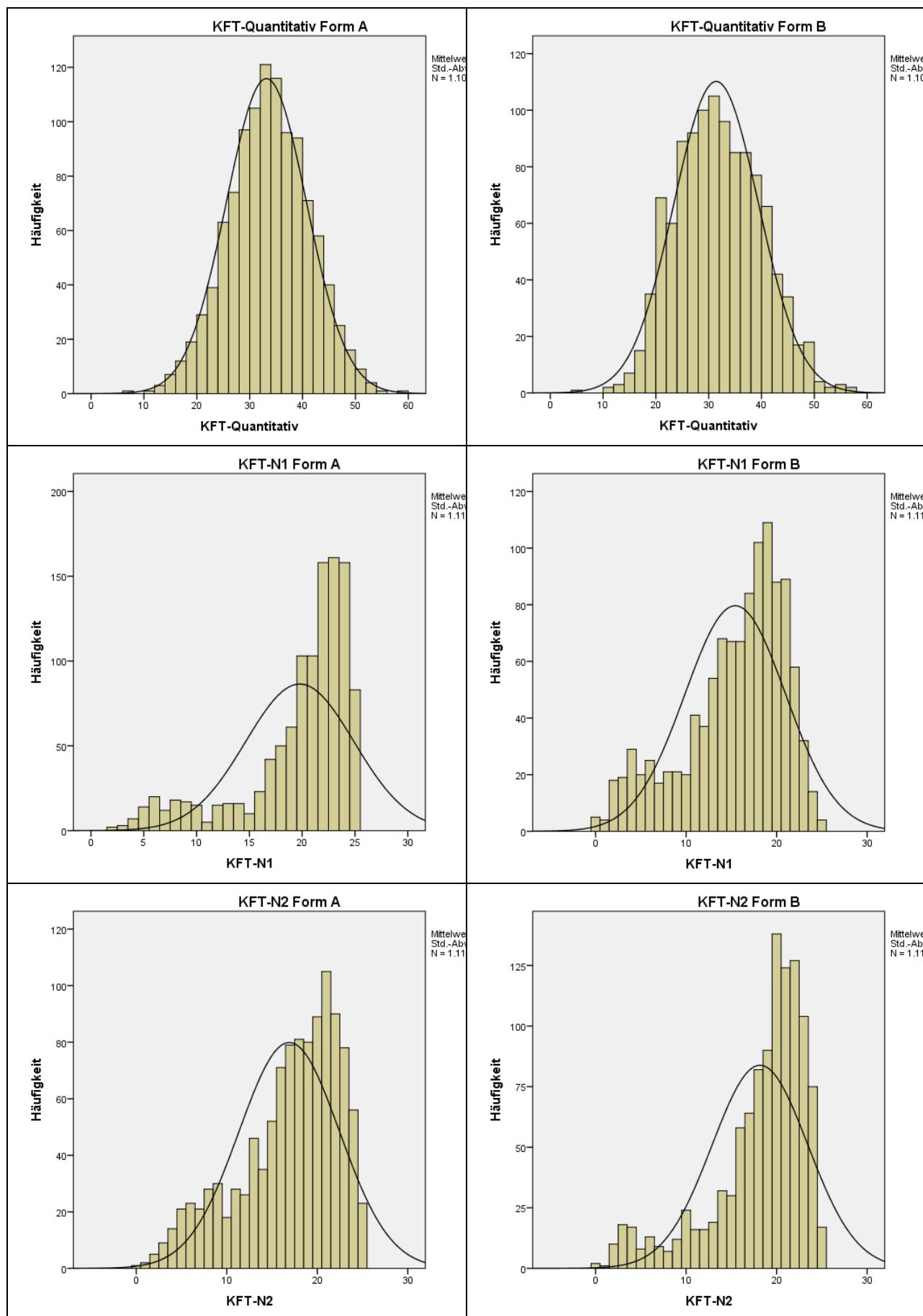
KFT	N5B	Mittel	StadAbw	Schiefe	Exzeß	Pro25	Pro50	Pro75
V1	1134	12,64	4,289	-,087	-,399	10	13	16
V2	1134	12,72	4,805	-,134	-,736	9	13	16
V3	1134	8,97	3,797	,161	-,706	6	9	12
V	1134	34,33	10,293	,041	-,494	27	34	42
Q1	1135	11,64	3,169	,038	-,182	9	12	14
Q2	1131	11,08	4,287	-,187	-,695	8	11	15
Q3	1135	6,40	2,590	,218	-,216	5	6	8
Q	1130	29,17	7,778	,075	-,405	24	29	35
N1	1135	11,91	5,394	,247	-,951	7	11	16
N2	1135	15,84	6,794	-,441	-1,223	9	18	22
N3	1134	9,71	3,423	-,536	-,487	7	10	12
N	1134	37,48	12,552	-,125	-,893	27	38	47
GL	1127	101,10	23,726	,013	-,512	84	101	118

Die folgende Abbildung enthält die Verteilungsmerkmale der Skalen des KFT 4-12+R für Klassenstufe 7.









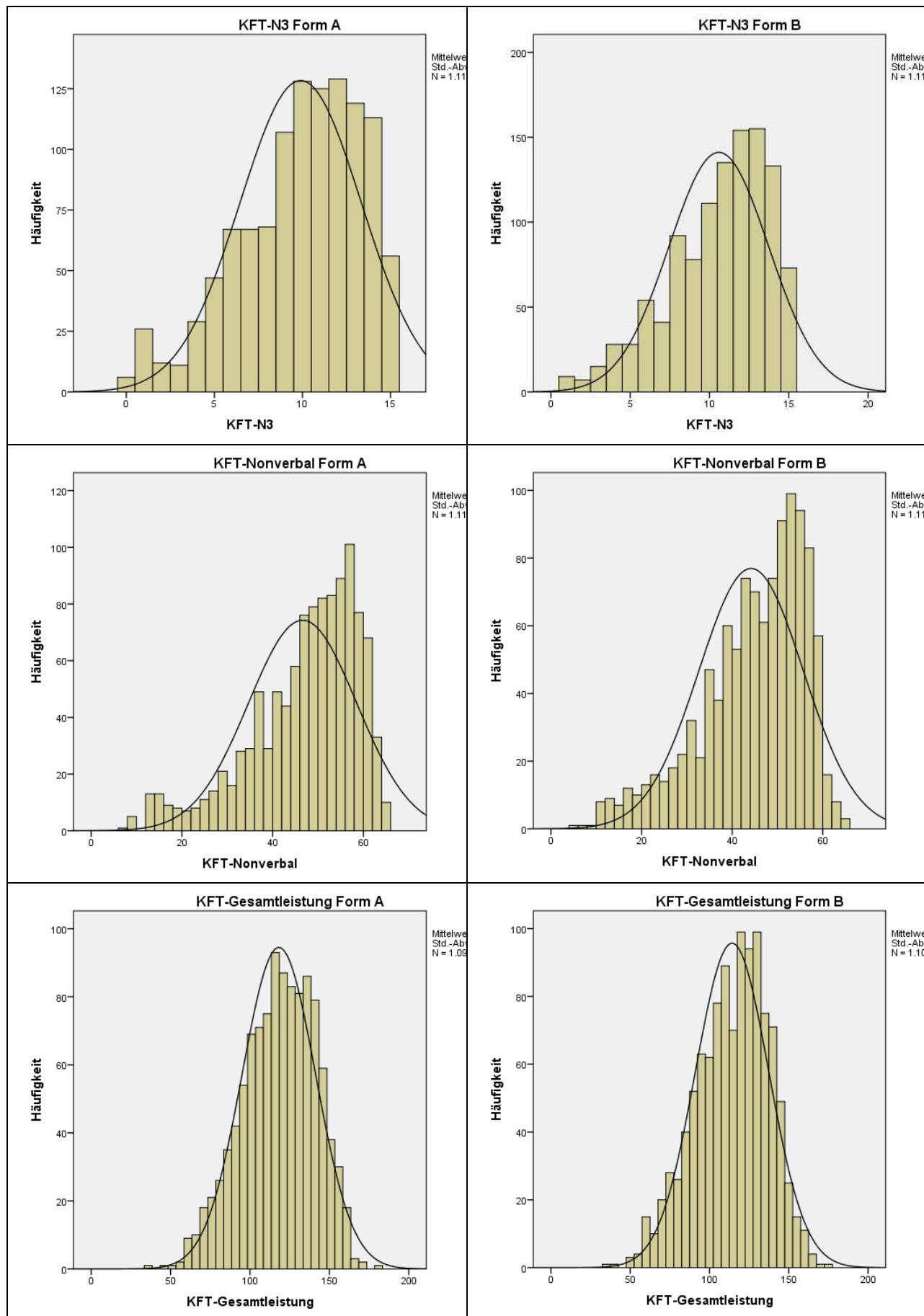


Abbildung 5.3.1.2: Verteilungsmerkmale der KFT-Subtests, Testdimensionen und der Gesamtleistung für Klasse 7 Form A und B.

Tabelle 5.3.1.31: Verteilungsmerkmale der KFT-Subtests, Testdimensionen und der Gesamtleistung für Klasse 7 Form A

KFT	N7A	Mittel	StadAbw	Schiefe	Exzeß	Pro25	Pro50	Pro75
V1	1103	12,57	4,032	-,207	-,336	10	13	16
V2	1103	16,02	3,482	-,549	,499	14	16	19
V3	1103	9,76	3,616	,089	-,653	7	10	12
V	1103	38,35	9,436	-,145	-,329	32	39	45
Q1	1104	10,41	3,529	,401	,041	8	10	13
Q2	1104	15,81	3,404	-1,163	1,508	14	17	18
Q3	1102	6,93	2,937	,217	-,441	5	7	9
Q	1102	33,16	7,588	-,061	-,084	28	33	38
N1	1112	19,77	5,132	-1,540	1,650	18	22	23
N2	1111	16,91	5,548	-,795	-,163	14	18	21
N3	1110	9,90	3,449	-,683	-,069	8	10	13
N	1110	46,60	11,926	-1,014	,630	40	49	56
GL	1095	118,18	23,109	-,325	-,300	102	120	136

Tabelle 5.3.1.32: Verteilungsmerkmale der KFT-Subtests, Testdimensionen und der Gesamtleistung für Klasse 7 Form B

KFT	N7B	Mittel	StadAbw	Schiefe	Exzeß	Pro25	Pro50	Pro75
V1	1110	13,10	3,869	-,055	-,326	10	13	16
V2	1111	14,31	3,421	-,439	,064	12	15	17
V3	1111	11,18	3,499	-,183	-,473	9	11	14
V	1110	38,58	8,674	-,138	-,421	32,75	39	45
Q1	1113	12,51	3,599	,194	-,074	10	12	15
Q2	1112	12,69	4,023	-,261	-,506	10	13	16
Q3	1110	6,19	2,671	,208	-,340	4	6	8
Q	1109	31,42	8,030	,163	-,326	25	31	37
N1	1113	15,43	5,574	-,821	-,084	13	17	20
N2	1113	18,17	5,297	-1,379	1,430	16	20	22
N3	1113	10,59	3,147	-,771	,041	9	11	13
N	1113	44,19	11,546	-,885	,301	38	46	53
GL	1106	114,24	23,053	-,372	-,222	99	117	132

5.3.2 Reliabilität

Um die Frage zu beantworten, wie hoch die Reliabilitätskoeffizienten der Skalen des KFT 4-12 + R sind, wurde die Reliabilität der Skalen für die beiden Testformen A und B mit drei unterschiedlichen Methoden (Testhalbierung „Split-half“, innere Konsistenz, Testwiederholung) bestimmt. Diese Methoden wurden getrennt auf die Datensätze der Klassenstufen 5 und 7

angewendet: Klassenstufe 5 (N5 = 2090) und Klassenstufe 7 (N7 = 2225). Die Berechnungen wurden für jede Testform getrennt durchgeführt.

Testhalbierungsmethode. Für diese Methode wurde der KFT-4-12 R an jeder der beiden Klassenstufen durchgeführt. Diese Hälften müssen so parallel wie möglich sein. Da die Aufgaben des KFT nach dem Schwierigkeitsgrad geordnet sind, wurden die Aufgaben in jeder Klassenstufe nach dem ***Odd-Even-Prinzip*** in zwei Hälften geteilt. Die erste Hälfte besteht somit aus den Aufgaben mit ungeraden Ordnungsnummern und die zweite Hälfte aus jenen mit geraden Nummern. Danach wurde die Korrelation zwischen den Testergebnissen der beiden Stichprobenhälften berechnet, welche den Split-Half-Reliabilitätskoeffizienten des Tests ausdrückt. Um die Reliabilität der Skalen zu ermitteln bzw. diese auf ganze Testlänge hochzurechnen, wurden die Formeln von ***Spearman-Brown*** und ***Guttman*** verwendet.

Innere Konsistenz. Bei dieser Methode werden die Skalen des KFT 4-12 +R nicht in zwei Teile wie bei der Testhalbierungsmethode, sondern in so viele Teile aufgeteilt wie sie über Items verfügt. Die Formel zur Bestimmung von ***Cronbachs Alpha*** wurde verwendet um die innere Konsistenz zu ermitteln. Die Konsistenzkoeffizienten der beiden Testformen in beiden Klassenstufen 5 und 7 sind in den Tabellen 5.3.2.1 bis 5.3.2.4 angegeben.

Tabelle 5.3.2.1: Konsistenzkoeffizienten der Subtests und der Gesamtleistung des KFT 4-12+ R von Klasse 5 für Testform A

	N5A	Cronbachs Alpha	Spearman Brown	Guttman
KFT-V1	954	,77	,78	,78
KFT-V2	955	,80	,80	,80
KFT-V3	954	,73	,76	,75
KFT-Q1	955	,42	,49	,49
KFT-Q2	953	,81	,83	,83
KFT-Q3	952	,65	,69	,69
KFT-N1	951	,85	,87	,87
KFT-N2	953	,92	,93	,93
KFT-N3	953	,77	,77	,77
KFT-V	954	,88	,88	,88
KFT-Q	951	,78	,82	,82
KFT-N	951	,92	,94	,94
KFT-GL	946	,93	,94	,94

Legende: V1: Wortschatz, V2: Wortklassifikationen, V3: Wortanalogien, Q1: Mengenvergleiche, Q2: Zahlenreihen, Q3: Gleichungenbilden, N1: Figurenklassifikation, N2: Figurenanalogien, N3: Faltaufgaben

Tabelle 5.3.2.2: Konsistenzkoeffizienten der Subtests und der Gesamtleistung des KFT 4-12+ R von Klasse 5 für Testform B

	N5B	Cronbachs Alpha	Spearman Brown	Guttman
KFT-V1	1134	,74	,76	,76
KFT-V2	1134	,79	,82	,81
KFT-V3	1134	,73	,73	,72
KFT-Q1	1133	,58	,60	,60
KFT-Q2	1130	,81	,83	,83
KFT-Q3	1135	,62	,65	,64
KFT-N1	1135	,84	,84	,84
KFT-N2	1135	,92	,92	,92
KFT-N3	1134	,79	,77	,76
KFT-V	1134	,87	,87	,86
KFT-Q	1127	,82	,86	,85
KFT-N	1134	,92	,93	,93
KFT-GL	1124	,93	,94	,94

Legende: V1: Wortschatz, V2: Wortklassifikationen, V3: Wortanalogien, Q1: Mengenvergleiche, Q2: Zahlenreihen, Q3: Gleichungenbilden, N1: Figurenklassifikation, N2: Figurenanalogien, N3: Faltaufgaben

Tabelle 5.3.2.3: Konsistenzkoeffizienten der Subtests und der Gesamtleistung des KFT 4-12+ R von Klasse 7 für Testform A

	N7A	Cronbachs Alpha	Spearman Brown	Guttman
KFT-V1	1103	,74	,73	,72
KFT-V2	1103	,66	,70	,70
KFT-V3	1103	,70	,72	,72
KFT-Q1	1104	,59	,62	,62
KFT-Q2	1104	,77	,81	,81
KFT-Q3	1102	,67	,71	,71
KFT-N1	1112	,88	,87	,87
KFT-N2	1111	,86	,87	,87
KFT-N3	1110	,79	,79	,78
KFT-V	1103	,86	,87	,87
KFT-Q	1102	,81	,83	,83
KFT-N	1110	,93	,93	,93
KFT-GL	1095	,93	,94	,94

Legende: V1: Wortschatz, V2: Wortklassifikationen, V3: Wortanalogien, Q1: Mengenvergleiche, Q2: Zahlenreihen, Q3: Gleichungenbilden, N1: Figurenklassifikation, N2: Figurenanalogien, N3: Faltaufgaben

Tabelle 5.3.2.4: Konsistenzkoeffizienten der Subtests und der Gesamtleistung des KFT 4-12+ R von Klasse 7 für Testform B

	N7B	Cronbachs Alpha	Spearman Brown	Guttman
KFT-V1	1110	,70	,69	,69
KFT-V2	1111	,59	,62	,62
KFT-V3	1110	,70	,72	,72
KFT-Q1	1109	,60	,60	,60
KFT-Q2	1111	,79	,81	,80
KFT-Q3	1110	,59	,61	,60
KFT-N1	1111	,86	,85	,85
KFT-N2	1113	,87	,87	,87
KFT-N3	1113	,77	,76	,75
KFT-V	1109	,82	,84	,84
KFT-Q	1104	,81	,82	,82
KFT-N	1111	,92	,92	,92
KFT-GL	1098	,93	,94	,93

Legende: V1: Wortschatz, V2: Wortklassifikationen, V3: Wortanalogien, Q1: Mengenvergleiche, Q2: Zahlenreihen, Q3: Gleichungenbilden, N1: Figurenklassifikation, N2: Figurenanalogien, N3: Faltaufgaben

Wenn man die Konsistenzkoeffizienten der einzelnen Skalen des KFT4-12+R für beide Testformen, die in Tabellen 5.3.2.1 und 5.3.2.2 stehen, mit den Konsistenzkoeffizienten der einzelnen Skalen des KFT 4-12+R für beide Testformen von Klassenstufe 5, die im Manual (Heller & Perleth, 2000, S. 20) stehen, vergleicht, findet man Folgendes: Einerseits stimmen die beiden Konsistenzkoeffizienten in Hinblick auf die Subtests KFT-Q1 und KFT-Q3 darin überein, dass die beiden Subtests am niedrigsten ausfallen, jedoch zeigen sich für diese Subtests in der hier berichteten Untersuchung noch niedrigere Werte. Dazu kommt, dass die Subtests KFT-V1, KFT-V2, und KFT-V3 in der soeben dargestellten Analyse niedriger ausfallen. Andererseits sind die Werte für die übrigen Subtests, Testdimensionen und die Gesamtleistung sehr ähnlich zu jenen im Manual.

Die Konsistenzkoeffizienten der Subtests KFT-Q1, KFT-Q3 und KFT-V2 von Klassenstufe 7, die in Tabelle 5.3.2.3 und 5.3.2.4 dargestellt sind, sind hier die niedrigsten für beiden Testformen. Das bestätigt teilweise die Angaben im Manual, nämlich dass die Untertests KFT-Q1 und KFT-Q3 am niedrigsten in der Testform A ausfallen. In Bezug auf die Testform B fallen die Untertests KFT-V2 und KFT-Q3 am niedrigsten aus. Insgesamt stimmen die Konsistenzkoeffizienten der Skalen des KFT-4-12+R (Untertests, Testdimensionen und Gesamtleistung) von Klassenstufe 7 für die beiden Testformen gut mit den Angaben im Manual überein.

Darüber hinaus ergibt sich für den KFT 4-12+R über alle Subtests und über beide Testformen hinweg in beiden Klassenstufen ein Median von $Rel = .77$, über alle Testdimensionen und über beide Testformen hinweg in beiden Klassenstufen ein Median von $Rel = .86$, über die beiden Gesamtleistungen und über beide Testformen hinweg in beiden Klassen ein Median von $Rel = .93$. Diese Mediane zeigen, dass die Reliabilitätskoeffizienten des KFT 4-12+R als sehr gut (Gesamtleistung und Testdimensionen) bis ausreichend (Untertests) zu bezeichnen sind, d.h. die Homogenität des KFT 4-12+R kann als ausreichend für die Subtests und als sehr gut für die Testteile sowie die Gesamtleistung bezeichnet werden.

Retest-Reliabilität. Für diese Methode wurde ausgenutzt, dass der KFT 4-12+ R zweimal in einem zeitlichen Abstand von zwei Jahren mit der gleichen Stichprobe durchgeführt wurde. Da außerdem in der Regel die Testform in jedem Jahr gewechselt wird, bearbeiten die Schüler der 5. und 7 Klassenstufe normalerweise dieselbe Testform. Daher wurde für die Bestimmung des Reliabilitätskoeffizient die Korrelation zwischen den Ergebnissen der Durchführungen in Klassenstufe 5 und 7 berechnet.

Tabelle 5.3.2.5: Stabilität der Subtests, Testteile und der Gesamtleistung des KFT 4-12+R

	V1	V2	V3	V	Q1	Q2	Q3	Q	N1	N2	N3	N	Gl
Test-form-A	,64	,48	,51	,73	,27	,44	,42	,55	,38	,42	,43	,53	,70
Test-form-B	,54	,45	,44	,67	,29	,45	,22	,50	,43	,47	,51	,61	,69

Legende: V1: Wortschatz, V2: Wortklassifikationen, V3: Wortanalogien, Q1: Mengenvergleiche, Q2: Zahlenreihen, Q3: Gleichungenbilden, N1: Figurenklassifikation, N2: Figurenanalogien, N3: Faltaufgaben. Testform A: Schüler, die ausschließlich Testform A bei den beiden Testdurchführungen bearbeitet haben (NA = 771); Testform B: Schüler, die nur Testform B zu den beiden Messzeitpunkten bearbeitet haben (NB = 764).

Die Retest- Reliabilitätskoeffizienten der Skalen von KFT 4-12+R, die in Tabelle 5.3.2.5. zusammengestellt sind, zeigen, dass die Ergebnisse der beiden Messzeitpunkte niedrig miteinander korrelieren, besonders jene der Subtests KFT-Q1 und KFT-Q3, die ähnlich niedrig wie die innere Konsistenz ausfallen. Das bedeutet, dass diese Befunde keine ausreichende Messgenauigkeit zumindest dieser Skalen des KFT 4-12+R bestätigen können, obwohl die meisten der Konsistenzkoeffizienten von den Skalen (Subtests KFT-N1, KFT-N2, Testdimensionen und Gesamt-

leistung) des KFT 4-12+R auf mittlere bis hohe Reliabilität hinweisen. Dies kann auch dadurch erklärt werden, dass das Zeitintervall von zwei Jahren zwischen den beiden Anwendung ziemlich groß ist. Das stimmt mit den Angaben im theoretischen Teil darin überein, dass die Reliabilität unterschätzt werden kann, wenn der Zeitabstand zwischen den beiden Messzeitpunkten groß ist. Daneben können differenzielle Fördereffekte an der Schule für unterschiedliche Schülergruppen (z.B. aufgrund der Sprachförderung) oder auch Testleitereffekte verantwortlich sein.

5.3.3 Aspekte der Validität

Um der Frage nachzugehen, ob der KFT 4-12+R misst, was er messen soll, wurden Validitätsaspekte anhand verschiedener Methoden (faktorielle Validität, Kriteriumsvalidität) berechnet.

5.3.3.1 Faktorielle Validität

Es wurden Faktorenanalysen zum KFT 4-12+ R für die Testformen A und B nach der Hauptkomponentenanalyse mit anschließender Varimax-Rotation und Kaisernormalisierung für die beiden Klassenstufen 5 und 7 ($N_5 = 2090$, $N_7 = 2225$) durchgeführt. Jede Klassenstufe wurde bezüglich der Testform in zwei Teil-Stichproben ($N_{5A} = 955$, $N_{5B} = 1135$, $N_{7A} = 1112$, $N_{7B} = 1113$) aufgeteilt. Die Faktorenanalyse ergab für die erste Teilstichprobe der Fünftklässler ($N_5 = 2090$) für die beiden Testformen drei Faktoren (vgl. auch die Abbildungen 5.3.3.1 und 5.3.3.2 zu den Scree-Testen), die zusammen 60,24 % (Testform A) bzw. 62,54 % (Testform B) der Gesamtvarianz aufklären.

Der **erste Faktor (1)**, der 36,98 % (Testform A) und 36,44 % (Testform B) der Gesamtvarianz aufklärt, kann als Faktor des nonverbalen-figuralen Denkens (Reasoning mit Space nach Thurstone) bezeichnet werden. Die Subtests KFT-Figurenklassifikation, KFT-Figurenanalogien und KFT-Faltauflagen laden am höchsten auf diesem Faktor. Zusätzlich laden die Untertests KFT-Zahlenreihen und KFT-Wortanalogien auch hoch auf diesem Faktor. Die Ladungen von KFT-Figurenklassifikation und Figurenanalogien sind jedoch höher als jene der anderen Subtests.

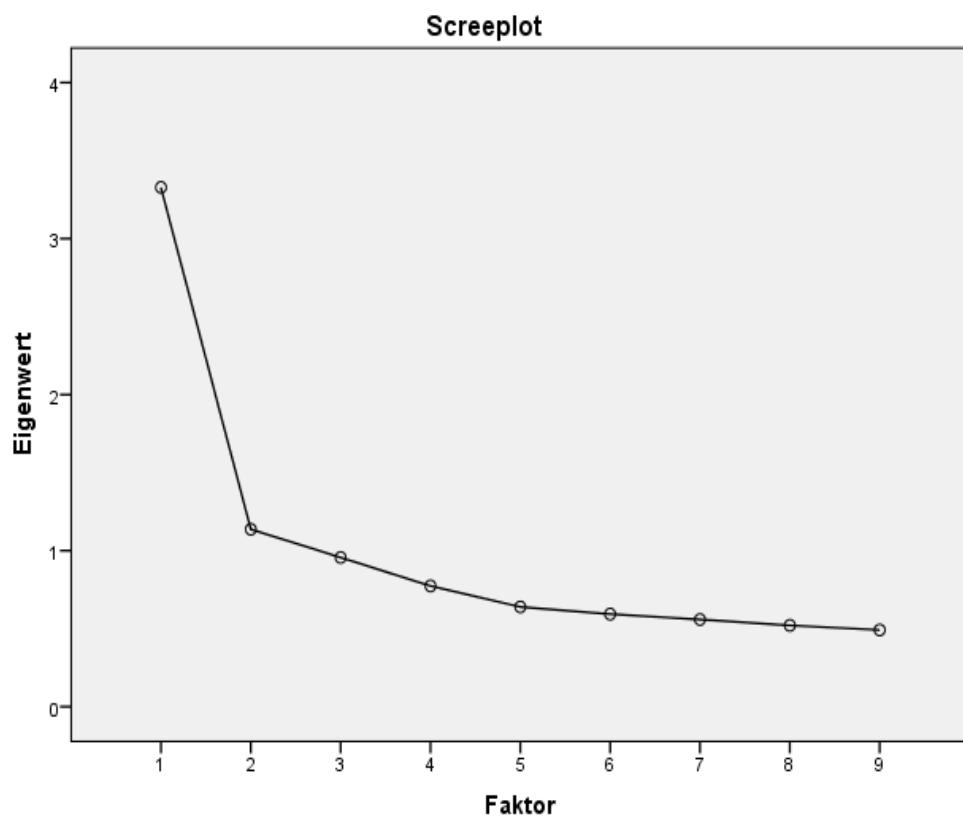


Abbildung 5.3.3.1: Scree-Plot zur Faktorenanalyse von Klassenstufe 5, Testform A

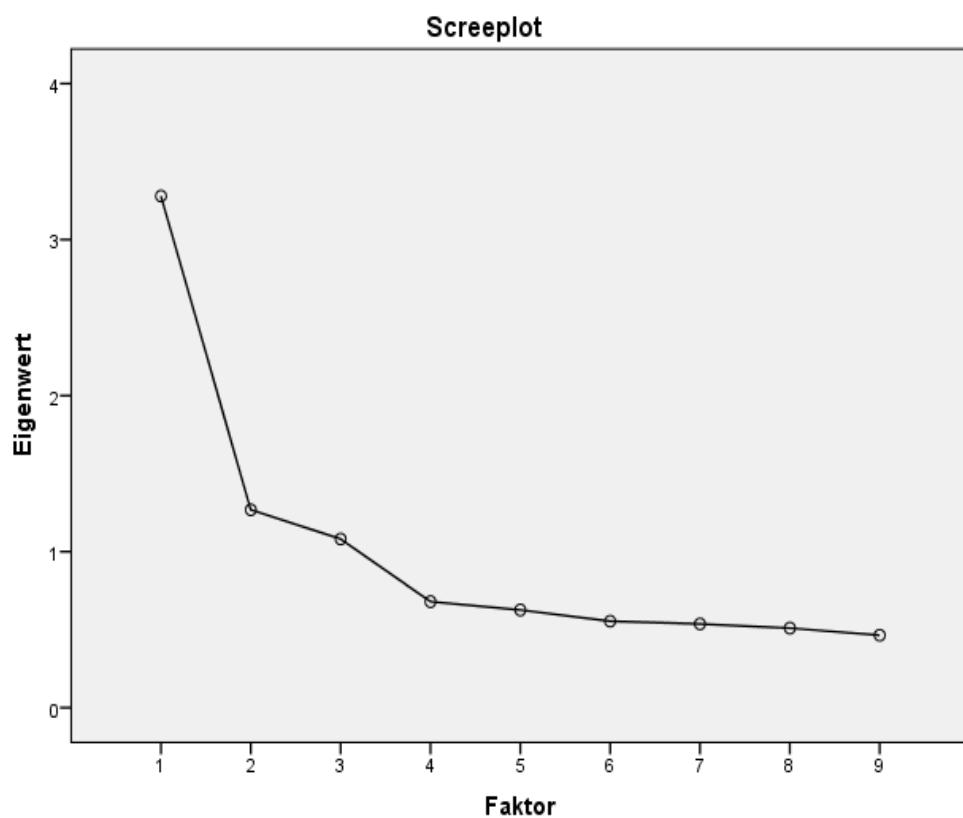


Abbildung 5.3.3.2: Scree-Plot zur Faktorenanalyse von Klassenstufe 5, Testform B

Tabelle 5.3.3.1: Faktorenstruktur des KFT4-12+ R in Klassestufe 5 Testform A

	N5A	Faktor (1)	Faktor (2)	Faktor (3)	h^2
KFT-Wortschatztest	954	,072	,807	,179	,688
KFT-Wortklassifikationen	955	,172	,771	,070	,629
KFT-Wortanalogien	953	,319	,709	,091	,613
KFT-Mengenvergleiche	955	,059	,083	,759	,586
KFT-Zahlenreihen	953	,482	,157	,558	,568
KFT- Gleichungenbilden	952	,137	,117	,721	,552
KFT-Figurenklassifikation	951	,812	,155	-,009	,684
KFT-Figurenanalogien	952	,739	,183	,167	,608
KFT-Faltaufgaben	951	,616	,183	,286	,494
Eigenwerte		3,328	1,137	,956	
Prozent der unrotierten Lösung		36,982%	12,633%	10,625%	
Prozentualer Anteil der kumulierten Varianz an Gesamtvarianz			60,240%		

Legende: Extraktionsmethode: Hauptkomponente; Rotationsmethode: Varimax mit Kaiser – Normalisierung.

Tabelle 5.3.3.2: Faktorenstruktur des KFT4-12+ R in Klassestufe 5 Testform B

	N5B	Faktor (1)	Faktor (2)	Faktor (3)	h^2
KFT-Wortschatztest	1134	,032	,790	,261	,693
KFT-Wortklassifikationen	1133	,149	,798	,040	,660
KFT-Wortanalogien	1131	,402	,666	,017	,605
KFT-Mengenvergleiche	1135	,064	,082	,790	,635
KFT-Zahlenreihen	1128	,398	,084	,654	,593
KFT- Gleichungenbilden	1130	,081	,125	,765	,607
KFT-Figurenklassifikation	1134	,795	,042	,107	,645
KFT-Figurenanalogien	1134	,800	,191	,117	,689
KFT-Faltaufgaben	1132	,612	,289	,208	,502
Eigenwerte		3,279	1,268	1,081	
Prozent der unrotierten Lösung		36,437%	14,093%	12,014%	
Prozentualer Anteil der kumulierten Varianz an Gesamtvarianz			62,544%		

Legende: Extraktionsmethode: Hauptkomponente; Rotationsmethode: Varimax mit Kaiser – Normalisierung.

Der **zweite Faktor (2)**, der 12,63 % (Testform A) bzw. 14,09 % (Testform B) der Gesamtvarianz erklärt, ist der Faktor des verbalen Denkens (Verbal Comprehension nach Thurstone). Die Untertests KFT-Wortschatz, KFT-Wortklassifikation und KFT-Wortanalogien zeigen die höchsten Ladungen auf diesem Faktor. Die Untertests KFT-Wortschatztest und KFT-Wortklassifikationen weisen höhere Ladung als der Subtest Wortanalogien auf.

Der **dritte Faktor (3)**, der 10,63 % (Testform A) bzw. 12,01 % (Testform B) der Gesamtvarianz besitzt, ist als Faktor des quantitativen Denkens (Number mit Reasoning nach Thurstone) zu charakterisieren. Die Subtests KFT-Mengenvergleiche, KFT-Zahlenreihen und KFT-Gleichungenbilden haben hohe Ladung auf diesem Faktor. Der Subtest KFT-Mengenvergleiche hat dabei eine höhere Ladung als die anderen beiden quantitativen Subtests. Die Ergebnisse der Faktorenanalysen der Untertests des KFT4-12+ R für die Klassenstufe 5 sind in den Tabellen 5.3.3.1 und 5.3.3.2 zusammengestellt.

Die Faktorenanalyse bezüglich der zweiten Teilstichprobe ($N7=2225$; Extraktionsmethode: Hauptkomponentenanalyse; Rotationsmethode: Varimax mit Kaiser-Normalisierung) für die beiden Testformen A ($N7A = 1112$) und B ($N7B = 1113$) erbrachte ebenso drei Faktoren, die zusammen 67,79 % der Gesamtvarianz für die Testform A und 64,48 % der Gesamtvarianz für die Testform B aufklären (vgl. auch die Abbildungen 5.3.3.3 und 5.3.3.4 zu den Screeplots). Während der erste Faktor 43,39 % (Testform A) bzw. 41,50 % (Testform B) der Gesamtvarianz erklärt, verfügen die zwei übrigen Faktoren über 13,41 % (Testform A) bzw. 12,39 % (Testform B) und 10,98 % (Testform A) bzw. 10,59 % (Testform B) der Gesamtvarianz.

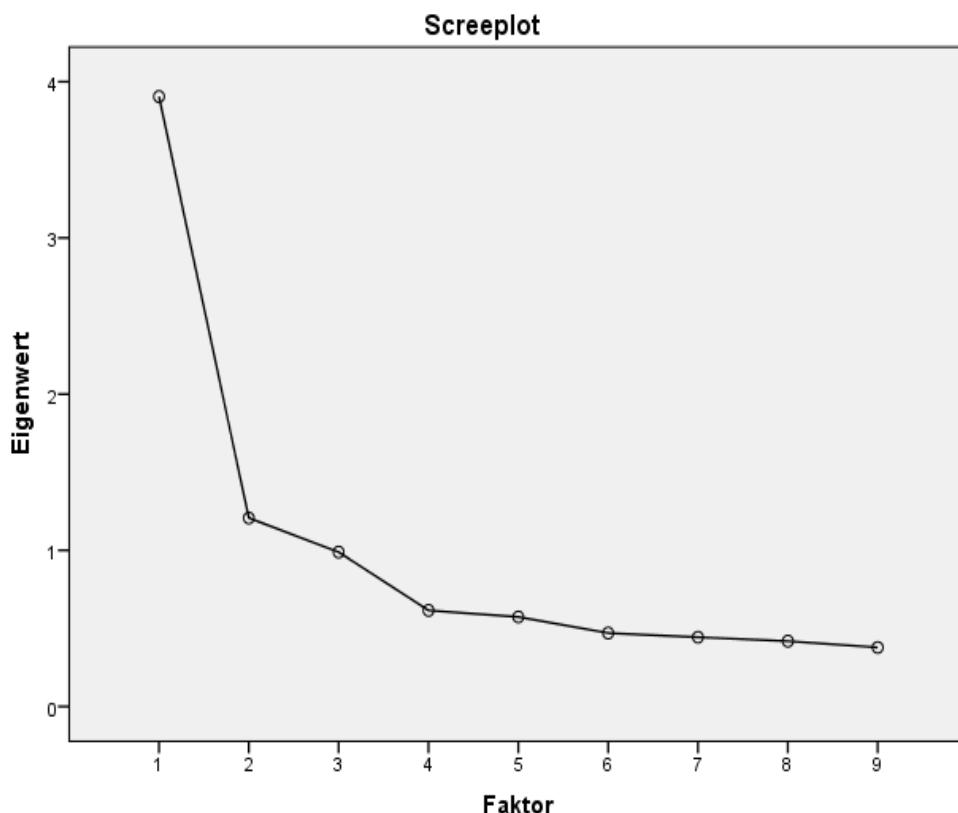


Abbildung 5.3.3.3: Scree-Plot zur Faktorenanalyse von Klassenstufe 7, Testform A

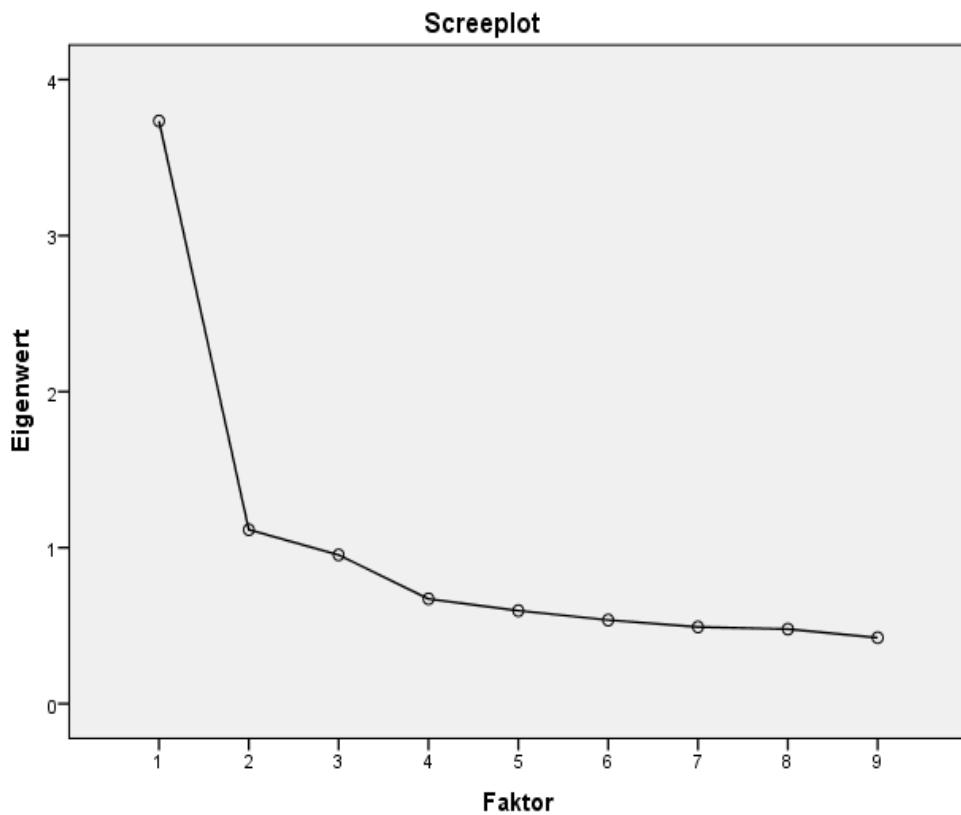


Abbildung 5.3.3.4: Scree-Plot zur Faktorenanalyse von Klassenstufe 7, Testform B

Tabelle 5.3.3.3: Faktorenanalyse des KFT4-12+ R in Klassenstufe 7 Testform A

	N7A	Faktor(1)	Faktor(2)	Faktor(3)	h^2
KFT-Wortschatztest	1103	,095	.823	,203	,728
KFT-Wortklassifikationen	1103	,167	.817	,132	,714
KFT-Wortanalogien	1102	,269	.778	,183	,711
KFT-Mengenvergleiche	1104	,142	,204	.747	,620
KFT-Zahlenreihen	1104	.370	,276	.585	,555
KFT- Gleichungenbilden	1100	,117	,093	.813	,683
KFT-Figurenklassifikation	1112	.836	,115	,111	,725
KFT-Figurenanalogien	1111	.768	,253	,209	,697
KFT-Faltaufgaben	1104	.782	,147	,185	,667
Eigenwerte		3,905	1,207	,989	
Prozent der unrotierten Lösung		43,391%	13,412%	10,984%	
Prozentualer Anteil der kumulierten Varianz an Gesamtvarianz				67,786%	

Legende: Extraktionsmethode: Hauptkomponente; Rotationsmethode: Varimax mit Kaiser – Normalisierung.

Tabelle 5.3.3.4: Faktorenstruktur des KFT4-12+ R in Klassestufe 7 Testform B

	N7B	Faktor(1)	Faktor(2)	Faktor(3)	h²
KFT-Wortschatztest	1110	,145	,786	,202	,679
KFT-Wortklassifikationen	1111	,145	,824	,050	,703
KFT-Wortanalogien	1111	,364	,639	,200	,580
KFT-Mengenvergleiche	1113	,109	,215	,758	,633
KFT-Zahlenreihen	1112	,502	,060	,630	,653
KFT- Gleichungenbilden	1104	,131	,112	,767	,619
KFT-Figurenklassifikation	1108	,727	,212	,157	,598
KFT-Figurenanalogien	1113	,799	,238	,100	,704
KFT-Faltaufgaben	1113	,755	,138	,210	,633
Eigenwerte		3,735	1,115	,953	
Prozent der unrotierten Lösung		41,500%	12,388%	10,592%	
Prozentualer Anteil der kumulierten Varianz an Gesamtvarianz			64,480%		

Legende: Extraktionsmethode: Hauptkomponente; Rotationsmethode: Varimax mit Kaiser – Normalisierung.

Der **erste Faktor (1)** ist als Faktor des nonverbalen-figuralen Denkens für beide Testformen A und B zu interpretieren. Die nonverbalen Subtests (KFT-Figurenklassifikation, KFT-Figurenanalogien und KFT-Faltaufgaben) haben die höchsten Ladungen auf dem ersten extrahierten Faktor. Außerdem laden auch der KFT-Subtest KFT-Zahlenreihen in mittlerer Höhe auf diesem Faktor für Testform A und die KFT-Untertests KFT-Wortanalogien und KFT-Zahlenreihen für Testform B. Die Ladung von KFT-Figurenklassifikation ist höher als jene der anderen Subtests für Testform A und KFT-Figurenanalogien für Testform B.

Der **zweite Faktor (2)** ist als Faktor des verbalen-sprachlichen Denkens für beide Testformen A und B zu erkennen. Auf diesem Faktor laden die verbalen Subtests (KFT-Wortschatztest, KFT-Wortklassifikationen und KFT-Wortanalogien) am höchsten. Die Ladungen von den Subtests KFT-Wortschatztest und KFT-Wortklassifikationen sind die höchsten für Testform A und KFT-Wortklassifikationen für Testform B.

Der **dritte Faktor (3)** ist für die beiden Testformen als Faktor des mathematischen und quantitativen Denkens zu bezeichnen, auf dem die Subtests KFT-Mengenvergleiche, KFT-Zahlenreihen, und KFT-Gleichungenbilden am höchsten laden. Die Ladung vom Subtest KFT-Gleichungenbilden ist höher als jene der anderen beiden quantitativen Subtests für Testform A und die Untertests KFT-Mengenvergleiche sowie KFT-Gleichungenbilden für Testform B. Die Ergebnisse der

Faktorenanalyse zu den Subtests des KFT4-12+ R für die beiden Testformen A und B von Klassenstufe 7 sind in den Tabellen 5.3.3.3 und 5.3.3.4 zusammengefasst.

Aus diesen faktorenanalytischen Ergebnissen kann man schließen, dass die Faktorenstruktur des KFT 4-12+R für die beiden Testformen bezüglich der Klassenstufen 5 und 7 jeweils drei Faktoren beinhaltet. Der verbale-sprachliche Faktor wird durch den V-Teil, der quantitative Faktor durch den Q-Teil, und der nonverbale Faktor durch den N-Teil repräsentiert. Das bestätigt die Gültigkeit des KFT 4-12+ R zur Messung differentieller intellektueller Fähigkeiten von Schülern/innen der fünften und siebten Klassenstufe.

5.3.3.2 Kriteriumsvalidität

Für die **Übereinstimmungsvalidität** wurden Korrelationen zwischen den Ergebnissen der Schüler/innen in den KFT 4-12+ R Untertests, Dimensionen (KFT-V, KFT-Q, KFT-N) und der Gesamtleistung in Klassenstufe 7 mit den Schulnoten (Zensuren in Deutsch, Englisch und Mathematik), die als Validitätskriterium dienten und dem Zwischenzeugnis der 7. Klasse entnommen werden konnten, ermittelt (Tabelle 5.3.3.5). Darüber hinaus konnten auch Aspekte der **prognostischen Validität** durch die Korrelation zwischen den Ergebnissen der Schüler/innen in den KFT 4-12 + R-Subtests, Testteilen und der Gesamtleistung, die in Klasse 5 erhoben wurde, ebenfalls mit den Schulzensuren im Halbjahreszeugnis der 7. Klasse bestimmt werden (Tabelle 5.3.3.6).

Wie Tabelle 5.3.3.5 zeigt, korreliert der Subtest KFT-V1 (Wortschatztest) mit der Deutschzensur signifikant auf dem Niveau von 0,01 sowie mit den Englisch- und Mathematiknoten auf dem Niveau von 0,05. Die Korrelation mit der Deutschzensur ist am höchsten mit $r = .286$. Die Untertests KFT-V2 (Wortklassifikationen) und KFT-V3 (Wortanalogien) korrelieren mit allen Schulnoten auf dem Niveau von 0,01. Die Korrelationen der beiden Untertests KFT-V2 sowie KFT-V3 mit der Deutschnote fallen am höchsten aus: $r = .492$ und $r = .413$. Die Korrelationen des Untertests KFT-Q1 (Mengenvergleiche) mit allen Schulnoten sind signifikant auf dem 0,01- und 0,05-Niveau. Die Korrelation zwischen dem Untertest KFT-Q1 und der Mathematikzensur ist am höchsten mit $r = .328$.

Tabelle 5.3.3.5: Korrelationsergebnisse der Skalen des KFT-4-12+R der Siebklässeler mit den Schulnoten aus Klasse 7 (Übereinstimmungsvalidität)

	N	Deutsch	Englisch	Mathematik
KFT-V1	146	<u>-,286**</u>	<u>-,199*</u>	<u>-,186*</u>
KFT-V2	146	<u>-,492**</u>	<u>-,261**</u>	<u>-,248**</u>
KFT-V3	146	<u>-,413**</u>	<u>-,251**</u>	<u>-,353**</u>
KFT-Q1	146	<u>-,252**</u>	<u>-,192*</u>	<u>-,328**</u>
KFT-Q2	146	-,117	-,024	<u>-,265**</u>
KFT-Q3	145	-,083	,040	-,119
KFT-N1	154	-,023	-,031	-,096
KFT-N2	154	-,061	-,137	-,144
KFT-N3	148	,006	-,130	-,108
KFT-V	146	<u>-,458**</u>	<u>-,281**</u>	<u>-,302**</u>
KFT-Q	145	<u>-,242**</u>	-,117	<u>-,343**</u>
KFT-N	154	-,021	-,125	-,136
KFT-GL	141	<u>-,276**</u>	<u>-,193*</u>	<u>-,303**</u>

Legende: KFT-V1: Wortschatz, KFT-V2: Wortklassifikationen, KFT-V3: Wortanalogien, KFT-Q1: Mengenvergleiche, KFT-Q2: Zahlenreihen, KFT-Q3: Gleichungenbilden, KFT-N1: Figurenklassifikation, KFT-N2: Figurenanalogien, KFT-N3: Faltaufgaben.
 ** = Die Korrelation ist auf dem Niveau von 0,01 (2-seitig) signifikant. * = Die Korrelation ist auf dem Niveau von 0,05 (2-seitig) signifikant.

Subtest KFT-Q2 (Zahlenreihen) zeigt einen signifikanten Zusammenhang mit der Mathematikzensur mit $r = ,265$ auf dem Niveau von 0,01 und keine signifikanten Zusammenhänge mit den anderen Schulnoten. Untertest KFT-Q3 (Gleichungenbilden) zeigt keine signifikanten Zusammenhänge mit den erfassten Zensuren. Die Skalen des N Teils (KFT-N1-Figurenklassifikation, KFT-N2-Figurenanalogien und KFT-N3-Faltaufgaben) korrelieren nicht signifikant mit den Schulnoten.

In Bezug auf die Testteile sind die Korrelationen des Verbalteil (KFT-V) signifikant auf dem 0,01- Niveau mit allen Noten. Die höchste Korrelation zeigt sich mit der Deutschnote mit $r = .458$. Dazu kommt, dass der quantitative Teil (KFT-Q) mit den Deutsch- und Mathematiknoten auf dem Niveau von 0,01 signifikant korreliert und keinen signifikanten Zusammenhang zur Englischnote zeigt. Am höchsten ist die Korrelation mit der Mathematikzensur mit $r = ,343$. Bezuglich des Nonverbalen Teils (KFT-N) zeigen sich keine signifikanten Korrelationen. Das gilt nicht für die Gesamtleistung (KFT-GL), hier sind die Korrelationen auf dem Niveau von 0,01 und 0,05 signifikant. Der höchste Zusammenhang zeigt sich mit der Mathematiknote: $r = ,303$.

Insgesamt kann man sagen, dass, obwohl die Stichprobe klein ist, sich viele signifikante Übereinstimmungskoeffizienten zeigen. Deutsch korreliert mit den Untertests des V Teils, KFT-Q1, den Testteilen KFT-V und KFT-Q sowie mit der Gesamtleistung KFT-GL. Am höchsten ist die Korrelation mit dem Subtest KFT-V2 ($r = .492$). Außerdem korreliert auch Englisch mit allen Untertests des V Teils, KFT-Q1, dem Testteil KFT-V und der Gesamtleistung. Die höchste Korrelation zeigt sich mit dem KFT-V Test, der als Faktor des Sprachgebundenen Denken bezeichnet wird. Darüber hinaus korreliert die Mathematiknote mit allen Untertests des V Teils, KFT-Q1, KFT-Q2, den Testdimensionen KFT-V, KFT-Q und der Gesamtleistung KFT-GL. Die Korrelation ist mit KFT-V3 und KFT-Q am höchsten mit $r = .353$ und $r = .343$, letzterer gilt als Faktor des zahlengebundenen Denkens. Keine der Korrelationen hinsichtlich N Teil und Subtest KFT-Q3 mit Zensuren sind signifikant.

Tabelle 5.3.3.6: Korrelationsergebnisse der Skalen des KFT-4-12+R von Klasse 5 mit den Schulnoten von Klasse 7 (prognostische Validität)

	N	Deutsch	Englisch	Mathe
KFT-V1	133	<u>-.379**</u>	<u>-.350**</u>	<u>-.323**</u>
KFT-V2	133	<u>-.337**</u>	<u>-.263**</u>	<u>-.313**</u>
KFT-V3	133	<u>-.343**</u>	<u>-.181*</u>	<u>-.196*</u>
KFT-Q1	133	-.139	-.074	-.132
KFT-Q2	132	-.155	-.050	-.175*
KFT-Q3	132	-.118	-.098	-.155
KFT-N1	131	<u>-.177*</u>	-.038	<u>-.238**</u>
KFT-N2	132	<u>-.176*</u>	-.131	<u>-.255**</u>
KFT-N3	132	<u>-.210*</u>	<u>-.261*</u>	<u>-.279**</u>
KFT-V	133	<u>-.430**</u>	<u>-.329**</u>	<u>-.329**</u>
KFT-Q	132	<u>-.177*</u>	-.108	<u>-.211*</u>
KFT-N	131	<u>-.240**</u>	-.112	<u>-.330**</u>
KFT-GL	130	<u>-.276**</u>	<u>-.193*</u>	<u>-.303**</u>

Legende: KFT-V1: Wortschatz, KFT-V2: Wortklassifikationen, KFT-V3: Wortanalogien, KFT-Q1: Mengenvergleiche, KFT-Q2: Zahlenreihen, KFT-Q3: Gleichungenbilden, KFT-N1: Figurenklassifikation, KFT-N2: Figurenanalogien, KFT-N3: Faltaufgaben. ** = Die Korrelation ist auf dem Niveau von 0,01 (2-seitig) signifikant. * = Die Korrelation ist auf dem Niveau von 0,05 (2-seitig) signifikant.

Tabelle 5.3.3.6 enthält die Zusammenhänge zwischen den KFT-Ergebnissen, die in Klasse 5 erhoben wurden und den Noten im Zwischenzeugnis der Klasse 7, informiert also über Aspekte der **prognostischen Validität** des KFT 4-12+ R. Es bestehen signifikante Zusammenhänge auf

dem Niveau von 0,01 und 0,05 zwischen den Skalen des V-Teils (V1-Wortschatz, V2-Wortklassifikationen und V3-Wortanalogien) mit allen Zensuren. Am höchsten sind die Korrelationen mit der Deutschzensur: $r = .379$, $r = .337$ und $r = -343$.

Es zeigten sich keine signifikanten Korrelationen der Skalen des quantitativen Teils (Q1-Mengenvergleiche, Q3-Gleichungenbilden und Q2-Zahlenreihen) mit den Schulnoten, eine Ausnahme bildet die Korrelation zwischen KFT-Q2 und der Mathematiknote. Die Untertests KFT-N1 Figurenklassifikation und KFT-N2 Figurenanalogien korrelieren mit den Deutsch- und Mathematikzensuren signifikant auf dem Niveau von 0,01 sowie 0,05, jedoch nicht signifikant mit der Englischnote. Die Korrelationen mit den Mathematiknoten sind am höchsten mit $r = .238$ und $r = .255$. Der Subtest N3 Faltaufgaben zeigt signifikante Zusammenhänge mit allen Noten. Die höchste Korrelation zeigt sich mit der Mathematiknote mit $r = .279$.

Bezüglich der Testteile sind die Korrelationen des V-Teils mit allen Noten signifikant, am höchsten korreliert er mit der Deutschzensur ($r = .430$). Die Korrelationen des Q Teils mit allen Noten sind auf dem Niveau von 0,05 signifikant, mit Ausnahme jener mit der Englischzensur. Die höchste Korrelation zeigt sich mit der Mathematiknote ($r = .211$).

In Bezug auf den N Teil sind die Korrelationen auf dem Niveau von 0,01 signifikant, mit Ausnahme jener mit der Englischzensur. Am höchsten ist die Korrelation mit der Mathematiknote mit $r = .330$. Bezüglich der Gesamtleistung GL sind die Korrelationen auf dem Niveau von 0,01 und 0,05 mit allen Zensuren signifikant. Die Korrelation mit der Mathematikzensur ist mit $r = .303$ am höchsten.

Aus den Korrelationsergebnissen, die in Tabelle 5.3.3.6 zusammengestellt sind, kann man schließen, dass der KFT 4-12+ R gute prognostische Validität besitzt. Im Hinblick auf die Korrelationen der Skalen des V-Teils ergeben sich eine gute Prognose ($r = .430$) für die Deutschnote und ausreichende bis gute Prognosen für die Englischnote (bis $r = .350$) und für die Mathematiknote ($r = .329$). Die Skalen des Q-Teils bzw. KFT-Q1, KFT-Q3 und KFT-Q2 weisen keine Zusammenhänge mit den Schulnoten auf, auch nicht mit der Mathematiknote (außer KFT-Q2 mit der Mathematiknote mit $r = .175$). Außerdem zeigt die Testdimension KFT-Q ausreichende prognostische Validität für die Deutsch- und Mathematikzensuren. Die Skalen des N Teils (KFT-N1, KFT-N2 und KFT-N3) und die zusammengefasste Testdimension KFT-N zeigen ausreichende bis gute Validität hinsichtlich der Deutschnote ($r = .240$) und gute Validität in Bezug auf die Mathematiknote $r = .330$. Außerdem ergibt sich ein ausreichender Zusammenhang

zwischen dem Untertest KFT-N3 und der Englischzensur ($r = .216$). Die Korrelationskoeffizienten mit der KFT-GL zeigen ausreichende prognostische Validität hinsichtlich der Englischnote ($r = .193$) und gute prognostische Validität hinsichtlich der späteren Deutsch- ($r = .276$) sowie der Mathematiknote ($r = .303$).

5.4 Weitere Befunde zur Intelligenzsentwicklung, Geschlecht, Migration und Schulleistung für den KFT 4-12+ R in der vorliegenden Stichprobe

5.4.1 Kohorteneffekte intellektueller Kompetenzen für die Klassenstufen 5 und 7 in Bezug auf den KFT 4-12 R

5.4.1.1 Kohorteneffekte für die Klassenstufe 5

Für die Analyse der Kohorten- bzw. Jahrgangseffekte, also der Veränderung der Schülerschaft über den betrachteten Zeitraum hinweg, wurden immer 3-4 Jahrgänge zusammengefasst, um zufällige Schwankungen auszugleichen und das statistische Material sowie die Interpretation zu vereinfachen. Die Unterschiede in den KFT-Leistungen der Kohorten (Veränderungen der Schülerschaft über den betrachteten Zeitraum) wurden mittels einfaktorieller Varianzanalysen (SPSS-Prozedur „ONEWAY“) berechnet.

Tabelle 5.4.1.1: Varianzanalysen zu den Unterschieden der Jahrgangsgruppen in den KFT 4-12+R Dimensionen (KFT-V, KFT-Q, KFT-N) und der Gesamtleistung für die Klassenstufe 5

KFT-V		Quadratsumme	df	Mittel der Quadrate	F	Signifikanz
	Zwischen den Gruppen	862,917	3	287,639	5,384	,001
	Innerhalb der Gruppen	111278,966	2083	53,422		
	Gesamt	112141,883	2086			
KFT-Q	Zwischen den Gruppen	715,944	3	238,648	3,948	,008
	Innerhalb der Gruppen	125554,740	2077	60,450		
	Gesamt	126270,684	2080			
KFT-N	Zwischen den Gruppen	686,906	3	228,969	3,249	,021
	Innerhalb der Gruppen	146644,334	2081	70,468		
	Gesamt	147331,240	2084			
KFT-GL	Zwischen den Gruppen	752,959	3	250,986	5,193	,001
	Innerhalb der Gruppen	99997,374	2069	48,331		
	Gesamt	100750,333	2072			

In der vorliegenden Untersuchung liefert die Varianzanalyse auf dem 5%-Niveau signifikante Ergebnisse für KFT-Verbal ($F = 5.38, p = 0.001$), KFT-Quantitativ ($F = 3.95, p = 0.008$), KFT-Nonverbal ($F = 3.25, p = 0.021$) und die KFT- Gesamtleistung ($F = 5.19, p = 0.001$) (siehe Tabelle 5.4.1.1). Die Post-Hoc-Tests (nach Scheffé) erbrachten auf dem 5%-Niveau bezüglich der KFT-Dimensionen (KFT-V, KFT-Q, KFT-N) und der KFT-Gesamtleistung jeweils zwei homogene Untergruppen (siehe Tabellen 5.4.1.2, 5.4.1.4, 5.4.1.6 und 5.4.1.8; die Mittelwerte und weitere Verteilungsmerkmale der Kohorten bzw. Jahrgangsgruppen finden sich in den Tabellen 5.4.1.3, 5.4.1.5, 5.4.1.7 und 5.4.1.9).

Tabelle 5.4.1.2: Post-Hoc-Test „homogene Untergruppen“ der verschiedenen Jahrgangsgruppen hinsichtlich KFT-V für Klassestufe 5

Jahrgänge	N	Scheffé-Prozedur	
		Untergruppe für Alpha = ,050.	
		1	2
2001-2003	491	43,7312	
2004-2006	469	44,3134	44,3134
2007-2010	628	44,6306	44,6306
1998-2000	499	,273	45,5551
Signifikanz			,060

Tabelle 5.4.1.3: Mittelwerte von KFT-Verbal der Schüler für Klassenstufe 5 nach Jahrgängen

	N	Mittelwert	Standard-abweichung	Standard-fehler	95%-Konfidenzintervall für den Mittelwert		Minimum	Maximum
					Untergrenze	Obergrenze		
1998-2000	499	45,5551	6,99243	,31302	44,9401	46,1701	22,00	65,00
2001-2003	491	43,7312	7,16518	,32336	43,0958	44,3665	22,00	68,00
2004-2006	469	44,3134	7,13049	,32926	43,6664	44,9604	25,00	69,00
2007-2010	628	44,6306	7,78277	,31057	44,0207	45,2404	22,00	75,00
Gesamt	2087	44,5688	7,33207	,16050	44,2540	44,8835	22,00	75,00

Tabelle 5.4.1.4: Post-Hoc-Test „homogene Untergruppen“ der verschiedenen Jahrgangsgruppen hinsichtlich KFT-Q für Klassestufe 5

Jahrgänge	N	Scheffé-Prozedur	
		Untergruppe für Alpha = ,050.	
		1	2
2001-2003	489	44,4806	
2004-2006	468	44,8590	44,8590
2007-2010	624	45,6314	45,6314
1998-2000	500	,131	45,9780
Signifikanz			,150

Tabelle 5.4.1.5: Mittelwerte von KFT-Quantitativ der Schüler für Klassenstufe 5 nach Jahrgängen

	N	Mittelwert	Standard-abweichung	Standard-fehler	95%-Konfidenzintervall für den Mittelwert		Minimum	Maximum
					Untergrenze	Obergrenze		
1998-2000	500	45,9780	7,71859	,34519	45,2998	46,6562	20,00	72,00
2001-2003	489	44,4806	7,52891	,34047	43,8116	45,1495	20,00	64,00
2004-2006	468	44,8590	7,82089	,36152	44,1486	45,5694	20,00	66,00
2007-2010	624	45,6314	7,97259	,31916	45,0047	46,2582	20,00	75,00
Gesamt	2081	45,2705	7,79147	,17080	44,9356	45,6055	20,00	75,00

In Bezug auf KFT-V und KFT-Q lieferten die Post-Hoc-Tests jeweils zwei homogene Untergruppen, die sich voneinander signifikant unterscheiden. Außerdem weisen die betrachteten Jahrgänge pro Gruppe untereinander keinen signifikanten Unterschied auf. Die erste Gruppe besteht aus den Jahren 2001-2003, 2004-2006 und 2007-2010, die sich von der zweiten Gruppe unterscheidet, welche die Jahrgänge 2004-2006, 2007-2010 und 1998-2000 umfasst (siehe Tabelle 5.4.1.2 und 5.4.1.4, jeweils Spalte 1). Die jeweils zweite Spalte in den Tabellen 5.4.1.2 und 5.4.1.4 zeigt, dass sich die Jahrgänge 1998-2000 (zweite Gruppe) von den Jahren 2001-2003 (erste Gruppe) signifikant unterscheiden.

Tabelle 5.4.1.6: Post-Hoc-Test „homogene Untergruppen“ der verschiedenen Jahrgangsgruppen hinsichtlich KFT-N für Klassenstufe 5

Jahrgänge	N	Scheffé-Prozedur	
		Untergruppe für Alpha = ,050.	
		1	2
2001-2003	490	48,5143	
2004-2006	470	49,5596	49,5596
1998-2000	500	49,7540	49,7540
2007-2010	625		50,0320
Signifikanz		,132	,846

Tabelle 5.4.1.7: Mittelwerte des KFT-Nonverbal der Schüler für Klassenstufe 5 nach Jahrgangsgruppen

	N	Mittelwert	Standard-abweichung	Standard-fehler	95%-Konfidenzintervall für den Mittelwert		Minimum	Maximum
					Untergrenze	Obergrenze		
1998-2000	500	49,7540	7,73052	,34572	49,0748	50,4332	22,00	77,00
2001-2003	490	48,5143	8,02888	,36271	47,8016	49,2269	22,00	70,00
2004-2006	470	49,5596	8,41435	,38813	48,7969	50,3223	22,00	77,00
2007-2010	625	50,0320	9,13708	,36548	49,3143	50,7497	22,00	82,00
Gesamt	2085	49,5022	8,40811	,18414	49,1410	49,8633	22,00	82,00

Der Post-Hoc-Test (Scheffé) erbrachte auf dem 5%-Niveau zwei homogene Untergruppen bezüglich des KFT-N. Systematischen Differenzen ergaben sich zwischen der ersten Gruppe

(Jahrgänge 2001-2003, 2004-2006 und 1998-2000) und der zweiten Gruppe, die aus den Jahrgängen 2004-2006, 1998-2000 und 2007-2010 besteht (Tabelle 5.4.1.6, Spalte 1 und 2). Der signifikante Unterschied ist auf die unterschiedlichen Leistungen der Jahrgänge 2007-2010 (zweite Gruppe) bzw. Jahrgänge 2001-2003 (erste Gruppe, siehe Tabelle 5.4.1.6, Spalte 2) zurückzuführen. Die Jahrgänge pro Gruppe zeigen ansonsten untereinander keinen signifikanten Unterschied.

Tabelle 5.4.1.8: Post-Hoc-Test „homogene Untergruppen“ der verschiedenen Jahrgangsgruppen hinsichtlich KFT-GL für Klassenstufe 5

Jahrgänge	N	Scheffé-Prozedur	
		1	2
2001-2003	488	45,3299	
2004-2006	464	46,3211	46,3211
2007-2010	621		46,5974
1998-2000	500		46,9940
Signifikanz		,158	,494

Tabelle 5.4.1.9: Mittelwerte der KFT-Gesamtleistung der Schüler für Klassenstufe 5 nach Jahrgangsgruppen

	N	Mittelwert	Standard-abweichung	Standard-fehler	95%-Konfidenzintervall für den Mittelwert		Minimum	Maximum
					Untergrenze	Obergrenze		
1998-2000	500	46,9940	6,60221	,29526	46,4139	47,5741	20,00	64,00
2001-2003	488	45,3299	6,87941	,31142	44,7180	45,9418	20,00	66,00
2004-2006	464	46,3211	6,51119	,30227	45,7271	46,9151	27,00	65,00
2007-2010	621	46,5974	7,57429	,30395	46,0005	47,1943	20,00	79,00
Gesamt	2073	46,3329	6,97314	,15315	46,0325	46,6332	20,00	79,00

In Bezug auf die KFT-GL lieferte der Post-Hoc-Test (Scheffe) auf dem 5%-Niveau ebenfalls zwei homogene Untergruppen. Die erste Gruppe umfasst die Jahre 2001-2003 und 2004-2006, die Jahrgänge 2004-2006, 2007-2010 und 1998-2000 bilden die zweite Gruppe (Tabelle 5.4.1.8, Spalten 1 und 2). Die beiden Gruppen unterscheiden sich voneinander signifikant und zeigen untereinander keine signifikante Differenz. Darüber hinaus unterscheiden sich die Jahrgänge 2007-2010 und 1998-2000 signifikant von den Jahren 2001-2003 (Tabelle 5.4.1.8, Spalte 2).

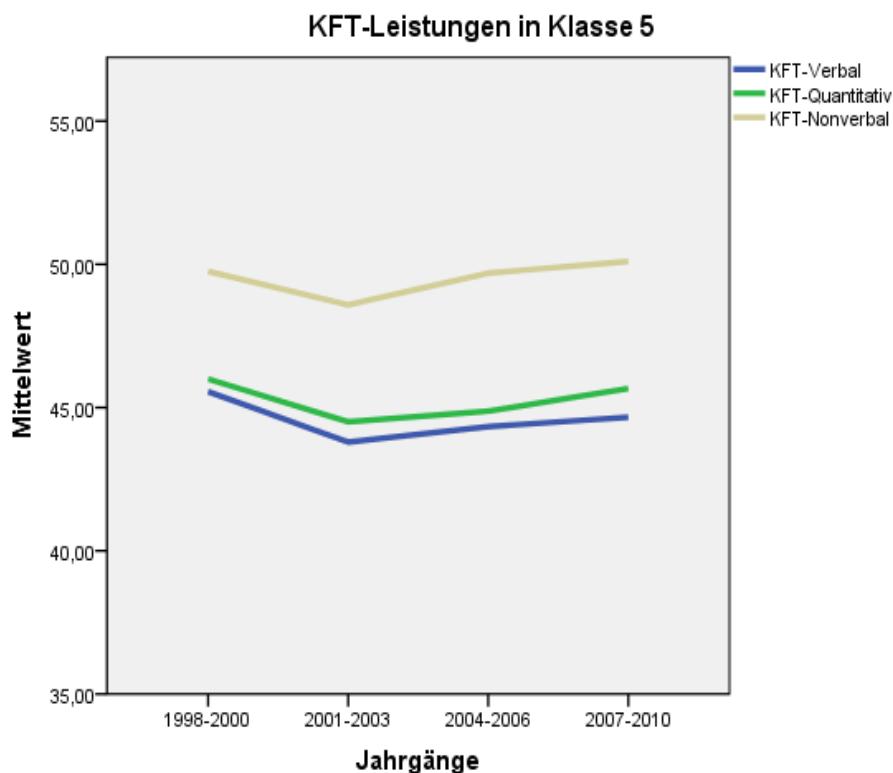


Abbildung 5.4.1.1: Veränderung der KFT-Dimensionen (KFT-V, KFT-Q, KFT-N) in Klassenstufe 5

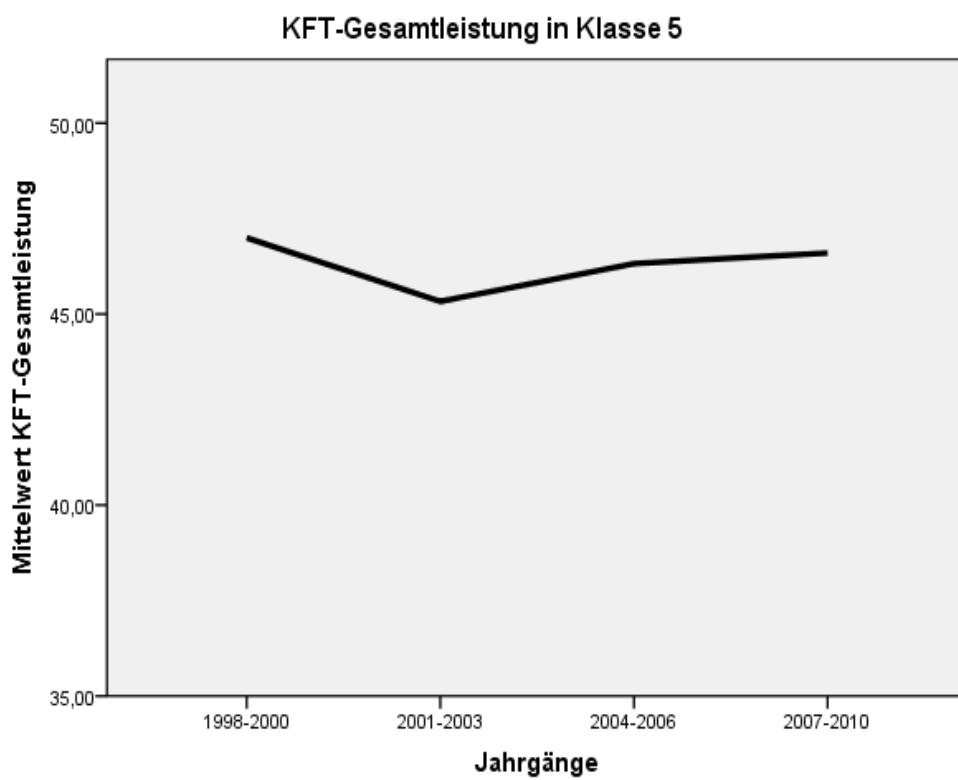


Abbildung 5.4.1.2: Veränderung der KFT-Gesamtleistung in Klassenstufe 5

Wenn man die Abbildung 5.4.1.1 betrachtet, findet man, dass die Schüler/innen der Klassenstufe 5 bessere Leistungen im KFT-N als im KFT-V und KFT-Q vom Jahr 1998 bis zum Jahr 2010 erbracht haben. Außerdem zeigen sich Verschlechterungen in den KFT-Leistungen (verbale, quantitative und nonverbale Leistungen) bis 2001-2003, danach zeigen sich kaum Veränderungen. Abbildung 5.4.1.2 zeigt, dass sich die Gesamtleistung bis 2001-2003 verschlechtert. Danach verbessert sie sich wieder und bleibt hinsichtlich der Summe aus der verbalen, quantitativen und nonverbalen Leistungen stabil bis 2007-2010.

5.4.1.2 *Kohorteneffekte für die Klassenstufe 7*

Tabelle 5.4.1.10: Varianzanalysen zu den Unterschieden der Jahrgangsgruppen in den KFT 4-12+R Dimensionen (KFT-V, KFT-Q, KFT-N) und der Gesamtleistung für die Klassenstufe 7

KFT-V		Quadratsumme	df	Mittel der Quadrate	F	Signifikanz
	Zwischen den Gruppen	3743,016	3	1247,672	20,038	,000
	Innerhalb der Gruppen	137544,765	2209	62,266		
	Gesamt	141287,780	2212			
KFT-Q	Zwischen den Gruppen	3087,770	3	1029,257	16,300	,000
	Innerhalb der Gruppen	139364,130	2207	63,146		
	Gesamt	142451,900	2210			
KFT-N	Zwischen den Gruppen	5568,978	3	1856,326	23,937	,000
	Innerhalb der Gruppen	172082,772	2219	77,550		
	Gesamt	177651,750	2222			
KFT-GL	Zwischen den Gruppen	4961,038	3	1653,679	30,174	,000
	Innerhalb der Gruppen	120404,170	2197	54,804		
	Gesamt	125365,208	2200			

Die Anwendung der einfaktoriellen Varianzanalyse auf die zweite Teil-Stichprobe (Klassenstufe 7) auf dem 5%-Niveau ergab, dass sich die kognitiven Fähigkeiten in den Dimensionen des KFT 4-12+R und der KFT-Gesamtleistung signifikant verändern (KFT-V: $F = 20,04, p = 0,000$, KFT-Q: $F = 16,30, p = 0,000$, KFT-N: $F = 23,94, p = 0,000$ und KFT-GL: $F = 30,17, p = 0,000$); Tabelle 5.4.1.10). Die Post-Hoc-Tests (Tabellen 5.4.1.11, 5.4.1.13, 5.4.1.15 und 5.4.1.17) verdeutlichen die Unterschiede (Mittelwerte und weitere Verteilungsstatistiken in den Tabellen 5.4.1.12, 5.4.1.14, 5.4.1.16 und 5.4.1.18).

Tabelle 5.4.1.11: Post-Hoc-Test „homogene Untergruppen“ der verschiedenen Jahrganggruppen hinsichtlich KFT-V für Klassenstufe 7

Scheffé-Prozedur

Testjahrgänge	N	Untergruppe für Alpha = ,050.		
		1	2	3
2007-2011	760	43,6776		
2004-2006	472	43,8284		
2001-2003	491		45,4969	
1998-2000	490			46,8714
Signifikanz		,992	1,000	1,000

Tabelle 5.4.1.12: Mittelwerte von KFT-Verbal der Schüler für Klassenstufe 7 nach Jahrgängen

	N	Mittelwert	Standard-abweichung	Standard-fehler	95%-Konfidenzintervall für den Mittelwert		Minimum	Maximum
					Untergrenze	Obergrenze		
1998-2000	490	46,8714	8,20605	,37071	46,1430	47,5998	21,00	69,00
2001-2003	491	45,4969	7,24903	,32714	44,8542	46,1397	25,00	67,00
2004-2006	472	43,8284	7,81706	,35981	43,1214	44,5354	21,00	66,00
2007-2011	760	43,6776	8,12339	,29467	43,0992	44,2561	21,00	72,00
Gesamt	2213	44,8206	7,99208	,16989	44,4874	45,1538	21,00	72,00

Die Post-Hoc-Tests (Scheffe) lieferten auf dem 5%-Niveau drei homogene Untergruppen bezüglich des KFT-V und der KFT-Gesamtleistung. Die erste Gruppe besteht aus den Jahrgängen 2007-2011 und 2004-2006, die sich von sämtlichen anderen Gruppen signifikant unterscheidet, aber diese Jahrgänge weisen keinen signifikanten Unterschied untereinander auf und das gilt auch für die anderen Jahrgänge pro Gruppe (Tabellen 5.4.1.11 und 5.4.1.17 unten, Spalte 1). Außerdem zeigen die Jahre 2001-2003 (zweite Gruppe) eine signifikante Differenz zu den Jahrgängen 2007-2011 und 2004-2006 (erste Gruppe), sowie zu den Jahren 1998-2000 (dritte Gruppe) (Tabelle 5.4.1.11 und 5.4.1.17, Spalte 2). Darüber hinaus unterscheiden sich die Jahrgänge 1998-2000 (dritte Gruppe) signifikant von der zweiten Gruppe (Jahrgänge 2001-2003) und auch von den Jahren 2007-2011 und 2004-2006 (erste Gruppe) (Tabelle 5.4.1.11 und 5.4.1.17, Spalte 3).

Tabelle 5.4.1.13: Post-Hoc-Test „homogene Untergruppen“ der verschiedenen Jahrgangsgruppen hinsichtlich KFT-Q für Klassenstufe 7

Scheffé-Prozedur

Testjahrgänge	N	Untergruppe für Alpha = ,050.		
		1	2	3
2007-2011	759	46,0804		
2004-2006	472	47,2097	47,2097	
2001-2003	491		47,6578	
1998-2000	489			49,2740
Signifikanz		,147	,838	1,000

Tabelle 5.4.1.14: Mittelwerte von KFT-Quantitativ der Schüler für Klassenstufe 7 nach Jahrgängen

	N	Mittelwert	Standard-abweichung	Standard-fehler	95%-Konfidenzintervall für den Mittelwert		Minimum	Maximum
					Untergrenze	Obergrenze		
1998-2000	489	49,2740	8,03951	,36356	48,5597	49,9884	21,00	74,00
2001-2003	491	47,6578	8,28701	,37399	46,9230	48,3927	21,00	77,00
2004-2006	472	47,2097	7,49497	,34498	46,5318	47,8876	28,00	72,00
2007-2011	759	46,0804	7,93394	,28798	45,5150	46,6457	21,00	77,00
Gesamt	2211	47,3781	8,02857	,17074	47,0433	47,7129	21,00	77,00

Anhand der Post-Hoc-Analyse (Scheffé) auf dem 5%-Niveau können bezüglich des KFT-Q drei homogene Untergruppen unterschieden werden, die sich signifikant voneinander unterscheiden und die untereinander keinen signifikanten Unterschied aufweisen. Die signifikante Veränderung lässt sich einerseits auf die systematischen Differenzen zwischen der ersten Gruppe, welche die Jahrgänge 2007-2011 und 2004-2006 bilden, und den anderen zwei Gruppen bzw. die Jahre 2001-2003 (zweite Gruppe) und 1998-2000 (dritte Gruppe) (Tabelle 5.4.1.13, Spalte 1) zurückführen. Andererseits finden sich Unterschiede zwischen den Jahrgängen 2001-2003 (zweite Gruppe), den Jahrgängen 2007-2011 (erste Gruppe) und den Jahren 1998-2000 (dritte Gruppe) (Tabelle 5.4.1.13, Spalte 2). Weiter sind die Differenzen zwischen den Jahrgängen 1998-2000 (dritte Gruppe), den Jahren 2007-2011, 2004-2006 (erste Gruppe) sowie 2001-2003 (zweite Gruppe) (Tabelle 5.4.1.13, Spalte 3) signifikant. Die Jahrgänge pro Gruppe weisen untereinander keinen signifikanten Unterschied auf.

Die Durchführung der Post-Hoc-Analyse (Scheffé) auf dem 5%-Niveau erbrachte zwei homogene Untergruppen bezüglich des KFT-N. Die erste Gruppe besteht aus den Jahrgängen 2007-2011 und 2004-2006, die sich von der zweiten Gruppe signifikant unterscheidet (Tabelle 5.4.1.15, Spalte 1). Dazu kommt, dass die Jahrgänge 2001-2003 und 1998-2000 (zweite Gruppe) eine signifikante Differenz zu den Jahren 2007-2011 und 2004-2006 (erste Gruppe) (Tabelle

5.4.1.15, Spalte 2) aufweisen. Darüber hinaus weisen die Jahrgänge pro Gruppe untereinander keinen signifikanten Unterschied auf.

Tabelle 5.4.1.15: Post-Hoc-Test „homogene Untergruppen“ der verschiedenen Jahrgangsgruppen hinsichtlich KFT-N für Klassenstufe 7

Testjahrgänge	N	Scheffé-Prozedur	
		1	2
2007-2011	769	49,2445	
2004-2006	473	49,4989	
2001-2003	491		52,0407
1998-2000	490		52,9020
Signifikanz		,974	,466

Tabelle 5.4.1.16: Mittelwerte des KFT-Nonverbal der Schüler für Klassenstufe 7 nach Jahrgangsgruppen

	N	Mittelwert	Standard-abweichung	Standard-fehler	95%-Konfidenzintervall für den Mittelwert		Minimum	Maximum
					Untergrenze	Obergrenze		
1998-2000	490	52,9020	8,53997	,38580	52,1440	53,6601	29,00	80,00
2001-2003	491	52,0407	8,13473	,36712	51,3194	52,7620	21,00	75,00
2004-2006	473	49,4989	8,72366	,40111	48,7108	50,2871	21,00	75,00
2007-2011	769	49,2445	9,41478	,33951	48,5780	49,9109	25,00	80,00
Gesamt	2223	50,7224	8,94155	,18965	50,3505	51,0943	21,00	80,00

Tabelle 5.4.1.17: Post-Hoc-Test „homogene Untergruppen“ der verschiedenen Jahrgangsgruppen hinsichtlich KFT-GL für Klassenstufe 7

Scheffé-Prozedur

Testjahrgänge	N	Untergruppe für Alpha = ,050.		
		1	2	3
2007-2011	752	45,9495		
2004-2006	471	46,5648		
2001-2003	489		48,2986	
1998-2000	489			49,7321
Signifikanz		,608	1,000	1,000

Tabelle 5.4.1.18: Mittelwerte der KFT-Gesamtleistung der Schüler für Klassenstufe 7 nach Jahrgangsgruppen

	N	Mittelwert	Standard-abweichung	Standard-fehler	95%-Konfidenzintervall für den Mittelwert		Minimum	Maximum
					Untergrenze	Obergrenze		
1998-2000	489	49,7321	7,63833	,34542	49,0534	50,4108	25,00	72,00
2001-2003	489	48,2986	6,67059	,30165	47,7059	48,8913	22,00	68,00
2004-2006	471	46,5648	7,46649	,34404	45,8887	47,2408	25,00	70,00
2007-2011	752	45,9495	7,65571	,27918	45,4014	46,4975	21,00	77,00
Gesamt	2201	47,4434	7,54879	,16090	47,1279	47,7590	21,00	77,00

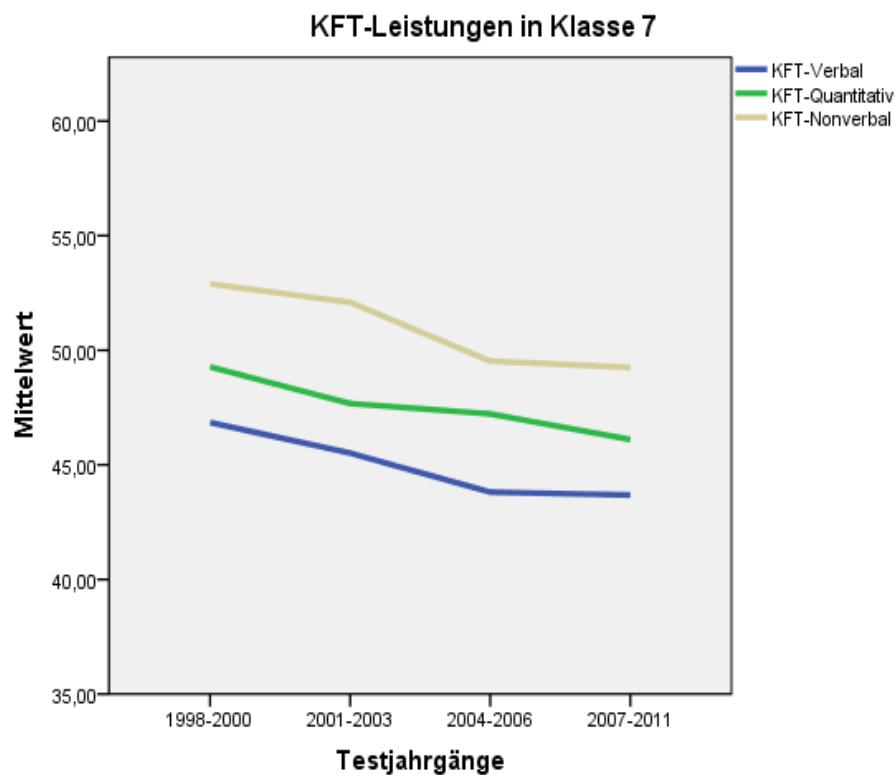


Abbildung 5.4.1.3: Veränderung der KFT-Dimensionen (KFT-V, KFT-Q, KFT-N) in Klassenstufe 7

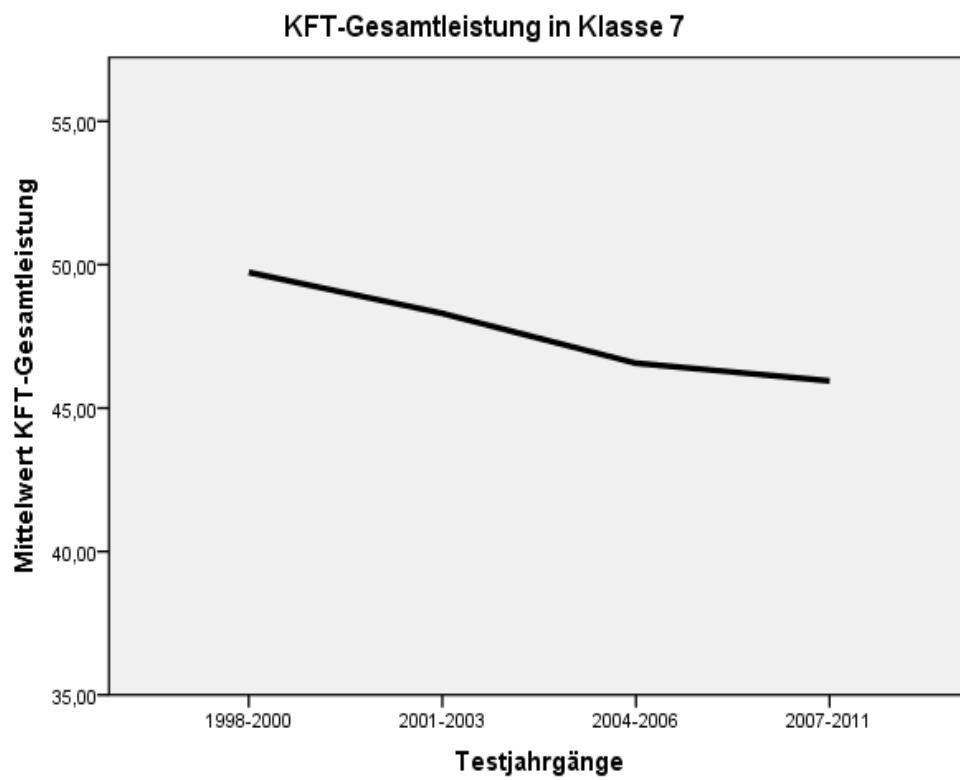


Abbildung 5.4.1.4: Veränderung der KFT-Gesamtleistung in Klassenstufe 7

Abbildung 5.4.1.3 zeigt, dass die nonverbalen Leistungen über jenen der anderen Test-Dimensionen (verbale und quantitative Leistungen) liegen. Außerdem war die verbale Leistung von den Schülern/innen in Klassenstufe 7 von 1998 bis zu 2011 am schlechtesten. Das bestätigt einerseits die Ergebnisse zu Klassenstufe 5. Andererseits verändern sich die kognitiven Fähigkeiten (KFT-V, KFT-Q und KFT-N) von 1998 zu 2011 signifikant. Die nonverbale Leistung bleibt kontinuierlich bzw. konstant bis 2001-2003. Dann sinkt sie bis 2004-2006 ab. Danach zeigt sie kaum Veränderung bis 2007-2011. Die quantitative Leistung verschlechtert sich bis 2001-2003, danach bleibt sie ohne Veränderung bis 2004-2006 und dann nimmt sie wieder bis 2007-2011 ab. In Bezug auf die verbale Leistung verschlechtern sich die Leistungen bis 2007-2011 und es zeigen sich ab dann kaum oder keine Änderungen. Abbildung 5.4.1.4 schließlich zeigt, dass sich die Gesamtleistung bis 2007-2011 verschlechtert. Dieses Ergebnis ist schlüssig, weil die Leistungen in allen KFT-Dimensionen (verbale, quantitative und nonverbale Leistungen) vom Jahr 1998 bis zum Jahr 2011 absinken.

Zusammenfassend kann man sagen dass, sich die verbalen, quantitativen und nonverbalen Kompetenzen sowie die Gesamtleistung des KFT 4-12+R in beiden Klassenstufen 5 und 7 vom Jahr 1998 bis zum Jahr 2010 und 2011 signifikant verändern. Darüber hinaus stimmen die Ergebnisse von den beiden Klassenstufen 5 und 7 darin überein, dass die nonverbale Leistung der Schüler/innen in den beiden Klassen 5 und 7 am besten und die verbale Leistung am schlechtesten ist. Die Gründe dafür könnten sein, dass bei einem großen Teil der Stichprobe, fast 46,6% der Stichprobe in Klassenstufe 5 und 43,6% der Stichprobe in Klassenstufe 7, die Muttersprache nicht Deutsch ist bzw. die am häufigsten zu Hause gesprochene Sprache nicht Deutsch ist. Da sich die Schule außerdem in einem Münchener Brennpunktviertel befindet, könnte man dieses Ergebnis mit vergleichsweise geringerer häuslicher Förderung erklären. Die Messung der nonverbalen bzw. fluiden Intelligenz ist also wichtig, um die Intelligenz der Schüler/innen nicht zu unterschätzen, wie es durch einen Intelligenztests passieren könnte, der überwiegend verbale oder auch quantitative Kompetenzen erfasst.

5.4.2 Geschlechtsunterschiede im Untersuchungszeitraum

Geschlechtsunterschiede bezüglich der kognitiven Fähigkeiten im Erhebungszeitraum der letzten Jahre wurden an einer Stichprobe der Fünft- und Siebtklässler untersucht ($N_5 = 2091$ und $N_7 = 2229$). Während 1113 (53,23%) Jungen und 978 (46,77%) Mädchen aus Klassenstufe 5 an dieser Untersuchung teilgenommen haben, waren es 1165 (52,27%) Jungen und 1064 (47,73%) Mäd-

chen aus Klassenstufe 7. Zur Varianzanalyse der Unterschiede wurde die SPSS-Prozedur „ANOVA“ verwendet.

Tabelle 5.4.2.1: Varianzanalysen der KFT 4-12+R Dimensionen (KFT-V, KFT-Q, KFT-N) und der Gesamtleistung von Klassenstufe 5 nach Geschlecht und Jahrgangsgruppen

			Experimentelle Methode				
			Quadratsumme	df	Mittel der Quadrate	F	Sig.
KFT- Verbal	2-Weg- Wechselwirkungen Modell Residuen Insgesamt	(Kombiniert)	1080,691	4	270,173	5,059	,000
		Geschlecht	175,129	1	175,129	3,279	,070
		Jahrgänge	919,631	3	306,544	5,740	,001
		Geschlecht * Jahrgänge	390,813	3	130,271	2,439	,063
		Modell	1471,504	7	210,215	3,936	,000
		Residuen	110969,207	2078	53,402		
KFT- Quantitativ	2-Weg- Wechselwirkungen Modell Residuen Insgesamt	Insgesamt	112440,711	2085	53,928		
		(Kombiniert)	6955,454	4	1738,864	30,121	,000
		Geschlecht	6203,743	1	6203,743	107,464	,000
		Jahrgänge	842,319	3	280,773	4,864	,002
		Geschlecht * Jahrgänge	197,331	3	65,777	1,139	,332
		Modell	7152,786	7	1021,827	17,701	,000
KFT- Nonverbal	2-Weg- Wechselwirkungen Modell Residuen Insgesamt	Residuen	119613,502	2072	57,729		
		Insgesamt	126766,288	2079	60,975		
		(Kombiniert)	785,638	4	196,409	2,785	,025
		Geschlecht	74,784	1	74,784	1,060	,303
		Jahrgänge	708,791	3	236,264	3,350	,018
		Geschlecht * Jahrgänge	406,277	3	135,426	1,920	,124
KFT- Gesamt- leistung	2-Weg- Wechselwirkungen Modell Residuen Insgesamt	Modell	1191,915	7	170,274	2,414	,018
		Residuen	146427,062	2076	70,533		
		Insgesamt	147618,976	2083	70,868		
		(Kombiniert)	1766,067	4	441,517	9,167	,000
		Geschlecht	955,287	1	955,287	19,834	,000
		Jahrgänge	816,203	3	272,068	5,649	,001
		Geschlecht * Jahrgänge	259,216	3	86,405	1,794	,146
		Modell	2025,283	7	289,326	6,007	,000
		Residuen	99409,209	2064	48,163		
		Insgesamt	101434,492	2071	48,979		

5.4.2.1 Klassenstufe 5

Im Zuge der Varianzanalyse wurde zunächst KFT-V als abhängige Variable, Geschlecht und die Jahrgangsgruppen (wie in Abschnitt 5.4.1) als Faktoren einzubeziehen. Tabelle 5.4.2.1 zeigt, dass es keinen signifikanten Unterschied ($p > .05$) beim Faktor Geschlecht ($F = 3,28, p = .070$) und bei

der Wechselwirkung ($F = 2,44, p = .063$) gibt. Das gilt nicht für den Faktor Jahrgänge ($F = 5,74, p = .001$).

Bezüglich des KFT-Q (KFT-Q als abhängige Variable, Geschlecht und Jahrgänge als Faktoren) gibt Tabelle 5.4.2.1 an, dass es Signifikanz bei den Haupteffekten ($p < .05$), beim Faktor Geschlecht ($F = 107,46, p = .000$) und Jahrgangsgruppen ($F = 4,86, p = .002$) gibt, aber keine Signifikanz bei der Wechselwirkung ($F = 1,14, p = .332$), d.h. die Faktoren Geschlecht und Jahrgänge beeinflussen signifikant die quantitative Leistung.

Hinsichtlich KFT-N (KFT-N abhängige Variable, Jahrgangsgruppen und Geschlecht Faktoren) liefert Tabelle 5.4.2.1 uns Informationen darüber, dass es keinen signifikanten Unterschied für den Haupteffekt ($p > .05$) vom Faktor Geschlecht ($F = 1,06, p = .303$) und für die Wechselwirkung ($F = 1,92, p = .124$) gibt. Das gilt nicht für den Faktor Jahrgangsgruppen ($F = 3,35, p = .018$), d.h. nur der Faktor Jahrgangsgruppe hat einen Einfluss auf die nonverbale Leistung.

In Bezug auf die KFT-Gesamtleistung (KFT-GL als abhängige Variable, Geschlecht und Jahrgangsgruppen als Faktoren) zeigen sich in Tabelle 5.4.2.1 signifikante Unterschiede bei den Haupteffekten ($p < .05$), also bei beiden Faktoren Geschlecht ($F = 19,83, p = .000$) und Jahrgangsgruppen ($F = 5,65, p = .001$), aber nicht bei der Wechselwirkung ($F = 1,79, p = .146$). d.h. beide Faktoren üben einen gemeinsamen Einfluss auf die Gesamtleistung aus.

Um die Ergebnisse besser interpretieren zu können, werden die Befunde durch Graphiken verdeutlicht: Die statistischen Ergebnisse zeigen, dass der Unterschied zwischen Mädchen und Jungen bezüglich ihren verbalen Leistung nicht signifikant ist. Die Abbildung 5.4.2.1 zeigt nun, dass Mädchen meist schlechter als Jungen im KFT-V abschneiden. Die verbale Leistung von Mädchen verschlechtert sich bis 2001-2003, danach verbessert sie sich wieder bis 2007-2010. Die verbale Leistung von Jungen verändert sich kaum.

In Abbildung 5.4.2.2 fällt auf, dass die quantitativen Leistungen der Mädchen unter den quantitativen Leistungen der Jungen liegen. Das bestätigt die statistischen Ergebnisse, dass sich Mädchen und Jungen signifikant hinsichtlich ihrer Leistungen im KFT-Quantitativ unterscheiden. Zusätzlich verschlechtert sich die quantitative Leistung der Mädchen bis 2001-2003. Dann zeigt sie kaum Veränderung bis 2007-2010. Das gilt für die quantitative Leistung der Jungen, aber die Verschlechterung ist bei den Mädchen stärker als bei den Jungen.

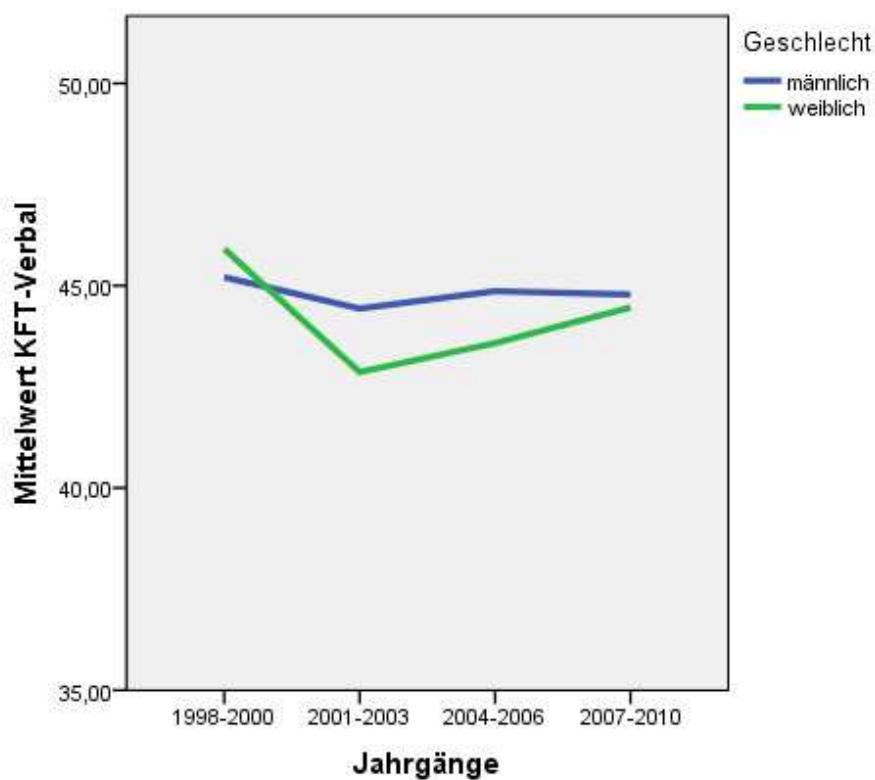


Abbildung 5.4.2.1: Entwicklung der verbalen KFT-Leistung nach Geschlecht und Jahrgangsgruppen in Klassenstufe 5

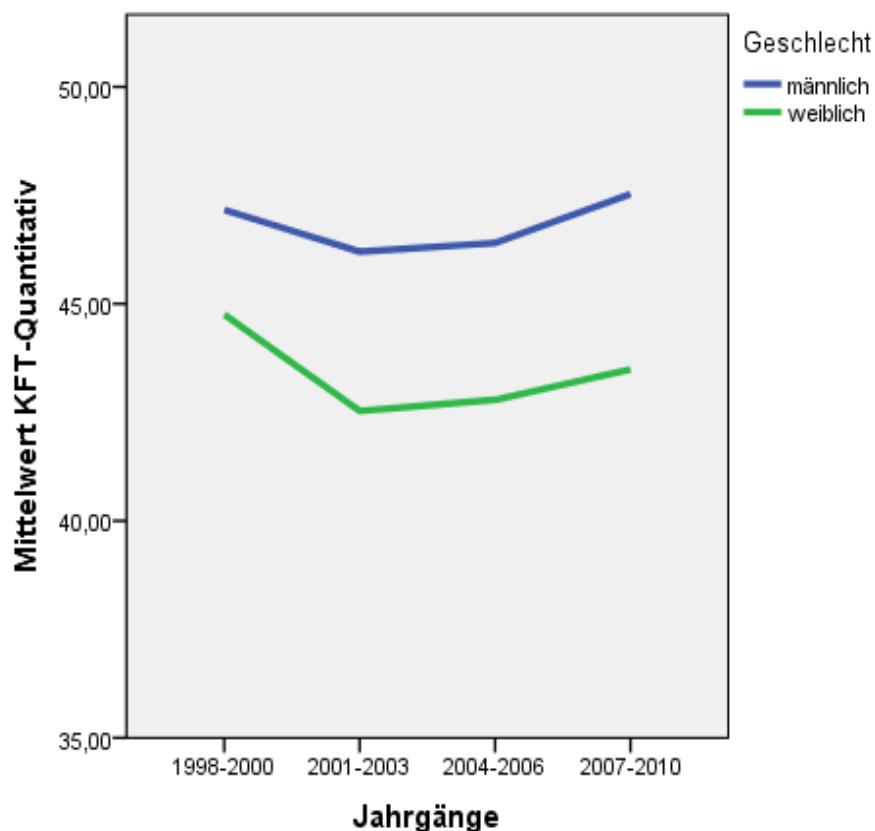


Abbildung 5.4.2.2: Entwicklung der quantitativen KFT-Leistung nach Geschlecht und Jahrgangsgruppen in Klassenstufe 5

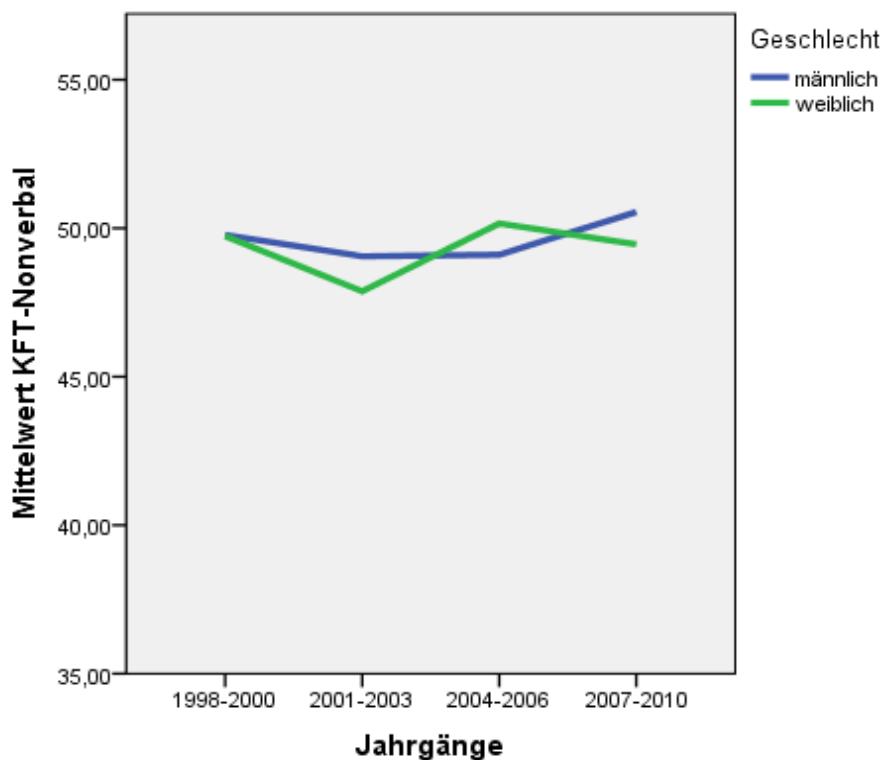


Abbildung 5.4.2.3: Entwicklung der nonverbalen KFT-Leistung nach Geschlecht und Jahrgangsgruppen in Klassenstufe 5

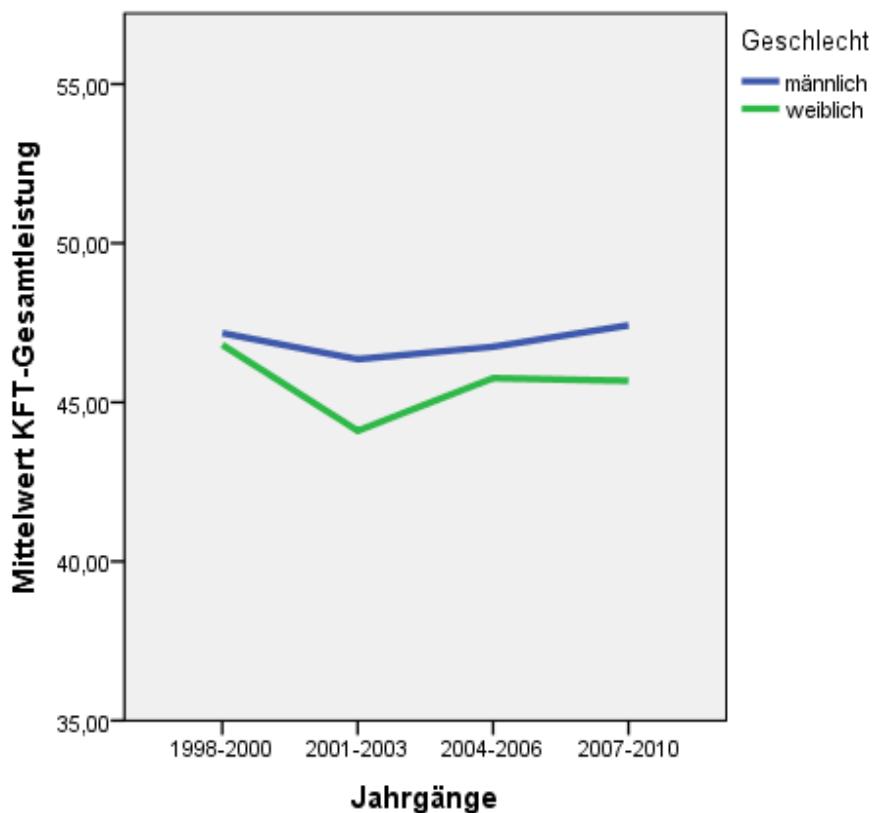


Abbildung 5.4.2.4: Entwicklung der KFT-Gesamtleistung nach Geschlecht und Jahrgangsgruppen in Klassenstufe 5

Wenn man Abbildung 5.4.2.3 betrachtet, fällt auf, dass sich Mädchen und Jungen in der nonverbalen Leistung nicht signifikant unterscheiden und das stimmt mit den statistischen Ergebnissen überein. Darüber hinaus zeigen die nonverbalen Leistungen von Jungen kaum Veränderung bis 2004-2006 und dann steigen sie wieder bis 2007-2010 an. Die nonverbalen Leistungen von Mädchen schwanken, es gibt eine Verschlechterung bis 2001-2003, danach eine Verbesserung bis 2004-2006 und einen Abstieg 2007-2010.

Abbildung 5.4.2.4 bestätigt die statistischen Ergebnisse, dass sich Mädchen und Jungen signifikant bezüglich der KFT-Gesamtleistung unterscheiden. Außerdem schneiden Mädchen deutlich schlechter als Jungen in der KFT-Gesamtleistung ab. Ihre Gesamtleistung verschlechtert sich bis 2001-2003. Dann nimmt ihre Gesamtleistung bis 2004-2006 zu, danach gibt es kaum Veränderungen. Im Gegensatz zu Mädchen bleibt die KFT-Gesamtleistung von Jungen stabil ohne oder mit geringen Veränderungen von den Jahrgängen 1998-2000 bis zu den Jahrgängen 2007-2010.

5.4.2.2 Klassenstufe 7

In Bezug auf KFT-V als abhängige Variable und Geschlecht und Testjahrgänge als Faktoren zeigt Tabelle 5.4.2.2, dass es einen signifikante Einflüsse der Haupteffekte insgesamt ($p < .05$) sowie von Geschlecht ($F = 20,86, p = .000$) und Testjahrgängen ($F = 21,16, p = .000$) gibt, d.h. die Faktoren Geschlecht und Jahrgangsgruppen haben einen Einfluss auf die verbale Leistung. Hinsichtlich der Wechselwirkung ($F = 0,36, p = .783$) liegt keine Signifikanz vor.

Im KFT-Q (KFT-Q abhängige Variable, Geschlecht und Jahrgangsgruppen Faktoren) liefert Tabelle 5.4.2.2 uns Informationen darüber, dass es eine Signifikanz bei den Haupteffekten ($p < .05$) sowie den Faktoren Geschlecht ($F = 64,42, p = .000$) und Testjahrgänge ($F = 17,92, p = .000$) gibt. Hinsichtlich der Wechselwirkung ($F = 1,22, p = .301$) ergab sich keine Signifikanz.

Bezüglich des KFT-N (KFT-N abhängige Variable, Geschlecht und Jahrgangsgruppen Faktoren) beeinflusst der Faktor Geschlecht ($F = 4,19, p = .041$) die nonverbale Leistung ebenso wie der Faktor Testjahrgänge ($F = 24,59, p = .000$; für beide Haupteffekten: $p < 0,05$). Hinsichtlich der Wechselwirkung ($F = 0,49, p = .687$) liegt keine Signifikanz vor (Tabelle 5.4.2.2).

Im KFT-GL (KFT-GL abhängige Variable, Geschlecht und Testjahrgänge Faktoren) haben die beiden Faktoren Geschlecht ($F = 27,86, p = .000$) und Jahrgangsgruppen ($F = 31,67, p = .000$)

bei den Haupteffekten (insgesamt $p < 0,05$) einen Einfluss auf die Gesamtleistung. Hinsichtlich der Wechselwirkung ($F = 0,21, p = .893$) liegt ebenfalls keine Signifikanz vor (Tabelle 5.4.2.2).

Tabelle 5.4.2.2: Varianzanalysen der KFT 4-12+R Dimensionen (KFT-V, KFT-Q, KFT-N) und der Gesamtleistung von Klassenstufe 7 nach Geschlecht und Jahrgangsgruppen

			Experimentelle Methode				
			Quadratsumme	df	Mittel der Quadrate	F	Sig.
KFT- Verbal	2-Weg- Wechselwirkungen Modell Residuen Insgesamt	(Kombiniert)	5123,073	4	1280,768	20,601	,000
		Geschlecht	1297,047	1	1297,047	20,863	,000
		Testjahrgänge	3946,130	3	1315,377	21,158	,000
		Geschlecht *	66,960	3	22,320	,359	,783
		Testjahrgänge					
			5190,033	7	741,433	11,926	,000
			136959,990	2203	62,170		
KFT- Quantitativ	2-Weg- Wechselwirkungen Modell Residuen Insgesamt	(Kombiniert)	7076,705	4	1769,176	28,647	,000
		Geschlecht	3978,422	1	3978,422	64,419	,000
		Testjahrgänge	3320,484	3	1106,828	17,922	,000
		Geschlecht *	225,828	3	75,276	1,219	,301
		Testjahrgänge					
			7302,534	7	1043,219	16,892	,000
			135929,919	2201	61,758		
KFT- Nonverbal	2-Weg- Wechselwirkungen Modell Residuen Insgesamt	(Kombiniert)	5997,651	4	1499,413	19,270	,000
		Geschlecht	325,846	1	325,846	4,188	,041
		Testjahrgänge	5739,382	3	1913,127	24,587	,000
		Geschlecht *	115,292	3	38,431	,494	,687
		Testjahrgänge					
			6112,942	7	873,277	11,223	,000
			172197,300	2213	77,812		
KFT- Gesamt- leistung	2-Weg- Wechselwirkungen Modell Residuen Insgesamt	(Kombiniert)	6572,911	4	1643,228	30,008	,000
		Geschlecht	1525,372	1	1525,372	27,855	,000
		Testjahrgänge	5202,359	3	1734,120	31,667	,000
		Geschlecht *	33,643	3	11,214	,205	,893
		Testjahrgänge					
			6606,554	7	943,793	17,235	,000
			119979,976	2191	54,760		

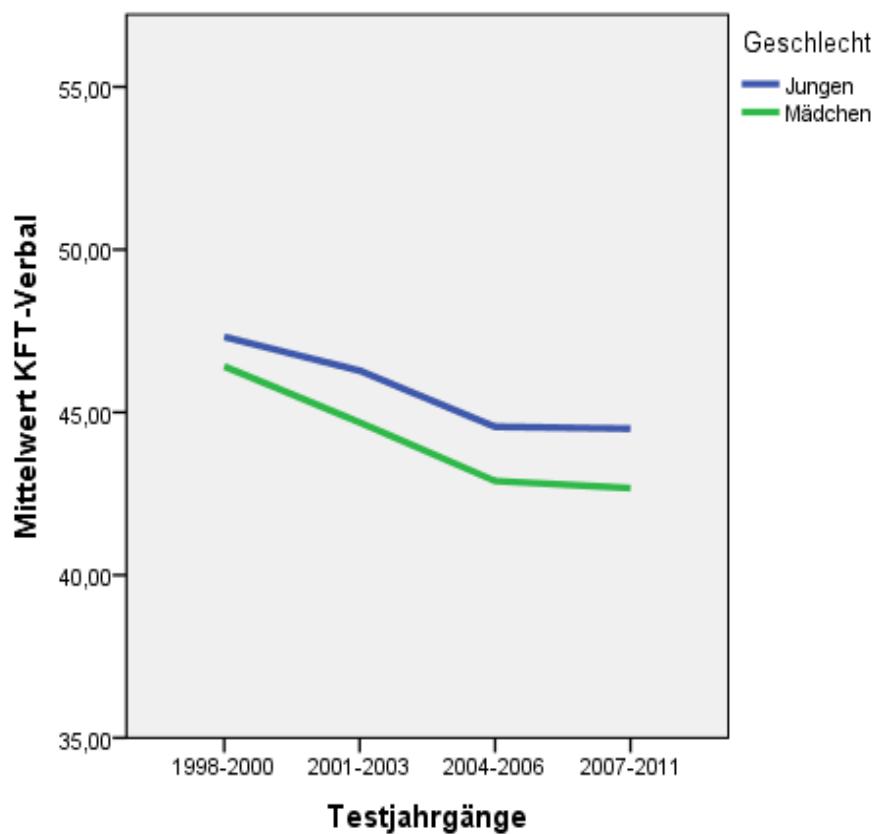


Abbildung 5.4.2.5: Entwicklung der verbalen KFT-Leistung nach Geschlecht und Jahrgangsgruppen in Klassenstufe 7

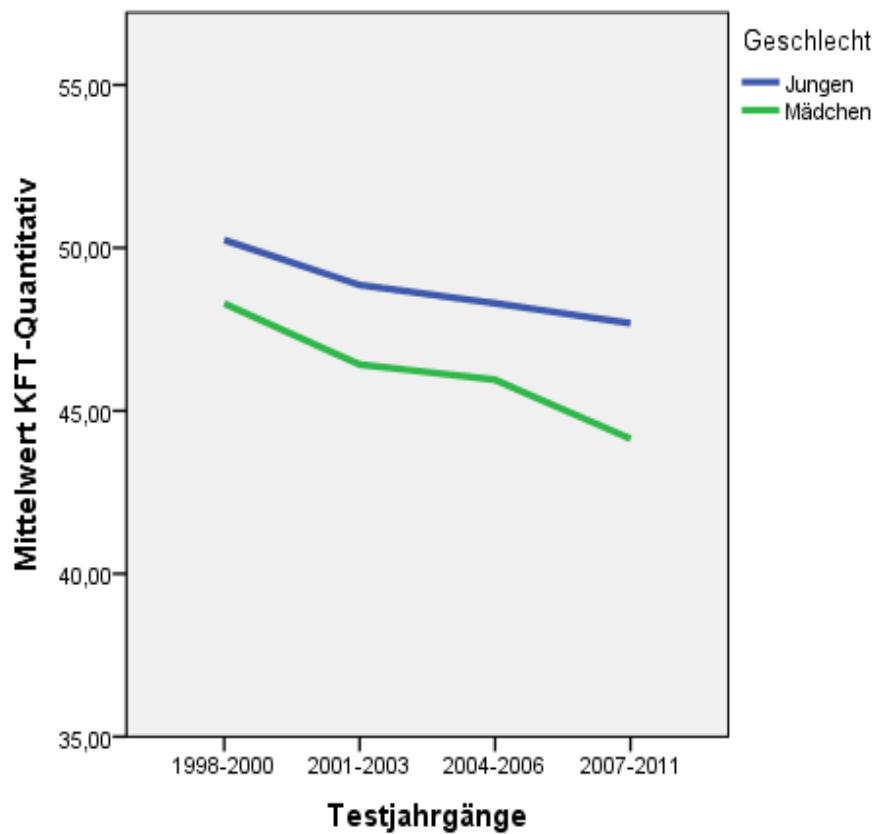


Abbildung 5.4.2.6: Entwicklung der quantitativen KFT-Leistung nach Geschlecht und Jahrgangsgruppen in Klassenstufe 7

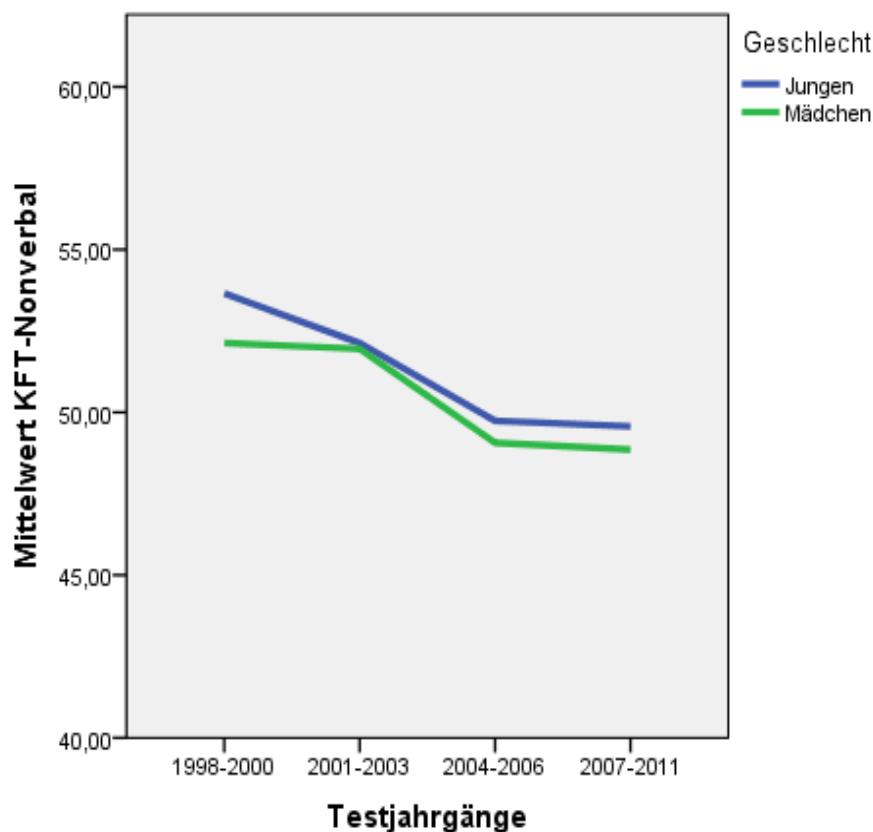


Abbildung 5.4.2.7: Entwicklung der nonverbalen KFT-Leistung nach Geschlecht und Jahrgangsgruppen in Klassenstufe 7

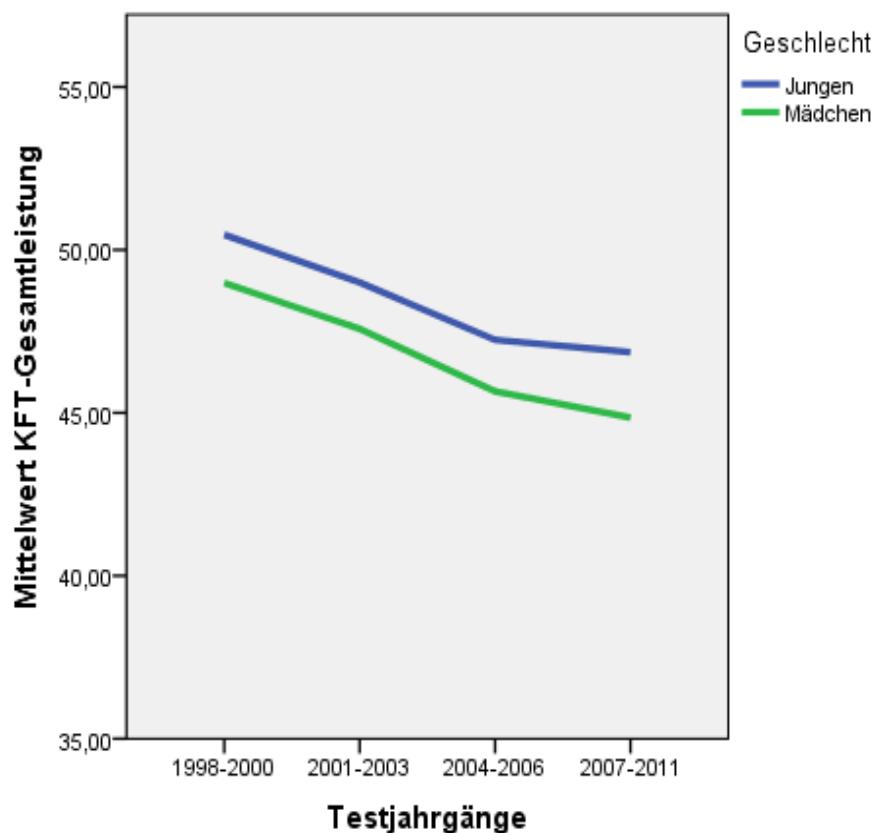


Abbildung 5.4.2.8: Entwicklung der KFT-Gesamtleistung nach Geschlecht und Jahrgangsgruppen in Klasse 7

Abbildung 5.4.2.5 zeigt, dass Jungen im KFT-V bessere Leistung als Mädchen erbracht haben. Das stimmt mit den statistischen Ergebnissen überein, dass Jungen und Mädchen sich in ihren Leistungen im KFT-V signifikant unterscheiden. Darüber hinaus gibt die Abbildung 5.4.2.5 an, dass die verbale Leistung von Mädchen und Jungen von den Jahren 1998-2000 bis zu den Jahrgängen 2007-2011 abnimmt. Dazu kommt, dass es am Anfang der Untersuchung kleine Unterschiede zwischen Jungen und Mädchen gibt, die Differenz zwischen den beiden Geschlechtern mit der Zeit aber immer größer wird.

Es fällt in der Abbildung 5.4.2.6 auf, dass Jungen deutlich besser als Mädchen im KFT-Q Teil abschneiden. Es fällt auch auf, dass die quantitative Leistung von den Jungen und Mädchen von den Jahren 1998-2000 bis zu den Jahrgängen 2007-2011 absinkt. Die quantitative Leistung von Mädchen sinkt bis 2001-2003 ab, danach bleibt sie ohne Veränderung bis 2004-2006 und nimmt dann wieder ab. Im Gegensatz zu Mädchen zeigt die quantitative Leistung von Jungen keine große Veränderung.

Die Abbildung 5.4.2.7 veranschaulicht den signifikanten Unterschied (Tabelle 5.4.2.2) zwischen den Jungen und Mädchen im KFT-N Teil. Die nonverbale Leistung von Jungen liegt über der nonverbalen Leistung der Mädchen, aber die nonverbale Leistung verschlechtert sich bei den Jungen und den Mädchen von den Jahren 1998-2000 bis zu den Jahrgängen 2007-2011. Darüber hinaus ist die Verschlechterung der nonverbalen Leistung der Jungen in den ersten Jahren 1998-2000 bis 2004-2006 größer als in den letzten Jahrgängen bzw. 2004-2006 bis 2007-2011. Bei Mädchen verändert sich die Leistung bis 2001-2003 nicht, dann nimmt sie bis 2004-2006 etwas ab, danach zeigt sie wie bei Jungen keine Veränderung bis 2007-2011.

Wenn man die Abbildung 5.4.2.8 betrachtet, fällt auf, dass Jungen in der Gesamtleistung besser als Mädchen abschneiden. Das bestätigt die statistischen Ergebnisse, dass sich Jungen und Mädchen signifikant unterscheiden. Darüber hinaus zeigt die Abbildung 5.4.2.8, dass die Gesamtleistung von Jungen und Mädchen von den Jahren 1998-2000 bis zu den Jahrgängen 2007-2011 absinkt. Dieser Befund ist plausibel, weil die verbale, quantitative und nonverbale Leistung bei den Jungen und Mädchen von 1998-2000 bis 2007-2011 abnimmt und sich die Gesamtleistung aus der Summe der KFT-Teile (KFT-V, KFT-Q und KFT-N) ergibt.

Zusammenfassend kann man zu dem Ergebnis kommen, dass Jungen in der Gesamtleistung besser als Mädchen in den beiden Altersstufen der Klassenstufen 5 und 7 abschneiden. In Bezug auf die Differenzen in den kognitiven verbalen, quantitativen und nonverbalen Fähigkeiten sind

die Ergebnisse in den beiden Klassenstufen nicht immer ähnlich. Während sich Jungen und Mädchen in Klassenstufe 5 in den verbalen und nonverbalen Teilen nicht signifikant voneinander unterscheiden, schneiden Jungen in Klassenstufe 7 in den verbalen und nonverbalen Teilen besser als Mädchen ab.

Dieses Ergebnis stimmt mit den meisten Studien insofern nicht überein, als dass Mädchen meist bessere Leistungen als Jungen im verbalen Bereich zeigen. Im KFT-Q gibt es signifikante Unterschiede zugunsten der Jungen in beiden Klassenstufen. Das bestätigt die Ergebnisse einiger Studien und stimmt mit Ergebnissen anderer Studien wiederum nicht überein (wie im theoretischen Teil gezeigt wurde). Zusammenfassend ist festzuhalten, dass die Frage von Geschlechtsunterschieden in kognitiven Fähigkeiten weiterhin umstritten bzw. in der Diskussion bleiben muss.

5.4.3 Muttersprache bzw. zu Hause überwiegend gesprochene Sprache

Hier werden die Unterschiede zwischen Kindern bezüglich der KFT 4-12+R Dimensionen und der Gesamtleistung untersucht, die unterschiedlicher Muttersprachen haben bzw. die zu Hause unterschiedliche Sprachen sprechen. Die Einteilung der zahlreichen Sprachen erfolgte dabei, wie in Abschnitt 5.2.1 erläutert, grob in regionale Sprachgruppen. Die statistischen Analysen (Varianzanalysen) wurden wieder mit der SPSS-Prozedur „ANOVA“ durchgeführt.

5.4.3.1 Klassenstufe 5

Im verbalen Teil gilt KFT-V als abhängige Variable, Jahrgänge und Sprachgruppe als Faktoren. Es zeigen sich in Tabelle 5.4.3.1 signifikante Unterschiede bei den Haupteffekten ($p < .05$), Sprachgruppe ($F = 151,09, p = .000$) und Jahrgangsgruppe ($F = 4,803, p = .002$), d.h. diese Faktoren beeinflussen die verbale Leistung. Bei der Wechselwirkung liegt keine Signifikanz ($F = 1,58, p = .114$) vor.

Im KFT-Q (KFT-Q als abhängige Variable, Jahrgangsgruppe und Sprachgruppe als Faktoren) gibt es einen signifikanten Unterschied bei den Haupteffekten ($p < .05$) hinsichtlich der beiden Faktoren Jahrgänge ($F = 3,79, p = .010$) und Sprachgruppe ($F = 4,56, p = .003$). Das bedeutet, dass Jahrgangsgruppe und Sprachgruppe einen Einfluss auf die quantitative Leistung haben. Bezuglich der Wechselwirkung ($F = 0,97, p = .459$) gibt es keine Signifikanz.

Tabelle 5.4.3.1: Varianzanalysen der KFT 4-12+R Dimensionen (KFT-V, KFT-Q, KFT-N) und der Gesamtleistung von Klassenstufe 5 nach Sprach- und Jahrgangsstufen

		Experimentelle Methode				
		Quadratsumme	df	Mittel der Quadrate	F	Sig.
KFT-Verbal	(Kombiniert)	20733,035	6	3455,506	78,828	,000
	Haupteffekte	19870,118	3	6623,373	151,095	,000
	Jahrgänge	631,592	3	210,531	4,803	,002
	2-Weg-Wechselwirkungen	624,910	9	69,434	1,584	,114
	Modell	21357,944	15	1423,863	32,482	,000
	Residuen	90783,939	2071	43,836		
KFT-Quantitativ	Insgesamt	112141,883	2086	53,759		
	(Kombiniert)	1539,249	6	256,541	4,265	,000
	Haupteffekte	823,305	3	274,435	4,563	,003
	Jahrgänge	684,111	3	228,037	3,791	,010
	2-Weg-Wechselwirkungen	527,263	9	58,585	,974	,459
	Modell	2066,512	15	137,767	2,291	,003
KFT-Nonverbal	Residuen	124204,172	2065	60,147		
	Insgesamt	126270,684	2080	60,707		
	(Kombiniert)	5253,235	6	875,539	12,801	,000
	Haupteffekte	4566,328	3	1522,109	22,254	,000
	Jahrgänge	727,688	3	242,563	3,546	,014
	2-Weg-Wechselwirkungen	562,647	9	62,516	,914	,512
KFT-Gesamtleistung	Modell	5815,882	15	387,725	5,669	,000
	Residuen	141515,358	2069	68,398		
	Insgesamt	147331,240	2084	70,696		
	(Kombiniert)	9568,809	6	1594,801	36,156	,000
	Haupteffekte	8815,850	3	2938,617	66,622	,000
	Jahrgänge	674,674	3	224,891	5,099	,002
	2-Weg-Wechselwirkungen	449,040	9	49,893	1,131	,337
	Modell	10017,849	15	667,857	15,141	,000
	Residuen	90732,484	2057	44,109		
	Insgesamt	100750,333	2072	48,625		

In Bezug auf den nonverbalen Teil (KFT-N als abhängige Variable, Jahrgangsstufe und Sprachgruppe als Faktoren) zeigen die Haupteffekte ($p < .05$) der Faktoren Jahrgänge ($F = 3,55, p = .014$) und Sprachgruppe ($F = 22,25, p = .000$), dass diese Faktoren eine Wirkung auf die nonverbale Leistung haben. Die Wechselwirkung zeigt keine Signifikanz ($F = .91, p = .512$).

In der KFT-Gesamtleistung (KFT-GL als abhängige Variable, Jahrgangsstufe und Sprachgruppe als Faktoren) haben die beiden Faktoren Sprachgruppe ($F = 66,62, p = .000$) und Jahrgangsstufe ($F = 5,099, p = .002$) einen signifikanten Einfluss von $p < .05$ auf die Gesamtleistung. Das gilt aber nicht für die Wechselwirkung ($F = 1,131, p = .337$).

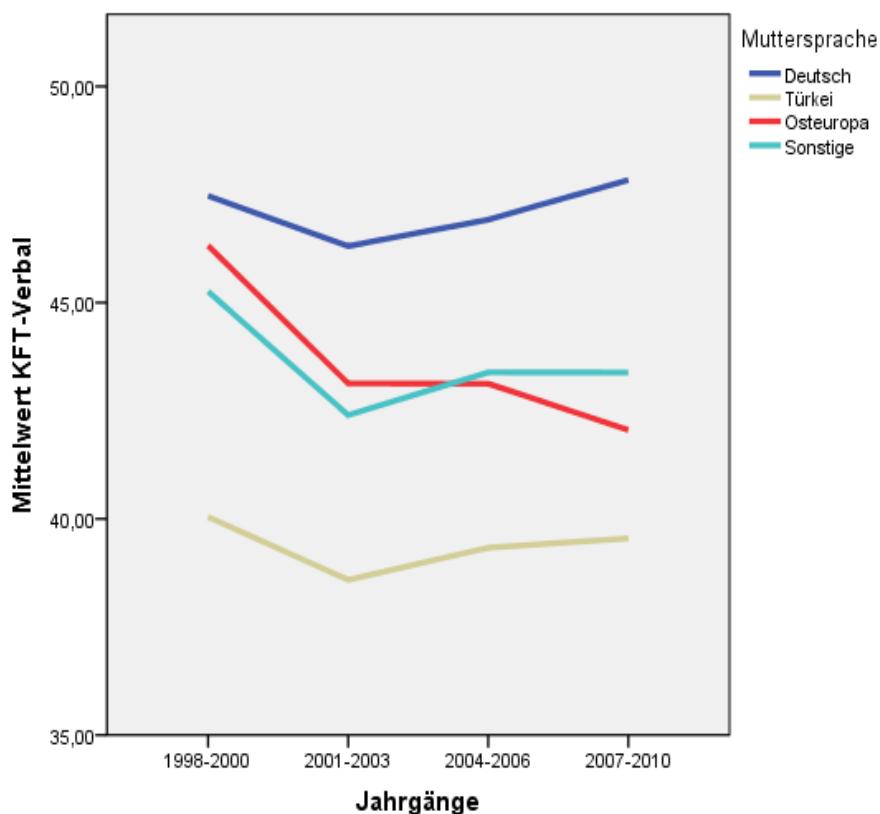


Abbildung 5.4.3.1: Verbale KFT-Leistung nach Sprach- und Jahrgangsgruppe in Klassenstufe 5

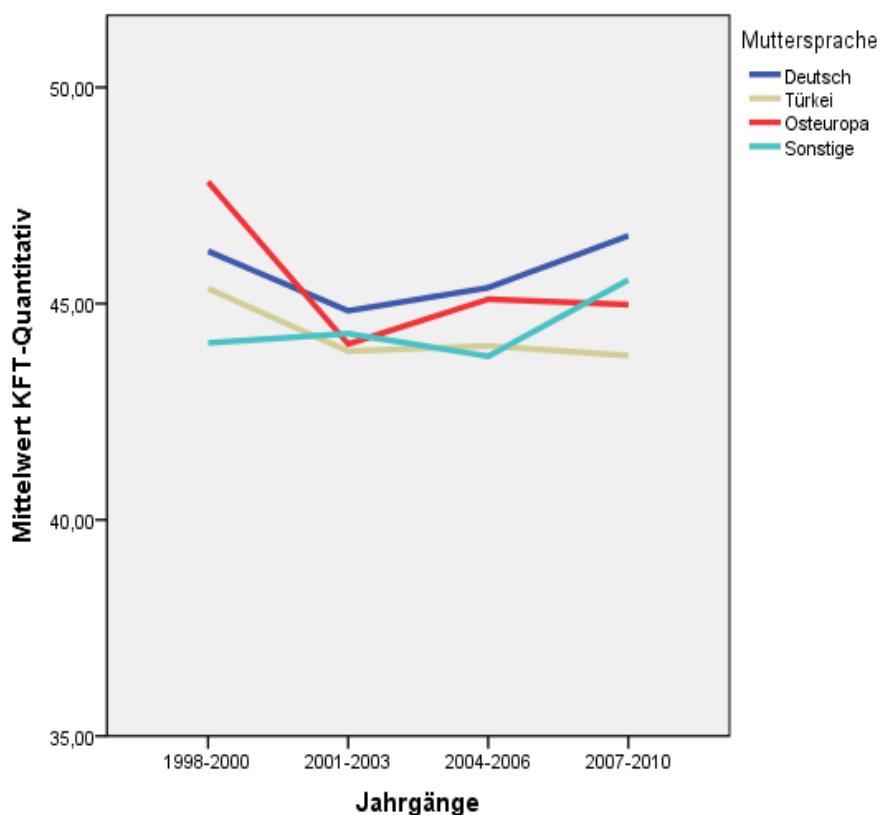


Abbildung 5.4.3.2: Quantitative KFT-Leistung nach Sprach- und Jahrgangsgruppe in Klassenstufe 5

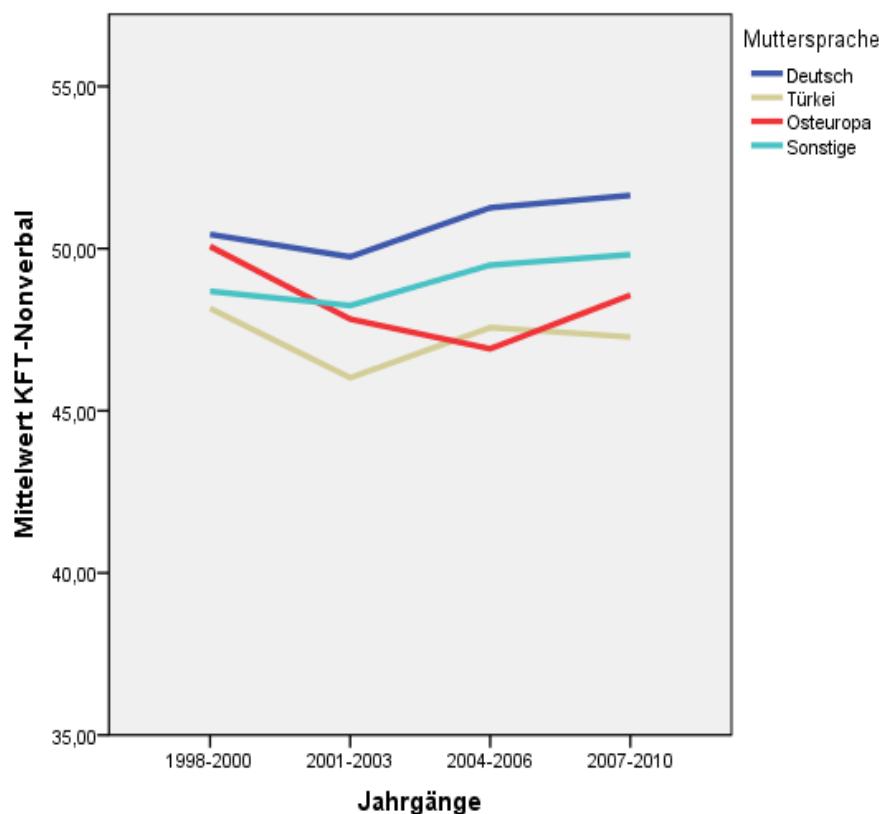


Abbildung 5.4.3.3: Nonverbale KFT-Leistung nach Sprach- und Jahrgangsgruppe in Klassenstufe 5

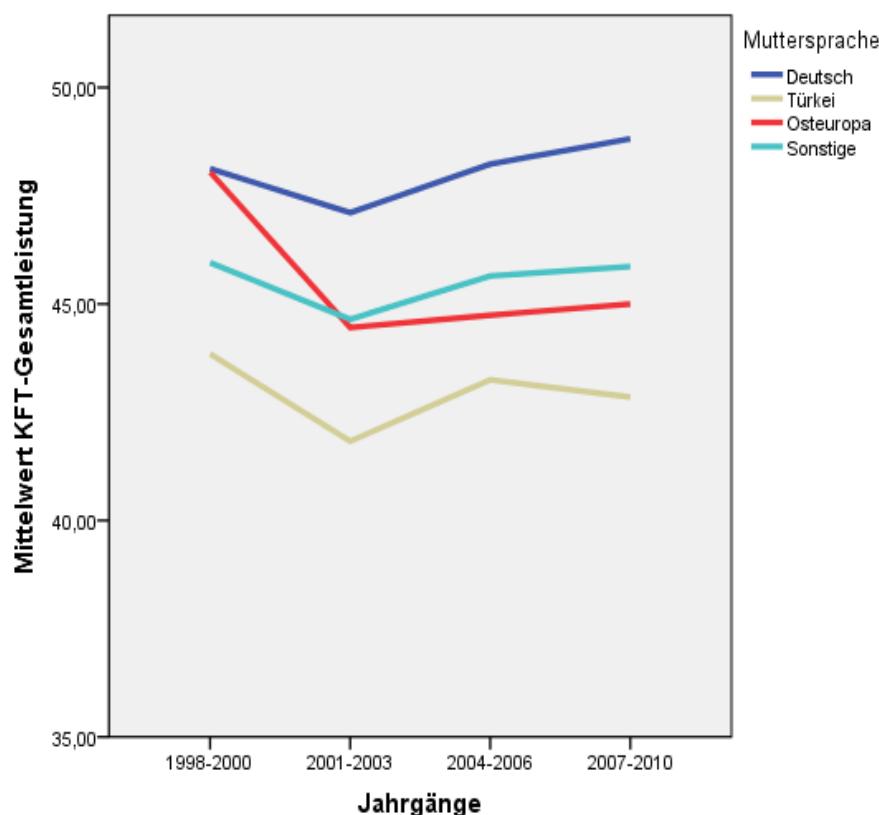


Abbildung 5.4.3.4: KFT-Gesamtleistung nach Sprach- und Jahrgangsgruppe in Klassenstufe 5

Die Ergebnisse werden durch die obigen Graphiken veranschaulicht. Abbildung 5.4.3.1 veranschaulicht die statistischen Ergebnisse, dass sich die Sprachgruppen im verbalen Teil signifikant voneinander unterscheiden. In dieser Abbildung fällt auf, dass die verbale Leistung der deutschsprachigen Schüler/innen über jenen der anderen Sprachgruppen liegt. Außerdem schneiden die türkisch-sprachigen Schüler/innen schlechter als die anderen Sprachgruppen ab. Es fällt auch auf, dass die verbalen Leistungen von deutsch- und türkischsprachigen Schüler/innen von 1998-2000 bis 2001-2003 abnehmen, danach verbessern sie sich bis 2007-2010. In Bezug auf die übrigen Sprachgruppen bzw. der Gruppe „Osteuropa“ verschlechtert sich die verbale Leistung von 1998- 2000 bis 2001-2003, dann zeigt sie kaum Veränderung bis 2004-2006 und danach wieder einen Abstieg. Bezuglich der Restgruppe „Sonstige“, die aus Sprachen verschiedener Ländern (Süd-europa, Afrika, Arabien, und Asien) besteht, sinkt die verbale Leistung von 1998-2000 bis 2001- 2003 wie bei allen anderen Sprachgruppen ab, danach nimmt sie bis 2004-2006 zu und dann zeigt sich keine Veränderung bis 2007-2010

Wenn man Abbildung 5.4.3.2 betrachtet, fällt der (signifikante, siehe Tabelle 5.4.3.1) Unterschied zwischen den Sprachgruppen bezüglich der quantitativen Leistungen auf. Die quantitativen Leistungen der deutschsprachigen Schüler/innen liegen meist über jenen der anderen Sprachgruppen und jene der türkischsprachigen Schüler/innen unter den Leistungen aller anderen Sprachgruppen. Außerdem fällt auf, dass die Differenzen zwischen den Sprachgruppen im quantitativen Teil nicht so groß wie jene im verbalen und nonverbalen Teil sind.

Die Abbildung 5.4.3.3 zeigt, dass die deutschsprachigen Schüler/innen die beste Leistung im nonverbalen Teil im Vergleich zu den anderen Gruppen erbracht haben. Außerdem kann man der Abbildung 5.4.3.3 entnehmen, dass die türkischsprachigen Schüler/innen meist am schlechtesten abschneiden. Das stimmt mit den statistischen Ergebnissen darin überein, dass die Sprachgruppe die nonverbale Leistung signifikant beeinflusst. Es fällt in der Abbildung 5.4.3.3 auch auf, dass sich die nonverbale Leistung über alle Jahrgänge verändert (signifikant nach Tabelle 5.4.3.1). Die nonverbalen Leistungen von deutschsprachigen Schüler/innen und der Restgruppe „Sonstige“ ist im Vergleich zu jenen der anderen Sprachgruppen stabiler. Bei der Gruppe Osteuropa verschlechtert sich die nonverbale Leistung bis 2004/06 und verbessert sich wieder, bei türkischsprachigen Schüler gibt es einen Abstieg bis 2001-2003, dann einen Anstieg.

Abbildung 5.4.3.4 gibt an, dass sich die Sprachgruppen bezüglich der Gesamtleistung von den Jahren 1998-2000 bis zu den Jahrgängen 2007-2010 signifikant voneinander unterscheiden. Die Gesamtleistung von deutschsprachigen Schüler/innen liegt über jenen der anderen Sprach-

gruppen und die Gesamtleistung von türkischsprachigen Schüler/innen unter jenen der anderen Sprachgruppen. Zusätzlich zeigt die Gesamtleistung von allen Sprachgruppen eine Verschlechterung von 1998-2000 bis 2001-2003, dann gibt es eine Verbesserung fast bei allen Gruppen, besonders bei deutschsprachigen Schüler/innen und der Gruppe Sonstige bis 2007-2010. Bei türkischsprachigen Schüler/innen schwankt die Gesamtleistung.

Tabelle 5.4.3.2: Varianzanalysen der KFT 4-12+R Dimensionen (KFT-V, KFT-Q, KFT-N) und der Gesamtleistung von Klassenstufe 7 nach Sprach- und Jahrgangsstufen

		Experimentelle Methode				
		Quadratsumme	df	Mittel der Quadrate	F	Sig.
KFT-Verbal	(Kombiniert)	30592,461	6	5098,743	101,636	,000
	Haupteffekte	26849,445	3	8949,815	178,402	,000
	Testjahrgänge	2525,467	3	841,822	16,781	,000
	2-Weg-Wechselwirkungen	479,456	9	53,273	1,062	,388
	Modell	31071,917	15	2071,461	41,292	,000
	Residuen	110215,863	2197	50,167		
KFT-Quantitativ	Insgesamt	141287,780	2212	63,873		
	(Kombiniert)	5855,325	6	975,887	15,701	,000
	Haupteffekte	2767,554	3	922,518	14,842	,000
	Testjahrgänge	2779,426	3	926,475	14,906	,000
	2-Weg-Wechselwirkungen	167,920	9	18,658	,300	,975
	Modell	6023,244	15	401,550	6,461	,000
KFT-Nonverbal	Residuen	136428,656	2195	62,154		
	Insgesamt	142451,900	2210	64,458		
	(Kombiniert)	8921,123	6	1486,854	19,543	,000
	Haupteffekte	3352,145	3	1117,382	14,687	,000
	Testjahrgänge	5068,136	3	1689,379	22,205	,000
	2-Weg-Wechselwirkungen	822,686	9	91,410	1,201	,289
KFT-Gesamtleistung	Modell	9743,809	15	649,587	8,538	,000
	Residuen	167907,941	2207	76,080		
	Insgesamt	177651,750	2222	79,951		
	(Kombiniert)	15281,963	6	2546,994	50,796	,000
	Haupteffekte	10320,925	3	3440,308	68,612	,000
	Testjahrgänge	4105,503	3	1368,501	27,293	,000
	2-Weg-Wechselwirkungen	524,078	9	58,231	1,161	,316
	Modell	15806,040	15	1053,736	21,015	,000
	Residuen	109559,168	2185	50,141		
	Insgesamt	125365,208	2200	56,984		

5.4.3.2 Klassenstufe 7

In Tabelle 5.4.3.2 zeigt sich, dass es einen signifikanten Unterschied bei den Haupteffekten ($p < .05$) von den beiden Faktoren Testjahrgänge und Sprachgruppe bei allen abhängigen Variablen

(KFT-V, KFT-Q, KFT-N und KFT-GL) gibt. d.h. die Faktoren Testjahrgänge und Sprachgruppe haben einen Einfluss auf die verbalen, quantitativen und nonverbalen Fähigkeiten sowie die Gesamtleistung im KFT-4-12+R. Darüber hinaus ergaben sich hinsichtlich der Wechselwirkungen der beiden Faktoren bei keiner der abhängigen Variablen signifikante Ergebnisse.

Im KFT-V ergaben sich beim Faktor Jahrgangsgruppe ein F-Wert von $F = 16,781$ ($p = .000$), beim Faktor Sprachgruppe ein F-Wert von $F = 178,402$ ($p = .000$) und bei der Wechselwirkung ein F-Wert von $F = 1,06$ ($p = .388$). Im KFT-Q und beim Faktor Jahrgangsgruppe ist $F = 14,91$, $p = .000$, beim Faktor Sprachgruppe ist $F = 14,84$, $p = .000$, und bei der Wechselwirkung ist $F = 0,300$, $p = .975$. Bezuglich des KFT-N und beim Faktor Jahrgangsgruppe ergibt sich ein F-Wert von $F = 22,21$, $p = .000$, beim Faktor Sprachgruppe einer von $F = 14,69$, $p = .000$, und bei der Wechselwirkung ein F-Wert von $F = 1,20$, $p = .289$. In Bezug auf die KFT-Gesamtleistung KFT-GL ist beim Faktor Jahrgangsgruppe $F = 27,29$, $p = .000$, beim Faktor Sprachgruppe ist $F = 68,61$, $p = .000$, und bei der Wechselwirkung $F = 1,16$, $p = .316$.

Um die Ergebnisse besser zu verdeutlichen, werden im Folgenden die Befunde grafisch präsentiert.

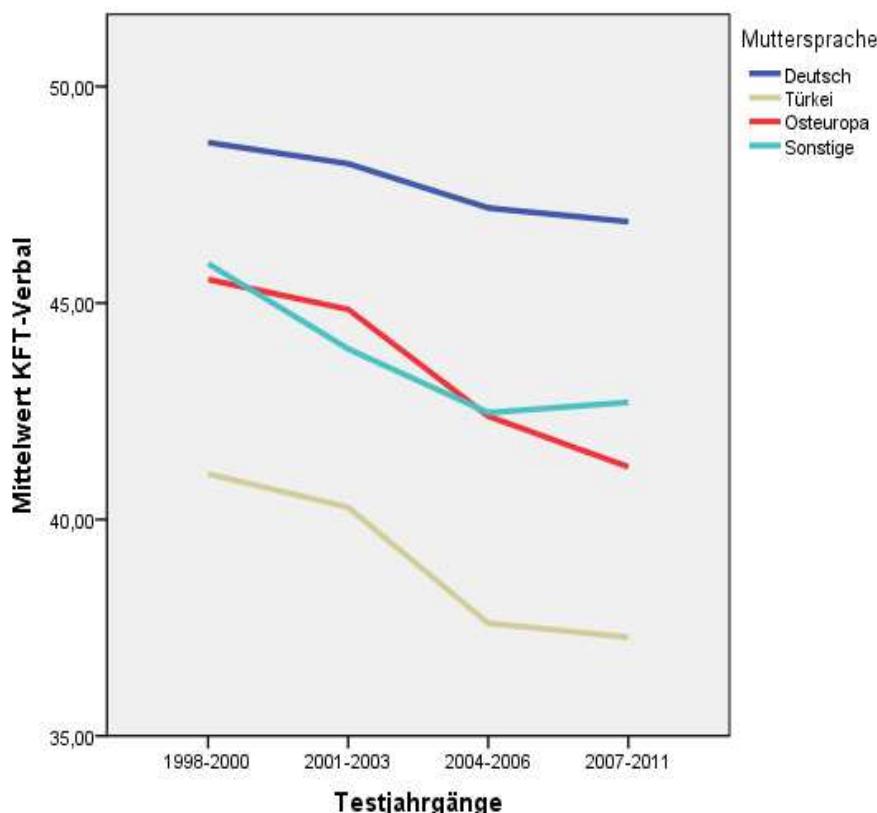


Abbildung 5.4.3.5: Verbale KFT-Leistung nach Sprach- und Jahrgangsgruppe in Klassenstufe 7

Es fällt in der Abbildung 5.4.3.5 auf, dass die verbale Leistung von deutschsprachigen Schüler/innen über jener von allen anderen Gruppen liegt. Darüber hinaus schneiden die türkischsprachigen Schüler/innen schlechter als die übrigen Sprachgruppen ab. Das steht mit den statistischen Ergebnissen in Übereinstimmung, dass sich die Sprachgruppen im verbalen Teil signifikant unterscheiden. Außerdem verschlechtert sich die verbale Leistung in allen Sprachgruppen von den Jahren 1998-2000 bis zu den Jahrgängen 2007-2011.

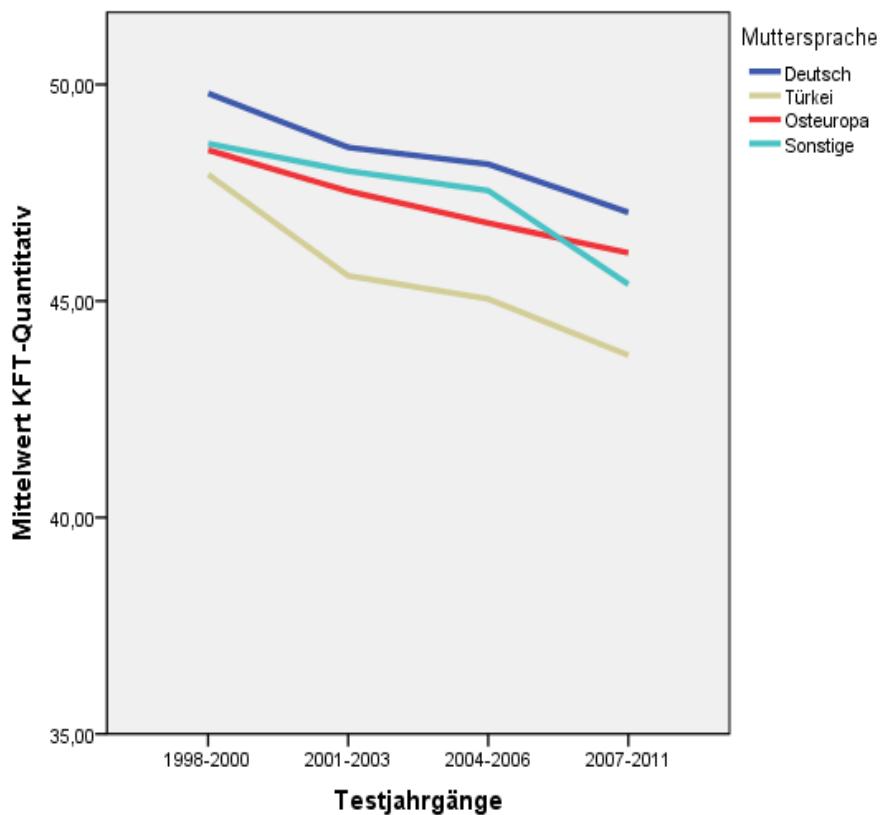


Abbildung 5.4.3.6: Quantitative KFT-Leistung nach Sprach- und Jahrgangsgruppe in Klassenstufe 7

Die Abbildung 5.4.3.6 zeigt, dass die deutschsprachigen Schüler/innen bessere quantitative Leistung als die anderen Sprachgruppen erbracht haben. Die quantitative Leistung von türkischsprachigen Schüler/innen liegt unter jener aller übrigen Gruppen. Dazu kommt, dass die quantitative Leistung von allen Sprachgruppen von 1998-2000 bis 2007-2011 absteigt, d.h. die Sprachgruppe und die Testjahrgänge haben einen Einfluss auf die quantitative Leistung.

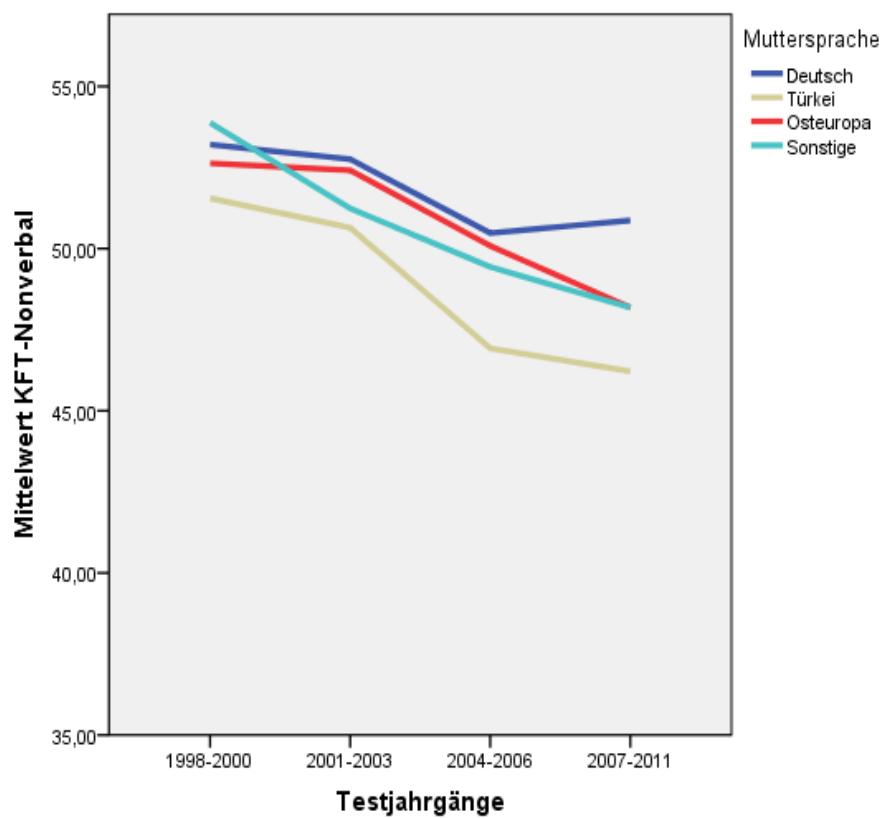


Abbildung 5.4.3.7: Nonverbale KFT-Leistung nach Sprach- und Jahrgangsgruppe in Klassenstufe 7

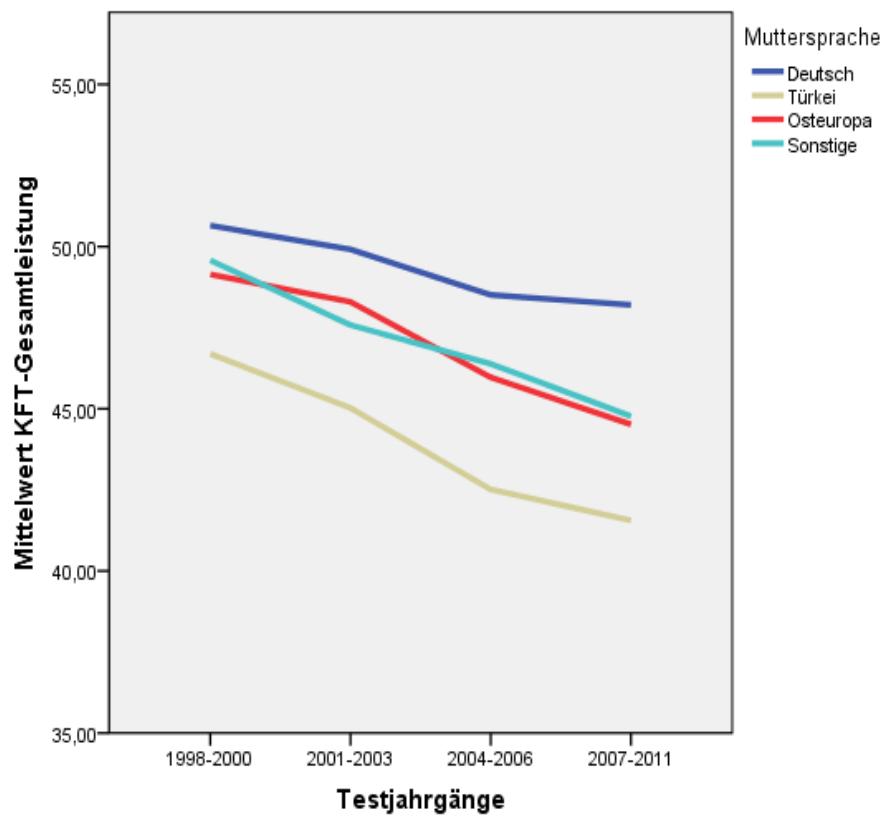


Abbildung 5.4.3.8: KFT-Gesamtleistung nach Sprach- und Jahrgangsgruppe in Klassenstufe 7

In der Abbildung 5.4.3.7 fällt auf, dass die Schüler aller Sprachgruppen bessere Leistung im nonverbalen Teil als im verbalen und quantitativen Teil erbracht haben. Es fällt auch auf, dass die nonverbale Leistung von deutschsprachigen Schülern/innen meist über jener der anderen Sprachgruppen liegt. Außerdem schneiden die türkischsprachigen Schüler/innen deutlich schlechter als die anderen Gruppen ab, wie in der verbalen und quantitativen Leistung auch. Darüber hinaus verschlechtert sich die nonverbale Leistung von allen Sprachgruppen von den Jahren 1998-2000 bis 2007-2011. Das bestätigt die statistischen Ergebnisse, dass die Differenzen zwischen den Sprachgruppen von 1998-2000 bis 2007-2011 signifikant sind.

Die deutschsprachigen Schüler/innen erbrachten bessere Leistungen im verbalen, quantitativen und nonverbalen Teil als die anderen Sprachgruppen. Außerdem schnitten die türkischsprachigen Schüler/innen deutlich schlechter als die anderen Gruppen im verbalen, quantitativen und nonverbalen Teil ab. Darüber hinaus haben die KFT-Leistungen bei allen Sprachgruppen über alle Jahrgänge hinweg abgenommen. Diese Ergebnisse gelten auch hier für die Gesamtleistung von allen Gruppen, wie Abbildung 5.4.3.8 zeigt.

Aus den statistischen Ergebnissen der fünften und siebten Klassenstufen, die in den Tabellen 5.4.3.1 und 5.4.3.2. zusammengefasst sind und in den Abbildungen 5.4.3.1 bis 5.4.3.8 veranschaulicht werden, kann man schließen, dass die Sprachgruppe bzw. die Sprache, die die Schüler/innen zu Hause mit ihren Eltern am häufigsten sprechen, die verbale, nonverbale, quantitative und die Gesamtleistung im KFT 4-12+ R von den Jahrgängen 1998-2000 bis zu den Jahren 2007-2011 in beiden Klassenstufen 5 und 7 beeinflusst. Außerdem schneiden die Sprachgruppen in den beiden Teilstichproben der Fünft- und Siebtklässler bezüglich der KFT-Skalen im Vergleich schlechter als die deutschsprachigen Schüler/innen ab. Dazu kommt, dass die türkischsprachigen Schüler/innen die schlechtesten Leistungen im Vergleich zu den anderen Sprachgruppen erbracht haben.

Diese Befunde bestätigen was im theoretischen Teil über die ethnischen Unterschiede erläutert wurde, nämlich dass sich im deutschen Schulsystem unterdurchschnittlicher oder schlechtere Leistungen bei ausländischen Schülern/innen bzw. Schülern/innen mit Migrationshintergrund führt. Die Unterschiede zwischen deutschen (bzw. hier: deutschsprachigen) und Schülern mit Migrationshintergrund (hier: Kinder, die zu Hause nicht überwiegend Deutsch sprechen; nicht-deutsche Sprachgruppen) könnte man durch kulturelle Defizite (vgl. Diefenbach, 2007), wie im theoretischen Teil erläutert, erklären, wenn sich Unterschiede nur zwischen deutsch- und türkischsprachigen Schüler/innen zeigen würden. Es ergaben sich aber auch Unterschiede

zwischen deutschsprachigen Schülern/innen und solchen aus Sprachgruppe Europa, die zumindest teilweise über einen ähnlichen kulturellen Hintergrund verfügen müssten. Bei aller Vorsicht, weil die Einteilung in die Sprachgruppen auf groben, durch Lehrkräfte erhobenen Angaben der Kinder beruht und sicher nicht sehr belastbar bzw. valide sind, liegt der Schluss nahe, dass auch andere Faktoren diese Differenzen verursacht haben könnten, wie z.B. der sozioökonomische Status (Wohnort, Schulabschluss, Beruf und Einkommen der Eltern usw.), der nachweislich einen Einfluss auf den Schulerfolg hat.

5.4.4 Veränderungen in den kognitiven Fähigkeiten von Klassenstufe 5 zu Klassenstufe 7

5.4.4.1 Veränderungen der KFT-Leistungen für unterschiedliche Kohorten (Jahrgangskohorten)

Für die Analyse der Veränderung der intellektuellen Fähigkeiten von Klasse 5 zu Klasse 7, erfasst mit den Dimensionen (KFT-V, KFT-Q, KFT-N) und der Gesamtleistung des KFT 4-12+R, wurden Differenzwerte für die normierten Skalen berechnet (diese werden im Folgenden als DiffJG-V, DiffJG-Q, DiffJG-N und DiffJG-GL bezeichnet) die dann mittels einfaktorieller Varianzanalyse (SPSS-Prozedur „ONEWAY“) auf signifikante Unterschiede überprüft wurden. Dieses Vorgehen liefert im Fall von zwei Messzeitpunkten dieselben Ergebnisse wie eine Varianzanalyse mit Messwiederholung.

Tabelle 5.4.4.1: Varianzanalysen der Differenzen der Leistungen im KFT 4-12+R in den Dimensionen (KFT-V, KFT-Q, KFT-N) und der Gesamtleistung (KFT-GL) von Klassenstufe 5 zu Klassenstufe 7 nach Jahrgangskohorten

	Quadratsumme	df	Mittel der Quadrate	F	Signifikanz
DiffJG-V	Zwischen den Gruppen	2281,555	3	760,518	,000
	Innerhalb der Gruppen	61853,763	1578	39,198	
	Gesamt	64135,318	1581		
DiffJG-Q	Zwischen den Gruppen	2561,391	3	853,797	,000
	Innerhalb der Gruppen	121514,354	1581	76,859	
	Gesamt	124075,745	1584		
DiffJG-N	Zwischen den Gruppen	5654,684	3	1884,895	,000
	Innerhalb der Gruppen	105399,434	1589	66,331	
	Gesamt	111054,118	1592		
DiffJG-GL	Zwischen den Gruppen	5298,208	3	1766,069	,000
	Innerhalb der Gruppen	96603,912	1575	61,336	
	Gesamt	101902,120	1578		

Legende: Siehe Tabelle 5.4.4.2.

Tabelle 5.4.4.2: Mittelwerte der Differenzen bezüglich der KFT 4-12+R Dimensionen (V, Q, N) und der Gesamtleistung der Schüler von Klasse 5 zu Klasse 7 nach Jahrgängen

	N	Mittelwert	Standardabweichung	Standardfehler	95%-Konfidenzintervall für den Mittelwert		Minimum	Maximum	
					Untergrenze	Obergrenze			
DiffJG-V	1998-2000	405	1,5383	5,52731	,27465	,9983	2,0782	-15,00	20,00
	2001-2003	425	-,1624	6,54804	,31763	-,7867	,4620	-28,00	63,00
	2004-2006	384	-1,7708	7,28241	,37163	-2,5015	-1,0401	-62,00	52,00
	2007-2010	368	-,7582	5,47958	,28564	-1,3199	-,1964	-16,00	16,00
	Gesamt	1582	-,2560	6,36917	,16013	-,5701	,0581	-62,00	63,00
	1998-2000	406	3,2709	8,41313	,41754	2,4501	4,0917	-19,00	59,00
DiffJG-Q	2001-2003	425	1,4165	8,18669	,39711	,6359	2,1970	-53,00	48,00
	2004-2006	386	1,2824	9,69271	,49335	,3124	2,2524	-66,00	48,00
	2007-2010	368	-,3587	8,77522	,45744	-1,2582	,5408	-58,00	53,00
	Gesamt	1585	1,4467	8,85046	,22231	1,0106	1,8827	-66,00	59,00
	1998-2000	406	3,5222	7,65898	,38011	2,7749	4,2694	-18,00	45,00
	2001-2003	425	1,0612	7,51608	,36458	,3446	1,7778	-54,00	26,00
DiffJG-N	2004-2006	386	-,6269	7,62400	,38805	-1,3899	,1360	-50,00	22,00
	2007-2010	376	-1,4176	9,70381	,50044	-2,4016	-,4335	-60,00	21,00
	Gesamt	1593	,6943	8,35210	,20926	,2838	1,1047	-60,00	45,00
	1998-2000	406	3,2537	6,18778	,30709	2,6500	3,8574	-9,00	50,00
	2001-2003	425	,6729	7,71967	,37446	-,0631	1,4090	-58,00	68,00
	2004-2006	384	-1,0078	8,95871	,45717	-1,9067	-,1089	-69,00	51,00
DiffJG-GL	2007-2010	364	-1,4176	8,31384	,43576	-2,2745	-,5606	-67,00	50,00
	Gesamt	1579	,4459	8,03597	,20223	,0492	,8425	-69,00	68,00

Legende: DiffJG-V: Differenzen zwischen KFT-V von Klasse 7 und KFT-V von Klasse 5;
 DiffJG-Q: Differenzen zwischen KFT-Q von Klasse 7 und KFT-Q von Klasse 5;
 DiffJG-N: Differenzen zwischen KFT-N von Klasse 7 und KFT-N von Klasse 5;
 DiffJG-GL: Differenzen zwischen KFT-GL von Klasse 7 und KFT-GL von Klasse 5.

Die Durchführung der Varianzanalysen mit den Jahrgangsgruppen als unabhängigen Faktoren lieferte auf dem 5%-Niveau signifikante Ergebnisse für die verbalen KFT-Leistungen DiffJG-V ($F = 19,40, p = .000$), die quantitativen KFT-Leistungen DiffJG-Q ($F = 11,109, p = .000$), die nonverbalen KFT-Leistungen DiffJG-N ($F = 28,42, p = .000$) und die KFT-Gesamtleistung DiffJG-GL ($F = 28,79, p = .000$) (Tabelle 5.4.4.1, Tabelle 5.4.4.2 enthält die zugehörigen Differenzwerte in T-Wert-Einheiten). Darüber hinaus erbrachten die Post-Hoc-Tests (Scheffe)

auf dem 5%-Niveau bezüglich der Differenzen der KFT- Dimensionen (KFT-V, KFT-Q, KFT-N) und der KFT-Gesamtleistung drei homogene Untergruppen (Tabellen 5.4.4.3 bis 5.4.4.6 unten), d. h. die Post-Hoc-Analysen bestätigten die Ergebnisse der Varianzanalyse.

Tabelle 5.4.4.3: Post-Hoc-Test „homogene Untergruppen“ der verschiedenen Jahrgänge hinsichtlich der Differenzen im KFT-V von Klassenstufe 5 zu Klassenstufe 7

Jahrgänge	N	Scheffé-Prozedur		
		Untergruppe für Alpha = ,050.		
		1	2	3
2004-2006	384	-1,7708		
2007-2010	368	-,7582	-,7582	
2001-2003	425		-,1624	
1998-2000	405	,161	,618	1,5383
Signifikanz				1,000

In Bezug auf die Differenzen im KFT-V lieferte der Post-Hoc-Test drei homogene Untergruppen, die sich voneinander signifikant unterscheiden. Ansonsten wiesen die Jahrgänge pro Gruppe untereinander keinen signifikanten Unterschied auf. Die erste Gruppe besteht aus den Jahrgängen 2004-2006 und 2007-2010, die sich von den übrigen Gruppen signifikant unterscheiden (Tabelle 5.4.4.3, Spalte 1). Die Jahrgänge 2007-2010 und 2001-2003 bilden die zweite Gruppe. Die Jahre 2001-2003 (zweite Gruppe) unterscheiden sich von den Jahrgängen 2004-2006 (erste Gruppe) und 1998-2000 (dritte Gruppe) (Tabelle 5.4.4.3, Spalte 2). Die Tabelle 5.4.4.3, Spalte 3 gibt an, dass sich die Jahrgänge 1998-2000 (dritte Gruppe) von den Jahren 2004-2006, 2007-2010 (erste Gruppe), sowie 2001-2003 (zweite Gruppe) unterscheiden.

Tabelle 5.4.4.4: Post-Hoc-Test „homogene Untergruppen“ der verschiedenen Jahrgänge hinsichtlich der Differenzen im KFT-Q von Klassenstufe 5 zu Klassenstufe 7

Jahrgänge	N	Scheffé-Prozedur		
		Untergruppe für Alpha = ,050.		
		1	2	3
2007-2010	368	-,3587		
2004-2006	386	1,2824	1,2824	
2001-2003	425		1,4165	
1998-2000	406	,075	,997	3,2709
Signifikanz				1,000

Anhand der Post-Hoc-Analyse (Scheffé) auf dem 5%-Niveau können bezüglich der Differenzen im KFT-Q drei homogene Untergruppen unterschieden werden, die sich signifikant voneinander unterscheiden und die untereinander keinen signifikanten Unterschied aufweisen. Die signifikante Veränderung lässt sich zurückführen auf die systematischen Differenzen zwischen der ersten Gruppe, die die Jahrgänge 2007-2010 und 2004-2006 umfasst, und den anderen beiden

Gruppen bzw. die Jahrgänge 2004-2006 und 2001-2003 (zweite Gruppe), und 1998-2000 (dritte Gruppe) (Tabelle 5.4.4.4, Spalte 1), sowie die Unterschiede zwischen den Jahrgängen 2001-2003 (zweite Gruppe), den Jahrgängen 2007-2010 (erste Gruppe) und den Jahren 1998-2000 (dritte Gruppe) (Tabelle 5.4.4.4, Spalte 2) und auch auf die Differenzen zwischen den Jahrgängen 1998-2000 (dritte Gruppe), den Jahren 2007-2010 und 2004-2006 (erste Gruppe) sowie 2001-2003 (zweite Gruppe) (Tabelle 5.4.4.4, Spalte 3). Ansonsten weisen die Jahrgänge pro Gruppe untereinander keinen signifikanten Unterschied auf.

Tabelle 5.4.4.5: Post-Hoc-Test „homogene Untergruppen“ der verschiedenen Jahrgänge hinsichtlich der Differenzen im KFT-N von Klassenstufe 5 zu Klassenstufe 7

Jahrgänge	N	Scheffé-Prozedur		
		Untergruppe für Alpha = ,050.		
		1	2	3
2007-2010	376	-1,4176		
2004-2006	386	-,6269		
2001-2003	425		1,0612	
1998-2000	406	,599	1,000	3,5222
Signifikanz				1,000

Tabelle 5.4.4.6: Post-Hoc-Test „homogene Untergruppen“ der verschiedenen Jahrgänge hinsichtlich der Differenzen im KFT-GL von Klassenstufe 5 zu Klassenstufe 7

Jahrgänge	N	Scheffé-Prozedur		
		Untergruppe für Alpha = ,050.		
		1	2	3
2007-2010	364	-1,4176		
2004-2006	384	-1,0078		
2001-2003	425		,6729	
1998-2000	406	,910	1,000	3,2537
Signifikanz				1,000

Bezüglich der Differenzen im KFT-N und KFT-GL lieferten die Post-Hoc-Tests (Scheffé) auf dem 5%-Niveau ebenfalls drei homogene Untergruppen. Die erste Gruppe umfasst die Jahrgänge 2007-2010 und 2004-2006, die Jahrgänge 2001-2003 bilden die zweite Gruppe. Die dritte Gruppe besteht aus den Jahrgängen 1998-2000. Diese verschiedenen Gruppen unterscheiden sich voneinander signifikant und zeigen untereinander keine signifikante Differenz (Tabelle 5.4.4.5 und 5.4.4.6, Spalte 1). Darüber hinaus unterscheiden sich die Jahrgänge 2001-2003 (zweite Gruppe) von den Jahrgängen 2007-2010 und 2004-2006 (erste Gruppe), sowie von 1998-2000 (dritte Gruppe) (Tabelle 5.4.4.5 und 5.4.4.6, Spalte, 2). Zusätzlich unterscheiden sich die Jahrgänge 1998-2000 (dritte Gruppe) signifikant von den Jahrgängen 2007-2010, 2004-2006 (erste Gruppe) und 2001-2003 (zweite Gruppe) (Tabelle 5.4.4.5 und 5.4.4.6, Spalte 3).

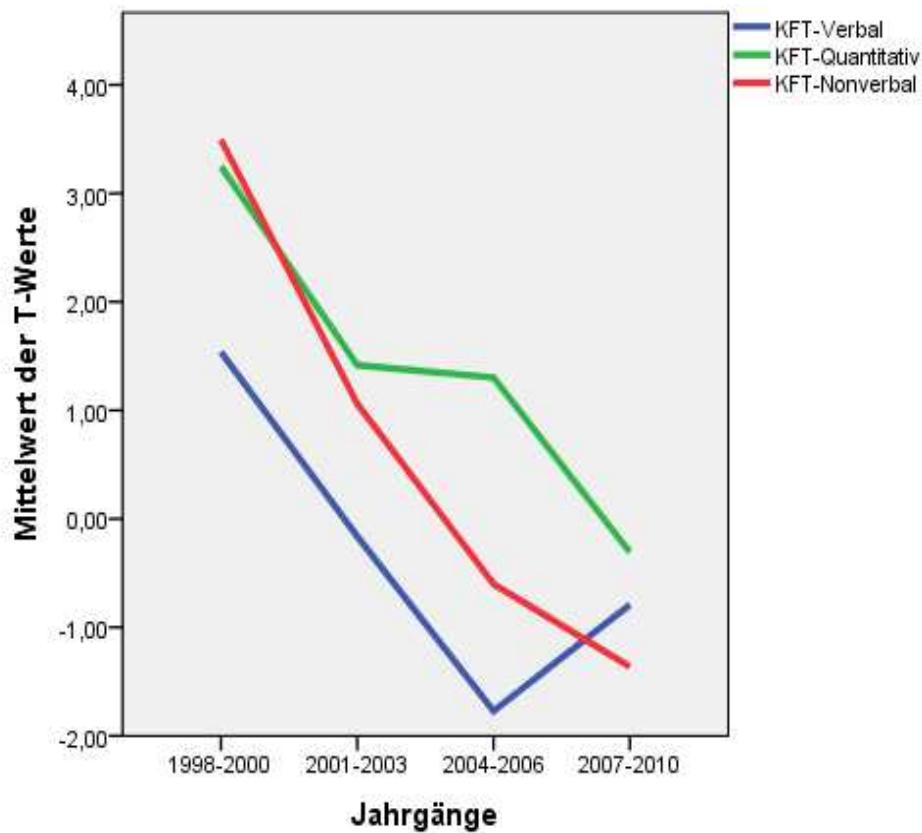


Abbildung 5.4.4.1: Veränderung der KFT-Leistungen (Dimensionen) von Klassenstufe 5 bis Klassenstufe 7 für die Jahrgangsgruppen

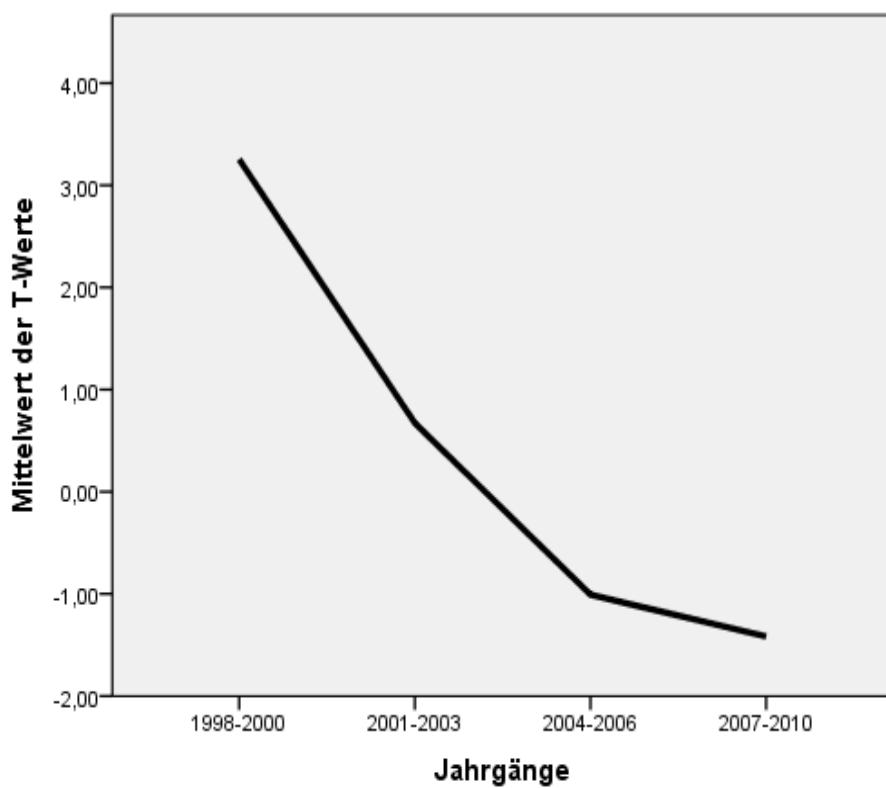


Abbildung 5.4.4.2: Veränderung der KFT-Gesamtleistung von Klassenstufe 5 bis Klassenstufe 7 für die Jahrgangsgruppen

Die Abbildung 5.4.4.1 bestätigt die statistischen Ergebnisse, dass sich die KFT-Leistungen signifikant von den Jahrgängen 1998-2000 bis zu den Jahren 2007-2010 unterscheiden. In der quantitativen Leistung steigern sich die Schülerinnen und Schüler ab 2001-2003 bis 2004-2006 relativ am meisten, die geringste Steigerung bzw. ein leichtes Absinken (2004-2006) ist dagegen für die verbale Leistung zu registrieren. Es fällt in der Abbildung 5.4.4.1 auf, dass die KFT-Leistungen bzw. die verbalen, quantitativen und nonverbalen Leistungen sich anfangs des Untersuchungszeitraums durchweg von Klasse 5 bis Klasse 7 um 3 bis 4 T-Werte verbessern, während ab Mitte des Untersuchungszeitraums praktisch keine Verbesserungen mehr, sondern eher geringe Verschlechterungen auftreten. Diese bleiben allerdings kleiner als 2 T-Wert-Punkte bzw. 0,2 Standardabweichungen und damit im Sinne von Cohen wenig bedeutend (vgl. Cohen, 1977; Cohen, 1992; Bortz, 1995). Abbildung 5.4.4.2 zeigt ein entsprechendes Bild für die Veränderung der Gesamtleistung von Klasse 5 bis Klasse 7. Es zeigt sich bereits in der Gruppe 2001-2003 praktisch keine nennenswerte Verbesserung der KFT-Gesamtleistung von Klasse 5 zu Klasse 7 mehr.

5.4.4.2 Veränderungen der KFT-Leistungen für unterschiedliche Sprachgruppen

Die Veränderungen hinsichtlich der KFT 4-12+R Dimensionen (KFT-V, KFT-Q, KFT-N) und der KFT-Gesamtleistung von Klasse 5 zu Klasse 7 wurden nicht nur nach Kohorten bzw. Jahrgangsgruppen, sondern auch nach Sprachgruppen (Muttersprache bzw. zu Hause überwiegend gesprochene Sprache) analysiert. Um diese Veränderungen oder Entwicklungen zu ermitteln, wurden wieder Varianzanalysen für Differenzwert mit Hilfe der SPSS-Prozedur „ONEWAY“ durchgeführt.

Es ergaben durchweg keine signifikanten Ergebnisse auf dem 5%-Niveau, weder bei den Differenzen der KFT 4-12+R Teile KFT-Q und KFTN, noch bei den Differenzen der KFT-Gesamtleistung. Eine Ausnahme bildete lediglich der verbale Teil. Im Einzelnen ergaben sich für DiffJG-V $F = 4,217, p = .006$, für DiffJG-Q $F = 2,002, p = .112$, für DiffJG-N $F = 0,830, p = .477$ und für DiffJG-GL $F = 1,154, p = .326$ (Tabelle 5.4.4.7 unten; Tabelle 5.4.4.8 enthält wieder die Mittelwerte der jeweiligen Differenzwerte).

Tabelle 5.4.4.7: Varianzanalysen der Differenzen der Leistungen im KFT 4-12+R in den Dimensionen (KFT-V, KFT-Q, KFT-N) und der Gesamtleistung (KFT-GL) von Klassenstufe 5 zu Klassenstufe 7 nach Sprachgruppen

		Quadratsumme	df	Mittel der Quadrate	F	Signifikanz
DiffJG-V	Zwischen den Gruppen	510,064	3	170,021	4,217	,006
	Innerhalb der Gruppen	63625,254	1578	40,320		
	Gesamt	64135,318	1581			
DiffJG-Q	Zwischen den Gruppen	469,521	3	156,507	2,002	,112
	Innerhalb der Gruppen	123606,224	1581	78,182		
	Gesamt	124075,745	1584			
DiffJG-N	Zwischen den Gruppen	173,719	3	57,906	,830	,477
	Innerhalb der Gruppen	110880,399	1589	69,780		
	Gesamt	111054,118	1592			
DiffJG-GL	Zwischen den Gruppen	223,421	3	74,474	1,154	,326
	Innerhalb der Gruppen	101678,700	1575	64,558		
	Gesamt	101902,120	1578			

Legende: DiffJG-V: Differenzen zwischen KFT-V von Klasse 7 und KFT-V von Klasse 5;
 DiffJG-Q: Differenzen zwischen KFT-Q von Klasse 7 und KFT-Q von Klasse 5;
 DiffJG-N: Differenzen zwischen KFT-N von Klasse 7 und KFT-N von Klasse 5;
 DiffJG-GL: Differenzen zwischen KFT-GL von Klasse 7 und KFT-GL von Klasse 5.

Tabelle 5.4.4.8: Mittelwerte der Differenzen bezüglich der KFT 4-12+R Dimensionen (V, Q, N) und der Gesamtleistung der Schüler von Klasse 5 zu Klasse 7 nach Sprachgruppen

	N	Mittelwert	Standardabweichung	Standardfehler	95%-Konfidenzintervall für den Mittelwert		Minimum	Maximum
					Untergrenze	Obergrenze		
DiffJG-V	Deutsch	,2506	6,44641	,21957	-,1804	,6815	-20,00	63,00
	Türkei	-,8353	6,03363	,32437	-1,4733	-,1973	-28,00	17,00
	Osteuropa	-,6383	5,11860	,37331	-1,3747	,0981	-12,00	14,00
	Sonstige	-1,1398	7,49283	,54940	-2,2237	-,0559	-62,00	24,00
	Gesamt	-,2560	6,36917	,16013	-,5701	,0581	-62,00	63,00
DiffJG-Q	Deutsch	1,9132	8,39539	,28562	1,3526	2,4738	-53,00	59,00
	Türkei	,7637	8,68934	,46647	-,1538	1,6812	-58,00	26,00
	Osteuropa	,6524	10,45071	,76423	-,8553	2,1601	-66,00	48,00
	Sonstige	1,3529	9,37301	,68542	,0007	2,7051	-48,00	36,00
	Gesamt	1,4467	8,85046	,22231	1,0106	1,8827	-66,00	59,00
DiffJG-N	Deutsch	,5998	8,13698	,27555	,0589	1,1406	-54,00	45,00
	Türkei	,8362	8,88701	,47639	-,1008	1,7732	-60,00	27,00
	Osteuropa	1,4278	8,36139	,61145	,2215	2,6341	-50,00	30,00
	Sonstige	,1344	8,31854	,60994	-1,0689	1,3378	-46,00	19,00
	Gesamt	,6943	8,35210	,20926	,2838	1,1047	-60,00	45,00
DiffJG-GL	Deutsch	,7889	7,67453	,26140	,2758	1,3019	-58,00	68,00
	Türkei	,0318	8,31633	,44709	-,8476	,9112	-67,00	22,00
	Osteuropa	,0486	9,78702	,71956	-1,3710	1,4683	-69,00	50,00
	Sonstige	,0215	7,15916	,52494	-1,0141	1,0571	-52,00	25,00
	Gesamt	,4459	8,03597	,20223	,0492	,8425	-69,00	68,00

Tabelle 5.4.4.9: Post-Hoc-Test „homogene Untergruppen“ der verschiedenen Sprachgruppen hinsichtlich der Differenzen im KFT-V von Klassenstufe 5 zu Klassenstufe 7

Muttersprache	N	Scheffé-Prozedur	
		Untergruppe für Alpha = ,050.	1
Sonstige	186		-1,1398
Türkisch	346		-,8353
Osteuropäisch	188		-,6383
Deutsch	862		,2506
Signifikanz			,090

Tabelle 5.4.4.10: Post-Hoc-Test „homogene Untergruppen“ der verschiedenen Sprachgruppen hinsichtlich der Differenzen im KFT-Q von Klassenstufe 5 zu Klassenstufe 7

Muttersprache	N	Scheffé-Prozedur	
		Untergruppe für Alpha = ,050.	1
Osteuropäisch	187		,6524
Türkisch	347		,7637
Sonstige	187		1,3529
Deutsch	864		1,9132
Signifikanz			,430

Tabelle 5.4.4.11: Post-Hoc-Test „homogene Untergruppen“ der verschiedenen Sprachgruppen hinsichtlich der Differenzen im KFT-N von Klassenstufe 5 zu Klassenstufe 7

Scheffé-Prozedur

Muttersprache	N	Scheffé-Prozedur	
		Untergruppe für Alpha = ,050.	1
Sonstige	186		,1344
Deutsch	872		,5998
Türkisch	348		,8362
Osteuropäisch	187		1,4278
Signifikanz			,355

Tabelle 5.4.4.12: Post-Hoc-Test „homogene Untergruppen“ der verschiedenen Sprachgruppen hinsichtlich der Differenzen im KFT-GL von Klassenstufe 5 zu Klassenstufe 7

Scheffé-Prozedur

Muttersprache	N	Scheffé-Prozedur	
		Untergruppe für Alpha = ,050.	1
Sonstige	186		,0215
Türkisch	346		,0318
Osteuropäisch	185		,0486
Deutsch	862		,7889
Signifikanz			,746

Darüber hinaus ergaben die Post-Hoc-Tests (Scheffé) auf dem 5%-Niveau bezüglich der KFT-Dimensionen (KFT-V, KFT-Q, KFT-N) und der KFT-Gesamtleistung jeweils nur eine homogene Untergruppe (Tabellen 5.4.4.9 bis 5.4.4.12), d. h. es gibt keine Heterogenität zwischen den unterschiedlichen Sprachgruppen für die Veränderungen der KFT-Leistungen von Klasse 5 zu

Klasse 7, sondern alle verändern sich einheitlich, sind also in dieser Beziehung homogen. Insofern bestätigen die Post-Hoc-Analysen die Ergebnisse der Varianzanalyse.

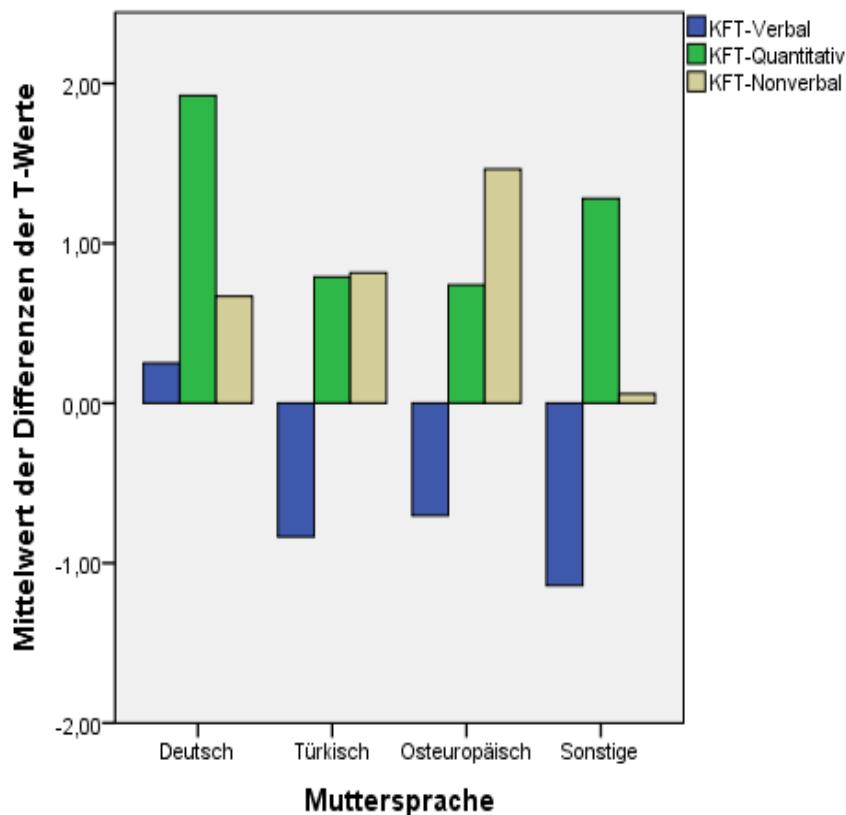


Abbildung 5.4.4.3: Veränderung der KFT-Leistungen (Dimensionen) von Klassenstufe 5 bis Klassenstufe 7 für die verschiedenen Sprachgruppen

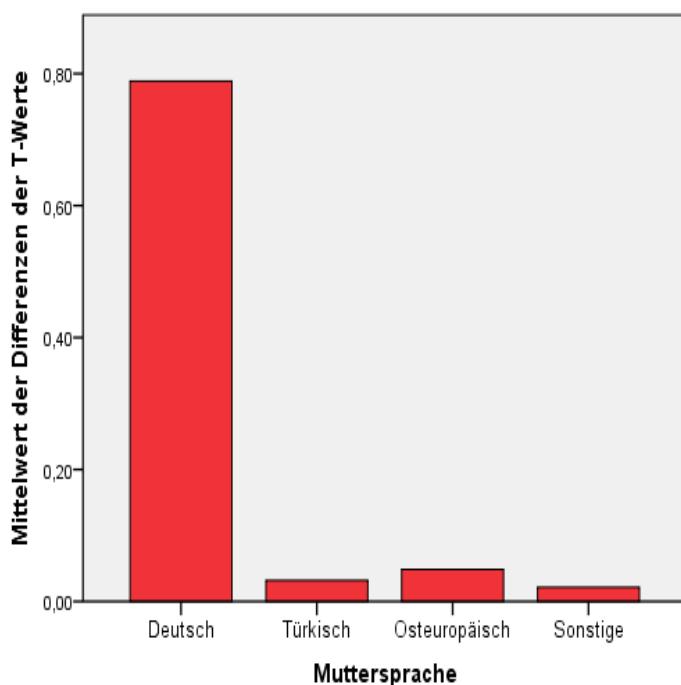


Abbildung 5.4.4.4: Veränderung der KFT-Gesamtleistung von Klassenstufe 5 bis Klassenstufe 7 für die verschiedenen Sprachgruppen

Die Abbildung 5.4.4.3 zeigt, dass es eine Verschlechterung bezüglich des KFT-V für alle Sprachgruppen gibt, deren Muttersprache nicht Deutsch ist bzw. die zu Hause nicht überwiegend Deutsch sprechen. Allerdings beträgt diese Verschlechterung gerade 1 T-Wertpunkt, also 0,1 Standardabweichungen, und ist damit sehr geringer Effekt zu bewerten und praktisch kaum bedeutsam. In den anderen KFT-Leistungen (KFT-Q und KFT-N) verbessern sich alle Sprachgruppen um bis zu zwei T-Werte (schwacher Effekt). Alle Sprachgruppen weisen von Klasse 5 bis Klasse 7 eine sehr geringe, annähernd vernachlässigbare Verbesserung hinsichtlich der KFT-Gesamtleistung auf, wie Abbildung 5.4.4.4 verdeutlicht. Die KFT-Gesamtleistung bleibt also über den betrachteten Zeitraum praktisch stabil.

5.4.4.3 Geschlechtsspezifische Veränderungen der KFT-Leistungen

Es wurde überprüft, ob sich die Mädchen und Jungen signifikant unterschiedlich bezüglich der KFT-Dimensionen (V, Q, N) und der KFT-Gesamtleistung von Klasse 5 zu Klasse 7 verändern. Diese Geschlechtsunterschiede wurden mittels t-Test für unabhängige Stichproben überprüft.

Tabelle 5.4.4.13: t-Tests der Differenzen der KFT-Leistungen in den KFT-Dimensionen und der KFT-Gesamtleistung von Klassenstufe 5 zu Klassenstufe 7 nach Geschlecht

	Levene-Test der Varianzgleichheit		T-Test für die Mittelwertgleichheit							
	F	Signifikanz	T	df	Sig. (2-seitig)	Mittlere Differenz	Standardfehler der Differenz	95% Konfidenzintervall der Differenz		
								Untere	Obere	
DiffJG-V	Varianzen sind gleich	,306	,580	1,880	1580	,060	,60287	,32076	-,02628	1,23203
	Varianzen sind nicht gleich			1,876	1539,895	,061	,60287	,32130	-,02735	1,23309
DiffJG-Q	Varianzen sind gleich	6,202	,013	-1,559	1583	,119	-,69439	,44543	-1,56808	,17930
	Varianzen sind nicht gleich			-1,581	1573,238	,114	-,69439	,43911	-1,55568	,16690
DiffJG-N	Varianzen sind gleich	,026	,872	-1,031	1591	,303	-,43239	,41943	-1,25507	,39030
	Varianzen sind nicht gleich			-1,034	1578,106	,301	-,43239	,41827	-1,25281	,38804
DiffJG-GL	Varianzen sind gleich	4,305	,038	-,894	1577	,372	-,36236	,40538	-1,15750	,43279
	Varianzen sind nicht gleich			-,912	1527,915	,362	-,36236	,39719	-,14145	,41674

Legende: Siehe Tabelle 5.4.4.2.

Tabelle 5.4.4.14: Mittelwerte der Differenzen bezüglich der KFT 4-12+R Dimensionen (V, Q, N) und der Gesamtleistung der Schüler von Klasse 5 zu Klasse 7 nach Geschlecht

	Geschlecht	N	Mittelwert	Standardabweichung	Standardfehler des Mittelwertes
DiffJG-	Jungen	845	,0249	6,29079	,21641
V	Mädchen	737	-,5780	6,44708	,23748
DiffJG-	Jungen	846	1,1229	9,66867	,33242
Q	Mädchen	739	1,8173	7,79934	,28690
DiffJG-	Jungen	849	,4923	8,51291	,29216
N	Mädchen	744	,9247	8,16434	,29932
DiffJG-	Jungen	842	,2767	9,09486	,31343
GL	Mädchen	737	,6391	6,62320	,24397

Legende: DiffJG-V: Differenzen zwischen KFT-V von Klasse 7 und KFT-V von Klasse 5;
 DiffJG-Q: Differenzen zwischen KFT-Q von Klasse 7 und KFT-Q von Klasse 5;
 DiffJG-N: Differenzen zwischen KFT-N von Klasse 7 und KFT-N von Klasse 5;
 DiffJG-GL: Differenzen zwischen KFT-GL von Klasse 7 und KFT-GL von Klasse 5.

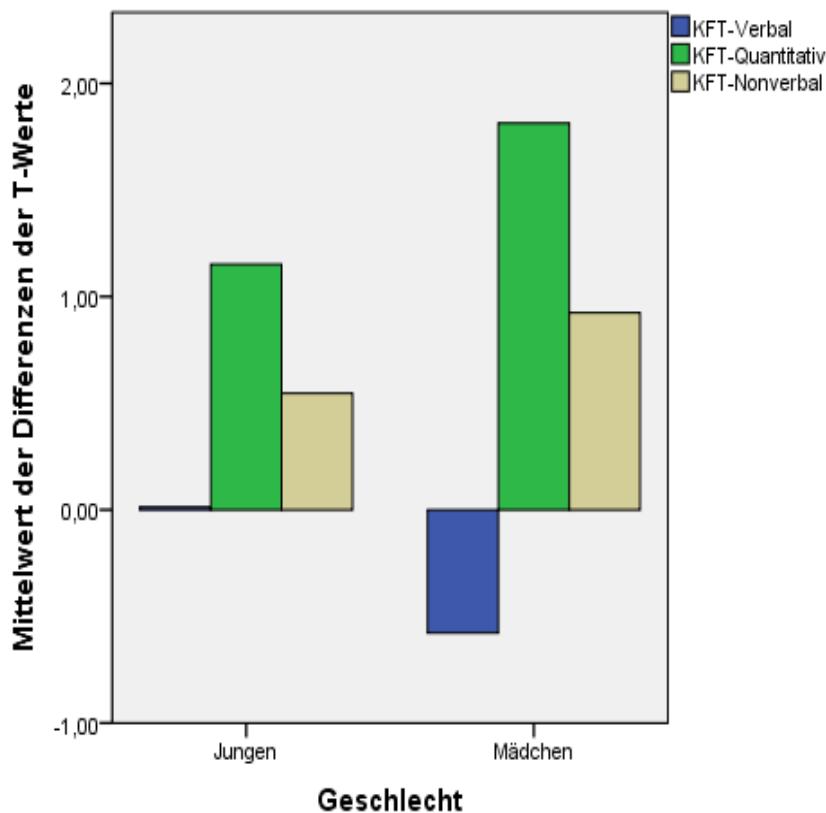


Abbildung 5.4.4.5: Veränderung der KFT-Leistungen (Dimensionen) von Klassenstufe 5 bis Klassenstufe 7 nach Geschlecht

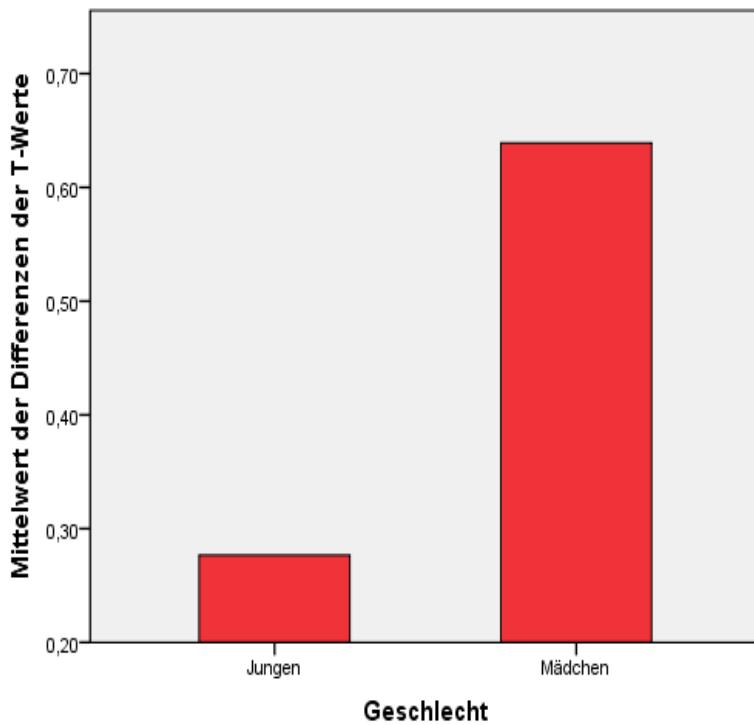


Abbildung 5.4.4.6: Veränderung der KFT-Gesamtleistung von Klassenstufe 5 bis Klassenstufe 7 nach Geschlecht

Der Levene-Test der Varianzgleichheit ergab p -Werte $> .05$ bezüglich der Differenzen der KFT-V und KFT-N. Das gilt aber nicht für KFT-Q und KFT-GL. Für DiffJG-V $p = .580$, für DiffJG-Q $p = .013$, für DiffJG-N $p = .872$ und für DiffJG-GL $p = .038$. Die Varianzen können also als homogen nur bei KFT-V und KFT-N angesehen werden (Tabelle 5.4.4.13; die Mittelwerte der Differenzen finden sich in Tabelle 5.4.4.14). Darüber hinaus zeigt Tabelle 5.4.4.13, dass es keinen signifikanten Unterschied zwischen den Geschlechtern bezüglich der KFT-Dimensionen (KFT-V, KFT-Q, KFT-N) und der KFT-Gesamtleistung gibt. Die p -Werte des t-Tests für 2-seitige Signifikanz sind $p = .060$ bei DiffJG-V, $p = .114$ bei DiffJG-Q, $p = .303$ bei DiffJG-N und $p = .362$ bei DiffJG-GL.

Abbildung 5.4.4.5 zeigt eine sehr geringe Verschlechterung bei Mädchen und Jungen in Bezug auf den KFT-V (0,5 T-Wert-Punkte bzw. 0,05 Standardabweichungen). Gleichzeitig gibt die Abbildung an, dass sich beide Geschlechter hinsichtlich ihrer quantitativen und nonverbalen Leistungen von Klasse 5 zu Klasse 7 in minimalem Umfang verbessern (maximal 2 T-Wert-Punkte oder 0,2 Standardabweichungen). Wenn man die Abbildung 5.4.4.6 betrachtet, wird deutlich, dass sich die KFT-Gesamtleistung bei Jungen und Mädchen von Klasse 5 bis Klasse 7 in sehr geringem Umfang verbessert (ca. 0,5 T-Wert-Punkte bzw. 0,05 Standardabweichungen).

Zusammenfassend kann man festhalten, dass die KFT-Leistungen von Klasse 5 bis Klasse 7 über alle Jahrgänge hinweg absinken. Ein großer Teil der Stichprobe (43,3%) spricht mit ihren Eltern zu Hause kein oder kaum Deutsch. Das hat negative Auswirkung auf die Entwicklung der KFT-Leistungen, besonders auf die verbale Leistung, da, wie die Ergebnisse dieser Untersuchung gezeigt haben, die Muttersprache besonders die verbalen KFT-Leistungen beeinflusst. Darüber hinaus kann man diese Verschlechterung im verbalen Teil damit erklären, dass der KFT nur Wortschatz und nicht andere Kompetenzen wie z.B. Grammatik erfasst. Außerdem könnte die Verschlechterung auf geringe Förderung zu Hause, an der Schule oder auf die persönlichen Merkmale wie Interesse und Motivation zurückzuführen sein.

6. Zusammenfassung und Interpretation der Untersuchungs-ergebnisse zum Kognitiven Fähigkeitstests KFT 4-12+ R für die Klassen 5 und 7

Zur Bewertung der Qualität des KFT 4-12+ R wurden anhand von Daten von Fünft- und Siebklässlern einer Gesamtschule, die über mehr als 10 Jahre von 1998 bis 2011 gesammelt wurden, Itemanalysen durchgeführt bzw. Schwierigkeits- und Trennschärfeindizes analysiert sowie Aspekte der Gütekriterien Reliabilität und Validität überprüft. Darüber hinaus wurden die Einflüsse von Faktoren wie Geschlecht und Muttersprache auf die KFT-Dimensionen und die KFT-Gesamtleistung in den Klassenstufen 5 und 7 untersucht. Weiter wurde geprüft, ob an Veränderungen der Leistungen in den KFT-Dimensionen und der KFT-Gesamtleistung in den untersuchten Klassenstufen im letzten Jahrzehnt Kohorten oder Generationeneffekte ablesbar sind. Schließlich wurden Aspekte der Intelligenzentwicklungen anhand der (individuellen) Veränderungen der KFT-Leistungen (Dimensionen und Gesamtleistung) von Klasse 5 zu Klassenstufe 7 analysiert.

6.1. Zusammenfassung der empirischen Untersuchungsergebnisse

6.1.1 Itemanalysen und Verteilungsmerkmale zu den KFT-Subtests, Testdimensionen und der Gesamtleistung

Zur Überprüfung, ob der KFT 4-12+ R für Klassen 5 und 7 genügend Differenzierungsfähigkeit besitzt und ob die Testpunktswerte annährend normal verteilt sind, wurden Itemanalysen, hier vor allem über die Berechnung von Itemschwierigkeits- und Trennschärfeindizes, und Skalenanalysen durchgeführt.

Es wurden hierfür Datensätze der Klassenstufen 5 und 7 herangezogen, die sich jeweils in etwa zu gleichen Teilen auf die beiden Testformen (TF) verteilten (N5A = 955, N5B = 1135, N7A = 1112, N7B = 1113). Die Ergebnisse der Itemanalysen zum KFT 4-12+ R in beiden Klassen 5 und 7 zeigten, dass einige einzelne Items mit einem Schwierigkeitsindex von $p > .80$ oder $p < .20$ eigentlich den Anforderungen nicht genügen (Tabellen 5.3.1.3 bis 5.3.1.28). Im Einzelnen betrifft dies bei Testform A in Klasse 5 48 Items und in Klasse 7 50 Items und bei Testform B 27 Items in Klasse 5 und 39 in Klasse 7. Weiter ist festzuhalten, dass die durchschnittlichen

Itemschwierigkeiten für die Skalen des KFT 4-12+ R für die beiden Testformen in beiden Klassen 5 und 7 recht unterschiedlich ausfallen und zwischen $p = .41$ und $p = .79$ liegen (Tabelle 5.3.1.1).

Im Gegensatz zu den Itemschwierigkeiten fallen die Trennschärfeindizes von einzelnen Items für die meisten Skalen des KFT 4-12+ R mit $r < .30$ niedrig aus (Tabellen 5.3.1.3 bis 5.3.1.28). Das gilt auch für die durchschnittlichen Trennschärfen der meisten Skalen des KFT 4-12+ R für die beiden Testformen in den Klassenstufen 5 und 7. Die durchschnittliche Trennschärfekoeffizienten der KFT 4-12+ R Untertests, Dimensionen und der Gesamtleistung für die beiden Testformen in beiden Klassen 5 und 7 variieren von $r = .10$ bis $r = .54$ (Tabelle 5.3.1.2).

Für die verbalen Untertests KFT-V1, KFT-V3 und die Dimension KFT-V, die quantitativen Subtests KFT-Q1, KFT-Q3 und die Dimension KFT-Q in Klasse 5 für beide Testformen sowie für KFT-N1, KFT-N, KFT-V2 (Form B) und die KFT-Gesamtleistung für beide Testformen der Klassenstufe 5 ergeben sich mittlere Schwierigkeitsgrade bzw. normale Schwierigkeitsverteilungen (vgl. Abbildung 5.3.1.1 oben). Das gilt analog auch für den V-Teil von Klasse 7 in beiden Testformen, die quantitativen Untertests KFT-Q1, KFT-Q2 (nur Form B) und KFT-Q3 sowie die Dimension KFT-Q und schließlich auch für die KFT-Gesamtleistung für beide Testformen, aber nicht für den N-Teil (vgl. Abbildung 5.3.1.2). Zusätzlich wirken sich die schiefen Verteilungen in den Untertests und Testteilen bei beiden Klassenstufen 5 und 7 auf die Differenzierungsfähigkeit des KFT 4-12+R aus. Außerdem zeigen einige Skalen des KFT 4-12+R für beide Testformen in Klassenstufen 5 (KFT-V2, KFT-Q2 und KFT-N1) sowie 7 (Subtests des KFT-N-Teils und KFT-Q2) sehr ungleiche Schwierigkeitsgrade (vgl. die Abbildungen 5.3.1.1 und 5.3.1.2).

6.1.2 Zur Reliabilität des KFT 4-12+ R

Zur Überprüfung der Reliabilität des KFT 4-12+ R wurden Analysen nach drei unterschiedlichen Methoden der Reliabilitätsschätzung angewandt: Testhalbierungsreliabilität „Split-half“, innere Konsistenz und Retest-Reliabilität. Dafür wurde auf dieselben Stichproben wie für die Item- und Trennschärfefanalysen zurückgegriffen (N5A = 955, N5B = 1135, N7A = 1112, N7B = 1113).

In Bezug auf die **Testhalbierungsreliabilität** nach den Formeln von Spearman- Brown und Guttman liegen die Reliabilitätskoeffizienten für die KFT-Untertests der Fünftklässler (Testform A) zwischen $Rel = .49$ und $Rel = .94$ bzw. bei $Rel = .94$ für die Gesamtleistung, sowie für Testform B zwischen $Rel = .60$ und $Rel = .93$ für die KFT-Subtests und bei $Rel = .94$ für die KFT-Gesamtleistung (Tabellen 5.3.2.1 und 5.3.2.2). Für die Siebklässler ergeben sich entsprechend Reliabilitätskoeffizienten nach Spearman- Brown und Guttman zwischen $Rel = .62$ und $Rel = .93$ für die KFT-Untertests und von $Rel = .94$ für die Gesamtleistung (Testform A) sowie zwischen $Rel = .60$ und $Rel = .92$ für die KFT- Untertests und von $Rel = .94$ für die Gesamtleistung (Testform B) , sowie zwischen $Rel = .60$ und $Rel = .92$ für die Subtests und $Rel = .93$ für die Gesamtleistung (Form B) aus (Tabellen 5.3.2.3 und 5.3.2.4).

Bezüglich der **inneren Konsistenz** der KFT-Skalen nach Cronbach-Alpha ergaben sich für die Fünftklässler Konsistenzkoeffizienten zwischen $Rel = .42$ und $Rel = .92$ und $Rel = .93$ für die KFT- Gesamtleistung (Testform A) bzw. zwischen $Rel = .58$ und $Rel = .92$ (KFT-Subtests) und $Rel = .93$ für die KFT-Gesamtleistung (Testform B) (Tabellen 5.3.2.1 und 5.3.2.2). Darüber hinaus variieren die Konsistenzkoeffizienten in Bezug auf die Siebklässler von $Rel = .59$ bis $Rel = .93$ für die KFT-Untertests und $Rel = .93$ für die Gesamtleistung (Testform A) bzw. für Testform B von $Rel = .59$ bis $Rel = .92$ für die KFT-Subtests und $Rel = .93$ für die KFT- Gesamtleistung (Tabellen 5.3.2.3 und 5.3.2.4).

Für die **Retest-Reliabilität** des KFT 4-12+ R (Retestung nach zwei Jahren) sind die Stabilitätskoeffizienten erheblich niedriger ausgefallen. Die Koeffizienten reichen für Testform A von $Rel = .27$ bis zu $Rel = .73$ für die KFT-Untertests und $Rel = .70$ für die KFT- Gesamtleistung), für Testform B ergaben sich Wert von $Rel = .22$ bis zu $Rel = .67$ für die KFT- Subtests und $Rel = .69$ für die KFT- Gesamtleistung (Tabelle 5.3.2.5).

Insgesamt kann man sagen, dass die Reliabilitätskoeffizienten des KFT 4-12+ R für die beiden Testformen nach der Halbierungsmethode und die Konsistenzkoeffizienten im befriedigenden (Untertests) bis sehr guten Bereich (Gesamtleistung und Testdimensionen) und die Stabilitätskoeffizienten im niedrigen bis mittleren Bereich liegen (nach der Klassifikation von Perleth & Sierwald, 2000), wobei allerdings zu berücksichtigen ist, dass es sich hier einerseits um ein sehr langes Intervall handelt und die Schule sich um die (Begabungs-)Förderung von Schülern/innen vor allem aus benachteiligten Familien bemüht.

6.1.3 Zur Validität des KFT 4-12+ R

Hinsichtlich der Validität des KFT 4-12+ R wurden die Konstruktvalidität durch Faktorenanalysen und die Kriteriumsvalidität durch Korrelationen mit Schulzensuren überprüft.

6.1.3.1 Konstruktvalidität

In Bezug auf die Konstruktvalidität wurden die Untertests des KFT 4-12+ R faktorenanalysiert. Die Faktorenanalysen des KFT 4-12+ R für die Testformen A und B wurden nach der Hauptkomponentenanalyse mit der Varimax-Rotationsmethode (mit Kaisernormalisierung) für die Analysestichproben von Klasse 5 und 7 durchgeführt. Die Stichproben wurden wie für die anderen Analysen auch bezüglich der Testform aufgeteilt (hier N5A = 955, N5B = 1135, N7A = 1112, N7B = 1113). Die Faktorenanalysen des KFT 4-12+ R ergaben für beide Testformen von Klasse 5 drei Faktoren (Tabellen 5.3.3.1 und 5.3.3.2). Die extrahierten Faktoren lassen sich wie folgt interpretieren:

Faktor (1) wird als Faktor des nonverbalen-figuralen Denkens (Reasoning mit Space nach Thurstone) bezeichnet, auf dem die Subtests KFT-Figurenklassifikation (KFT-N1), KFT-Figurenanalogien (KFT-N2) und KFT-Faltaufgaben (KFT-N3) ihre höchsten Ladungen haben. Zusätzlich haben auch die Untertests KFT-Zahlenreihen KFT-Q2 und KFT-Wortklassifikationen KFT-V2 auch hohe Ladung auf diesem Faktor. Aber die Ladungen von KFT Figurenklassifikation und Figurenanalogien sind höher als die der anderen Subtests.

Faktor (2) ist deutlich als Faktor des verbalen- sprachlichen Denkens (Verbal Comprehension nach Thurstone) zu erkennen. Die Untertests KFT-Wortschatz (KFT-V1), KFT-Wortklassifikationen (KFT-V2) und KFT-Wortanalogien (KFT-V3) haben auf diesem Faktor die höchsten Ladungen. Die Untertests KFT-Wortschatztest (KFT-V1) und KFT-Wortklassifikationen (KFT-V2) weisen dabei höhere Ladung auf als der Subtest Wortanalogien (KFT-V3).

Faktor (3) lässt sich als Faktor des quantitativen Denkens (Number mit Reasoning nach Thurstone) interpretieren. Die Subtests KFT-Mengenvergleiche (KFT-Q1), KFT-Zahlenreihen (KFT-Q2) und KFT- Gleichungenbilden (KFT-Q3) haben auf diesem Faktor die höchsten

Ladungen. Der Subtest KFT-Mengenvergleiche hat dabei eine höhere Ladung als die anderen beiden quantitativen Subtests.

Die faktorenanalytischen Befunde zum KFT 4-12+ R für die beiden Testformen, die an der ersten Analysestichprobe der Fünftklässler gefunden wurden, stimmen mit den Ergebnissen der Faktorenanalysen an der zweiten Teil-Stichprobe der Siebklässler hinsichtlich der drei gefundenen Faktoren überein (Tabellen 5.3.3.3 und 5.3.3.4):

Faktor (1) kann als Faktor des nonverbalen-figuralen Denkens für beide Testformen A und B gelten. Die nonverbalen Untertests bzw. KFT-Figurenklassifikation (KFT-N1), KFT-Figurenanalogen (KFT-N2) und KFT-Faltaufgaben (KFT-N3) haben hohe Ladung auf diesem ersten Faktor. Darüber hinaus laden für Testform B auch die KFT-Wortanalogen (KFT-V3) und KFT-Zahlenreihen (KFT-Q2) in mittlerer Höhe auf diesem Faktor, für Testform A aber nur KFT-Zahlenreihen. Da es sich hierbei um Untertests mit bekanntermaßen hoher Sättigung an induktivem Denken (Reasoning) handelt, widersprechen diese Ladungen der Interpretation des Faktors nicht. Für Testform A ist die Ladung des Subtests KFT-Figuren-Klassifikationen (KFT-N1) dabei deutlich höher als die der übrigen Untertests und die der KFT-Figurenanalogen (KFT-N2) für Testform B.

Faktor (2) kann für beide Testformen A und B als Faktor des verbal-sprachlichen Denkens bezeichnet werden. Auf diesem Faktor finden sich die höchsten Ladungen für die verbalen Untertests (KFT-Wortschatztest V1, KFT-Wortklassifikationen V2 und KFT-Wortanalogen V3), wobei die Ladungen der Subtests KFT-Wortklassifikationen (V2) und KFT-Wortschatztest (V1) höher sind als die anderen verbalen Untertests für Testform A und die der KFT-Wortklassifikationen (V2) für Testform B.

Faktor (3) lässt sich für beide Testformen als Faktor des mathematischen und quantitativen Denkens interpretieren. Auf diesem Faktor weisen die Subtests KFT-Mengenvergleiche (KFT-Q1), KFT-Zahlenreihen (KFT-Q2) und KFT- Gleichungenbilden (KFT-Q3) hohe bzw. die höchsten Ladungen auf. Untertest KFT-Gleichungenbilden lädt dabei am höchsten von allen drei quantitativen Subtests bei Testform A und zusätzlich auch KFT-Mengenvergleiche bei Testform B.

Zusammenfassend kann festgehalten werden, dass für die Fünft- und Siebklässler und für beide Testformen eine Faktorenstruktur des KFT 4-12+ R von drei Faktoren gefunden

wurde. Der verbale-sprachliche Faktor wird durch die Tests des KFT-V-Teils, der quantitative Faktor durch die des KFT-Q-Teils, und der nonverbale Faktor durch die KFT-N-Tests bestimmt. Das bestätigt recht deutlich die Gültigkeit des KFT 4-12+ R als ein differentieller Intelligenztest zur Erfassung der intellektuellen Fähigkeiten von Schülern/innen der fünften und siebten Klassen in Bezug auf verbale, quantitative und nonverbale Fähigkeiten.

6.1.3.2 Kriteriumsvalidität

Zur Überprüfung der Kriteriumsvalidität des KFT 4-12+ R wurden in dieser Untersuchung Übereinstimmungsvalidität und prognostische Validität berechnet:

Übereinstimmungsvalidität. Zur Einschätzung der Übereinstimmungsvalidität wurden die Ergebnisse der KFT 4-12+ R Untertests, Testteile (Dimensionen) und der Gesamtleistung von ca. 154 Schüler/innen von Klasse 7 mit ihren Schulzensuren in Deutsch, Englisch und Mathematik korreliert. Die Koeffizienten zur Übereinstimmungsvalidität zeigen, dass die sprachlichen bzw. verbalen Untertests des verbalen Teil (KFT-V) signifikante Zusammenhänge zu allen Schulnoten haben, mit $r = .46$ am höchsten fällt die mit Deutschzensur aus. Darüber hinaus korrelieren die quantitativen Subtests des KFT-Q-Teils bzw. KFT-Q1 (Mengenvergleiche) mit allen Schulnoten signifikant auf dem 5%-Niveau, besonders mit der Mathematikzensur mit $r = .33$, Untertest KFT-Q2 (Zahlenreihen) korreliert nur mit der Mathematikzensur ($r = .27, p < .05$), der Subtest KFT-Q3 (Gleichungenbilden) hingegen zeigt keine signifikante Zusammenhänge mit den Zensuren und der KFT-Q-Gesamt nur mit der Deutsch- und Mathematiknote ($r = .34$). Außerdem wiesen die Tests des KFT-N Teils (Figurenklassifikation KFT-N1, Figurenanalogien KFT-N2 und Faltaufgaben KFT-N3) keine Beziehungen zu den Schulnoten auf. Schließlich ist festzuhalten, dass die KFT-Gesamtleistung signifikante Zusammenhänge zu allen Zensuren, besonders zur Mathematikzensur mit $r = .30$ zeigt (Tabelle 5.3.3.5).

Prognostische Validität. Um die prognostische Validität zu überprüfen, wurden die Ergebnissen von ca. 133 Schülern und Schülerinnen in den KFT 4-12-Subtests, Dimensionen (Testteilen) und der Gesamtleistung, die in Klasse 5 erhoben wurden, mit den Schulzensuren, die später in Klasse 7 erzielt bzw. erhoben wurden, in Beziehung gesetzt. Die entsprechenden Koeffizienten zeigen, dass die Skalen des KFT-V-Teils bzw. KFT-V1 (Wortschatztest), KFT-V2 (Wortklassifikationen), KFT-V3 (Wortanalogien) sowie KFT-V-Gesamt Beziehungen zu allen

Schulnoten aufweisen, besonders mit der Deutschnote mit $r = .43$. Das gilt so nicht für die Tests des KFT-Q-Teils (Mengenvergleiche KFT-Q1, Gleichungenbilden KFT-Q3 und Zahlenreihen KFT-Q2), deren Korrelationen mit allen Noten nicht signifikant ausfielen, mit Ausnahme der Korrelation zwischen KFT-Q2 und der Mathematiknote ($r = .18$). Zusätzlich zeigt die KFT-Q-Dimension signifikante Zusammenhänge zu der Deutsch- und Mathematikzensur, am höchsten ist dabei die Korrelation mit der Mathematiknote ($r = .21$).

Die Korrelationen der Skalen des KFT-N-Teils (Figurenklassifikation KFT-N1, Figurenanalogien KFT-N2, Faltaufgaben KFT-N3) sowie die Dimension KFT-N fallen mit allen Zensuren außer Englisch signifikant aus, wobei allerdings Untertest N3 mit der Englischnote immerhin mit $r = .26$ korreliert. Die höchste Korrelation von KFT-N-Dimension ergibt sich mit der Mathematikzensur ($r = .33$). Die Korrelationsanalysen zur KFT-Gesamtleistung (KFT-GL) zeigen signifikante Zusammenhänge mit allen Zensuren auf. Die Gesamtleistung korreliert mit der Mathezensur am höchsten mit $r = .30$ (Tabelle 5.3.3.6).

Zusammenfassung. Die Ergebnisse zur Übereinstimmungsvalidität belegen auch, dass mit dem KFT 4-12+ R gut verbale und quantitative Kompetenzen erfasst werden können, d.h. dass sie mit Leistungen im verbalen und quantitativen Bereich übereinstimmen. Für die nonverbalen Fähigkeiten trifft das so nicht zu, auch weil diesem Bereich nicht direkt Schulleistungen zugeordnet werden können. Letztlich wird das bestätigt, was die KFT-Autoren Heller und Perleth (2000) schreiben, nämlich dass der KFT 4-12+ R zur Erfassung der verbal-sprachlicher, quantitativer und nonverbaler Fähigkeiten konzipiert und geeignet ist. Darüber hinaus zeigten die Befunde zur prognostischen Validität, dass der KFT 4-12+ R gute prognostische Beziehungen zu Schulzensuren aufweist, d.h. der KFT 4-12+ R kann zur Vorhersage der schulischen Leistung benutzt werden.

6.1.4 Einflussfaktoren auf die Intelligenzleistung

6.1.4.1 Geschlecht

Anhand zweifaktorieller Varianzanalysen wurden die Einflüsse von Geschlecht und Schuleintritts-Jahrgang auf die KFT- Untertests, Dimensionen und die KFT-Gesamtleistung für die beiden Klassenstufen 5 und 7 berechnet. Die verbalen, quantitativen, nonverbalen Fähigkeiten und die Gesamtleistung stellen die abhängigen Variablen dar, Geschlecht und

Schuleintritts-Jahrgang fungieren als Faktoren. Die Ergebnisse der Varianzanalysen für Klasse 5 zeigten, dass das Geschlecht keinen Einfluss auf die verbalen und nonverbalen Fähigkeiten ausübt. Dafür beeinflusst das Geschlecht die quantitative Kompetenz sowie die Gesamtleistung. Zusätzlich hat der Faktor Jahrgänge einen Einfluss auf alle KFT-Leistungen (alle diese Einflüsse sind signifikant). Jungen schneiden im quantitativen Teil und der Gesamtleistung besser ab als Mädchen.

In Bezug auf die Analysestichprobe der Siebklässler zeigte sich, dass die beiden Faktoren bzw. Geschlecht und Schuleintritts-Jahrgang sich auf die verbalen, quantitativen und nonverbalen Fähigkeiten, sowie die Gesamtleistung auswirken, d.h. Jungen und Mädchen unterscheiden sich bezüglich ihrer verbalen, quantitativen und nonverbalen Fähigkeiten sowie ihrer Gesamtleistung signifikant, wobei die Unterschiede zugunsten der Jungen ausfallen.

Aus den unterschiedlichen Befunden der Varianzanalysen für die beiden Teil-Stichproben der Fünft- und Siebklässler kann man nicht ableiten, dass sich Jungen und Mädchen bezüglich der kognitiven Fähigkeiten unterscheiden. Die Ergebnisse betreffen zunächst diese spezielle Stichprobe. In der vorliegenden Untersuchung schneiden Jungen nicht nur im quantitativen Teil und der Gesamtleistung, sondern auch im verbalen Teil ab besser als Mädchen.

6.1.4.2 Muttersprache

Für die beiden Teil-Stichproben der Fünft- und Siebklässler wurden die Einflüsse der "Muttersprache" und des Schuleintritts-Jahrgang auf die KFT-Teile und die KFT-Gesamtleistung untersucht. Die Variable "Muttersprache" unterscheidet dabei Kinder, bei denen zu Hause überwiegend Deutsch gesprochen wird (Muttersprache Deutsch) von solchen, die zu Hause mit ihren Eltern kaum oder kein Deutsch sprechen. Dabei wurden grob folgende (Sprach-)Gruppen unterschieden: Deutsch, Türkisch, Osteuropäische Sprache und Sonstige (Tabelle 5.2.1.2). Für diese Analysen wurden die Unterschiede anhand zweifaktorieller Varianzanalysen untersucht.

Die KFT-Dimensionen KFT-V, KFT-Q, KFT-N und die KFT-Gesamtleistung wurden als abhängige Variablen sowie Jahrgang und die Muttersprache als Faktoren für diese Analysen herangezogen. Die Ergebnisse der Varianzanalyse für die beiden Klassen stimmen darin überein, dass die Faktoren Muttersprache und Jahrgang einen Einfluss auf die KFT-Leistungen in beiden Klassenstufen 5 und 7 haben. Zusammenfassend kann man festhalten, dass in der

untersuchten Stichprobe alle Sprachgruppen in der Analyse schlechter als die mit Deutsch als Muttersprache abschneiden, wobei die türkischsprachigen Schüler/innen die schlechteste Leistung im Vergleich zu den übrigen Sprachgruppen aufweisen.

6.1.5 Veränderungen der Leistungen in den KFT-Dimensionen und der KFT-Gesamtleistung in den untersuchten Klassenstufen im letzten Jahrzehnt

Die Veränderung der KFT 4-12+R-Leistungen (Dimensionen und Gesamtleistung) von 1998 bis 2011 für die Klassenstufen 5 und 7 (Kohorten- bzw. Jahrgangseffekte) wurden anhand von einfaktoriellen Varianzanalysen überprüft. Die Dimensionen KFT-V, -Q, -N und die Gesamtleistung KFT-GL wurden dabei als abhängige Variablen und der Jahrgang als unabhängiger Faktor in die Untersuchungen einbezogen.

Die Ergebnisse der Varianzanalyse von Klassenstufe 5 ließen erkennen, dass sich die KFT-Leistungen signifikant von den Jahrgängen 1998-2000 bis zu den Jahrgängen 2007-2010 verändern. Die Analysen mittels Post-hoc-Tests (Scheffé) zeigten Verschlechterungen in den KFT-Dimensionen und der KFT-Gesamtleistung von den Jahrgängen 1998-2000 bis 2001-2003 und dann kaum mehr Veränderungen bis zu den Jahrgängen 2007-2010. Die statistischen Befunde der Varianzanalyse für Klassenstufe 5 gelten analog auch für die Klassenstufe 7. Auch hier für Klassenstufe 7 zeigten die KFT-Dimensionen und die KFT-Gesamtleistung signifikante Veränderungen bzw. Verschlechterung von 1998-2000 bis 2007-2011 zeigen. Die Post-hoc-Tests bestätigten die signifikanten Ergebnisse.

Insgesamt kann man somit aber festhalten, dass sich die einzelnen Einschulungs-Jahrgänge über den analysierten Zeitraum hinweg nur wenig hinsichtlich ihrer kognitiven Fähigkeiten bzw. der mit dem KFT 4-12+ R erfassten verbalen, quantitative und nonverbalen Fähigkeiten sowie der Gesamtleistung unterscheiden. Die gefundenen Veränderungen scheinen eher auf unsystematische Schwankungen von Jahrgang zu Jahrgang zurückzuführen zu sein.

6.1.6 Muster der Intelligenzsentwicklung anhand der KFT-Leistungen in der Eingangsstufe der Gesamtschule

Um die (individuellen) Veränderungen der Leistungen in den KFT-Teile (KFT-V, KFT-Q und KFT-N) sowie der KFT-Gesamtleistung von Klasse 5 zu Klasse 7 in den Einschulungs-Jahrgängen von 1998 bis 2010 zu bestimmen, wurden zunächst die Differenzen der KFT-Dimensionen und der KFT-Gesamtleistung zwischen den Klassen 7 und 5 bestimmt. Danach wurden für die Differenzwerte einfaktorielle Varianzanalysen und Post-hoc-Tests (nach Scheffé) durchgeführt. Als Faktoren dienten dabei die (gruppierten) Einschulungs-Jahrgänge sowie die Sprachgruppen und Geschlecht. Dieses Vorgehen entspricht einer Varianzanalyse mit Messwiederholung für zwei Messzeitpunkte.

Für die (gruppierten) Einschulungs-Jahrgänge ergaben die Varianzanalysen signifikante Ergebnisse sowohl für die KFT-Dimensionen wie die KFT-Gesamtleistung, wobei die Post-hoc-Tests diese Ergebnisse noch bestätigten und genauer fassbar machen: Die Leistungen in den KFT-Dimensionen (KFT-V, KFT-Q, KFT-N) und die KFT-Gesamtleistung unterscheiden sich im betrachteten Zeitraum (1998 bis 2010) signifikant. Für die Schüler/innen der Jahrgänge 1998-2000 finden sich Verbesserungen in allen KFT-Dimensionen und der KFT-Gesamtleistung um 3 bis 4 T-Werte. Während sich im KFT-V eine kleine Verschlechterung für die Jahrgänge 2004-2006 zeigt, verbessert sich die quantitative Leistung am meisten für die Schüler/innen der Jahrgänge 2001-2003 bis 2004-2006. Für KFT-N und die KFT-Gesamtleistung lassen sich für die Jahrgänge ab 2004-2006 keine Steigerungen mehr nachweisen. Jedoch kommt diesen Veränderungen insgesamt nur eine sehr geringe Bedeutung zu, weil sie sehr klein ausfallen (kleiner als 2-T Werte-Punkte).

Darüber hinaus wurden Veränderungen der KFT-Dimensionen und der KFT-Gesamtleistung von Klassenstufe 5 zu 7 bei 4 Sprachgruppen ("Muttersprache" Deutsch, Türkisch, Osteuropäisch und sonstige) inspiziert. Die Ergebnisse zeigen im einzelnen, dass bei den Schülern/innen mit nichtdeutscher Muttersprache (d.h. bei denen, die zu Hause mit ihren Eltern überwiegend nicht Deutsch sprechen) ein (relatives) Absinken ihrer verbalen Leistung festzuhalten ist. Das gilt nicht für die übrigen KFT-Dimensionen bzw. den quantitativen und den nonverbalen Teil und auch nicht für die Gesamtleistung, wo es Verbesserung bis zu zwei T-Werte bei den einzelnen Sprachgruppen gibt. Diese Ergebnisse sind allerdings nicht signifikant.

Zusätzlich wurden noch Veränderung der Leistungen in den KFT-Dimensionen und der KFT-Gesamtleistung zwischen Jungen und Mädchen anhand des t-Tests für unabhängige Stichproben analysiert. Diese Untersuchungen erbrachten keine signifikanten Ergebnisse, d.h. die Jungen und Mädchen der Münchner Gesamtschule scheinen sich in ihrer Intelligenzsentwicklung bzw. der Entwicklung ihrer kognitiven Fähigkeiten von Klasse 5 zu Klasse 7 nicht zu unterscheiden. Beide Geschlechter zeigen nur sehr kleine Verschlechterung in ihren verbalen Leistungen, während sich ihre quantitative, nonverbale und die KFT-Gesamtleistung bis 2 T-Werten geringfügig verbessern.

6.2 Interpretation

Hauptziele der vorliegenden Studie sind die Bewertung der Qualität des KFT 4-12+ R anhand der beiden Klassenstufen 5 und 7 der Willy-Brandt-Gesamtschule München sowie die Untersuchung der Entwicklungen der KFT-Leistungen in den untersuchten Klassenstufen im letzten Jahrzehnt (Kohorteneffekte) sowie die Veränderung der kognitiven Fähigkeiten von Klasse 5 zu Klasse 7 in Abhängigkeit von Einflussfaktoren wie Geschlecht und Muttersprache.

6.2.1 Bewertung der Qualität des KFT 4-12+ R

Um die Bewertung des KFT 4-12+ R zu ermitteln, wurden anhand von Daten der Fünft- und Siebtklässler/innen der Willy-Brandt-Gesamtschule Itemanalysen durchgeführt und die Reliabilität sowie die Validität des KFT 4-12+ R überprüft.

Itemanalysen und Verteilungsmerkmale zu den KFT-Subtests, den Testdimensionen und der KFT-Gesamtleistung. Die Ergebnisse der vorliegenden Untersuchung zeigen, dass die Rohwerte einiger Skalen des KFT 4-12+ R für beide Testformen A und B, sowie in beiden Klassenstufen 5 und 7 in etwa normalverteilt sind. Jedoch zeigen einige andere Untertests und Testdimensionen in jeder der beiden Klassenstufe schiefe Verteilungen in einer Testform oder in beiden Testformen: KFT-V2, KFT-Q2, KFT-N1 und KFT-N bei Testform A, sowie KFT-N2 und KFT-N3 bei beiden Formen in Klasse 5 und die Skalen KFT-N sowie KFT-Q1 und KFT-Q2 für Testform A. Das hat Auswirkungen auf die Höhe der Trennschärfe der Items und die individuellen Unterschiede (vgl. Lienert & Raatz, 1998). Darüber hinaus haben manche

Skalen des KFT 4-12+ R für beide Testformen in Klassenstufen 5 (KFT-V2, KFT-Q2 und KFT-N1) sowie 7 (Subtests des KFT-N-Teils und KFT-Q2) keine gleichen Schwierigkeitsgrade.

Ein Testverfahren sollte Items mit mittleren Schwierigkeitsgraden $p > .20$, $p < .80$ und mit Trennschärfen $r > .30$ enthalten (Bühner, 2004). Die vorliegenden Befunde zeigen, dass diese Bedingungen für den KFT 4-12+ R teilweise erfüllt sind, weil einerseits viele Items der Skalen für beide Testformen A und B in beiden Klassenstufen 5 und 7 mittlere Schwierigkeiten aufweisen, andererseits zeigen die einzelnen Items entweder zu niedrige oder zu hohe Schwierigkeiten. Außerdem sind die durchschnittlichen Itemtrennschärfen in den Untertests KFT-Q1, KFT-Q3, den Testteilen KFT-V, KFT-Q, und der KFT-Gesamtleistung für beide Testformen von Klassen 5 und 7 zu niedrig. Das gilt auch für die Subtests KFT-V1, KFT-V3 in Klasse 5 Testform B sowie KFT-V1 und KFT-V2 für beide Testformen in Klasse 7.

Reliabilität des KFT 4-12+ R. Die Konsistenzkoeffizienten der einzelnen Skalen von KFT4- 12+R für beide Testformen in Klassenstufe 5 fallen für die beiden Subtests KFT-Q1 und KFT-Q3 am niedrigsten aus. Das bestätigt teilweise die Befunde von Heller und Perleth (2000), die Koeffizienten fallen aber in der vorliegenden Untersuchung noch niedriger aus. Dazu kommt, dass sich hier auch für die Subtests KFT-V1, KFT-V2 und KFT-V3 niedrigere Koeffizienten finden. Jedoch sind die Statistiken für die übrigen KFT-Subtests, Testdimensionen und die KFT-Gesamtleistung sehr ähnlich.

Die Konsistenzkoeffizienten der Subtests KFT-Q1, KFT-Q3 und KFT-V2 in Klassenstufe 7 fallen am niedrigsten für beide Testformen aus. Das bestätigt teilweise die Befunde von Heller & Perleth (2000), nach denen die Werte für die Untertests KFT-Q1 und KFT-Q3 am niedrigsten bei Testform A ausfallen, aber in Bezug auf Testform B fallen die Koeffizienten der Untertest KFT-V2 und KFT-Q3 am niedrigsten aus. Insgesamt stimmen die Konsistenzkoeffizienten der Skalen des KFT 4-12+ R (Subtests, Testdimensionen und Gesamtleistung) von Klassenstufe 7 für beide Testformen aber gut mit den von Heller und Perleth (2000) berichteten Konsistenzkoeffizienten überein.

Die Retest-Reliabilitätskoeffizienten der Skalen des KFT 4-12+ R (Retestung nach zwei Jahren) fallen sehr niedrig aus, besonders die der Subtests KFT-Q1 und KFT-Q3 für beide Testformen, die so niedrig wie die der inneren Konsistenz sind. Als Erklärung hierfür kann auf das große Zeitintervall (zwei Jahre) zwischen den Durchführungen hingewiesen werden. Das stimmt mit

der Literatur von Perleth und Sierwald (2000) dahingehend überein, dass die Reliabilität unterschätzt werden kann, wenn der Zeitabstand zwischen den Testerhebungen groß ist.

Validität des KFT 4-12+R. Die Faktorenstruktur des KFT 4-12+ R für die beiden Testformen bezüglich der beiden Klassenstufen 5 und 7 ergab jeweils drei Faktoren. Der verbale-sprachliche Faktor wird durch die Untertests des V-Teils, der quantitative Faktor durch die des Q-Teils und der nonverbale Faktor durch die des N-Teils bestimmt. Das bestätigt die faktorenanalytischen Ergebnisse der KFT-Autoren Heller und Perleth (2000; vgl. auch die Ausführungen in Kapitel 3 dieser Arbeit) und den Zweck der Konstruktion des KFT 4-12+ R als ein differentieller Intelligenztest zur Erfassung der kognitiven Fähigkeiten von Schülern/innen der fünften und siebten Klassen in Bezug auf verbale, quantitative und nonverbale Fähigkeiten.

Die Übereinstimmungsvalidität und die prognostische Validität sind zwei von vier Aspekten (siehe Kapitel 3) zur Abschätzung der Kriteriumsvalidität und konnten hier für die Skalen des KFT 4-12+ R anhand einer Stichprobe von Klassenstufe 7 überprüft werden. Die Übereinstimmungskoeffizienten des KFT 4-12+ R zeigen viele signifikante Zusammenhänge zu den Kriteriumsvariablen (Schulnoten). Der KFT 4-12+ R ist also gut zur Messung der verbalen und quantitativen Fähigkeiten geeignet ist. Diese Kompetenzen stimmen also mit den verbalen und quantitativen Schulleistungen überein, was aber so nicht für die nonverbalen Fähigkeiten gilt.

Die prognostischen Validitätskoeffizienten in der vorliegenden Arbeit zeigen, dass der KFT 4-12+ R gute Vorhersage-Beziehungen zu Schulzensuren aufweist, d.h. der KFT 4-12+ R kann zur Prognose schulischer Leistungen verwendet werden. Das bestätigt auch die Rolle der Intelligenz als wichtiger Prädiktor für die Vorhersage der Schulleistung (Heller, 2008; Perleth, 2008; vgl. auch Kapitel 4).

6.2.2 Einflussfaktoren

In der vorliegenden Studie wurde der Einfluss von Muttersprache- und Geschlechtsunterschieden auf KFT-Leistungen zusammen mit Kohorten- und Entwicklungseffekten überprüft.

Geschlechtsunterschiede. Die Ergebnisse bezüglich der Kohorteneffekte sind in beiden Klassenstufen nicht einheitlich: Einerseits unterscheiden sich Jungen und Mädchen in ihren

verbalen und nonverbalen Leistungen in Klassenstufe 5 nicht und in Bezug auf die quantitative- und Gesamtleistung schneiden Jungen besser als Mädchen ab. Andererseits gibt es in Klassenstufe 7 signifikante Differenzen zwischen Jungen und Mädchen zugunsten der Jungen in verbalen, quantitativen und nonverbalen Leistungen sowie in der Gesamtleistung. Diese Befunde stimmen teilweise mit Lynn (1994, zit. n. Hirnstein & Hausmann, 2010; Halpern, 2004, zit. n. Woolfolk, 2008; Rost, 2009; vgl. auch Kapitel 4) drin überein, dass es einen Unterschied zwischen Jungen und Mädchen in Intelligenz- und kognitiven Fähigkeitstests gibt

Zusätzlich bestätigen die vorliegenden Befunde auch das Ergebnis von Lynn (1994, zit. n. Hirnstein & Hausmann, 2010), dass Jungen bessere Ergebnisse in Intelligenztests als Mädchen erzielen. Darüber hinaus sind die Befunde der vorliegenden Untersuchung von daher unerwartet und überraschend, als Jungen nicht nur in der quantitativen Leistung besser abschneiden, sondern auch in der sprachlichen Leistung. Dass stimmt etwa mit den Resultaten von Halpern (2004, zit. n. Woolfolk, 2008; vgl. auch Kapitel 4) nicht überein, dass Mädchen bessere Noten als Jungen in verbalen bzw. sprachlichen Fächern erhalten. Im Hinblick auf Geschlechtsunterschiede in den mathematischen Fähigkeiten bestätigen die Ergebnisse der vorliegenden Untersuchung nicht immer die Befunde in der Literatur, weil widersprüchliche Ergebnisse aus verschiedenen Studien vorlegen, ob diese Unterschiede überhaupt vorhanden und/oder zugunsten der Jungen ausfallen (Woolfolk, 2008; vgl. auch Kapitel 4).

Woher diese Geschlechtsunterschiede in Klassenstufe 5 röhren, muss im vorliegenden Fall Gegenstand von Spekulation bleiben. In Frage kommen etwa auch geschlechtsspezifische Bevorzugung bestimmter Schulen durch die Eltern und nicht unbedingt die kognitiven Fähigkeiten. Die Unterschiede in beiden Klassenstufen könnten auch durch weitere verschiedene Faktoren verursacht sein, etwa biologische Faktoren (Sexualhormone), soziale Faktoren (Rollenmodelleffekte und Stereotype), psychologische Faktoren (Motivation und emotionale Persönlichkeitsmerkmale) und biopsychosoziale Faktoren (hohe Leistung ergibt sich aus einem Komplex unterschiedlicher Einflussfaktoren; siehe auch die Erklärungsansätze in Kapitel 4).

In Bezug auf die Entwicklungseffekte zeigen die vorliegenden Befunde, dass sich Mädchen und Jungen in den KFT-Dimensionen (KFT-V, KFT-Q und KFT-N) und der KFT-Gesamtleistung nicht unterscheiden. Bei beiden Geschlechtsgruppen findet sich eine sehr kleine Verschlechterung in ihrer verbalen Leistung sowie sehr kleine Verbesserung in ihren nonverbalen, quantitativen und der Gesamtleistung d.h. man kann keine Aussage machen, ob

Mädchen oder Jungen bezüglich ihrer kognitiven Fähigkeiten sich eingangs der Sekundarstufe unterschiedlich entwickeln.

Muttersprache. In der vorliegenden Untersuchung wurde festgestellt, dass sich Schüler mit unterschiedlicher "Muttersprache" in den KFT-Leistungen in beiden Analysestichproben in Klassenstufen 5 und 7 unterscheiden, wobei vor allem Differenzen zwischen deutschsprachigen und Schülern/innen, die zu Hause nicht überwiegend Deutsch sprechen, auftreten (siehe Kapitel 5, Tabelle 5.2.1.2). Weiter finden sich für unterschiedliche Kohorten (Jahrgangsgruppen) jeweils etwas unterschiedliche Verhältnisse.

Außerdem zeigen die Befunde, dass die deutschsprachigen Schüler/innen besser als alle anderen Sprachgruppen abschneiden und die türkischsprachigen Schüler/innen die schlechteste Leistung im Vergleich zu den übrigen Sprachgruppen aufweisen. Darüber hinaus zeigen die Ergebnisse, dass sich alle Sprachgruppen von den Jahrgangsgruppen 1998-2000 bis 2007-2011 bezüglich ihrer kognitiven Fähigkeiten in Klasse 7 verschlechtern. Darüber hinaus zeigen die Befunde der Sprachgruppen in Klasse 5, dass es eine Verschlechterung der Leistungen fast bei allen Sprachgruppen in den KFT-Dimensionen von den Jahrgängen 1998-2000 bis 2001-2003 zu verzeichnen ist. Die Sprachgruppe „Osteuropa“ verschlechtert sich hierbei am stärksten. Ab Mitte des Untersuchungszeitraums gibt es sogar eine kleine Steigerung bei den übrigen Sprachgruppen in einigen Skalen. So verbessern sich die deutschsprachigen Schüler/innen in den KFT-Dimensionen und der KFT-Gesamtleistung, die türkischsprachigen im KFT-V, KFT-N und der KFT-Gesamtleistung bis zu den Jahrgängen 2004-2006, die Restgruppe („Sonstige“) im KFT-N, im KFT-V und ebenfalls in der KFT-Gesamtleistung.

Diese Ergebnisse bestätigen die Resultate der oben in Kapitel 4 von Kristen (2006) zitierten Studie, nach der deutsche Schüler/innen bessere Noten als ausländische Schüler/innen bzw. türkische und italienische Schüler/innen erzielt haben. Als Begründung für diese Befunde könnte auf kulturelle Defizite verwiesen werden, wie es etwa Diefenbach (2007; vgl. ausführlicher in Kapitel 4) vorschlägt, wenn die Unterschiede nur zwischen deutsch und türkischsprachigen Schüler/innen auftreten würden, aber es gibt auch Unterschiede innerhalb der europäischen Sprachgruppen, deren Kultur im Vergleich mit der türkischen größere Ähnlichkeiten zur deutschen aufweist. Deswegen müssten andere Faktoren diese Differenzen verursachen wie etwa der sozioökonomische Status (Wohnort, Schulabschluss, Beruf und Einkommen der Eltern usw.; siehe Woolfolk, 2008; Tickly, 2006; vgl. ausführlicher Kapitel 4). Darüber hinaus könnten die Unterschiede auf das Messinstrument bzw. den KFT 4-12+ R besonders im verbalen und

sprachlichen Teil zurückzuführen sein, da der KFT 4-12+ R nur Wortschatz und keine andere Deutschkompetenzen wie Grammatikkenntnisse u.ä. erfasst.

Bezüglich der Entwicklungseffekte zeigen die bestehenden Ergebnisse, dass die Veränderungen aller Sprachgruppen, also sowohl die Verschlechterung aller Sprachgruppen, deren Muttersprache kein Deutsch ist, im KFT-V als auch die Verbesserung aller Sprachgruppen in den Dimensionen KFT-Q, KFT-N und der KFT-Gesamtleistung eher unbedeutend sind, weil diese Veränderungen nur 1 bis 2 T-Werte betragen (Bortz, 1995; Cohen, 1977; Cohen, 1992).

6.2.3 Veränderungen der Leistungen in den KFT-Dimensionen und der KFT-Gesamtleistung in den untersuchten Klassenstufen im letzten Jahrzehnt

Die Befunde der hier vorliegenden Untersuchung zeigen deutlich, dass die KFT-Leistungen bei Fünfklässlern/innen von den Jahrgängen 1998-2000 bis zu den Jahrgängen 2001-2003 abnehmen, danach gibt es kaum Veränderung. Das gilt teilweise auch für die Siebtklässler/innen, bei denen die KFT-Leistungen von 1998-2000 bis zu 2007-2011 abfallen. Diese Ergebnisse bestätigen jedenfalls den Befund von Flynn (1987; hiernach ist der sogenannten Flynn-Effekt benannt) nicht, der anhand der Daten von 14 Industrieländern zu dem Ergebnis gekommen ist, dass die IQ-Werte um 5 bis 25 Punkte pro Generation zunehmen. Man kann den Befund aber auch nicht so interpretieren, dass diese Veränderungen auf einen Rückgang der kognitiven Fähigkeiten oder der Intelligenz bei der Schülerschaft der Schule zurückzuführen sind, sondern sie könnten auch auf unsystematische Schwankungen von Jahrgang zu Jahrgang zurückzuführen sein.

6.2.4 Veränderung kognitiver Fähigkeiten von Klasse 5 zu Klasse 7

Die Ergebnisse dieser Untersuchung zeigen, dass sich die KFT-Leistungen von Klasse 5 zu Klasse 7 in den Jahrgängen 1998-2000 um 3 bis 4 T-Werte verbessern. Außerdem fallen die Leistungen im KFT-V ein bisschen in den Jahrgängen 2004-2006 ab. Das gilt aber nicht für den quantitativen Teil, wo die größten Steigerungen bei den ältesten Kohorten und geringere bei den Kohorten 2004-2006 und 2007-2010 zu verzeichnen sind. In Bezug auf die KFT-N-Leistung und die KFT-Gesamtleistung sind Verschlechterungen ab 2004-2006 zu verzeichnen, aber diese

Verschlechterungen bleiben kleiner als 2-T-Wert-Punkte und sind damit als praktisch gering einzuschätzen (im Sinne etwa von Bortz, 1995; Cohen, 1977; Cohen, 1992).

6.3 Folgerungen und Ausblick

Einsatz des kognitiven Intelligenztests KFT 4-12+ R in der Willy Brandt-Gesamtschule in Klassenstufen 5 und 7. Die vorliegenden Ergebnisse zeigen auf, dass einige Items im KFT 4-12+ R anhand der Itemanalyse ungültig sind, da ihre Schwierigkeitsindices entweder $p < .20$ oder $p > .80$ sowie ihre Trennschärfe $r < .30$ sind. Außerdem sollten die Reliabilitätskoeffizienten der Intelligenztests und Fähigkeitstests nach Perleth und Sierwald (2000) von $Rel = .85$ bis $Rel = .95$ (innere Konsistenz) sowie von $Rel = .80$ bis $Rel = .90$ (Stabilität) variieren. Die Konsistenz-Reliabilitätskoeffizienten der KFT 4-12+ R-Untertests liegen für die beiden Testformen nur im befriedigenden Bereich. Das gilt auch für die Stabilitätskoeffizienten der KFT 4-12+ R Subtests. Darüber hinaus weisen die Ergebnisse zur Übereinstimmungsvalidität darauf hin, dass die nonverbalen Fähigkeiten, so wie sie mit dem KFT 4-12+ R gemessen werden können, zur Vorhersage der Schulleistungen nicht so relevant sind. Außerdem wäre zu überlegen, ob der KFT 4-12+ R im verbalen Teil nicht nur Wortschatz-aufgaben, sondern auch andere Deutschkompetenzen wie Grammatik enthalten sollte.

Aus diesen Gründen empfiehlt sich deswegen der Einsatz des KFT 4-12+ R nur als Teil eines umfassenden diagnostischen Verfahrens und nicht als alleiniges Kriterium für Empfehlungen in der Schulberatung und darauf gegründete Entscheidungen, welche Kurse Schüler/innen benötigen, oder für die Empfehlung, welche Schullaufbahn die Schüler/innen anstreben sollten. Zusätzlich wäre es sinnvoll, wenn es für KFT 4-12+ R mehrere Versionen in unterschiedlichen Sprachen gäbe wie etwa Türkisch oder Arabisch. Das könnte bei der Testung von Kindern und Jugendlichen mit nicht deutscher Muttersprache, die in Deutschland leben, sehr helfen. Jedes Kind könnte in der Sprache (Deutsch, Türkisch usw.) getestet werden, das es am besten beherrscht.

Passende Förderkurse in Willy-Brandt-Gesamtschule. Die verbalen, quantitativen und nonverbalen Kompetenzen der Fünfklässler und Siebtklässler sind nach den hier präsentierten Auswertungen in der Willy-Brandt-Gesamtschule in den letzten Jahren etwas abgesunken. Außerdem beeinflusst „Muttersprache“ die kognitiven Fähigkeiten. Dementsprechend sollten Schüler/innen der Willy-Brandt-Gesamtschule durch entsprechende Kurse bzw. sprachliche und

kulturelle Kurse gefördert werden, besonders im verbalen Teil, weil ein Teil der Stichprobe Ausländer ist, die zu Hause kaum Deutsch sprechen. Zusätzlich ist es notwendig, den Unterricht durch gute Kommunikation und Beziehung Lehrer-Schüler oder regelmäßige Bewertung und Materialwechsel zu optimieren. Außerdem sind nicht nur kognitive Faktoren für den Schulerfolg einzubeziehen, sondern auch nicht kognitive Faktoren wie Persönliche Faktoren wie Motivation und Interesse, soziale Faktoren wie Erziehungsstil sowie Einstellungen der ausländischen Eltern über die Schulleistung usw.

Außerdem können die Ergebnisse dieser Studie nicht nur als Hilfsmittel für die Schulberatung in der Willy-Brandt-Gesamtschule verwendet werden, sondern auch für Forschung und Entwicklung theoretischen Modelle bzw. Intelligenz- und Leistungsmodelle fruchtbar gemacht werden.

Tabellenverzeichnis

Tabelle 3.1:	Skalen des HAWIK-IV	46
Tabelle 3.2:	Subtests des AID2	48
Tabelle 3.3:	Skalen und Untertests des K-ABC	52
Tabelle 3.4:	Untertests des PSB-R 6-13	56
Tabellen 3.5:	Die Skalen des CFT 1-R	59
Tabelle 3.6:	Die Untertests des I-S-T 2000 R	70
Tabelle 3.7:	Übersicht über Testverfahren zur Messung der Intelligenz und kognitiver Fähigkeiten	74
Tabelle 5.2.1.1:	Verteilung der Stichprobe pro Jahrgang vom Schuljahr 1998/1999 bis 2010/2011 bzw. 2011/2012 in den beiden Klassenstufen 5 und 7	99
Tabelle 5.2.1.2:	Verteilung der Teil-Stichproben der einzelnen Sprachgruppen in den Klassenstufen 5 und 7	100
Tabelle 5.2.2.1:	Testaufbau und Zeitbedarf des KFT 4-12+R	101
Tabelle 5.3.1.1:	Durchschnittliche Itemschwierigkeiten der Skalen des KFT 4-12+R für beide Testformen und die Klassenstufen 5 und 7	107
Tabelle 5.3.1.2:	Durchschnittliche Itemtrennschärfen der Skalen des KFT 4-12+R für beide Testformen und die Klassenstufen 5 und 7	108
Tabelle 5.3.1.3:	Schwierigkeits- und Trennschärfeindizes für den Untertest V1 des KFT 4-12+R in Klasse 5 für beide Testformen A und B	108
Tabelle 5.3.1.4:	Schwierigkeits- und Trennschärfeindizes für den Untertest V2 des KFT 4-12+R in Klasse 5 für beide Testformen A und B	109
Tabelle 5.3.1.5:	Schwierigkeits- und Trennschärfeindizes für den Untertest V3 des KFT 4-12+R in Klasse 5 für beide Testformen A und B	109
Tabelle 5.3.1.6:	Schwierigkeits- und Trennschärfeindizes für den Untertest Q1 des KFT 4-12+R in Klasse 5 für beide Testformen A und B	110
Tabelle 5.3.1.7:	Schwierigkeits- und Trennschärfeindizes für den Untertest Q2 des KFT 4-12+R in Klasse 5 für beide Testformen A und B	111
Tabelle 5.3.1.8:	Schwierigkeits- und Trennschärfeindizes für den Untertest Q3 des KFT 4-12+R in Klasse 5 für beide Testformen A und B	111
Tabelle 5.3.1.9:	Schwierigkeits- und Trennschärfeindizes für den Untertest N1 des KFT 4-12+R in Klasse 5 für beide Testformen A und B	112
Tabelle 5.3.1.10:	Schwierigkeits- und Trennschärfeindizes für den Untertest N2 des KFT 4-12+R in Klasse 5 für beide Testformen A und B	112
Tabelle 5.3.1.11:	Schwierigkeits- und Trennschärfeindizes für den Untertest N3 des KFT 4-12+R in Klasse 5 für beide Testformen A und B	113
Tabelle 5.3.1.12:	Schwierigkeits- und Trennschärfeindizes für die Dimension V des KFT 4-12+R in Klasse 5 für beide Testformen A und B	114
Tabelle 5.3.1.13:	Schwierigkeits- und Trennschärfeindizes für die Dimension Q des KFT 4-12+R in Klasse 5 für beide Testformen A und B	115

Tabelle 5.3.1.14: Schwierigkeits- und Trennschärfeindizes für die Dimension N des KFT 4-12+R in Klasse 5 für beide Testformen A und B	117
Tabelle 5.3.1.15: Schwierigkeits- und Trennschärfeindizes für die Gesamtleistung GL des KFT 4-12+R in Klasse 5 für beide Testformen A und B	118
Tabelle 5.3.1.16: Schwierigkeits- und Trennschärfeindizes für den Untertest V1 des KFT 4-12+R in Klasse 7 für beide Testformen A und B	123
Tabelle 5.3.1.17: Schwierigkeits- und Trennschärfeindizes für den Untertest V2 des KFT 4-12+R in Klasse 7 für beide Testformen A und B	123
Tabelle 5.3.1.18: Schwierigkeits- und Trennschärfeindizes für den Untertest V3 des KFT 4-12+R in Klasse 7 für beide Testformen A und B	124
Tabelle 5.3.1.19: Schwierigkeits- und Trennschärfeindizes für den Untertest Q1 des KFT 4-12+R in Klasse 7 für beide Testformen A und B	125
Tabelle 5.3.1.20: Schwierigkeits- und Trennschärfeindizes für den Untertest Q2 des KFT 4-12+R in Klasse 7 für beide Testformen A und B	126
Tabelle 5.3.1.21: Schwierigkeits- und Trennschärfeindizes für den Untertest Q3 des KFT 4-12+R in Klasse 7 für beide Testformen A und B	126
Tabelle 5.3.1.22: Schwierigkeits- und Trennschärfeindizes für den Untertest N1 des KFT 4-12+R in Klasse 7 für beide Testformen A und B	127
Tabelle 5.3.1.23: Schwierigkeits- und Trennschärfeindizes für den Untertest N2 des KFT 4-12+R in Klasse 7 für beide Testformen A und B	127
Tabelle 5.3.1.24: Schwierigkeits- und Trennschärfeindizes für den Untertest N3 des KFT 4-12+R in Klasse 7 für beide Testformen A und B	128
Tabelle 5.3.1.25: Schwierigkeits- und Trennschärfeindizes für die Dimension V des KFT 4-12+R in Klasse 7 für beide Testformen A und B	129
Tabelle 5.3.1.26: Schwierigkeits- und Trennschärfeindizes für die Dimension Q des KFT 4-12+R in Klasse 7 für beide Testformen A und B	130
Tabelle 5.3.1.27: Schwierigkeits- und Trennschärfeindizes für die Dimension N des KFT 4-12+R in Klasse 7 für beide Testformen A und B	132
Tabelle 5.3.1.28: Schwierigkeits- und Trennschärfeindizes für die Gesamtleistung GL des KFT 4-12+R in Klasse 7 für beide Testformen A und B	133
Tabelle 5.3.1.29: Verteilungsmerkmale der KFT-Subtests, Testdimensionen und der Gesamtleistung für Klasse 5 Form A.....	142
Tabelle 5.3.1.30: Verteilungsmerkmale der KFT-Subtests, Testdimensionen und der Gesamtleistung für Klasse 5 Form B	143
Tabelle 5.3.1.31: Verteilungsmerkmale der KFT-Subtests, Testdimensionen und der Gesamtleistung für Klasse 7 Form A.....	148
Tabelle 5.3.1.32: Verteilungsmerkmale der KFT-Subtests, Testdimensionen und der Gesamtleistung für Klasse 7 Form B	148
Tabelle 5.3.2.1: Konsistenzkoeffizienten der Subtests und der Gesamtleistung des KFT 4-12+ R von Klasse 5 für Testform A	149
Tabelle 5.3.2.2: Konsistenzkoeffizienten der Subtests und der Gesamtleistung des KFT 4-12+ R von Klasse 5 für Testform B	150

Tabelle 5.3.2.3: Konsistenzkoeffizienten der Subtests und der Gesamtleistung des KFT 4-12+ R von Klasse 7 für Testform A	150
Tabelle 5.3.2.4: Konsistenzkoeffizienten der Subtests und der Gesamtleistung des KFT 4-12+ R von Klasse 7 für Testform B	151
Tabelle 5.3.2.5: Stabilität der Subtests, Testteile und der Gesamtleistung des KFT 4-12+R.....	152
Tabelle 5.3.3.1: Faktorenstruktur des KFT4-12+ R in Klassenstufe 5 Testform A	155
Tabelle 5.3.3.2: Faktorenstruktur des KFT4-12+ R in Klassenstufe 5 Testform B	155
Tabelle 5.3.3.3: Faktorenstruktur des KFT4-12+ R in Klassenstufe 7 Testform A.....	157
Tabelle 5.3.3.4: Faktorenstruktur des KFT4-12+ R in Klassenstufe 7 Testform B	158
Tabelle 5.3.3.5: Korrelationsergebnisse der Skalen des KFT-4-12+R der Siebklässeler mit den Schulnoten aus Klasse 7 (Übereinstimmungsvalidität)	160
Tabelle 5.3.3.6: Korrelationsergebnisse der Skalen des KFT-4-12+R von Klasse 5 mit den Schulnoten von Klasse 7 (prognostische Validität).....	161
Tabelle 5.4.1.1: Varianzanalysen zu den Unterschieden der Jahrgangsgruppen in den KFT 4-12+R Dimensionen (KFT-V, KFT-Q, KFT-N) und der Gesamtleistung für die Klassenstufe 5	163
Tabelle 5.4.1.2: Post-Hoc-Test „homogene Untergruppen“ der verschiedenen Jahrgangsgruppen hinsichtlich KFT-V für Klassenstufe 5	164
Tabelle 5.4.1.3: Mittelwerte von KFT-Verbal der Schüler für Klassenstufe 5 nach Jahrgängen.....	164
Tabelle 5.4.1.4: Post-Hoc-Test „homogene Untergruppen“ der verschiedenen Jahrgangsgruppen hinsichtlich KFT-Q für Klassenstufe 5	164
Tabelle 5.4.1.5: Mittelwerte von KFT-Quantitativ der Schüler für Klassenstufe 5 nach Jahrgängen.....	165
Tabelle 5.4.1.6: Post-Hoc-Test „homogene Untergruppen“ der verschiedenen Jahrgangsgruppen hinsichtlich KFT-N für Klassenstufe 5	165
Tabelle 5.4.1.7: Mittelwerte des KFT-Nonverbal der Schüler für Klassenstufe 5 nach Jahrgangsgruppen	165
Tabelle 5.4.1.8: Post-Hoc-Test „homogene Untergruppen“ der verschiedenen Jahrgangsgruppen hinsichtlich KFT-GL für Klassenstufe 5	166
Tabelle 5.4.1.9: Mittelwerte der KFT-Gesamtleistung der Schüler für Klassenstufe 5 nach Jahrgangsgruppen	166
Tabelle 5.4.1.10: Varianzanalysen zu den Unterschieden der Jahrgangsgruppen in den KFT 4-12+R Dimensionen (KFT-V, KFT-Q, KFT-N) und der Gesamtleistung für die Klassenstufe 7	168
Tabelle 5.4.1.11: Post-Hoc-Test „homogene Untergruppen“ der verschiedenen Jahrgangsgruppen hinsichtlich KFT-V für Klassenstufe 7	169
Tabelle 5.4.1.12: Mittelwerte von KFT-Verbal der Schüler für Klassenstufe 7 nach Jahrgängen.....	169
Tabelle 5.4.1.13: Post-Hoc-Test „homogene Untergruppen“ der verschiedenen Jahrgangsgruppen hinsichtlich KFT-Q für Klassenstufe 7	170

Tabelle 5.4.1.14: Mittelwerte von KFT-Quantitativ der Schüler für Klassenstufe 7 nach Jahrgängen.....	170
Tabelle 5.4.1.15: Post-Hoc-Test „homogene Untergruppen“ der verschiedenen Jahrgangsgruppen hinsichtlich KFT-N für Klassenstufe 7	171
Tabelle 5.4.1.16: Mittelwerte des KFT-Nonverbal der Schüler für Klassenstufe 7 nach Jahrgangsgruppen.....	171
Tabelle 5.4.1.17: Post-Hoc-Test „homogene Untergruppen“ der verschiedenen Jahrgangsgruppen hinsichtlich KFT-GL für Klassenstufe 7	171
Tabelle 5.4.1.18: Mittelwerte der KFT-Gesamtleistung der Schüler für Klassenstufe 7 nach Jahrgangsgruppen	171
Tabelle 5.4.2.1: Varianzanalysen der KFT 4-12+R Dimensionen (KFT-V, KFT-Q, KFT-N) und der Gesamtleistung von Klassenstufe 5 nach Geschlecht und Jahrgangsgruppen.....	174
Tabelle 5.4.2.2: Varianzanalysen der KFT 4-12+R Dimensionen (KFT-V, KFT-Q, KFT-N) und der Gesamtleistung von Klassenstufe 7 nach Geschlecht und Jahrgangsgruppen.....	179
Tabelle 5.4.3.1: Varianzanalysen der KFT 4-12+R Dimensionen (KFT-V, KFT-Q, KFT-N) und der Gesamtleistung von Klassenstufe 5 nach Sprach- und Jahrgangsgruppen	184
Tabelle 5.4.3.2: Varianzanalysen der KFT 4-12+R Dimensionen (KFT-V, KFT-Q, KFT-N) und der Gesamtleistung von Klassenstufe 7 nach Sprach- und Jahrgangsgruppen	188
Tabelle 5.4.4.1: Varianzanalysen der Differenzen der Leistungen im KFT 4-12+R in den Dimensionen (KFT-V, KFT-Q, KFT-N) und der Gesamtleistung (KFT-GL) von Klassenstufe 5 zu Klassenstufe 7 nach Jahrgangsgruppen.....	193
Tabelle 5.4.4.2: Mittelwerte der Differenzen bezüglich der KFT 4-12+R Dimensionen (V, Q, N) und der Gesamtleistung der Schüler von Klasse 5 zu Klasse 7 nach Jahrgängen.....	194
Tabelle 5.4.4.3: Post-Hoc-Test „homogene Untergruppen“ der verschiedenen Jahrgänge hinsichtlich der Differenzen im KFT-V von Klassenstufe 5 zu Klassenstufe 7	195
Tabelle 5.4.4.4: Post-Hoc-Test „homogene Untergruppen“ der verschiedenen Jahrgänge hinsichtlich der Differenzen im KFT-Q von Klassenstufe 5 zu Klassenstufe 7	195
Tabelle 5.4.4.5: Post-Hoc-Test „homogene Untergruppen“ der verschiedenen Jahrgänge hinsichtlich der Differenzen im KFT-N von Klassenstufe 5 zu Klassenstufe 7	196
Tabelle 5.4.4.6: Post-Hoc-Test „homogene Untergruppen“ der verschiedenen Jahrgänge hinsichtlich der Differenzen im KFT-GL von Klassenstufe 5 zu Klassenstufe 7	196
Tabelle 5.4.4.7: Varianzanalysen der Differenzen der Leistungen im KFT 4-12+R in den Dimensionen (KFT-V, KFT-Q, KFT-N) und der Gesamtleistung (KFT-GL) von Klassenstufe 5 zu Klassenstufe 7 nach Sprachgruppen	199

Tabelle 5.4.4.8: Mittelwerte der Differenzen bezüglich der KFT 4-12+R Dimensionen (V, Q, N) und der Gesamtleistung der Schüler von Klasse 5 zu Klasse 7 nach Sprachgruppen	199
Tabelle 5.4.4.9: Post-Hoc-Test „homogene Untergruppen“ der verschiedenen Sprachgruppen hinsichtlich der Differenzen im KFT-V von Klassenstufe 5 zu Klassenstufe 7	200
Tabelle 5.4.4.10: Post-Hoc-Test „homogene Untergruppen“ der verschiedenen Sprachgruppen hinsichtlich der Differenzen im KFT-Q von Klassenstufe 5 zu Klassenstufe 7	200
Tabelle 5.4.4.11: Post-Hoc-Test „homogene Untergruppen“ der verschiedenen Sprachgruppen hinsichtlich der Differenzen im KFT-N von Klassenstufe 5 zu Klassenstufe 7	200
Tabelle 5.4.4.12: Post-Hoc-Test „homogene Untergruppen“ der verschiedenen Sprachgruppen hinsichtlich der Differenzen im KFT-GL von Klassenstufe 5 zu Klassenstufe 7	200
Tabelle 5.4.4.13: t-Tests der Differenzen der KFT-Leistungen in den KFT-Dimensionen und der KFT-Gesamtleistung von Klassenstufe 5 zu Klassenstufe 7 nach Geschlecht	202
Tabelle 5.4.4.14: Mittelwerte der Differenzen bezüglich der KFT 4-12+R Dimensionen (V, Q, N) und der Gesamtleistung der Schüler von Klasse 5 zu Klasse 7 nach Geschlecht.....	203

Abbildungsverzeichnis

Abbildung 2.1:	Spearmans Modell	14
Abbildung 2.2:	Intelligenzstrukturmodell von Guilford	18
Abbildung 2.3:	Berliner Intelligenzstrukturmodell von Jäger.....	19
Abbildung 5.3.1.1:	Verteilungsmerkmale der KFT-Subtests, Testdimensionen und der Gesamtleistung für Klasse 5 Form A und B	142
Abbildung 5.3.1.2:	Verteilungsmerkmale der KFT-Subtests, Testdimensionen und der Gesamtleistung für Klasse 7 Form A und B.	147
Abbildung 5.3.3.1:	Scree-Plot zur Faktorenanalyse von Klassenstufe 5, Testform A.....	154
Abbildung 5.3.3.2:	Scree-Plot zur Faktorenanalyse von Klassenstufe 5, Testform B	154
Abbildung 5.3.3.3:	Scree-Plot zur Faktorenanalyse von Klassenstufe 7, Testform A.....	156
Abbildung 5.3.3.4:	Scree-Plot zur Faktorenanalyse von Klassenstufe 7, Testform B	157
Abbildung 5.4.1.1:	Veränderung der KFT-Dimensionen (KFT-V, KFT-Q, KFT-N) in Klassenstufe 5	167
Abbildung 5.4.1.2:	Veränderung der KFT-Gesamtleistung in Klassenstufe 5	167
Abbildung 5.4.1.3:	Veränderung der KFT-Dimensionen (KFT-V, KFT-Q, KFT-N) in Klassenstufe 7.....	172
Abbildung 5.4.1.4:	Veränderung der KFT-Gesamtleistung in Klassenstufe 7	172
Abbildung 5.4.2.1:	Entwicklung der verbalen KFT-Leistung nach Geschlecht und Jahrgangsgruppen in Klassenstufe 5	176
Abbildung 5.4.2.2:	Entwicklung der quantitativen KFT-Leistung nach Geschlecht und Jahrgangsgruppen in Klassenstufe 5	176
Abbildung 5.4.2.3:	Entwicklung der nonverbalen KFT-Leistung nach Geschlecht und Jahrgangsgruppen in Klassenstufe 5	177
Abbildung 5.4.2.4:	Entwicklung der KFT-Gesamtleistung nach Geschlecht und Jahrgangsgruppen in Klassenstufe 5	177
Abbildung 5.4.2.5:	Entwicklung der verbalen KFT-Leistung nach Geschlecht und Jahrgangsgruppen in Klassenstufe 7	180
Abbildung 5.4.2.6:	Entwicklung der quantitativen KFT-Leistung nach Geschlecht und Jahrgangsgruppen in Klassenstufe 7	180
Abbildung 5.4.2.7:	Entwicklung der nonverbalen KFT-Leistung nach Geschlecht und Jahrgangsgruppen in Klassenstufe 7	181
Abbildung 5.4.2.8:	Entwicklung der KFT-Gesamtleistung nach Geschlecht und Jahrgangsgruppen in Klasse 7	181
Abbildung 5.4.3.1:	Verbale KFT-Leistung nach Sprach- und Jahrgangsgruppe in Klassenstufe 5	185
Abbildung 5.4.3.2:	Quantitative KFT-Leistung nach Sprach- und Jahrgangsgruppe in Klassenstufe 5	185
Abbildung 5.4.3.3:	Nonverbale KFT-Leistung nach Sprach- und Jahrgangsgruppe in Klassenstufe 5	186

Abbildung 5.4.3.4: KFT-Gesamtleistung nach Sprach- und Jahrgangsgruppe in Klassenstufe 5	186
Abbildung 5.4.3.5: Verbale KFT-Leistung nach Sprach- und Jahrgangsgruppe in Klassenstufe7	189
Abbildung 5.4.3.6: Quantitative KFT-Leistung nach Sprach- und Jahrgangsgruppe in Klassenstufe 7	190
Abbildung 5.4.3.7: Nonverbale KFT-Leistung nach Sprach- und Jahrgangsgruppe in Klassenstufe 7	191
Abbildung 5.4.3.8: KFT-Gesamtleistung nach Sprach- und Jahrgangsgruppe in Klassenstufe 7	191
Abbildung 5.4.4.1: Veränderung der KFT-Leistungen (Dimensionen) von Klassenstufe 5 bis Klassenstufe 7 für die Jahrgangsgruppen.....	197
Abbildung 5.4.4.2: Veränderung der KFT-Gesamtleistung von Klassenstufe 5 bis Klassenstufe 7 für die Jahrgangsgruppen	197
Abbildung 5.4.4.3: Veränderung der KFT-Leistungen (Dimensionen) von Klassenstufe 5 bis Klassenstufe 7 für die verschiedenen Sprachgruppen	201
Abbildung 5.4.4.4: Veränderung der KFT-Gesamtleistung von Klassenstufe 5 bis Klassenstufe 7 für die verschiedenen Sprachgruppen.....	201
Abbildung 5.4.4.5: Veränderung der KFT-Leistungen (Dimensionen) von Klassenstufe 5 bis Klassenstufe 7 nach Geschlecht	203
Abbildung 5.4.4.6: Veränderung der KFT-Gesamtleistung von Klassenstufe 5 bis Klassenstufe 7 nach Geschlecht	204

Literaturverzeichnis

- Asendorpf, J. B. (2004). *Psychologie der Persönlichkeit* (3., überarb. und aktualisierte Aufl.). Berlin: Springer.
- Amthauer, R., Brocke, B., Liepmann, D. & Beauducel, A. (2001). *I-S-T 2000 R. Intelligenz-Struktur-Test 2000 R*. Göttingen: Hogrefe.
- Beerman, L., Heller, K., A., & Menacher, P. (1992). *Mathe: nichts für Mädchen? Begabung und Geschlecht am Beispiel von Mathematik, Naturwissenschaft und Technik* (1. Aufl.). Bern: Huber.
- Bortz, J., & Döring, N. (1995). *Forschungsmethoden und Evaluation*. (Zweite, vollständig überarbeitete und aktualisierte Auflage). Berlin: Springer.
- Bühl, A. (2010). *PASW 18. Einführung in die moderne Datenanalyse*. (12., aktualisierte Aufl.). München: Pearson Studium.
- Bühner, M. (2004). *Einführung in die Test- und Fragebogenkonstruktion*. München: Pearson Studium.
- Bühner, M. (2006). *Einführung in die Test- und Fragebogenkonstruktion* (2., aktualisierte und erweiterte Aufl.). München: Pearson Studium.
- Cohen, J. (1977). *Statistical Power Analysis for the Behavioral Sciences*. 2 New York. Academic Press.
- Cohen, J. (1992): A power primer. *Psychological Bulletin*, 112, 155-159.
- Diefenbach, H. (2007). *Kinder und Jugendliche aus Migrantenfamilien im deutschen Bildungssystem. Erklärungen und empirische Befunde*. Wiesbaden: VS Verlag Für Sozialwissenschaften/GWV Fachverlag GmbH.
- Ericsson, K. A. & Charness, N. (1994). Expert performance: Its structure and acquisition. *American Psychologist*, 49, 725-747.
- Friedman, H. S. & Schustack, M. W. (2004). *Persönlichkeitspsychologie und differentielle Psychologie* (2., aktualisierte Aufl.). München: Person Studium.
- Flynn, J. R. (1987). Massive IQ gains in 14 nations: What IQ tests really measure. *Psychological Bulletin*, 101(2), 171-191.
- Gardner; H. (1994). *Abschied vom IQ. Die Rahmen-Theorie der vielfachen Intelligenzen*. Stuttgart: Klett-Cotta.

- Grob, A., Hagmann-von Arx, P., & Bodmer, N. (2009). Entwicklungsdiagnostik. In S. Schneider & J. Margraf (Hrsg.), *Lehrbuch der Verhaltenstherapie* (S. 145-158). Heidelberg: Springer.
- Heinze, A. (2005). *Lösungsverhalten mathematisch begabter Grundschulkinder- aufgezeigt an ausgewählten Problemstellungen* (1. Aufl.). Münster, Westf: LIT.
- Heller, K.A. (1973) *Intelligenzmessung, zur Theorie und Praxis der Begabungsdiagnostik in Schule und Sonderpädagogik unter Berücksichtigung neuer Forschungsergebnisse*. Villingen-Schwenningen: Neckar Verlag.
- Heller, K. A. (Hrsg.). (2000). *Begabungsdiagnostik in der Schul- und Erziehungsberatung* (2., vollst. überarb. Aufl.). Bern: Huber.
- Heller, K. A. (2000). Einführung in den Gegenstandsbereich der Begabungsdiagnostik. In K. A. Heller (Hrsg.), *Begabungsdiagnostik in der Schul- und Erziehungsberatung*. (2., vollst. überarb. Aufl., S. 13-40) Bern: Huber.
- Heller, K. A. (2008). *Von der Aktivierung der Begabungsreserven zur Hochbegabtenförderung, Forschungsergebnisse aus vier Dekaden*. Berlin: LIT-Verlag.
- Heller, K. & Geisler, H. J. (1983). *Kognitiver Fähigkeitstest (Grundschulform)*. KFT 1-3. Weinheim: Beltz.
- Heller, K. & Geisler, H. J. (1983). *Kognitiver Fähigkeitstest (Kindergartenform)* KFT-K. Weinheim: Beltz.
- Heller, K.A. & Perleth, Ch. (2000). Diagnostik. In *Lexikon der Psychologie*, Band 1 (S. 311-314). Heidelberg: Spektrum Akademischer Verlag.
- Heller, K. A. & Perleth, Ch. (2000). Informationsquellen und Messinstrumente. In K. A. Heller (Hrsg.), *Begabungsdiagnostik in der Schul- und Erziehungsberatung*. (2., vollst. überarb. Aufl., S. 96-216) Bern: Huber.
- Heller, K.A. & Perleth, Ch. (2000). *Kognitiver Fähigkeitstest für 4. bis 12. Klassen. Revision (KFT 4-12+ R)*. Göttingen: Beltz.
- Hirnstein, M. & Hausmann, M. (2010). Kognitive Geschlechtsunterschiede. In G. Steins (Hrsg.), *Handbuch Psychologie und Geschlechterforschung* (1. Aufl. S.69-85). Wiesbaden: VS Verlag für Sozialwissenschaften/GWV Fachverlag GmbH.
- Holling, H. & Kanning, U. P. (1999). *Hochbegabung, Forschungsergebnisse und Fördermöglichkeiten*. Göttingen: Hogrefe .
- Holling, H.; Preckel, F.; & Vock, M. (2004): *Intelligenzdiagnostik*. Göttingen: Hogrefe

- Horn, W., Lukesch, H., Kormann, A., & Mayrhofer, S. (2002). *Prüfsystem für Schul- und Bildungsberatung für 4. bis 6. Klassen, Revidierte Fassung (PSB- R 4-6)*. Göttingen: Hogrefe.
- Horn, W., Lukesch, H., Mayrhofer, S., & Kormann, A. (2004). *Prüfsystem für Schul- und Bildungsberatung für 6. bis 13. Klassen, Revidierte Fassung (PSB- R 6-13)*. Göttingen: Hogrefe.
- Hülsheger, U. R., Maier, G. W., Stumpp, T. & Muck, P. M. (2006). Vergleich kriteriumsbezogener Validitäten verschiedener Intelligenztests zur Vorhersage von Ausbildungserfolg in Deutschland. *Zeitschrift für Personalpsychologie*, 5 (4), 145-162.
- Ingenkamp, K. (1997). *Lehrbuch der Pädagogischen Diagnostik* (Studienausgabe., 4., neu ausgestattete Aufl.). Weinheim: Beltz.
- Ingenkamp, K. & Lissmann, U. (2008). *Lehrbuch der pädagogischen Diagnostik* (6., neu ausgestattete Aufl.). Weinheim: Beltz.
- Joswig, H. (Hrsg) (2000). *Begabungen erkennen- Begabte fördern. Beiträge anlässlich der Wissenschaftlichen Arbeitstagung des ABB e. V. in Rostock vom 22-24.10.1999*. Universität Rostock: ABB-Drucke 4.
- Kail, R. & Pellegrino, J. W. (1989). *Menschliche Intelligenz, die drei Ansätze der Psychologie* (2. Aufl.). Heidelberg: Spektrum der Wissenschaft.
- Koop, C., Schenker, I., Müller, G., Welzien, S. & Karg-Stiftung. (Hrsg.). (2010). *Begabung wagen. Ein Handbuch für den Umgang mit Hochbegabung in Kindertagesstätten*. Weimar: Verlag das Netz.
- Kristen, C. (2006). Ethnische Diskriminierung in der Grundschule? Die Vergabe von Noten und Bildungsempfehlungen. *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 58 (1), 79-97.
- Kubinger, K. D & Wurst, E. (2000). *Adaptives Intelligenz Diagnostikum 2 (AID 2)*. Göttingen: Beltz.
- Lienert, G. A. (1961). *Testaufbau und Testanalyse*. Weinheim: Beltz.
- Lienert, G. A. & Raatz, U. (1998). *Testaufbau und Testanalyse* (6. Aufl.). Weinheim: Beltz.
- Lindenberger, U. & Schaefer, S. (2008). Erwachsenenalter und Alter. In R. Oerter & L. Montada (Hrsg.), *Entwicklungspsychologie*. (6., vollst. überarb. Aufl.). (S. 366-409). Weinheim: Beltz.

- Maltby, J., Day, L. & Macaskill, A. (2011). *Differentielle Psychologie, Persönlichkeit und Intelligenz*. (2., aktualisierte Aufl.). München: Pearson Studium.
- Mecheril, P. & Quehl, T. (Hrsg.). (2006). *Die Macht der Sprachen - Englische Perspektiven auf die mehrsprachige Schule*. Münster: Waxmann.
- Melchers, P. & Preuß, U. (1991). *Kaufman-Assessment Battery for Children (K-ABC) von Kaufman, A. S. & Kaufman, N. L, Durchführungs- und Auswertungshandbuch*. Amsterdam: Swets & Zeitlinger.
- Melchers, P. & Preuß, U. (1997). *Kaufman-Assessment Battery for Children (K-ABC) von Kaufman, A. S. & Kaufman, N. L Interpretationshandbuch* (3., teilweise ergänzte Aufl.). Amsterdam: Swets & Zeitlinger.
- Myers, D. G. (2008). *Psychologie*. (2., erw. und aktualisierte Aufl.). Heidelberg: Springer.
- Neubauer, A. & Stern, E (2007). *Lernen macht intelligent. Warum Begabung gefördert werden muss* (1Aufl.). München: DVA.
- Perleth, Ch. (1997). *Zur Rolle von Begabung und Erfahrung bei der Leistungsgenese. Ein Brückenschlag zwischen Begabungs- und Expertiseforschung* (Habilitationsschrift). München: LMU.
- Perleth, Ch. (2000). Neue Tendenzen und Ergebnisse in der Begabungs- und Intelligenzdiagnostik. In H. Joswig (Hrsg.), *Begabungen erkennen - Begabte fördern* (S. 35-64). Rostock: Universität Rostock.
- Perleth, Ch. (2008). Intelligenz und Kreativität. In W. Schneider & M. Hasselhorn (Hrsg.), *Handbuch der pädagogischen Psychologie* (S.15-27). Göttingen: Hogrefe.
- Perleth, Ch. & Heller, K.A. (1995). Testkompendium und Inventar. In L. v. Rosenstiel, M. Hockel & W. Molt (Hrsg.), *Handbuch der angewandten Psychologie* (S. 1-27). München: ecomed.
- Perleth, Ch. & Sierwald, W. (2000). Testtheoretische Konzepte der Begabungsdiagnostik. In K. A. Heller. (Hrsg.), *Begabungsdiagnostik in der Schul- und Erziehungsberatung*, (2. vollst. überarb. Aufl.). (S. 41-95). Bern: Huber.
- Perleth, C., Schauer, S., Hofmann, U., Volk, H., & Wernberger, H. (1989). Intelligence testing in a Bavarian Comprehensive School. Paper presented at the Paper presented at the XII. International School Psychology Colloquium, Ljubljana.

- Petermann, F. & Petermann, U. (Hrsg.). (2008). *HAWIK-IV, Hamburg-Wechsler-Intelligenztest für Kinder- IV - Übersetzung und Adaptation der WISC-IV von David Wechsler* (2. ergänzte Aufl.). Bern: Huber.
- Preckel, F. (2010). Intelligenztests in der Hochbegabtendiagnostik. In F. Preckel, W. Schneider & H. Holling (Hrsg.), *Diagnostik von Hochbegabung* (= Tests und Trends N.F., Bd. 8). (S. 19-32). Göttingen: Hogrefe.
- Preckel, F., Schneider, W. & Holling, H. (Hrsg.). (2010). *Diagnostik von Hochbegabung* (= Tests und Trends N.F., Bd. 8). Göttingen: Hogrefe.
- Rindermann, H. (2004). Intelligenz. In H. S. Friedman & M. W. Schustack, *Persönlichkeitspsychologie und differentielle Psychologie* (S. 372-389) (2., aktualisierte Aufl.). München: Person Studium.
- Rohrmann, S. (2010). Hochbegabung - Was ist das? In Koop, C., Schenker, I., Müller, G., Welzien, S. & Karg-Stiftung, *Begabung wagen. Ein Handbuch für den Umgang mit Hochbegabung in Kindertagesstätten.* (S.155-173). Weimar: Verlag das Netz.
- Rosenstiel, L., V., Hockel, M., & Molt, W. (Hrsg.), *Handbuch der angewandten Psychologie*. München: ecomed.
- Rost, D. H. (2009). *Intelligenz Fakten und Mythen*. Weinheim: Beltz.
- Roth, M., Schmidt, V., & Herzberg, Py. (2010). Psychologische Diagnostik in der Praxis: Ergebnisse einer Befragung unter BDP- Mitgliedern. *Report Psychologie*, 3, 118-128.
- Schmidt-Atzert, L., Deter, B. & Jaeckel, S. (2004). Prädiktion von Ausbildungserfolg: Allgemeine Intelligenz (g) oder spezifische kognitive Fähigkeiten? *Zeitschrift für Personalpsychologie*, 3(4), 147-158.
- Schneider, S. & Margraf, J. (Hrsg.) (2009). *Lehrbuch der Verhaltenstherapie, Bd. 3: Störungen im Kindes- und Jugendalter*. Berlin Heidelberg: Springer.
- Schneider, W. & Hasselhorn, M. (Hrsg.). (2008). *Lehrbuch der pädagogischen Psychologie*. Göttingen: Hogrefe.
- Schön-Gaedike, A.-K. (1978). *Intelligenz und Intelligenzdiagnostik: eine Einführung in theoretische und praktische Probleme der Intelligenzforschung*. Weinheim: Beltz.
- Schweizer, K. (2006). Intelligenz. In K. Schweizer. (Hrsg.), *Leistung und Leistungsdiagnostik* (S. 2-15). Berlin Heidelberg: Springer Medizin Verlag.

- Schweizer, K. (Hrsg.). (2006). *Leistung und Leistungsdiagnostik*. Berlin Heidelberg: Springer Medizin Verlag.
- Steins, G. (Hrsg.). (2010). *Handbuch Psychologie und Geschlechterforschung*. Wiesbaden: VS Verlag für Sozialwissenschaften/GWV Fachverlag GmbH.
- Stern, E. (2001). Intelligenz, Wissen, Transfer, und der Umgang mit Zeichensystemen. In E. Stern & J. Guthke (Hrsg.), *Perspektiven der Intelligenzforschung* (S. 163-204). Lengerich: Pabst Publisher.
- Stern, E. & Guthke, J. (Hrsg.) (2001). *Perspektiven der Intelligenzforschung*. Lengerich: Pabst Publisher.
- Tikly, L. (2006). Die Schulleistungen von Schülern ethnischer Minderheiten in Großbritannien – eine kritische Analyse der bildungspolitischen Maßnahmen zur Überwindung unterdurchschnittlicher Schulleistungen. In P. Mecheril & T. Quehl (Hrsg.), *Die Macht der Sprachen-Englische Perspektiven auf die mehrsprachige Schule* (S.97-119). Münster: Waxmann.
- Tücke, M. (2005). *Psychologie in der Schule-Psychologie für die Schule. Eine themenzentrierte Einführung in die pädagogische Psychologie für (zukünftige) Lehrer*. (4., überarb. und erw. Aufl.). Münster: LIT.
- Wechsler, D. (1961). *Die Messung der Intelligenz Erwachsener. Textband zum Hamburg-Wechsler-Intelligenztest für Erwachsene (HAWIE)*. (2. Aufl.). Bern: Huber.
- Weiβ, R. & Osterland, J. (1980). *Grundintelligenztest Skala 1 (CFT 1)* (4. Aufl.). Braunschweig: Westermann.
- Weiβ, R. H. & Osterland, J. (2013). *Grundintelligenztest, Skala 1 Revision (CFT 1-R)*. Göttingen: Hogrefe.
- Weiβ, R. H. & Weiβ, B. (2006). *Grundintelligenztest Skala 2- Revision (CFT 20-R)*. Göttingen: Hogrefe.
- Woolfolk, A. (2008). *Pädagogische Psychologie* (10 Aufl., bearbeitet und übersetzt von Ute Schönpflug). München: Pearson Studium.

Abstract zur Dissertation

Reliabilitäts- und Validitätsuntersuchungen zum kognitiven Fähigkeitstest KFT 4-12+R an einer Stichprobe von Fünf- und Siebklässlern der Willy-Brandt-Gesamtschule München

von Shaza Skender

Hintergrund und Ausgangspunkte der Studie:

- Die Willy-Brandt-Gesamtschule und die Universität München (seit 1999: Universität Rostock) arbeiten seit den 80er Jahren gemeinsam an einem Projekt angewandter Forschung über die Nützlichkeit des kognitiven Fähigkeits-Test KFT (ab 1997 wurde die aktuelle Version KFT 4-12+R verwendet; Heller & Perleth, 2000), der deutschen Version des Cognitive Abilities Tests (CAT, Thorndike & Hagen, 1971), für die speziellen Anforderungen der schulpsychologischen Arbeit an dieser besonderen Schule (vgl. Perleth, Schauer, Hofmann, Volk & Wernberger, 1989). Aus dieser Kooperation standen für diese Arbeit KFT-Daten für Fünf- und Siebklässler/innen der Kohorten bzw. Schuljahre 1998/99 bis 2011/12 zur Verfügung.
- Intelligenz scheint eine große Rolle zur Vorhersage von der Schulleistung (Perleth, 2008; Heller 2008) sowie von Berufserfolg (Hülsheger, Maier, Stumpp & Muck, 2006; Schmidt-Atzert, Deter & Jaeckel, 2004) zu spielen.
- Im Rahmen der Erforschung von Geschlechtsunterschieden bezüglich der Intelligenz finden sich in der Literatur stark unterschiedliche Ergebnisse (vgl. Hirnstein & Hausmann, 2010; Woolfolk, 2008).
- Migratenkinder schneiden bezüglich ihrer Schulleistung schlechter als die einheimischen Kinder im Schulsystem ab (Kristen, 2006; Tikly, 2006).

Fragestellung. Ziele der vorliegenden Studie waren es, die Qualität des KFT 4-12+R in den untersuchten Klassenstufen 5 und 7 der Münchner Gesamtschule zu bewerten, sowie die Veränderungen der Leistungen in den KFT-Dimensionen und der KFT-Gesamtleistung in den untersuchten Klassenstufen in der letzten Dekade von 1998 bis 2011 (Kohorteneffekte in Bezug auf die Jahrgänge) zu überprüfen und die Intelligenzsentwicklung anhand der KFT-Leistungen der Schüler/innen der Gesamtschule im Untersuchungszeitraum in Abhängigkeit von Einflussfaktoren wie Geschlecht und Muttersprache zu untersuchen (Entwicklungseffekte).

Methodik. In die vorliegenden Studie wurden die Datensätze der Fünft- und Siebtklässler/innen der Münchner Willy-Brandt-Gesamtschule im letzten Jahrzehnt von insgesamt mehr als 2200 Schülern/innen (Klassenstufe 5: N5=2091, davon 53,2% Jungen und 46,8% Mädchen; Klassenstufe 7: N7=2229, davon sind 52,3% Jungen und 47,7% Mädchen) einbezogen, die im Alter von etwa 10 und 12 Jahren sind. Außerdem wurden Schulzensuren in Deutsch, Englisch und Mathematik an einer Stichprobe von ca. 154 Schülern/innen der Klassenstufe 7 erhoben, um Aspekte der Kriteriumsvalidität zu bestimmen. Aufgrund fehlender Daten konnten aber nie die Ergebnisse aller Schüler/innen in die statistische Auswertung miteinbezogen werden. Die KFT-Daten wurden mit einem Scanner eingelesen und mit dem Programmpaket „SPSS“ analysiert. Außerdem wurden die eingelesenen Daten durch studentische und Hilfskräfte kontrolliert. Um Unterschiede zwischen deutschen und ausländischen Kindern zu analysieren, wurden die Datensätze für beide Klassenstufen 5 und 7 in „Sprachgruppen“ bzw. Herkunftsregionen unterteilt: Deutschland, Türkei, Osteuropa, Südeuropa, Westeuropa/USA, Asien und Afrika/Arabien.

Ergebnisse. Die Ergebnisse zur Bewertung der Qualität des KFT 4-12+ R in beiden Klassen 5 und 7 ergaben, dass einige einzelne Items mit einem Schwierigkeitsindex von $p > .80$ oder $p < .20$ zu entfernen oder zu modifizieren sind. Dies betrifft bei Testform A in Klassenstufe 5 48 Items und in Klassenstufe 7 50 Items und bei Testform B 27 Items in Klassenstufe 5 und 39 in Klassenstufe 7. Zusätzlich ist zu erwähnen, dass die durchschnittlichen Itemschwierigkeiten für die Skalen des KFT 4-12+ R für die beiden Testformen in beiden Klassenstufen 5 und 7 recht unterschiedlich ausfallen und zwischen $p = .41$ und $p = .79$ liegen. Das gilt nicht für die Trennschärfeindizes sowohl von einzelnen Items für die meisten Skalen des KFT 4-12+ R mit $r < .30$, als auch für die durchschnittlichen Trennschärfen der meisten Skalen des KFT 4-12+ R für die beiden Testformen in den Klassenstufen 5 und 7. Die durchschnittlichen Trennschärfekoeffizienten der KFT 4-12+ R Untertests, Dimensionen und der Gesamtleistung für die beiden Testformen in beiden Klassenstufen 5 und 7 variieren insgesamt von $r = .10$ bis $r = .54$.

Die Ergebnisse zu Verteilungsmerkmalen der Schwierigkeitsindizes der KFT-Untertests, KFT-Dimensionen und der KFT-Gesamtleistung in beiden untersuchten Klassenstufen für beide Testformen zeigten, dass einige Untertests und Testdimensionen schiefe Verteilungen in einer der beiden oder in beiden Testformen in jeder der untersuchten Klassenstufe aufweisen (KFT-V2, KFT-Q2, KFT-N1 und KFT-N bei Testform A, sowie KFT-N2 und KFT-N3 bei beiden Testformen in Klassenstufe 5 und die Skalen KFT-N sowie KFT-Q1 und KFT-Q2 für Testform A). Darüber hinaus wiesen einzelne Skalen des KFT 4-12+ R für beide Testformen in den

Klassenstufen 5 (KFT-V2, KFT-Q2 und KFT-N1) sowie 7 (Subtests des KFT-N-Teils und KFT-Q2) keine gleichen Schwierigkeitsgrade auf.

Die Befunde zur Reliabilität des Testverfahrens zeigten, dass die Reliabilitätskoeffizienten des KFT 4-12+ R für die beiden Testformen nach der Halbierungsmethode sowie die Konsistenzkoeffizienten im befriedigenden (Untertests) bis sehr guten Bereich (Gesamtleistung und Testdimensionen) und die Stabilitätskoeffizienten im niedrigen bis mittleren Bereich liegen

Zur Validität des KFT 4-12+R (faktorielle Validität) zeigten die Ergebnisse, dass sich für die beiden Klassenstufen und für die beiden Testformen eine Faktorenstruktur des KFT 4-12+ R mit drei Faktoren ergibt. Der verbale-sprachliche Faktor wird durch die Tests des KFT-V-Teils, der quantitative Faktor durch die des KFT-Q-Teils, und der nonverbale Faktor durch die KFT-N-Tests bestimmt. Bezuglich der Kriteriumsvalidität (Übereinstimmungsvalidität) bestätigen die Ergebnisse dieser Arbeit, dass die verbalen und quantitativen Dimensionen des KFT 4-12+R gut zur Prognose von Schulleistungen Fähigkeiten einsetzbar sind. Das konnte durch viele Zusammenhänge zwischen den Ergebnissen der Siebtklässler/innen in den verbalen und quantitativen KFT-Skalen mit den entsprechenden Schulnoten (den Kriteriumsvariablen) belegt werden. Zusätzlich wurde dies auch dadurch bestätigt, dass sich gute prognostische Beziehungen zwischen den Ergebnissen der in Klasse 5 erhobenen KFT-Skalen mit den Zensuren in Klasse 7 ergaben.

Die Ergebnisse zur den Kohorteneffekte der intellektuellen Kompetenzen im letzten Jahrzehnt bestätigen den sogenannt Flynn-Effekt nicht, weil die KFT-Leistungen bei den Fünftklässlern/innen von den Jahrgängen 1998-2000 bis zu den Jahrgängen 2001-2003 hin eher abnehmen und danach gibt es kaum Veränderung bis zu den Jahrgängen 2007-2010 zu verzeichnen sind. Das gilt teilweise auch für die Siebtklässler/-innen, bei denen die KFT-Leistungen von 1998-2000 bis 2007-2011 etwas abfallen. Bezuglich der Entwicklungseffekte haben die Befunde dieser Arbeit aufgewiesen, dass sich alle KFT-Leistungen von Klasse 5 zu Klasse 7 für die zwischen 1998-2000 eingeschulten Schüler/innen um 3 bis 4 T-Werte verbessern. Bei den späteren Jahrgängen verändern sich alle KFT-Leistungen ungleichmäßig. So finden sich etwa leichte Verschlechterung im verbalen Teil bei den Jahrgängen 2004-2006 sowie in den nonverbalen- und der Gesamtleistungen ab den Jahrgängen 2004-2006 oder Leistungssteigerungen im quantitativen Teil bis zu den Jahrgängen 2004-2006. Aber diese Veränderungen sind alle praktisch eher unbedeutend, weil sie kleiner als 2-T-Werteinheiten bleiben.

Die bestehenden Ergebnisse zu den Geschlechtsunterschiede in den untersuchten Klassenstufen 5 und 7 sowie in beiden Kohorten- und in Bezug auf Entwicklungseffekte sind nicht einheitlich. Während sich Jungen und Mädchen bezüglich ihren verbalen, quantitativen, nonverbalen und Gesamt-KFT-Leistungen in Klasse 7 signifikant unterscheiden, zeigten die Analysen für die Klassenstufe 5 nur in den quantitativen und der Gesamtleistung signifikante Differenzen. Die Unterschiede in beiden Klassenstufen fallen zugunsten der Jungen aus. In Bezug auf die Entwicklungseffekte haben die Befunde keine signifikanten Unterschiede zwischen den beiden Geschlechtergruppen aufgewiesen.

In Bezug auf den Faktor Muttersprache zeigten die vorliegenden Ergebnisse bezüglich der Kohorteneffekte, dass sich bei den Siebklässlern die Leistungen aller Sprachgruppen der Jahrgänge bzw. Kohorten 1998-2000 bis 2007-2011 bezüglich ihrer kognitiven Fähigkeiten verschlechtern. Darüber hinaus zeigen die Befunde für die Klassenstufe 5, dass die Jahrgänge bzw. Kohorten 2001-2003 fast bei allen Sprachgruppen etwas schwächere Leistungen in den KFT-Dimensionen erzielen als die Jahrgänge bzw. Kohorten 1998-2000. Besonders deutlich fallen die Unterschiede für die Sprachgruppe „Osteuropa“ aus. Ab Mitte des Untersuchungszeitraums gibt es sogar einen kleinen Anstieg bei den übrigen Sprachgruppen in manchen Skalen des KFT 4-12+R. So verbessern sich etwa die Leistungen der deutschsprachigen Schüler/innen in den KFT-Teilen und der KFT-Gesamtleistung, die der türkischsprachigen im KFT-V, KFT-N und der KFT-Gesamtleistung bis zu den Jahrgängen 2004-2006 und die der Restgruppe („Sonstige“) im KFT-N, KFT-V und auch in der KFT-Gesamtleistung. Die Veränderungen bzw. Verschlechterung oder Verbesserung aller Sprachgruppen in einigen Skalen bezüglich der Entwicklungseffekte sind allerdings nicht signifikant und praktisch nur wenig bedeutsam, weil sie mit 1 bis 2-T-Wert-Punkten sehr klein ausfallen.

Schlussfolgerungen. Die Ergebnisse der vorliegenden Arbeit zum KFT 4-12+R zeigen, dass mit dem KFT 4-12+ R die verbalen, nonverbalen und quantitativen Kompetenzen sowie das kognitive Gesamtleistungsniveau von Fünft- und Siebklässlern/innen erfasst werden können. Außerdem können aufgrund der hier berichteten Befunde die KFT-Leistungen als Hilfsmittel für Schulberater/innen bei Empfehlungen im Rahmen der Bildungs- bzw. Schullaufbahnberatung sowie zur Vorhersage der Schulleistung eingesetzt werden. Darüber hinaus eignet sich das Instrument für die Forschung in den genannten Bereichen. Darüber hinaus können die Ergebnissen zur Weiterentwicklung theoretische Modelle genutzt werden. Zusätzlich bestätigen die Ergebnisse der Fünft- und Siebklässler/innen die Wichtigkeit von Fördermaßnahmen durch entsprechende Kurse zur Verbesserung ihrer Schulleistungen in der Schule. Grundsätzlich sollten

allerding neben kognitive Fähigkeiten auch nicht-kognitiven Fähigkeiten für die Schulberatung und die Prognose des Schulerfolgs berücksichtigt werden.

Literaturverzeichnis

- Heller, K. A. (2008). *Von der Aktivierung der Begabungsreserven zur Hochbegabtenförderung, Forschungsergebnisse aus vier Dekaden*. Berlin: LIT-Verlag.
- Heller, K.A. & Perleth, Ch. (2000). Diagnostik. In *Lexikon der Psychologie*, Band 1 (S. 311-314). Heidelberg: Spektrum Akademischer Verlag.
- Hirnstein, M. & Hausmann, M. (2010). Kognitive Geschlechtsunterschiede. In G. Steins (Hrsg.), *Handbuch Psychologie und Geschlechterforschung* (1. Aufl. S.69-85). Wiesbaden: VS Verlag für Sozialwissenschaften/GWV Fachverlag GmbH.
- Hülsheger, U. R., Maier, G. W., Stumpp, T. & Muck, P. M. (2006). Vergleich kriteriumsbezogener Validitäten verschiedener Intelligenztests zur Vorhersage von Ausbildungserfolg in Deutschland. *Zeitschrift für Personalpsychologie*, 5 (4), 145-162.
- Kristen, C. (2006). Ethnische Diskriminierung in der Grundschule? Die Vergabe von Noten und Bildungsempfehlungen. *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 58 (1), 79-97.
- Perleth, Ch. (2008). Intelligenz und Kreativität. In W. Schneider & M. Hasselhorn (Hrsg.), *Handbuch der pädagogischen Psychologie* (S.15-27). Göttingen: Hogrefe.
- Perleth, C., Schauer, S., Hofmann, U., Volk, H., & Wernberger, H. (1989). Intelligence testing in a Bavarian Comprehensive School. Paper presented at the Paper presented at the XII. International School Psychology Colloquium, Ljubljana.
- Schmidt-Atzert, L., Deter, B. & Jaeckel, S. (2004). Prädiktion von Ausbildungserfolg: Allgemeine Intelligenz (g) oder spezifische kognitive Fähigkeiten? *Zeitschrift für Personalpsychologie*, 3(4), 147-158.
- Tikly, L. (2006). Die Schulleistungen von Schülern ethnischer Minderheiten in Großbritannien – eine kritische Analyse der bildungspolitischen Maßnahmen zur Überwindung unterdurchschnittlicher Schulleistungen. In P. Mecheril & T. Quehl (Hrsg.), *Die Macht der Sprachen-Englische Perspektiven auf die mehrsprachige Schule* (S.97-119). Münster: Waxmann.
- Woolfolk, A. (2008). *Pädagogische Psychologie* (10 Aufl., bearbeitet und übersetzt von Ute Schönpflug). München: Pearson Studium.

Persönliche Erklärung I

Hiermit versichere ich, dass ich die eingereichte Dissertation selbstständig und ohne fremde Hilfe verfasst, keine anderen als die von mir angegebenen Quellen und Hilfsmittel benutzt und die den benutzten Werken wörtlich oder inhaltlich entnommenen Stellen als solche kenntlich gemacht habe.

Shaza Skender

Rostock, den 11.10.2013

Persönliche Erklärung II

Ich habe mich nicht mit einer Arbeit zum gleichen Thema weder an der Universität Rostock noch einer anderen Hochschule erfolglos um den Doktorgrad beworben.

Shaza Skender

Rostock, den 11.10.2013