



universität
wien

Diplomarbeit

Titel der Arbeit

Allgemeine Entwicklungsdiagnostik in der frühen Kindheit

Verfasserin

Petra Bircsak

Angestrebter akademischer Grad

Magistra der Naturwissenschaften (Mag. rer. nat.)

Wien, im Dezember 2009

Studienkennzahl: 298

Studienrichtung: Psychologie

Betreuer: Ass.-Prof. Dr. Ursula Kastner-Koller

für Mara

INHALTSVERZEICHNIS

Einleitung	1
THEORETISCHER TEIL.....	5
1 Entwicklung als Gegenstand der Entwicklungsdiagnostik.....	6
1.1 Zur Problematik des Entwicklungsbegriffs	6
1.2 Modelle und Theorien der menschlichen Entwicklung	8
1.3 Annahmen über den Entwicklungsverlauf in der frühen Kindheit.....	13
1.3.1 Motorische Entwicklung	14
1.3.2 Entwicklung von Kognition und Gedächtnis.....	15
1.3.2.1 Die kognitive Entwicklung nach Piaget.....	16
1.3.2.2 Die Entwicklung der Informationsverarbeitung	19
1.3.2.3 Entwicklung mathematischer Fähigkeiten	25
1.3.3 Sprachentwicklung	27
1.3.4 Sozial-emotionale Entwicklung	29
2 Einführung in die Entwicklungsdiagnostik.....	33
2.1 Begriffsbestimmung Entwicklungsdiagnostik	33
2.2 Aufgaben und Ziele der psychologischen Entwicklungsdiagnostik	35
2.3 Historische Wurzeln der Entwicklungsdiagnostik	37
3 Methoden der Entwicklungsdiagnostik	43
3.1 Allgemein orientierte Entwicklungsscreenings.....	43
3.2 Allgemein orientierte Entwicklungstests.....	44
3.3 Verhaltensbeobachtung	46
3.4 Anamnese und Exploration.....	47
3.4.1 Befragung von Bezugspersonen.....	47
3.4.2 Befragung des Kindes	49
4 Qualitätsanforderungen an Entwicklungstestverfahren	51
4.1 Objektivität.....	52
4.2 Reliabilität	52
4.3 Validität	54
4.3.1 Inhaltsvalidität.....	54
4.3.2 Kriteriumsvalidität	55

4.3.3	Konstruktvalidität	55
4.4	Normierung	58
4.5	Skalierung	59
4.6	Ökonomie	60
4.7	Nützlichkeit	60
4.8	Zumutbarkeit	60
4.9	Unverfälschbarkeit	61
4.10	Fairness.....	61
4.11	Zur Auswahl und Durchführung eines Entwicklungstestverfahrens.....	61
5	Ausgewählte deutschsprachige Entwicklungstestverfahren: Darstellung und kritische Würdigung	63
5.1	Entwicklungsscreenings	63
5.1.1	Dortmunder Entwicklungsscreening für den Kindergarten (DESK 3-6)	63
5.1.2	Die diagnostischen Einschätzskalen (DES) zur Beurteilung des Entwicklungsstandes und der Schulfähigkeit	66
5.1.3	Erweiterte Vorsorgeuntersuchung (EVU).....	67
5.1.4	Vergleichende Bewertung der dargestellten Entwicklungsscreenings.....	70
5.2	Entwicklungstests	72
5.2.1	Bayley Scales of Infant Development – Second Edition (BSID II, Bayley II)	72
5.2.1.1	Bewertung des Verfahrens	74
5.2.2	Basisdiagnostik für umschriebene Entwicklungsstörungen im Vorschulalter (BUEVA).....	75
5.2.2.1	Bewertung des Verfahrens	79
5.2.3	Entwicklungstest sechs Monate bis sechs Jahre (ET 6-6).....	80
5.2.3.1	Bewertung des Verfahrens	84
5.2.4	Wiener Entwicklungstest (WET).....	85
5.2.4.1	Bewertung des Verfahrens	89
6	Resümee aus den theoretischen Ausführungen	92
	EMPIRISCHER TEIL.....	97
7	Zielsetzung und Fragestellung der Untersuchung	99
8	Untersuchungsinstrument	100
9	Durchführung der Untersuchung	104
10	Stichprobenbeschreibung	106

10.1	Die Kindergärten der Kinder.....	106
10.2	Alter und Geschlecht der Kinder.....	107
10.3	Familiärer Hintergrund der Kinder.....	107
10.4	Der sozioökonomische Status der Kinder	109
10.5	Die allgemeine Entwicklung der Kinder der Gesamtstichprobe.....	110
10.6	Entwicklungsunterschiede zwischen Subgruppen der Stichprobe	111
10.6.1	Kinder unterschiedlicher Muttersprache.....	111
10.6.2	Geschlechtsunterschiede.....	112
10.6.3	Leistungsunterschiede zwischen Kindern in Abhängigkeit vom Testleiter	113
10.7	Zusammenfassende Beurteilung der Stichprobe.....	115
11	Konstruktvalidität.....	116
11.1	Steigende Alterstrends.....	116
11.1.1	Methode.....	117
11.1.2	Unterschiede in den Testleistungen zwischen Vier- und Fünfjährigen	118
11.1.3	Unterschiede in den Testleistungen zwischen den Halbjahresgruppen	119
11.1.3.1	Leistungsunterschiede bei Subtests mit gleichen Aufgaben für Vier- und Fünfjährige..	119
11.1.3.2	Leistungsunterschiede bei Subtests mit altersangepassten Aufgaben.....	121
11.1.4	Zusammenfassende Beurteilung der Annahme steigender Alterstrends.....	124
11.2	Faktorielle Validität.....	125
12	Reliabilität der neu entwickelten Subtests	128
12.1	Reliabilität der Skala Muster Legen.....	128
12.2	Reliabilität der Skala Rechnen	130
13	Diskussion.....	133
	Zusammenfassung	139
	Abstract	140
	Literaturverzeichnis	141
	Tabellenverzeichnis	150
	Anhang.....	151

Einleitung

Die vorliegende Arbeit entstand im Zusammenhang mit der geplanten dritten Auflage des Wiener Entwicklungstests (WET; Kastner-Koller & Deimann, 1998, 2002, in Vorbereitung), wobei die darin enthaltene Sammlung von Normierungsdaten Bestandteil der notwendigen Vorarbeiten für die Neuauflage ist. Ebendiese Datensammlung war das vorrangige Ziel dieser Arbeit.

Der WET ist ein Verfahren der allgemein orientierten Entwicklungsdiagnostik und für den Einsatz in der frühen Kindheit vorgesehen. Dieser Sachverhalt führte zur Ableitung des inhaltlichen Ziels: Es sollen hier die Grundlagen, Zwecke und Methoden der Allgemeinen Entwicklungsdiagnostik dargelegt und solche allgemein orientierten psychometrischen Entwicklungstests und -screenings vorgestellt werden, die im deutschen Sprachraum vermehrt Anwendung finden. Zudem soll ein Beitrag zur Testanalyse des WET geliefert werden.

Warum diese Thematik von Interesse ist, wird im Folgenden kurz umrissen: Ausgeprägte Entwicklungsrückstände und Behinderungen werden zumeist schon in den ersten beiden Lebensjahren erkannt und behandelt. Weniger charakteristische Verzögerungen oder Teilleistungsstörungen bleiben jedoch häufig bis zum Eintritt in die Schule unbemerkt, wo dann die Schwierigkeiten des Kindes plötzlich augenscheinlich werden (Tröster, Flender & Reineke, 2004). Das ist insofern problematisch, weil Fördermaßnahmen, die auf früh erkannte Entwicklungsrückstände folgen, eine günstigere Prognose haben als Interventionen, die erst im Schulalter einsetzen. Auch in Hinblick auf die Motivation und das Selbstwertgefühl ist eine Frühförderung günstiger als eine Intervention, die mit dem Wissen des Kindes einhergeht, in einem oder mehreren Bereichen zu versagen (Esser & Wyszkon, 2002). Zudem ist Frühförderung auch aus politischer Perspektive notwendig und sinnvoll: Der deutsche Vizekanzler und Bundesminister für Arbeit und Soziales hat in diesem Zusammenhang am IV. Kölner Frühförderkongress betont: „Man spart nichts ein, wenn man Frühförderung nicht ordentlich erbringt; denn das, was da an Folgekosten durch verpasste Teilhabe auf die Kostenträger zukommt, ist ein Vielfaches dessen, was zunächst eingespart wird“ (Müntefering, 2008, S. 19). Diesen sozialwirtschaftlichen Aspekt bedenkend wurde in Österreich ein verpflichtendes Kindergartenjahr mit dem Ziel eingeführt, Kinder mit sprachlichen und sozial-emotionalen Schwierigkeiten bereits ein Jahr vor Schuleintritt gezielt zu fördern.

Um Frühförderung „ordentlich erbringen“ zu können, müssen Defizite erst einmal erkannt werden und notwendige Interventionen auf das Entwicklungsniveau des Kindes abgestimmt sein. Aber auch die Anregungen, die Kindern mit normaler oder akzelerierter Entwicklung zugute kommen sollen, müssen entwicklungskontingent erfolgen. Förderung muss also maßgeschneidert sein, damit Kinder weder über- noch unterfordert werden. Der Entwicklungsdiagnostik kommt also in der frühen Kindheit eine bedeutende Rolle zu, denn nur auf dem Hintergrund einer validen Beurteilung des Entwicklungsstandes kann die notwendige individuelle Förderung abgeleitet und sinnvoll erbracht werden.

Entwicklungsdiagnostik wird hier als multimethodales Vorgehen zur Beschreibung und zum Verständnis normaler und abweichender Entwicklung aufgefasst, welches auf die Ableitung von Interventionen und die Formulierung von Entwicklungsprognosen abzielt. Der multimethodale Ansatz wird vor allem deswegen vertreten, weil die ausschließlich psychometrische Erfassung der kindlichen Kompetenzen kaum ein realistisches Abbild von Entwicklungsstand und -verlauf geben kann. Die Befragung wichtiger Bezugspersonen und ein Gespräch mit dem Kind sind überdies wichtige Informationsquellen, die die Erhebung der Leistungsdaten untermauern. Die Beobachtung des Verhaltens des Kindes rundet den entwicklungsdiagnostischen Prozess ab. In vielen Fällen wird es zudem notwendig sein, einem Verfahren mit geringerem Aussagewert ein differenziertes Verfahren anzuschließen. Erst die Integration verschiedener multimethodal gewonnener Informationen ermöglicht eine präzise Diagnose.

Obwohl in Folge auf alle Methoden der Entwicklungsdiagnostik eingegangen wird, liegt der Schwerpunkt dieser Arbeit auf der psychometrischen Entwicklungsdiagnostik. Hierzu stehen Entwicklungsscreenings und Entwicklungstests zur Verfügung. Letztere stellen nach Deimann und Kastner-Koller (2007) die *via regia* zur Beurteilung des Entwicklungsstandes eines Kindes dar. Viele dieser Verfahren erfüllen die geltenden Standards für psychometrische Verfahren jedoch leider nur unzureichend und/oder sind entwicklungstheoretisch veraltet bzw. unzureichend begründet. Eine der wenigen Ausnahmen stellt der Wiener Entwicklungstest dar, welcher in Kürze in einer erweiterten und neu normierten Fassung publiziert werden wird.

Im theoretischen Teil dieser Arbeit wird zunächst der Entwicklungsbegriff einer näheren Betrachtung unterzogen und hierauf die verschiedenen Theorien der Entwicklung skizziert. Zudem wird der Entwicklungsverlauf in der frühen Kindheit beschrieben.

Ziel des zweiten Kapitels ist es, den Leser bzw. die Leserin in die Entwicklungsdiagnostik einzuführen, indem zunächst der Begriff bestimmt und die Zielsetzungen beschrieben werden und weiters auf die Geschichte der Entwicklungsdiagnostik eingegangen wird.

Im dritten Kapitel folgt die Beschreibung der Methoden der Entwicklungsdiagnostik.

Das vierte Kapitel beinhaltet die Erläuterung der Anforderungen, die psychometrische Verfahren und somit auch Entwicklungstestverfahren zu erfüllen haben.

Das letzte Kapitel des theoretischen Teils umfasst die Darstellung und kritische Würdigung verschiedener Entwicklungsscreenings und -tests, die im deutschen Sprachraum vermehrt Anwendung finden.

Im empirischen Teil wird die – an einer lokalen Stichprobe von 164 Wiener Kindern im Alter von 4;0 bis 5;11 Jahren durchgeführte – testanalytische Untersuchung der geplanten Neuauflage des Wiener Entwicklungstests beschrieben und deren Ergebnisse dargestellt und diskutiert.

THEORETISCHER TEIL

1 Entwicklung als Gegenstand der Entwicklungsdiagnostik

Bevor auf die Entwicklungsdiagnostik an sich eingegangen wird, gilt es, den Gegenstand der entwicklungspsychologischen Beurteilung zu thematisieren: Im Folgenden wird der Entwicklungsbegriff einer näheren Betrachtung unterzogen, unterschiedliche Entwicklungstheorien skizziert und auf den Entwicklungsverlauf in der frühen Kindheit eingegangen.

1.1 Zur Problematik des Entwicklungsbegriffs

Seit Entstehung der Entwicklungspsychologie gegen Ende des 19. Jahrhunderts wurden zahlreiche Versuche unternommen, den Entwicklungsbegriff in Worte zu fassen. Dennoch liefert die entwicklungspsychologische Literatur weder eine allgemein anerkannte Definition von Entwicklung noch besteht ein umfassender Konsens bezüglich der Inhalte des Entwicklungsbegriffs. Folgende Definitionen seien in chronologischer Reihenfolge als repräsentativ für die unterschiedlichen Auffassungen herausgegriffen:

W. Stern (1914, zitiert nach Schenk-Danzinger, 2006, S. 27): „Die seelische Entwicklung ist nicht bloß Hervortretenlassen angeborener Eigenschaften, sie ist aber auch nicht bloß ein Empfangen äußerer Einwirkungen, sondern das Ergebnis der Konvergenz innerer Angelegenheiten mit äußeren Entwicklungsbedingungen.“

K. Bühler (1918, zitiert nach Schenk-Danzinger, 2006, S. 27): „Zum Begriff der Entwicklung im ursprünglichen und engen Sinn des Wortes gehört zweierlei: nämlich Anlage im Anfangszustand und ein Plan, eine Zielrichtung des Werdens.“

H. Rempelin (1944, zitiert nach Schenk-Danzinger, 2006, S. 27): „Entwicklung ist eine nach immanenten Gesetzen sich vollziehende Differenzierung einander unähnlicher Teile bei zunehmender Strukturierung und Zentralisierung.“

E. Nagel (1957, zitiert nach Trautner, 1992, S. 45): „Entwicklung ist die regelhafte Abfolge von Veränderungen in einem in sich geschlossenen System mit einer klar umrissenen Struktur und einer bestimmten Menge von schon existierenden Fähigkeiten, die zu relativ dauerhaften Neuerwerbungen sowohl der Struktur wie der Funktionsweise des Systems führen.“

R. I. Watson (1959, zitiert nach Trautner, 1992, S. 45): „Wenn Veränderungen gesetzmäßig und aufeinander abgestimmt sind und die Fähigkeit des Organismus erhöhen, sich seiner Umgebung anzupassen, verdienen sie das Etikett Entwicklung.“

H. Thoma (1959, zitiert nach Trautner, 2003, S. 27) versteht Entwicklung „als eine Reihe von untereinander zusammenhängenden Veränderungen, die bestimmten Orten des zeitlichen Kontinuums zugeordnet sind.“

W. Kessen (1960, zitiert nach Trautner, 1992, S.46): „Ein Phänomen wird entwicklungsbedingt genannt, wenn es in regel- oder gesetzmäßiger Weise mit dem Alter in Beziehung gesetzt werden kann.“

H. D. Schmidt (1970, S. 20): „Wir bezeichnen solche psychophysischen Veränderungen als Entwicklung, deren Glieder existentiell auseinander hervortreten (d.h. in einem natürlichen inneren Zusammenhang stehen), sich Orten in einem Zeit-Bezugssystem zuordnen lassen und deren Übergänge von einem Anfangszustand in einen Endzustand mit Hilfe von Wertkriterien zu beschreiben sind.“

R. Oerter (1977, S. 22) definierte Entwicklung als „nicht zufällig erklärbare Veränderungsreihe, die mit dem individuellen Lebenslauf verknüpft ist.“

D. Ulich (1986, zitiert nach Flammer, 2003, S. 20): „Entwicklung ist lebensaltersbezogene, langfristige und geordnete unterschiedliche Veränderung unterschiedlicher Persönlichkeiten, in unterschiedlichen sich verändernden Umwelten.“

Die Auffassungen von Entwicklung, die in den vorgenannten Definitionen zum Ausdruck kommen, lassen sich grob in zwei Gruppen einteilen: Die erste (z. B. Bühler, 1918; Nagel, 1957) entspricht der traditionellen Sichtweise von Entwicklung. Diese wird maßgeblich an biologisches Wachstum und Reifung gebunden, der Umwelt lediglich eine auslösende bzw. hemmende Wirkung zugesprochen. Die zweite Gruppe von Definitionen (z. B. Schmidt, 1970; Thoma, 1959) ist weiter gefasst und zielt vorrangig auf Veränderungen innerhalb eines Zeitkontinuums ab. Letztere spiegelt die heute vorherrschende Auffassung wider und fokussiert auf die Veränderung individueller Lebensläufe. Zudem rückt die Eigenverantwortung der Person als auch die Dynamik der Umwelt in den Mittelpunkt. Entwicklung ist demnach bestimmt durch das Zusammenwirken endogener, exogener und autogener Faktoren, die sich wechselseitig bedingen. Eine zeitgemäße Erklärung von Entwicklung liefern Fuiko und Wurst (2003):

Entwicklung wird als lebenslanger Prozess, als eine Sequenz von Phasen des Aufbaus, der Stabilisierung und des Abbaus verstanden; diese Phasen laufen zum Teil parallel und haben Spezialisierung und Optimierung von Fähigkeiten und Eigenschaften zum Ziel. Entwicklung ist demnach das Produkt von Reifung, Umwelteinflüssen und individueller Gestaltung des Menschen. (S. 120)

Neuere entwicklungspsychologische Lehrbücher (z. B. Berk, 2004/2005; Oerter & Montada, 2008; Steinebach, 2000; Trautner, 2003) diskutieren die unterschiedlichen Sichtweisen von Entwicklung sowie die daraus resultierenden Theorien, liefern jedoch keinen Versuch einer präzisen Definition. Trautner (2006) meint hierzu treffend:

Man muss sich damit abfinden, dass Entwicklung kein klar abgrenzbarer Klassenbegriff ist, sondern ein an den Rändern unscharfer Typenbegriff oder prototypischer Begriff, der seine Bedeutung erst im Kontext spezifischer Theorien, d. h. auf dem Hintergrund eines bestimmten Welt- und Menschenbildes gewinnt. (S.85)

1.2 Modelle und Theorien der menschlichen Entwicklung

Grundlage jeder Entwicklungstheorie ist die Festlegung darauf, was unter dem Begriff Entwicklung zu verstehen ist. So mannigfaltig wie die Versuche, Entwicklung zu definieren, sind deshalb auch die Theorien zur menschlichen Entwicklung. Auch fokussieren die einzelnen Theorien oft nicht auf die Gesamtentwicklung, sondern auf unterschiedliche Teilaspekte. Die Darstellung der vielfältigen Theorien der menschlichen Entwicklung ist im Rahmen dieser Arbeit nicht möglich, sodass lediglich Metamodelle skizziert und charakteristische Vertreter angeführt werden können. Die folgenden Ausführungen sind infolgedessen stark komprimiert. Zur Vertiefung sei auf Flammer (2003), Miller (1993/1993), Trautner (2003) sowie Weinert und Weinert (2006) bzw. auf die Originalliteratur der jeweiligen Theoretiker verwiesen.

Endogenistische Entwicklungstheorien messen insbesondere der Reifung Bedeutung zu und sehen Entwicklung als universell, irreversibel und auf einen Reifezustand gerichtet an. Entwicklung erfolgt hier durch Reifungsschübe, welche aus einem inneren Reifungsplan hervorgehen. Aufgrund ihrer Phasenartigkeit ist sie in Entwicklungsstufen darstellbar. Demnach bleibt auch eine vorzeitige Förderung wirkungslos. Neben den oben angeführten Autoren der traditionellen Sichtweise gilt als typischer Vertreter Arnold

Gesell (1880-1961), welcher den Terminus Entwicklung oft synonym mit Wachstum verwendete (Flammer, 2003).

Eine weitere Extremposition nehmen *exogenistische Theorien* ein, die Entwicklung maßgeblich als Funktion exogener Faktoren und den Menschen als Produkt seiner Lebensumstände betrachten. Diese stark mechanistisch-materialistische Auffassung ist den Behavioristen zuzuschreiben, deren konsequentester Vertreter John B. Watson (1878-1958) war. Der Ansatz schreibt insbesondere Erzieherinnen und Erziehern sowie Sozialpolitikerinnen und Sozialpolitikern die Verantwortung für die Entwicklung eines Kindes zu. Dabei ist ein Eingreifen vornehmlich bei weniger günstigen Entwicklungsverläufen wesentlich. Das Kind handelt lediglich reaktiv. Mit zunehmendem Alter liegt die Verantwortung auch bei den Individuen selbst. Historisch gesehen ist der Philosoph John Locke (1632-1704) der wichtigste Vertreter dieser Richtung, dessen Ansicht nach ein Kind bei der Geburt eine „tabula rasa“ sei (Berk, 2005).

Auch die Tiefenpsychologie liefert Modelle der (kindlichen) Entwicklung. Den einflussreichen Theorien von Sigmund Freud (1856-1939) und Erik Erikson (1902-1994) ist die Annahme gemeinsam, dass das Individuum eine Reihe von Phasen durchläuft, in denen es entwicklungstypische Konflikte zu bewältigen hat und sich bei deren Lösung zwischen seinen biologischen Trieben und den Erwartungen seiner Umwelt entscheiden muss. Erikson übernahm die psychosexuellen Annahmen Freuds, betrachtete Entwicklung jedoch als lebenslangen Prozess und betonte die Rolle des sozialen und kulturellen Kontextes für die Krisenbewältigung. Psychoanalytische Entwicklungstheorien fokussieren auf die sozial-emotionale Entwicklung und weisen insbesondere für die Entwicklungspsychopathologie Relevanz auf (Habermas & Döll-Hentschker, 2007).

Humanistische Entwicklungstheorien sind subjektivistisch orientiert und gehen davon aus, dass sich der Mensch unter geeigneten Bedingungen spontan zum Besseren entwickelt. Entwicklungsförderung ist hier nur bei Fixierung nötig, wobei Förderung im Sinne von Vermittlung der Erfahrung des Akzeptierseins verstanden wird. Als Begründer der humanistischen Psychologie gelten Charlotte Bühler (1893-1974), Abraham H. Maslow (1908-1970) und Carl Rogers (1902-1987). Die Humanisten verstanden ihren Ansatz als Kontrapunkt zu den Annahmen der Behavioristen und Psychoanalytiker (Flammer, 2003).

Interaktionistischen Entwicklungstheorien ist die Annahme gemeinsam, dass sowohl die genetische Anlage des Individuums als auch die Umwelt die individuelle Entwicklung

bedingen. Da es überaus viele Ausprägungen dieser Theorien gibt, ist eine weitere Unterscheidung unumgänglich:

- *Konstruktivistische bzw. Strukturgenetische Theorien* sehen den Menschen in einer aktiven Auseinandersetzung mit seiner Umwelt, betonen aber die aktiv-konstruktivistische Rolle des Individuums als Problemlöser. Das Individuum konstruiert sein Wissen aufgrund von Erfahrungen und reguliert sich selbst. Das Menschenbild dieses Ansatzes kann demnach als organismisch-konstruktivistisch bezeichnet werden. Bedeutendster Vertreter dieses Ansatzes war Jean Piaget (1896-1980), dessen Theorie der kognitiven Entwicklung (z. B. Piaget, 1947/1966; 1950/1974, 1959/2003a, 1970/2003b) wohl eine der bekanntesten Entwicklungstheorien überhaupt ist. Nach Piaget durchläuft ein Kind, während sein Gehirn reift und es fortlaufend Erfahrungen im Umgang mit der Umwelt macht, vier diskontinuierliche Stadien der geistigen Entwicklung. Dabei strebt das Kind stets nach Gleichgewicht zwischen individuellen Zielen und Anforderungen der Umwelt. Zudem will es sich höher entwickeln. Eine Weiterentwicklung des Piagetschen Ansatzes sind die sogenannten Neo-Piaget-Theorien, welche die Stärken der Piagetschen Theorie mit den Stärken der Informationsverarbeitungstheorien zu verbinden versuchen. Die vermutlich einflussreichste neo-piagetische Theorie stammt von Robbie Case (1945-2000) (Case, 1985/1999).
- *Informationsverarbeitungstheorien* bedienen sich einer Computer-Metapher und beschreiben den Menschen als aktives, organisiertes, selbstmodifizierendes System, welches „Input“ zu „Output“ transformiert. Das menschliche Gehirn ist seine „Hardware“, erworbene Lern- und Gedächtnisstrategien seine „Software“ (Miller, 1993/1993). Entwicklung wird als Prozess kontinuierlicher Veränderung betrachtet. Veränderungen im Inhaltswissen, in der Metakognition und in den Informationsverarbeitungskapazitäten werden als treibende Kräfte der kognitiven Entwicklung angesehen. Oft zitierter Vertreter dieses Ansatzes ist Robert S. Siegler. Nach Siegler (1998/2001) verläuft Entwicklung nicht sprunghaft. Vielmehr bestehen verschiedene Denkweisen parallel und konkurrieren miteinander. Differenzierte Denkweisen werden nach und nach dominanter, Rückfälle in alte Denk- und Verhaltensmuster sind jedoch normal. Diese Annahme veranschaulicht er in einem Modell überlappender Wellen.

- *Dialektisch kontextualistische Entwicklungstheorien* berücksichtigen vor allem sozio-kulturelle Aspekte von Entwicklung. Im Vordergrund steht das mit Anderen interagierende, in einen gesellschaftlichen Kontext eingebundene aktive Individuum, das stets neuen Entwicklungsbedingungen unterworfen ist (Miller, 1993/1993). Entwicklung entsteht hier maßgeblich durch soziales Lernen, vermittelt durch mentale Werkzeuge, die von der Kultur bereitgestellt werden (z. B. Sprache, Symbole und Zeichen). Das Menschenbild ist demnach eher ein mechanistisches. Der wohl einflussreichste Vertreter der dialektischen Schule war Lev Wygotsky (1896-1934), welcher den Gedanken der „Zone der proximalen Entwicklung“ formulierte (Trautner, 2003). Diese verstand er als Distanz zwischen dem aktuellen Entwicklungsniveau eines Kindes, bestimmt durch dessen Fähigkeit, Probleme selbständig zu lösen, und einer höheren Ebene potentieller Entwicklung, bestimmt durch die Fähigkeit, Probleme unter Anleitung kompetenter sachkundiger Personen zu lösen.
- *Systemisch kontextualistische Theorien* gehen noch einen Schritt weiter, indem sie Mensch und Umwelt als Gesamtsystem (Ökosystem) betrachten und postulieren, dass die individuelle Entwicklung zirkulären Bedingungen unterworfen ist: Das Individuum wirkt verändernd auf seine Umwelt, welche wiederum das Individuum beeinflusst. Das Menschenbild der Systemtheoretiker kann als kontextualistisches bezeichnet werden. Bekanntester Vertreter war Uri Bronfenbrenner (1917-2005), auf welchen das Konzept der „ökologischen Validierung“ psychologischer Untersuchungen und die Mehrebenen-differenzierung von Entwicklungs-Umweltkontexten in Mikro-, Meso-, Exo- und Makrosystem zurückgehen (Flammer, 2003). Weiters spricht Bronfenbrenner vom Chronosystem, einer zeitlicher Dimension, die alle Systemebenen erfasst.
- *Handlungsorientierte Entwicklungstheorien* betrachten ebenfalls Umwelt und Individuum als Gesamtsystem, sehen jedoch die Handlungen eines Individuums als treibende Kraft der Entwicklung an. Zu den Handlungstheorien zählt auch das Konzept der „Entwicklungsaufgaben“, welches ursprünglich auf Robert J. Havighurst (1900-1991) zurückgeht (Flammer, 2003). Havighurst vertrat die Annahme, dass der jeweilige Lebensraum Anforderungen liefert, die sich in einer bestimmten Lebensperiode stellen und die das Individuum zu bewältigen hat. Diese Anforderungen können aus biologischen oder sozialen Gegebenheiten oder

aus persönlichen Zielsetzungen erwachsen. Insbesondere in Phasen des Übergangs ist Entwicklung hier durch eine aktive Bewältigung der Lebensumstände bestimmt.

Die einzelnen Ansätze zur Erklärung menschlicher Entwicklung haben im Laufe der letzten 100 Jahre an Einfluss gewonnen und wieder verloren. Die meisten Theorien entstanden aus der Kritik an bestehenden Modellen: Diese wurden entweder weiterentwickelt oder gänzlich verworfen und eine revolutionäre Erklärung für die menschliche Entwicklung aufgestellt (Miller, 1993/1993). Lange Zeit verhielten sich Reifungstheoretiker, Behavioristen, Psychoanalytiker, Konstruktivisten und deren Reformer, als sei ihr Standpunkt der einzig wahre. Heute sind die meisten Entwicklungspsychologen bereit, andere Sichtweisen anzuerkennen und Anregungen von diesen aufzugreifen, ohne dabei den eigenen Standpunkt zu vernachlässigen (Mietzel, 2002). Einem breiten Entwicklungsbegriff folgend wird Entwicklung gegenwärtig vorwiegend unter einer kontextuellen Perspektive betrachtet, und dem Individuum über die gesamte Lebensspanne hinweg aktiv gestaltender Einfluss zugesprochen. Die einseitigen Extrempositionen des Endogenismus und Exogenismus von Entwicklung besitzen heute jedenfalls nur noch historischen Wert.

Der Auffassung von Weinert und Weinert (2006) folgend, kann das Zeitalter der breiten abstrakten Metatheorien als beendet angesehen werden. Dominierend sind multikausal orientierte Mikromodelle und theoretische Aussagen mit begrenztem domänenspezifischen Geltungsbereich. Die Kenntnis der verschiedenen Ansätze kann jedenfalls davor bewahren, Entwicklung unter einer zu starren Perspektive zu betrachten und ermöglicht zudem die Integration von einzelnen Erkenntnissen zu einer eklektizistischen, auf eine spezifische Fragestellung anzuwendende Entwicklungstheorie (Miller, 1993/1993). Dies zeigt sich auch in der angewandten Entwicklungspsychologie: Kastner-Koller und Deimann (2002) orientierten sich beispielsweise bei der Konstruktion des Wiener Entwicklungstests nicht an einer einzelnen Theorie, sondern legten dem Verfahren verschiedene interaktionistische Theorien zugrunde.

Zusammenfassend ist in Hinblick auf das Thema dieser Arbeit festzuhalten, dass unterschiedliche Entwicklungsbegriffe und ihre zugrunde liegenden Welt- und Menschenbilder nicht nur zu unterschiedlichen Entwicklungstheorien führen, sondern auch noch unterschiedliche Fragestellungen, unterschiedliche Methoden und unterschiedliche Interpretationen bewirken (Trautner, 2006). Das einer entwicklungs-

diagnostischen Untersuchung zugrunde liegende Verständnis von Entwicklung beeinflusst demnach ihren Geltungsbereich und ihre Aussagekraft.

Da die vorliegende Arbeit die Allgemeine Entwicklungsdiagnostik in der frühen Kindheit thematisiert, werden im folgenden Abschnitt jene entwicklungsbedingten Veränderungen beschrieben, die vom dritten Lebensjahr bis zum Schuleintritt bei normaler Entwicklung auftreten.

1.3 Annahmen über den Entwicklungsverlauf in der frühen Kindheit

McCall (1987; vgl. auch Kopp & McCall, 1982; McCall, 1981) illustriert den Entwicklungsverlauf in der Kindheit mithilfe einer Metapher („scoop approach“): Dieser ist vergleichbar einer Kugel, die von dem höher liegenden engen Bereich der Fläche einer Schaufel zu deren tiefer liegendem breiten Flächenareal rollt. Während die Kugel in der Nähe des Schafts durch eine Rinne rollt, kann sie nur durch massive Einflüsse aus der Bahn gebracht werden. Auf dem breiten Flächenbereich hingegen eröffnen bereits geringe Einwirkungen eine Vielzahl an Möglichkeiten des Verlaufs. Die Kugel symbolisiert das sich entwickelnde Individuum, die Schaufel repräsentiert die möglichen individuellen Entwicklungspfade. Während die – durch die sich verbreiternde Schaufelform versinnbildlichte – Entwicklung in den ersten beiden Lebensjahren, in denen genetische und biologische Einflüsse dominieren, „kanalisiert“ und damit interindividuell einheitlich verläuft, gewinnen ab dem dritten Lebensjahr Umwelteinflüsse zunehmend an Bedeutung, was eine größere interindividuelle Varianz bewirkt. Daraus folgt, dass die kindliche Entwicklung ab dem dritten Lebensjahr gezielter beeinflussbar ist als zuvor, und auch Entwicklungsprognosen sodann an Aussagekraft gewinnen. Als optimale Phase für Entwicklungsinterventionen sieht McCall das Alter zwischen zwei und sechs Jahren an.

Den Annahmen des Schaufelmodells folgend erscheint in Bezug auf Entwicklungsdiagnostik daher eine Fokussierung auf Kinder zwischen zwei und sechs Jahren sinnvoll.

Über diese Altersgruppe hinaus verläuft die Entwicklung in den einzelnen Entwicklungsbereichen vergleichsweise kontinuierlich in Richtung zunehmender Kompetenz. Die bereits erworbenen Fähigkeiten werden ausdifferenziert und integriert,

sodass hieraus neue Fähigkeiten erwachsen können. Wird Entwicklung ganzheitlich betrachtet, sind in diesem Altersabschnitt keine ausgeprägten qualitativen Veränderungen zu erwarten (Kastner-Koller & Deimann, 2002).

1.3.1 Motorische Entwicklung

Im Bereich der Grobmotorik sind zwischen zwei und drei Jahren alle elementaren Bewegungen möglich. Das Kind kann laufen, hüpfen, klettern, steigen, tragen, einen Ball werfen und einiges mehr (Winter & Roth, 1994). In den Folgejahren werden die Grundbewegungen stetig verbessert. Die Ausdifferenzierung der Bewegungsformen wird häufig am Treppensteigen illustriert (z. B. Steinebach, 2000; Wendt, 1997): So kann ein normal entwickeltes Kind mit zwei Jahren eine Treppe ohne Anhalten im Nachschritt hinauf-, etwas später auch hinabsteigen. In der zweiten Hälfte des dritten Lebensjahres geht es Treppen im Wechselschritt hoch, muss sich jedoch hierbei festhalten. Kurze Zeit später wird dies auch treppab bewältigt. Im Alter von drei bis vier Jahren gelingt es dem Kind treppauf zu steigen auch ohne sich dabei anhalten zu müssen. Mit vier Jahren schafft es dies ebenso abwärts.

Mit der Verfestigung des Knochenbaus, dem Wachstum von Muskulatur und der Ausreifung des Zentralnervensystems verbessert sich zudem die Körperbeherrschung und psychomotorische Geschicklichkeit (Wendt, 1997): Mit zwei Jahren ist ein Kind noch nicht in der Lage, Bewegungen abrupt zu beenden ohne hinzufallen. Dies gelingt ihm erst im Laufe des vierten Lebensjahres. Damit Dreijährige einen Ball fangen können, muss dieser genau in die Arme geworfen werden – sie verbleiben mit ausgestreckten Armen in einer passiven Fanghaltung. Ein reifes Fangmuster, bei dem die Arme dem entgegenkommenden Ball entgegengeführt werden und die Flugbahn des Balles antizipiert wird, zeigt sich erst beim Übergang zum Schuleintritt (Scheid, 1994). Sie können nun zudem Rollschuh- und Schlittschuhlaufen sowie Radfahren (Wendt, 1997).

Die zunehmende Verfeinerung von Einzelbewegungen setzt Kinder am Ende der Vorschulzeit in die Lage, Grundbewegungen fließend zu verbinden, beispielsweise aus dem Lauf ein Hindernis zu durchkriechen (Winter & Roth 1994). Zudem kombinieren ältere Vorschulkinder Bewegungen und können nun Seilspringen oder einen Hampelmannsprung vollziehen (Berk, 2004/2005).

Auf Grundlage der elementaren Greifbewegungen bilden sich sowohl Bewegungen wie Werfen und Fangen als auch feinmotorische Fertigkeiten wie das manipulieren komplexer Objekte (z. B. Turmbau, Perlen auffädeln und Steckbrettspiele) und das Zeichnen und Schreiben (Scheid, 1994). Weitere Beispiele für feinmotorische Bewegungen, die sich im Verlauf des Vorschulalters entwickeln, sind nach Mietzel (2002) das Ausschneiden mit einer Schere, das Auf- und Zuknöpfen einer Bluse und das Ausmalen von Bildern. Die feinmotorische Kompetenzzunahme lässt sich gut am Ablauf der Entwicklung von Kinderzeichnungen veranschaulichen (Berk, 2004/2005): Kinder des westlichen Kulturkreises beginnen im Laufe des zweiten Lebensjahres zu zeichnen, wobei sie vorerst lediglich kritzeln. Um den dritten Geburtstag werden aus den Kritzeleien erkennbare Formen. Mit drei bis vier Jahren zeichnen Kinder ihre ersten Bilder von Menschen, wobei die Basisfigur zunächst kaum Details enthält (Kopffüßler). Zum Ende der Vorschulzeit produzieren Kinder komplexere, differenzierte Bilder. Werden Personen gezeichnet, sind Kopf, Körper und Gliedmaßen vorhanden.

Die Entwicklung der Motorik wird in engem Zusammenhang mit der perzeptuellen Entwicklung gesehen. Dies zeigt sich besonders deutlich in der Entwicklung der Auge-Hand-Koordination von Säuglingen. Man denke aber beispielsweise auch an den Umgang mit einer Schere, an das Werfen und Fangen von Bällen oder an das Balancieren auf einem Holzstamm. All diese Tätigkeiten stellen Anforderungen an die Fähigkeit zur Koordination von Auge und Hand. Wilkening und Krist (2008) schlagen deshalb vor, von perzeptiv-motorischen statt von motorischen Fertigkeiten zu sprechen. Auch wird ein wechselseitiger Einfluss zwischen der kognitiven und motorischen Entwicklung gesehen. Eine zielgerichtete Wurfbewegung erfordert nicht nur motorische Fertigkeiten, sondern beruht zudem auf differenzierten Planungs- und Steuerprozessen.

1.3.2 Entwicklung von Kognition und Gedächtnis

Die Veränderungen im Bereich der Kognition sind der wohl am häufigsten untersuchte Bereich der kindlichen Entwicklung, weswegen diese hier umfassender beschrieben werden. Nach Kastner-Koller und Deimann (2002) lassen sich zur Erklärung der kognitiven Entwicklung zwei grundsätzliche Positionen erkennen: Auf der einen Seite die von Piaget ausgehenden strukturalistischen Theorien, auf der anderen der Ansatz der Informationsverarbeitung. Diese beiden theoretischen Positionen weisen trotz unterschiedlicher Begriffe Gemeinsamkeiten auf. So formulieren alle Theorien die

Annahme, dass der Spracherwerb ein neues System der Repräsentation zur Verfügung stellt und somit eine neue Qualität des Denkens ermöglicht. Auch besteht Einigkeit darüber, dass mit circa sechs bis sieben Jahren wiederum eine qualitative Veränderung in den Denkstrukturen erfolgt. In der Zwischenzeit werden bestehende Kompetenzen ausdifferenziert, wodurch es dem Kind vermehrt gelingt, seine Fähigkeiten zu dezentrieren. Infolgedessen kann es zunehmend mehrere Aspekte eines Wahrnehmungsinhalts zur Problemlösung heranzuziehen.

1.3.2.1 Die kognitive Entwicklung nach Piaget

Piaget gilt als der vermutlich einflussreichste Entwicklungstheoretiker, der mit mehr als 50 Buchtiteln und mehreren hundert Aufsatzveröffentlichungen (z. B. Piaget, 1947/1966, 1950/1974, 1959/2003, 1970/2003b; Piaget & Inhelder, 1966/1972) die Forschung zur kindlichen kognitiven Entwicklung nachhaltig prägte. Gemäß seiner universal-konstruktivistischen Theorie, sind Kinder aktive Wesen, die sich entwickeln, indem sie in eine Auseinandersetzung mit der Welt eintreten, diese strukturieren und dabei die Welt und das Selbst verändern.

Nach Piaget wird die Phase zwischen zwei und sieben Jahren als Stadium des „präoperationalen anschaulichen Denkens“ bezeichnet. Das kindliche Denken ist in dieser Zeit von „Egozentrismus“ geprägt: Das Kind sieht sich mit seinen Bedürfnissen im Zentrum und ist der Meinung, dass jeder so denkt wie es selbst und seinen Standpunkt teilt. Somit wird die Wahrnehmung und Interpretation der Umwelt stets auf die eigene Person zentriert. Die egozentristische Sichtweise ist nach Piaget auch verantwortlich für den kindlichen „Animismus“ und sein von „Artifizialismus“ geprägtes Weltbild. Ein animistisch denkendes Kind geht davon aus, dass unbelebte Dinge Gedanken, Gefühle, Bedürfnisse und Absichten haben. Artifizialismus meint die Annahme, dass Naturerscheinungen von Menschenhand gemacht wurden. Insbesondere die Eltern werden als allmächtig empfunden.

Aufgrund des vorangegangenen Erwerbs der „sensumotorischen Intelligenz“ ist das präoperational denkende Kind in der Lage, aufeinanderfolgende Wahrnehmungen mit dementsprechenden ebenso aufeinanderfolgenden Bewegungen zu koordinieren. Hierfür muss es größere räumliche und zeitliche Entfernungen überbrücken, wofür symbolische Repräsentationen und zeitlicher Aufschub erforderlich sind. Einfach ausgedrückt, kann es nun über das Hier und Jetzt hinausgehen und über Ereignisse und Objekte nachdenken,

die in seiner unmittelbaren Umgebung nicht präsent sind. Das Kind neigt jedoch zu „Zentrierung“: Es kann nur einen Aspekt eines Objektes oder einer Situation beachten und vernachlässigt alle anderen. Zudem ist sein Umgang mit Symbolen und Repräsentationen geprägt von „fehlender Reversibilität“. Es ist demnach nicht in der Lage, konkrete Operationen in ihrem Ablauf mental rückgängig zu machen. Das voroperational denkende Kind hat auch Schwierigkeiten mit der Bildung von Klassen und mit hierarchischen Klassifikationen. Das Denken innerhalb dieses Stadiums ist folglich lediglich als prälogisch zu bezeichnen.

Nach Piaget ermöglicht der Erwerb mentaler Repräsentationen symbolische Spiele, durch welche wiederum neu erworbene, repräsentierende Schemata geübt würden. In diesen von Nachahmung geprägten Als-ob-Spielen werden Gegenständen andere Bedeutungen zuteil. Anfangs ist das Spiel noch selbstbezogen, richtet sich jedoch nach kurzer Zeit vermehrt auf andere Personen und entwickelt sich im Laufe der Zeit zum sozialen Rollenspiel. Sobald die Fähigkeit zur Rollen- und Perspektivenübernahme gegeben ist, gelingt es dem Kind zunehmend, den Egozentrismus zu überwinden.

Gegen Ende der Vorschulzeit befinden sich Kinder gemäß der piagetschen Theorie an der Schwelle zum „konkret-operationalen Denken“: Die Beweglichkeit im Denken wird immer größer, sodass Vorschulkinder mehrere Aspekte eines Wahrnehmungsinhalts berücksichtigen und für die Problemlösung nutzen können. Allerdings sind Kinder aus der Sicht Piagets erst zu Beginn der nächsten Phase zu logischem Denken fähig und können dann beispielsweise erkennen, dass bestimmte physikalische Merkmale von Objekten gleich bleiben, auch wenn sich ihre Erscheinung ändert.

Piaget nahm an, dass die Abfolge der Entwicklungsstadien universell sei, also die Entwicklung in einer Phase abgeschlossen sein müsse, bevor die nächste beginne. Dies wurde vielfach in Frage gestellt und neuere Forschungsergebnisse legen nahe, dass sich Veränderungen nicht so plötzlich vollziehen. Zwar konnte die Abfolge der von Piaget beschriebenen Entwicklungsschritte prinzipiell bestätigt werden, eine klare Abgrenzung verschiedener Denkformen in vier Stufen sowie eine eindeutige Zuordnung zu einzelnen Altersbereichen erscheint jedoch nicht angemessen (Nickel & Schmidt-Denter, 1995).

Case (1985/1999), der zu den sogenannten „Neo-Piagetianern“ zählt, weichte Piagets Stufenmodell auf, indem er nicht nur generelle, sondern auch differentielle Aspekte der Entwicklung untersuchte und Übergangsmechanismen beschrieb. Nach Case läuft

Entwicklung demnach zwar von Stufe zu Stufe ab, er betrachtet Entwicklung aber maßgeblich unter einem bereichs- bzw. aufgabenspezifischen Aspekt. Zudem verband er Piagets Theorie mit dem Informationsverarbeitungsansatz, indem er die biologisch fundierte Reifung des Kurzzeitgedächtnisses und die Automatisierung von Informationsverarbeitung als Determinanten der kognitiven Entwicklung betrachtete.

Berk (2004/2005) sieht diese flexible Auffassung von Entwicklungsstufen als Anerkennung der einzigartigen Qualität frühkindlichen Denkens an. Eine Struktur des Denkens verschwindet nicht einfach, indem sie durch eine höhere Stufe ersetzt wird, vielmehr wird sie überformt oder erweitert (Nickel & Schmidt-Denter, 1995). Entwicklung verläuft sonach bereichsspezifisch sowohl qualitativ als auch quantitativ. In manchen Bereichen erreicht sie früher als von Piaget postuliert ein höheres Niveau, bei unbekannten Problemstellungen gibt es Rückschritte auf ein tieferes Niveau. Eine eindeutige Zuordnung zu einzelnen Altersbereichen erscheint demzufolge als unangemessen. Schließlich können Kleinkinder ihren Egozentrismus in bestimmten Situationen überwinden. Umgekehrt verhalten sich auch ältere Kinder und Erwachsene mitunter egozentrisch. Neuere Forschungsergebnisse zeigen außerdem, dass Vorschulkinder durchaus bereits zu logischen Operationen fähig sind, sofern die zu bearbeitenden Aufgaben aus vertrauten Bereichen stammen: Einfache Analogieaufgaben können dann bereits von Dreijährigen gelöst werden, welche zudem Invarianz erkennen. Vierjährige sind bei entsprechend vertrauten Aufgaben in der Lage, syllogistische und transitive Schlüsse zu ziehen (vgl. Berk, 2004/2005; Goswami, 1998/2001). Fraglich ist, ob die in den jeweiligen Studien verwendeten modifizierten Aufgaben dieselben logischen Anforderungen enthalten wie die piagetschen Originalaufgaben (Goswami, 1998/2001).

Insgesamt wird heute davon ausgegangen, dass Piaget das logische Denken präoperationaler Kinder unterschätzte, zudem die Konsistenz des kindlichen Denkens überschätzte (Sodian, 2008). Bei aller Kritik bleibt die piagetsche Theorie dennoch aufschlussreich und liefert ein gutes Gefühl für das kindliche Denken (Siegler, 1998/2001).

Zusammenfassend kann festgehalten werden, dass ein Kind beim Schuleintritt in der Lage sein sollte, Denkopoperationen über konkrete Wahrnehmungsinhalte ohne unmittelbare Anschauungshilfen durchzuführen. Hierfür muss es realisieren, dass ein Sachverhalt aus mehreren Aspekten besteht. Weiters muss es imstande sein, diese Aspekte synchron im

Symbolsystem Sprache oder in der Vorstellung zu repräsentieren und für die Lösung eines Problems heranzuziehen. In den Jahren zuvor übt es diese Differenzierung (Kastner-Koller & Deimann, 2002).

1.3.2.2 Die Entwicklung der Informationsverarbeitung

Zu den Aspekten, die der piagetschen Theorie nicht zu entnehmen sind, zählen Aufschlüsse über die Entwicklung des Gedächtnisses. Hierzu liefern insbesondere die Informationsverarbeitungstheorien entscheidende Erkenntnisse. Diesen Theorien ist die Annahme gemeinsam, dass Denken Informationsverarbeitung ist. Sie sehen die kognitive Entwicklung als kontinuierlich an und rücken die Veränderungsmechanismen spezifischer kognitiver Bereiche in den Mittelpunkt des Interesses (Siegler, 1998/2001). Die Zunahme an Gedächtniskapazität und Informationsverarbeitungsgeschwindigkeit, die Verbesserung kognitiver Strategien und die zunehmende Vermehrung von Wissen sind für sie hierbei die Motoren der Entwicklung (Sodian, 2008).

Die auf diesen Annahmen basierenden Gedächtnismodelle sind durch Merkmale charakterisiert, die sich mit einer Computemetapher beschreiben lassen: Die Informationsverarbeitungskapazität und -geschwindigkeit werden mit der Hardware, erworbene Strategien und Fertigkeiten mit der Software eines Computers verglichen. Überdies wird angenommen, dass das Gedächtnis in verschiedene Untersysteme gegliedert ist, welche verschiedene Aufgaben zu erfüllen haben und miteinander interagieren (z. B. Mehrspeichermodell nach Atkinson und Shiffrin, 1968; Arbeitsgedächtnismodell nach Baddeley, 2009; Gedächtnismodell nach Case, 1985/1999). Demgemäß wird im Allgemeinen nicht die Entwicklung des Gesamtsystems untersucht, sondern die Entwicklung einzelner Subsysteme.

Die Gedächtniskapazität wird zumeist mit Aufgaben zur Gedächtnisspanne untersucht. Hierzu werden den Kindern Zahlen, Buchstaben oder Begriffe vorgegeben, die anschließend wiedergeben werden sollen. Bei derartigen Aufgaben lässt sich ein deutlicher Alterstrend feststellen: Zwei Jahre alte Kinder können sich Reihen mit zwei Zahlen merken und wiedergeben, Dreijährige reproduzieren Reihen von drei Zahlen, Vier- und Fünfjährige erinnern sich durchschnittlich an vier Zahlen, Neunjährige in etwa an sechs (Dempster, 1981, zitiert nach Siegler, 1998/2001). Auch in Versuchen zur Erfassung der Gedächtnisspanne für Wörter bzw. Handlungen zeigt sich ein Alterstrend: Case (1985/1999) berichtet, dass Kinder im Alter von ein- bis eineinhalb Jahren sich an

eine Serie von zwei Worten sowie von zwei Handlungen erinnern können, während Dreieinhalb- bis Fünfjährige bereits vier Worte bzw. Handlungen korrekt reproduzieren. Die Zunahme an Informationsgeschwindigkeit ändert sich mit dem Alter in ähnlicher Relation: Vier- bis Fünfjährige benötigen ein Drittel mehr Zeit als Acht- bis Neunjährige für die Verarbeitung grundlegender Operationen (Kail, 1991). Die stetige Verbesserung von Gedächtniskapazität und Informationsverarbeitungsgeschwindigkeit wird dabei in engem Zusammenhang mit Reifungsprozessen des Nervensystems gesehen, welche erst mit Erreichen der Adoleszenz abgeschlossen sind (Mietzel, 2002). Zudem bereitet es jüngeren Kindern noch Schwierigkeiten, Merkmale zu enkodieren und zu aktivieren. Dies sind jedoch Fähigkeiten, die bei Reproduktionsleistungen unabdingbar sind (Schneider & Büttner, 2008).

In Hinblick auf die Entwicklung der Gedächtniskapazität geht Case (1985/1999) davon aus, dass die Gesamtverarbeitungskapazität eines „exekutiven Verarbeitungsspeichers“ im Laufe des Lebens stabil bleibt. Nach Case wird ein Teil der Gesamtkapazität für den „Operationsspeicher“ benötigt, in welchem die kognitiven Prozesse aktiviert und ausgeführt werden. Der für diesen Speicher benötigte Raum wird im Laufe der Entwicklung geringer, da die Informationen reifungsbedingt zunehmend effizienter verarbeitet werden und somit weniger Kapazität beanspruchen. Diese frei werdende Kapazität steht sodann dem „Kurzzeitspeicher“ zur Verfügung, der der Aufrechterhaltung und/oder dem Abruf kürzlich abgelaufener kognitiver Aktivitäten dient. Demzufolge verbessert sich die Leistung des Kurzzeitspeichers, was den alterskorrelierten Zuwachs in der Gedächtnisspanne erklärt.

Ein weiteres Arbeitsgedächtnismodell wurde von Baddeley und Hitch (1974, zitiert nach Baddeley, 2009) entwickelt und in den folgenden 35 Jahren erweitert. Dieses betont weniger die strukturelle als vielmehr die funktionelle Komponente innerhalb der Informationsverarbeitung. Nach Baddeley (2009) verfügt das Arbeitsgedächtnis über eine Leit- und Koordinationsstelle („central executive“) und drei verschiedene Hilffsysteme. Die zentrale Exekutive wird als aufmerksamkeitsgestütztes System verstanden, welches der Selektion von aktuell im Arbeitsgedächtnis vorhandenen Informationen und der Steuerung entsprechender kognitiver Prozessen dient. Sie kommt beispielsweise zum Tragen, wenn die Aufmerksamkeit zwischen zwei oder mehr Aufgaben geteilt werden muss. Das phonologisch sprachliche Subsystem („phonological loop“) unterstützt die Bereithaltung und Verarbeitung akustischer oder sprachlicher Informationen. Die

Entwicklung des kindlichen Wortschatzes und der Erwerb von Vokabeln in einer Zweitsprache werden mit diesem Hilffssystem in Verbindung gebracht. Das visuell-räumliche Speichersystem („visuo-spatial sketchpad“) dient zur Aufrechterhaltung und Verarbeitung visueller und/oder räumlicher Informationen. Ein drittes multidimensionales Subsystem („episodic buffer“) bindet perzeptiven Input, Informationen aus den beiden Hilffssystemen und Inhalte des Langzeitgedächtnisses und speichert sie in integrierter Objektform oder als episodische Repräsentationen kurzfristig ab.

Untersuchungen legen die Vermutung nahe, dass das Arbeitsgedächtnis bis zum Alter von circa fünf Jahren vorwiegend mit visuell-räumlichen Codes arbeitet (z. B. Conrad, 1971 zitiert nach Goswami, 1998/2001; Hitch, Halliday, Schaafstal & Schraagen, 1988) und in Folge durch die Verwendung phonologischer Codes ergänzt wird. Die Ausnutzung des visuell-räumlichen Skizzenblocks scheint zudem im Laufe der Entwicklung zurückzugehen.

Infolge umfangreicher Forschungsergebnisse wird der Leistungsfähigkeit des Arbeitsgedächtnisses bzw. seiner Subsysteme prognostische Relevanz in Hinsicht auf das Erlernen der Kulturtechniken und der Qualität bei deren späterer Ausübung beigemessen (z. B. Baddeley & Lewis, 1981; Cain, Oakhill & Bryant, 2004; D'Amico & Guarnera, 2005; Gathercole, Tiffany, Briscoe, & Thorn, 2005; Grube & Hasselhorn, 2006; Lee & Kang, 2002; Schuchardt, Kunze, Grube, & Hasselhorn, 2006).

Überdies wird davon ausgegangen, dass die Kapazität des Ultrakurzzeitgedächtnisses (Sensorischer Speicher) bereits bei Vorschulkindern vollständig entwickelt ist (Siegler, 1998/2001). Allerdings wenden Kinder dieser Altersgruppe noch nicht genug Aufmerksamkeit auf, um relevante Informationen aus dem sensorischen Speicher auszuwählen und dem Arbeitsgedächtnis (Operationsspeicher) zur weiteren Verarbeitung zuzuleiten (Mietzel, 2002).

Während die Verbesserung von Gedächtnisleistungen in den ersten drei Jahren vorwiegend auf geirnhysiologische Veränderungen zurückzuführen ist, gewinnen in den Folgejahren Lern- und Gedächtnisstrategien an Bedeutung (Kastner-Koller & Deimann, 2002). Nach Siegler (1998/2001) weisen Kinder bereits ab einem Alter von zwei Jahren erste rudimentäre Gedächtnisstrategien auf, die beispielsweise in der Art und Weise erkennbar sind, wie kleine Kinder versteckte Objekte suchen. Überlegte kognitive Aktivitäten, die bewusst eingesetzt werden, um die Chancen des Erinnerns zu verbessern,

wie beispielsweise Wiederholen, Ordnen und das Fokussieren der Aufmerksamkeit, zeigen sich in der Regel jedoch erst ab dem fünften Lebensjahr. Komplizierte Elaborationstrategien sind erst ab der späten Kindheit bzw. frühen Adoleszenz beobachtbar (Schneider & Büttner, 2008). Prinzipiell tendieren sowohl Kindergarten- als auch Schulkinder dazu, äußere Erinnerungshilfen inneren vorzuziehen (Flavell, 1977/1979).

Da Drei- bis Vierjährige noch kein intentionales Memorierverhalten zeigen wenn explizite Behaltensinstruktionen vorgegeben werden, scheint das implizite gegenüber dem expliziten Gedächtnis eine wesentlich größere Rolle in dieser Altersgruppe zu spielen (Schneider & Büttner, 2008). Das implizite (oder nicht-deklarative) Gedächtnis zeigt sich in einem qualifizierten und präzisen Verhalten, welches auf vorangegangenen Erfahrungen basiert, jedoch unbewusst und automatisch ist. Es wird demnach mit dem perzeptuellen Lernen in Verbindung gebracht. Studien legen nahe, dass sich das implizite Gedächtnis früh entwickelt und es sich mit zunehmendem Alter auch nicht weiterentwickelt (Goswami, 1998/2001). Das explizite (oder deklarative) Gedächtnis ist bewusst und ermöglicht ein aktives Erinnern und Reproduzieren von Begriffen, Zeitpunkten, Orten und Ereignissen. Aufgrund der Ergebnisse ihrer neuropsychologischen Studie führt Bauer (2004) die Begrenzungen kleinerer Kinder im expliziten Gedächtnis darauf zurück, dass insbesondere der Frontalkortex noch nicht vollständig entwickelt ist. Im Gegensatz zum impliziten Gedächtnis entwickelt sich das explizite auch nach abgeschlossener Reifung weiter. Dies wird durch die zunehmende Leistungsfähigkeit anderer kognitiver Prozesse erklärt (Goswami, 1998/2001).

Als weiterer Motor der Gedächtnisentwicklung wurde die Zunahme an Wissen genannt. Hierbei wird zwischen inhaltlichem Vorwissen und dem Wissen über Gedächtnisvorgänge (Metakognition) unterschieden. Schneider und Büttner (2008) sehen das kindliche Inhaltswissen als eine Art mentales, hierarchisch strukturiertes Wörterbuch zu Objekten und Konzepten an. Aufgrund von Erfahrungen nimmt das Wissen über spezifische Objekte bzw. Konzepte zu, und die Anzahl der Konzepte (Knoten) und zugehörigen Verbindungen innerhalb einer solchen hierarchischen Struktur vergrößert sich. Während das Geflecht immer enger wird, steigt die Wahrscheinlichkeit, dass bei Aktivierung eines Knotens ein benachbarter Bereich zugänglich wird. Die Autoren folgern daraus, dass zwischen dem Ausmaß an Vorwissen in einem Bereich und der Gedächtnisleistung für Konzepte oder Objekte dieses Bereichs eine systematisch positive

Beziehung besteht. Unterschiede zwischen kleineren und größeren Kindern sowie zwischen Kindern und Erwachsenen sind demnach nicht nur auf Kapazitäts- und Strategiedefizite der Jüngeren zurückzuführen, sondern auch mit Effekten der Wissensbasis zu erklären. Allerdings kann nicht davon ausgegangen werden, dass größeres Inhaltswissen stets mit chronologischem Alter zusammenhängt. In einer Studie von Chi (1978, zitiert nach Goswami, 1998/2001) wurde die Bedeutung des bereichsspezifischen Wissens als Einflussfaktor für die Gedächtnisleistung eindrucksvoll gezeigt: Sechs- bis zehnjährige Kinder waren Erwachsenen im Merken kurz präsentierter Schachkonstellationen überlegen, was die Autorin darauf zurückführte, dass die Kinder allesamt hervorragende Schachspieler und die Erwachsenen Laien waren. Beim Wiedergeben vorgegebener Zahlenfolgen schnitten die Kinder hingegen schlechter ab als die Erwachsenen.

Eine Wissensart, die insbesondere für die Entwicklung des Langzeitgedächtnisses bei kleineren Kindern von Bedeutung ist, sind sogenannte Skripts, die eine Art schematisiertes Handlungswissen für prototypische Alltagsereignisse darstellen. Im Rahmen der Münchner Längsschnittstudie zur Genese individueller Kompetenzen (LOGIK) wurde festgestellt, dass Drei- bis Vierjährige eine Geschichte über einer Geburtstagsparty wesentlich akkurater und geordneter wiedergeben konnten als eine Geschichte über einen Spielnachmittag (Knopf & Schneider, 1998). Dies kann darauf zurückgeführt werden, dass die Kinder beim Nacherzählen der Geburtstagsgeschichte auf ein Skript bezüglich einer typischen Geburtstagsparty zurückgreifen konnten. Allerdings führen Skripte bei kleineren Kindern mitunter dazu, dass sie sich an etwas zu erinnern meinen, was nicht geschehen ist. Erst im Grundschulalter unterscheiden Kinder deutlicher zwischen dem, was typischerweise passiert und dem, was davon abweicht (Siegler, 1998/2001).

Im Laufe der Entwicklung bildet sich zudem ein Wissen über das Gedächtnis aus. Dieses spezifische Wissen wird nach Flavell (1971) Metagedächtnis genannt und lässt sich in deklaratives und prozedurales Gedächtniswissen gliedern (Schneider & Büttner, 2008). Das deklarative bzw. explizite Metagedächtnis umfasst das konkret verfügbare und verbalisierbare Wissen in Hinsicht auf das eigene Gedächtnis, die Schwierigkeit von Gedächtnisaufgaben und die Anwendbarkeit von Gedächtnisstrategien. Kreutzer, Leonard und Flavell (1975, zitiert nach Schneider & Büttner, 2008) kamen in ihrer Interviewstudie zu dem Ergebnis, dass rudimentäres Gedächtniswissen bereits bei Kindergartenkindern

vorhanden ist: Diese zeigten überwiegend Einsicht darin, dass man vergessen kann und es leichter ist, sich an Ereignisse zu erinnern, die gerade passiert sind als an solche, die länger zurückliegen. Allerdings zeigten sie im Gegensatz zu Schulkindern auch einen übertriebenen Optimismus hinsichtlich ihrer Gedächtniskapazität. Im Laufe der Schulzeit nimmt das deklarative metakognitive Wissen stark zu und wird expliziter und bewusster, erlangt jedoch erst gegen Ende der späten Adoleszenz seinen Höhepunkt (Schneider & Büttner, 2008). Unter prozeduraler bzw. impliziter Metakognition wird die unbewusste Überwachung und Regulation gedächtnisbezogener Aktivitäten verstanden. In Hinblick auf die Gedächtnisüberwachung zeigen bereits Kleinkinder beeindruckendes Wissen wenn sie beispielsweise ihren Sprachgebrauch kontrollieren, indem sie sich selbst korrigieren (Siegler, 1998/2001). Die Qualität derartiger Überwachungsprozesse bleibt über das Schulalter hinweg relativ stabil (Schneider & Büttner, 2008). Die Regulation gedächtnisbezogener eigener Aktivitäten ist hingegen einem deutlichen Alterstrend unterworfen. Eine mehrfach replizierte Studie von Dufresne und Kobasigawa (1989) kam zu dem Ergebnis, dass sich die Fähigkeit zur Selbstregulation erst im Schulalter ausbildet: Sechsjährige wandten für unterschiedlich schwierig zu lernende Paarassoziationen dieselbe Lernzeit auf, Achtjährige widmeten den schwierigen Aufgaben tendenziell mehr Zeit. Erst Kinder im Alter von 10 Jahren verbrachten signifikant mehr Zeit für das Lernen der schwierigeren Wortpaare.

Die Entwicklung des Gedächtnisses im Kleinkind- und Vorschulalter wird mit der zunehmend differenzierten Wahrnehmungsfähigkeit, der besseren kognitiven Leistungsfähigkeit, der Sprachentwicklung und dem Wissenserwerb in Verbindung gebracht. Umgekehrt wird davon ausgegangen, dass sowohl die Differenziertheit von Gedächtnisstrategien als auch die Qualität und Quantität des Inhaltswissens sowie die Ausprägung des Metagedächtnisses eine wichtige Rolle in Hinblick auf die kognitive Entwicklung spielen (Goswami, 1998/2001). Auch lernen Kinder, deren Informationsverarbeitungsgeschwindigkeit hoch ausgeprägt ist, mehr als Kinder mit geringerer Informationsverarbeitungsgeschwindigkeit (Kail, 1991). Weinert (1998) spricht treffend von Aufschaukelungseffekten zwischen einzelnen Domänen der kognitiven Entwicklung.

Da die bisherigen Erläuterungen zur kognitiven Entwicklung sehr allgemein gehalten waren, wird im Folgenden die Entwicklung einer spezifischen kognitiven Fähigkeit beschrieben: Der Erwerb erster mathematischer Kompetenzen. Zudem soll der Darstellung der oben genannten Aufschaukelungseffekte Rechnung getragen werden.

1.3.2.3 Entwicklung mathematischer Fähigkeiten

Vielfach wird davon ausgegangen, dass Menschen mit genetisch verankerten Grundlagen des Wissenserwerbs ausgestattet sind, und Neugeborene infolgedessen über ein intuitives mathematisches Wissen verfügen (Stern, 1998). In zahlreichen Habituationsstudien (z. B. Starkey, Spelke & Gelman, 1990; Wynn, 1995) wurde gezeigt, dass Säuglinge ein elementares Zahlenverständnis erkennen lassen: Sie sind imstande, ein Objekt von zweien und zwei Objekte von dreien zu unterscheiden, wenn sie zuvor auf eine gewisse Anzahl habituiert wurden. Derartige Ergebnisse legen nahe, dass Säuglinge subitieren, also den kardinalen Wert von ein bis drei oder vier Objekten ohne Abzählen auf einen Blick erfassen können (Siegler, 1998/2001). Andere Forscher (z. B. Xu, 2003) konnten jedoch zeigen, dass Säuglinge auch imstande sind, größere Anzahlen zu unterscheiden, wenn das Verhältnis der Anzahlen 1:2 beträgt (z. B. 4 versus 8 Objekte). Dies unterminiert die Annahme eines angeborenen Verständnisses für diskrete Mengen und deutet auf eine wahrnehmungsspezifische Sensibilität für unterschiedliche Quantitäten hin.

In etwa mit einem Jahr zeigen sich Anfänge des Begreifens von Ordinalität im Sinne eines Verständnisses für die Äquivalenz („gleich viel“) und Nichtäquivalenz („weniger“ oder „mehr“) von Mengen (Brannon, 2002; Siegler, 1998/2001). Ein wenig später drücken Kinder dieses Verständnis sprachlich aus und bezeichnen Mengen als viel oder wenig (Berk, 2004/2005). Die Übertragung dieser frühen Kompetenz auf etwas größere Zahlen fällt ihnen aber schwer. Fragt man Dreijährige beispielsweise „Was ist mehr: fünf Orangen oder sieben Orangen?“, so antworten nur wenige korrekt. Erst im Alter von fünf Jahren sind die meisten Kinder in der Lage, derartige Probleme des Ordnungsprinzips im Zahlenraum von 1 bis 9 korrekt lösen (Siegler & Robinson, 1982, zitiert nach Siegler, 1998/2001). Schwierigkeiten bei der Diskrimination zweier Mengen haben kleinere Kinder insbesondere dann, wenn die Zahlen nahe beieinander liegen (Rousselle, Palmers & Noël, 2004).

Aufbauend auf den frühen mathematischen Kompetenzen entwickeln sich die numerischen Leistungen mit dem Spracherwerb weiter. Indem Kinder mit Zahlwörtern und der Zählprozedur immer besser vertraut werden, entsteht ein zunehmendes Verständnis für Kardinalität und Ordinalität. Wenn Zwei- bis Dreijährige zu zählen beginnen, bringen sie die aufgesagten Zahlen aber noch nicht mit diskreten Mengen in Verbindung, sondern reihen sie nur aufeinander (Krajewski, 2008). Mit drei bis vier Jahren gelingt es ihnen, die Elemente einer etwas größeren Menge (bis circa 10 Objekte)

fehlerfrei durch Zählen zu bestimmen (Stern, 1998). Vier- bis Fünfjährige begreifen, dass die Anzahl der Objekte in der Menge dem letzten Zählschritt entspricht. Sie demonstrieren dies, indem sie die letzte Zahl besonders betonen (Siegler, 1998/2001). Mit etwa fünf Jahren erkennen die meisten Kinder weiters, dass es irrelevant ist, an welcher Position man zu zählen beginnt, solange nur alle Objekte gezählt werden (Siegler).

Vierjährige beginnen, das Zählen zur Lösung einfacher arithmetischer Probleme zu nutzen, wobei sie anfangs die Einzelmengen durch ihre Finger repräsentieren und die Lösung auszählen (Berk, 2004/2005). In den folgenden beiden Jahren entwickeln sie ein immer besseres Verständnis für die Beziehung von Zahlen im Sinne einer Zahlzerlegung und erkennen zudem, dass sich zwei Anzahlen durch eine dritte Anzahl unterscheiden (Krajewski, 2008).

Allerdings neigen auch ältere Vorschulkinder noch häufig zu Invarianz: Lediglich die Hälfte der fünf- bis sechsjährigen Teilnehmer der LOGIK-Studie erkannte, dass die Anzahl zweier Mengen gleich blieb, wenn eine Reihe von Objekten vor ihren Augen räumlich ausgedehnt wurde (Stern, 1998).

Als Mechanismen, die kleineren Kindern erste mathematische Leistungen ermöglichen, wurden bislang folgende angesprochen: (angeborenes) intuitives Wissen, (visuelle) perzeptive Fähigkeiten und der Spracherwerb. Des Weiteren kann das Arbeitsgedächtnis als unspezifischer Einflussfaktor für den Erwerb mathematischer Kompetenzen angesehen werden. So erleichtert das visuell-räumliche Subsystem die Wahrnehmung von unterschiedlichen Mengen und das innerliche Aufrechterhalten von Anzahlen. Die phonologische Schleife unterstützt das Zählen und das vokale Behalten von Zwischenergebnissen. Die Zentrale Exekutive ist für die Aufmerksamkeitssteuerung und die Koordination von visuellen und auditiven Informationen zuständig. Der episodische Puffer schließlich bindet gesprochene Zahlen mit geschriebenen Ziffern und arithmetischem Wissen. Berk (2004/2005) weist ferner darauf hin, dass Kinder einzelne der oben beschriebenen Fertigkeiten früher erwerben, wenn Erwachsene ihnen viele Gelegenheiten zum Zählen und Vergleichen von Mengen geben und sie hierzu auch auffordern.

1.3.3 Sprachentwicklung

Wenngleich auch mannigfaltige Beziehungen zwischen einzelnen (kognitiven) Funktionsbereichen der Entwicklung bestehen, sprechen Dissoziationen zwischen sprachlichen und anderen intellektuellen Fähigkeiten für eine Unabhängigkeit der Entwicklung von Sprache und Kognition (Weinert, 2007). Eine nicht unerhebliche Zahl von Kindern weist trotz altersentsprechender kognitiver Entwicklung dennoch gravierende Probleme beim Erwerb der Sprache auf. Andererseits zeigen auch etliche geistig retardierte Kinder mitunter bemerkenswerte sprachliche Kompetenzen (Grimm, 2003). Unter Berücksichtigung derartiger Beobachtungen wird dem Erwerb der Sprache ein eigenes Kapitel gewidmet. An dieser Stelle sei auch darauf hingewiesen, dass die Sprachentwicklung trotz einiger universeller Merkmale muttersprachspezifisch ist (Weinert & Grimm, 2008), hier jedoch nur auf die Entwicklung des Deutschen als Erstsprache eingegangen wird.

Nachdem ein Kind für gewöhnlich zwischen dem 10. und 14. Lebensmonat die ersten Wörter produziert, sollte es spätestens bis zum Alter von 24 Monaten 50 Wörter produktiv und 200 rezeptiv beherrschen (Grimm, 2003). Nach Überschreiten der „magischen“ 50-Wort-Marke mit durchschnittlich 18 Monaten kommt es in der Regel zu einem rapiden Anwachsen des aktiven Wortschatzes, welches sich über Jahre hinweg fortsetzt (Siegler, 1998/2001; Spangler & Schwarzer, 2008; Weinert & Grimm, 2008). Infolge dieser „Benennungsexplosion“ haben Kinder im Alter von 2;5 Jahren einen Wortschatz von circa 525 Wörtern (Menyuk, 2000). In der ersten Klasse, ist ihr Vokabularium auf 10 000 bis 14 000 Worte angestiegen (Anglin, 1993).

Für die schnell fortschreitende lexikalische und semantische Entwicklung werden verschiedene Lernmechanismen verantwortlich gemacht: In der Phase des frühen Worterwerbs sind insbesondere soziale Lernprozesse von Bedeutung, aufgrund derer das Kind assoziative Verknüpfungen zwischen Wort und Objekt herstellen kann (Weinert & Grimm, 2008). Nachdem ein Kind in etwa 50 Wörter aktiv benutzt, spielen zunehmend spezifische semantische Vorannahmen und das inzidenzielle Lernen eine Rolle (Spangler & Schwarzer, 2008). Diese sprachspezifischen Vorannahmen beinhalten Beschränkungen („constraints“), die die Anzahl möglicher Wortbedeutungen auf ein Minimum reduzieren. Inzidenzielles Lernen meint ein nicht beabsichtigtes, zufälliges und beiläufiges Lernen. Kinder sind dadurch in der Lage, ein lediglich einmal gehörtes Wort sinngemäß anzuwenden und in eine Begriffskategorie einzuordnen (Mietzel, 2002). In weiterer Folge

dient zudem der Satzrahmen als Ausgangspunkt für den Erwerb von Verben, Adverbien und syntaktischen Konnektiven (Weinert & Grimm, 2008). Überdies beeinflusst die zunehmende Leistungsfähigkeit des Arbeitsgedächtnisses den Spracherwerb, wobei insbesondere die phonologische Schleife für den Wortschatzerwerb von Bedeutung ist.

Das Erreichen der 50-Wort-Marke fällt weiters mit der Bildung erster Wortkombinationen zusammen, wodurch der Beginn der produktiven Grammatik erkennbar wird. Die anfänglichen Zwei- bis Dreiwortsätze muten „telegraphisch“ an, da grammatikalische Funktionswörter (wie Artikel, Hilfsverben, Konjunktionen) sowie Morpheme ausgelassen werden. Dennoch zeigt sich bereits in dieser Phase, dass die Kinder ein natürliches Gefühl für die Syntax besitzen, da die Worte in einer Bedeutungsrelation stehen und formale sprachliche Regularitäten (z. B. Verwendung von Verben nach einem Subjekt und vor einem Objekt) beachtet werden (Weinert & Grimm, 2008). Mit circa drei Jahren sprechen Kinder in multimorphen Sätzen (circa 4,5 Morpheme). In etwa ein Jahr später beherrschen sie die wesentlichen Satzkonstruktionen (Weinert, 2007). In den folgenden zwei Jahren verbessert sich die Kenntnis syntaktischer und morphologischer Regeln, sodass die Kinder zunehmend in der Lage sind, auch komplizierte Sachverhalte adäquat auszudrücken (Kastner-Koller & Deimann, 2002).

Im Vorschulalter treten, nach anfänglich korrekter Verwendung, oftmals morphologische Fehler auf, die durch Übergeneralisierung von Regeln zu erklären sind und auf einen Reorganisationsprozess der linguistischen Kompetenz hindeuten. Das Sprachwissen ist zu diesem Zeitpunkt dennoch implizit, erst mit circa acht Jahren können Kinder bewusst über Konstruktionsregeln der Sprache reflektieren (Weinert & Grimm, 2008).

Auch der angemessene Gebrauch von Sprache im sozialen Kontext, die pragmatische Kompetenz, zeigt einen ähnlichen bemerkenswerten Verlauf. Ungefähr zum Zeitpunkt des Erreichens der 50-Wort-Marke sind Kinder erstmals imstande, Intentionen sprachlich zum Ausdruck zu bringen (Spangler & Schwarzer, 2008; Weinert & Grimm, 2008). Wenn die Länge der Konversationseinheiten und die Anzahl der wechselseitigen Sätze zunehmen, sind Kinder imstande, Unterhaltungen zu führen. Entgegen der Egozentrismusannahme Piagets können bereits Drei- bis Vierjährige ihre Sprache an das Alter, Geschlecht und den sozialen Status ihrer Gesprächspartner anpassen (Berk, 2004/2005; Weinert & Grimm, 2008). In Situationen, in denen sie die Reaktionen des Zuhörers nicht sehen können, haben Vierjährige aber noch Schwierigkeiten, sich sprachlich anzupassen (Berk, 2004/2005). Beispielsweise deuten sie beim Telefonieren

auf Gegenstände oder vollführen Gesten anstatt Worte einzusetzen. Die pragmatischen Kompetenzen von Vorschulkindern sind demzufolge noch stark kontextabhängig. Unterstützt durch direkte oder indirekte soziale Unterweisung erlernen Kinder im Kleinkind- und Vorschulalter zudem kulturspezifische Höflichkeitsregeln (Spangler & Schwarzer, 2008).

Insgesamt nähern sich Kinder bis zum Ende der Vorschulzeit der Erwachsensprache an (Kastner-Koller & Deimann, 2002) und sind weiters in der Lage, effektive soziale Gespräche zu führen.

Wenn Eltern und Pädagogen Kindern aufmerksam zuhören, sie zum Sprechen anregen und in sanfter Weise subtile, indirekte Rückmeldungen über grammatikalische Fehler geben, ermutigen sie die Kinder zu einem aktiven, experimentellen Umgang mit Sprache und unterstützen und fördern so den Spracherwerb. Eine derartige Kombination aus Einfühlsamkeit und Ermutigung reifen Verhaltens ist nach Berk (2005) auch der Kern sozialer und emotionaler Entwicklung.

1.3.4 Sozial-emotionale Entwicklung

Obwohl Einjährige bereits in der Lage sind, basale (primäre) Emotionen wie Freude, Ärger, Trauer, Angst oder Überraschung auszudrücken, ist die Ausdrucksfähigkeit komplexer (sekundärer) Emotionen wie Stolz, Scham, Verlegenheit, Schuld, Neid und Mitleid erst im Laufe des zweiten und dritten Lebensjahrs beobachtbar. Derartige selbstbezogene oder soziale Emotionen setzen ein rudimentäres kognitives Selbstkonzept sowie das Wissen um sozial anerkannte Verhaltensstandards voraus (Schwarzer & Jovanovic, 2007; Petermann & Wiedebusch, 2003). Mit fortschreitender Sprachentwicklung gelingt es Kindern zunehmend besser, Emotionen verbal zu benennen. Zweijährige sind bereits dazu imstande, die primären Emotionen Freude, Trauer, Ärger und Angst zu verbalisieren und beziehen sich dabei auf sich selbst oder auf andere sowie weiters auch auf Vergangenheit, Zukunft oder Gegenwart. Vom dritten bis zum sechsten Lebensjahr nimmt der Emotionswortschatz stetig zu, wobei zunehmend ebenso komplexe Emotionen wie Langeweile und Aufregung sprachlich ausgedrückt werden können (Wellman, Harris, Banerjee & Sinclair, 1995). Hierdurch werden auch ausführliche Gespräche über Emotionen, beispielsweise über Ursachen und Konsequenzen von Emotionen, möglich (Petermann & Wiedebusch, 2003).

Die Entwicklung des Emotionsausdrucks geht Hand in Hand mit der Entwicklung des Emotionsverständnisses. Zweijährige nehmen positive Emotionen differenziert wahr, haben jedoch noch Schwierigkeiten, negative Emotionen wie Ärger und Trauer von einander abzugrenzen (Spangler & Schwarzer, 2008). Diese Fähigkeit verbessert sich in den Folgejahren deutlich, sodass Vorschulkinder die Gefühle anderer nicht nur richtig interpretieren und benennen, sondern auch vorhersehen können. Zudem kennen sie effektive Möglichkeiten, um auf die negativen Gefühle anderer verändernd einzuwirken und verstehen den Zusammenhang zwischen Denken und Fühlen sukzessive besser (Berk, 2004/2005). Das zunehmende Emotionsverständnis geht einher mit der Abnahme des kindlichen Egozentrismus und der Ausformung einer Theory of Mind und gilt als Voraussetzung für soziale Kompetenz und Kooperationsfähigkeit (Kastner-Koller & Deimann, 2002).

Eine weitere Voraussetzung für soziale Kompetenz ist in der Entwicklung der Emotionsregulation zu sehen. Während Kinder in den ersten zwei Lebensjahren vorwiegend darauf angewiesen sind, dass Erwachsene ihnen helfen, ihre Emotionen zu regulieren, entwickeln Kleinkinder bereits Methoden der Selbstregulation (Schwarzer & Jovanovic, 2007). Beispielsweise halten sie sich die Augen oder Ohren zu, um eintreffende Sinnesinformationen einzuschränken oder versuchen sich mit anderen Aktivitäten abzulenken. In den Folgejahren erweitert sich das Spektrum der verfügbaren Regulationsstrategien erheblich. Kindergartenkinder zeigen zunehmend bewusste Kontrollstrategien, wie beispielsweise strategische Aufmerksamkeitskontrolle. Dieser Fortschritt ist auf die Entwicklung des präfrontalen Kortex und der damit einhergehenden stetig sich verbessernden inhibitorischen Kontrolle zurückzuführen (Schwarzer & Jovanovic). Zudem lernen Drei- bis Vierjährige ihre kommunikativen Fähigkeiten zur Emotionsregulation einzusetzen (Petermann & Wiedebusch, 2003).

Neben dem kindlichen Temperament und der kognitiven und sprachlichen Leistungsfähigkeit ist insbesondere die Sozialisation für den Erwerb und die Qualität emotionaler Kompetenzen ursächlich. So beeinflusst der Emotionsstil der wichtigsten Bezugspersonen das emotionale Ausdrucksverhalten des Kindes (Petermann & Wiedebusch, 2003). Das kindliche Emotionsverständnis steht mit der Qualität der Mutter-Kind-Bindung in Zusammenhang: Sicher an ihre Mütter gebundene Kinder zeigen ein fortgeschrittenes Emotionsverständnis, insbesondere hinsichtlich der Unterscheidung negativer Emotionen (Laible & Thompson, 1998). Auch die Qualität und Quantität von

Gesprächen, die Kinder mit Erwachsenen über Emotionen führen, wirkt sich auf das Emotionswissen aus (Schwarzer & Jovanovic, 2007). Die Vorbildwirkung von Erwachsenen und der elterliche Erziehungsstil beeinflussen die kindliche Emotionsregulation (Berk, 2004/2005). Feinfühliges und akzeptierendes elterliches Reagieren auf den Ausdruck starker Gefühle wirkt sich auf den Erwerb aller emotionalen Fertigkeiten positiv aus.

Entwicklungsaufgaben der ersten drei Lebensjahre im sozialen Bereich sind der Aufbau erster Bindungen und die Entwicklung von Neugier- und Erkundungsverhalten sowie damit einhergehend die Entfaltung eines ersten Selbstkonzepts und die Entstehung von Selbstbewusstsein. Als wichtige Themen der folgenden drei Jahre gelten die Steuerung der eigenen Handlungsimpulse, der Umgang mit zunehmender Autonomie und die Fähigkeit zur Interaktion mit Gleichaltrigen (Mähler, 2008).

Der Kindertarteneintritt erfordert, dass das Kind sich von seinen primären Bezugspersonen für eine gewisse Zeit zu lösen hat und es lernen muss, damit umzugehen. Gleichzeitig erhält es die Gelegenheit, neue Beziehungen aufzubauen und hierdurch wichtige soziale Verhaltensweisen zu erlernen. Es muss nun in einer Gruppe zurechtkommen und seine eigene Position in dieser finden. Es gilt, Konkurrenz auszuhalten, mit anderen zu kooperieren, zu teilen und Regeln zu befolgen sowie Gefühle zu äußern und anderen Mitgefühl entgegenzubringen (Mähler, 2007, 2008).

Die Beziehung zu Peers verbessert sich mit steigendem Selbstbewusstsein, zunehmender Sprachkompetenz und immer besserem Verständnis der Emotionen anderer (Berk, 2004/2005). Die verstärkte Interaktion mit Gleichaltrigen führt dazu, dass das kooperative Spiel ab dem vierten Lebensjahr sprunghaft zunimmt. Gleichzeitig tritt mit der zunehmenden Unterscheidung von realer und mentaler Welt das Funktionsspiel in den Hintergrund und wird vom Fantasie- oder Rollenspiel abgelöst. Dieses erlaubt dem Kind, Erfahrungen zu verarbeiten und soziale Beziehungen zu erproben. Während zunächst nur bekannte Personen dargestellt werden, spielt das ältere Vorschulkind mehr allgemeine Rollen wie Krankenschwester oder Polizist (Nickel & Schmidt-Denter, 1995). Beim Übergang zum Schuleintritt zeigen Kinder ein hohes Verständnis von Rollenbezügen, sodass ihre Rollenspiele immer wirklichkeitsbezogener werden. Zudem streben sie eine detailgenaue Wiedergabe von Abläufen an (Mähler, 2007).

In Hinblick auf das Thema dieser Arbeit mögen die vorangegangenen Ausführungen zunächst einer zu umfangreichen Betrachtung unterworfen worden sein. Hierbei muss man jedoch bedenken, dass eine praktisch tätige Person, die ein Kind entwicklungsdiagnostisch begutachtet, über umfangreiches entwicklungspsychologisches Know-how zu verfügen hat. Ohne zu wissen, wie Entwicklung sich vollzieht und welchen Verlauf sie normalerweise nimmt, fehlen die Grundlagen für eine Beurteilung der individuellen Entwicklung. Dieses Wissen ist selbstverständlich auch unabdingbar für die Konstruktion eines standardisierten Verfahrens, wobei Konstrukteure ihr Verfahren zudem auch entwicklungstheoretisch zu begründen haben. Die einem Verfahren zugrunde liegende Entwicklungstheorie ist überdies ein Maß für dessen Güte. Dies impliziert, dass Personen, die derartige Verfahren bewerten, über die Qualität der einzelnen Entwicklungstheorien Bescheid wissen müssen. So und also eingehender betrachtet mag man nun zur Ansicht kommen (müssen), dass die vorangegangenen Ausführungen eher zu knapp erscheinen. Ein mehr an Information erscheint im Rahmen dieser Arbeit jedoch nicht angemessen, sodass zur Vertiefung auf die einschlägige Literatur verwiesen wird.

2 Einführung in die Entwicklungsdiagnostik

Im Folgenden wird erläutert, was unter Entwicklungsdiagnostik zu verstehen ist, wodurch sich die allgemeine von der spezifischen Entwicklungsdiagnostik abgrenzt, und welche Ziele sie verfolgt. Zudem wird auf die Geschichte der Entwicklungsdiagnostik eingegangen.

2.1 Begriffsbestimmung Entwicklungsdiagnostik

Entwicklungsdiagnostik ist sowohl ein Teilbereich der Psychologischen Diagnostik als auch der Angewandten Entwicklungspsychologie.

Als wissenschaftliche Disziplin befasst sich die Psychologische Diagnostik mit der Theorie, Entwicklung und Analyse diagnostischer Verfahren. Im Sinne einer Tätigkeit, dem psychologischen Diagnostizieren, ist sie der Fallbehandlung in der Praxis gewidmet und zielt auf das systematische Erfassen und Erklären psychischer Merkmale des in Betracht stehenden Menschen ab. Hierfür werden wissenschaftliche Methoden verwendet, welche durch die Forschung bereitgestellt werden. Der psychologisch-diagnostische Prozess erstreckt sich von der Klärung der Fragestellung und der Auswahl geeigneter Verfahren über die Anwendung und Auswertung dieser Verfahren zur Interpretation und Gutachtenerstellung, um schließlich im Festsetzen einer Intervention oder in einem Maßnahmenvorschlag zu münden (Kubinger, 2006). Dies impliziert, dass der Begriff Entwicklungsdiagnostik ebenso Theorie und Praxis verkörpert. Zum einen beschäftigt diese sich mit der Forschung in Hinsicht auf entwicklungsdiagnostische Verfahren, zum anderen bezweckt sie die Erfassung entwicklungspsychologischer Merkmale und wird in diesem Zusammenhang auch der Angewandten Entwicklungspsychologie zugeordnet. Eine diesem Zweck entsprechende Definition liefern Deimann und Kastner-Koller:

Entwicklungsdiagnostik als ein wesentlicher Bereich der Angewandten Entwicklungspsychologie beschäftigt sich mit der quantitativen und qualitativen Erfassung entwicklungsbedingter Kompetenzen im Hinblick auf die Beschreibung und Erklärung des aktuellen Kompetenzniveaus, die Prognose der zukünftigen Entwicklung sowie die Planung und Evaluation von Entwicklungsinterventionen. (2007, S. 558)

Diese Begriffsabgrenzung deutet auf drei wichtige Punkte hin: auf die zu erfassenden Merkmale, auf die anzuwendenden Methoden und auf die Ziele, die damit verfolgt werden.

1. Als Merkmale zu erfassen sind die entwicklungsbedingten Kompetenzen. Hierunter verstehen Deimann und Kastner-Koller (2007) Fähigkeiten und Fertigkeiten, „die auf der Transaktion zwischen genetischer Disposition und Umwelterfahrungen beruhen und sich über die Lebensspanne, also alterskorreliert, durch solche transaktionale Auf- und Abbauprozesse ausbilden“ (S. 558). Entwicklungsbedingte Kompetenzen können motorische, kognitive, sprachliche und sozial-emotionale Entwicklungsmerkmale betreffen, welche es altersabhängig zu erfassen gilt. Zeitgemäße Entwicklungsdiagnostik betrachtet Entwicklung unter einer interaktionistischen Perspektive über die Lebensspanne hinweg. Dennoch richtet sich Entwicklungsdiagnostik vor allem auf das Entwicklungsgeschehen im Säuglingsalter und in der frühen Kindheit.
2. Obwohl Entwicklungsdiagnostik im engeren Sinne zumeist unmittelbar mit dem Einsatz von Entwicklungstests verknüpft ist (Petermann & Macha, 2008), beschränkt sich das Methodenarsenal nicht nur auf Tests, sondern umfasst auch qualitative Verfahren wie Verhaltensbeobachtung und anamnestische Erhebungen. Auf die einzelnen Methoden der Entwicklungsdiagnostik wird im dritten Kapitel detailliert eingegangen. Es soll an dieser Stelle jedoch vorausgeschickt werden, dass Entwicklungsdiagnostik wissenschaftlich fundiert zu erfolgen hat. Standardisierte Entwicklungstestverfahren haben den Gütekriterien der psychologischen Diagnostik zu entsprechen und sind darüber hinaus entwicklungstheoretisch zu begründen.
3. Wie der obigen Definition bereits zu entnehmen ist, zielt moderne Entwicklungsdiagnostik auf die Beschreibung und Erklärung des aktuellen Entwicklungsstandes, die Prognose der weiteren Entwicklung, und die Planung und Evaluation von Entwicklungsinterventionen ab. Die Aufgaben und Ziele der Entwicklungsdiagnostik werden in weiterer Folge näher ausgeführt.

Da die vorliegende Arbeit die Allgemeine Entwicklungsdiagnostik zum Gegenstand hat, gilt es, diese von der Spezifischen abzugrenzen.

Die allgemein orientierte Entwicklungsdiagnostik befasst sich mit der Beschreibung und Erklärung entwicklungsbedingter Kompetenzen in mehreren grundlegenden Entwicklungsbereichen und hat ein Abbild der (kindlichen) Entwicklung in möglichst großer Bandbreite zum Ziel. In allgemein orientierten Entwicklungstests und -screenings findet sich eine weitgehend ähnliche Zusammenstellung von Kompetenzbereichen (Petermann & Macha, 2008):

- Motorik
- Kognition und Wahrnehmung
- Lernen und Gedächtnis
- Sprache
- sozial-emotionale Entwicklung sowie
- adaptive lebenspraktische Fertigkeiten

Allgemeine Entwicklungsdiagnostik liefert somit einen Überblick über die Gesamtentwicklung und eignet sich insbesondere zur Eingangsdiagnostik (Petermann & Macha, 2005).

Spezifische Entwicklungsdiagnostik erfasst Informationen zu einzelnen Kompetenzbereichen (z. B. Sprache, Motorik, Visuelle Wahrnehmung, Konzentration). Spezifische Entwicklungstests operationalisieren in mehreren Subtests verschiedene Dimensionen eines umschriebenen Entwicklungsbereichs (z. B. im Funktionsbereich Sprache: phonologisches Gedächtnis, expressive Sprache, rezeptive Sprache). Sie werden generell als Ergänzung zu einem allgemeinen Test eingesetzt.

2.2 Aufgaben und Ziele der psychologischen Entwicklungsdiagnostik

Entwicklungsdiagnostik zielt primär darauf ab, den aktuellen Entwicklungsstand zu beschreiben, zu bewerten und dessen Zustandekommen zu erklären. Diese Aufgaben wurden der Entwicklungsdiagnostik bereits vor mehr als zwei Jahrzehnten zugeschrieben (vgl. Montada, 1985; Rauh, 1985; Rennen-Allhoff & Allhoff, 1987, Kap. 1). Darüber hinaus finden sich in der Literatur und in den Manualen von Entwicklungstestverfahren

weitere, zumeist ähnlich beschriebene Aufgaben. Die Schwerpunktsetzung ist allerdings aufgrund der theoretischen und methodischen Zugänge der Autoren und Autorinnen mitunter unterschiedlich.

Tröster et al. (2004) betonen die Wichtigkeit der Früherkennung von Kindern mit Entwicklungsverzögerungen und die Identifikation von Risikogruppen. Esser und Wyschkon (2002), Grob, Hagmann-von Arx und Bodmer (2009) sowie Kastner-Koller und Deimann (2002) verfolgen einen förderdiagnostischen Ansatz und betonen die Bedeutung der Entwicklungsdiagnostik für die Planung notwendiger Entwicklungsinterventionen. Ettrich (2000), Petermann, Stein und Macha (2004) und Reuner, Rosenkranz, Pietz, und Horn (2007) fokussieren auf die Darstellung intraindividuelle Entwicklungsverläufe.

Darüber hinaus findet die entwicklungspsychologische Beurteilung ihren Einsatz beim Ausschluss von Krankheiten, wird zu Forschungszwecken angewandt (Reuner et al., 2007), dient der Evaluation von Entwicklungsinterventionen (Fuiko & Wurst, 2003) und steht allgemein im Dienste von Gesundheitsförderung und Prävention (Grob et al., 2009).

Entwicklungsdiagnostik kann damit hinsichtlich der unterschiedlichsten Zielsetzungen betrieben werden. Wie Johnson & Marlow (2006) treffend ausdrücken, ist „das ultimative Ziel“ der Entwicklungsdiagnostik jedoch stets die Ableitung einer Prognose. Dies gilt insbesondere für Entwicklungsscreenings (Deimann & Kastner-Koller, 2007).

Zusammenfassend können der allgemein orientierten Entwicklungsdiagnostik folgende Hauptaufgaben zugesprochen werden:

Sie dient

- der differenzierten Beschreibung des Entwicklungsstandes bzw. bei mehrmaliger Untersuchung des Entwicklungsverlaufs,
- der Diagnose von Entwicklungsabweichungen, indem die individuelle Entwicklung mit einer Entwicklungsnorm verglichen wird,
- der Darstellung von Stärken und Schwächen intraindividuelle Entwicklungs-kompetenzen,
- der Abklärung von individuellen, familienbezogenen, sozialen und versorgungsstrukturellen Entwicklungsbedingungen im Sinne einer Entwicklungs-psychopathologie,

- der Erstellung von Prognosen für den weiteren Entwicklungsverlauf,
- der Ableitung von Entwicklungsinterventionen,
- der Evaluation dieser Interventionen, wenn nach Abschluss der Intervention eine Wiederholungsmessung vorgenommen wird, und
- der Eingangsdiagnostik.

2.3 Historische Wurzeln der Entwicklungsdiagnostik

Nach Rennen-Allhoff und Allhoff (1987, Kap. 1) hat die Entwicklungsdiagnostik ihre Wurzeln in zahlreichen Forschungsrichtungen und -ergebnissen des 19. Jahrhunderts.

Zu jener Zeit zeigte sich ein zunehmendes psychiatrisches und pädagogisches Interesse an der Identifikation und Klassifikation von Personen mit kognitiven Beeinträchtigungen, zunächst insbesondere in Frankreich: Nachdem Jean Itard (1774-1838) sich gegen Ende des 18. Jahrhunderts der Erziehung eines Wolfskindes¹ angenommen hatte und in seiner fünfjährigen pädagogischen Unterweisung bescheidene, aber beobachtbare Erfolge erzielen konnte, kam die akademische Diskussion über Geisteskrankheiten, ihre Genese und ihre Prognose in Gang (Groffmann, 1983). So hob der französische Psychiater Jeanne Étienne Dominique Ésquirol (1772-1840) in seinem 1838 publizierten Hauptwerk „Des maladies mentales“ geistige Behinderungen (seinerzeit „Schwachsinn“) von psychischen Störungen ab. Zudem differenzierte er zwischen angeborenem und erworbenem Schwachsinn und unterschied auch noch zwei Grade von Schwachsinnigkeit, die „Idiotie“ und die „Imbezillität“ - Termini die bis in die zweite Hälfte des 20. Jahrhunderts gebräuchlich waren. Als Unterscheidungsmerkmal führte er das Ausmaß der Sprachfähigkeit an. Édouard Séguin (1812-1880), ein Schüler Ésquirols und Itards, war – im Gegensatz zu Itard selbst – von seinen Erfolgen mit dem Wolfskind beeindruckt und kam zu dem Schluss, dass auch Schwachsinnige erziehbar und trainierbar seien. In Folge gründete er 1840 eine Schule für geistesschwache Kinder. Seine 1850 erfolgte Emigration nach Amerika trug auch dort zu einer Ausbreitung des sonderpädagogischen

¹ Der „Knabe von Aveyron“ wurde 1798 im Alter von etwa 12 Jahren von Jägern im Wald von Aveyron aufgefunden und war offenbar isoliert und ohne menschliche Fürsorge aufgewachsen. Er konnte nicht reden und zeigte die Gewohnheiten von Tieren.

Anstaltswesens bei. Gegen Ende des 19. Jahrhunderts wurde die pädagogische Sonderstellung geistig retardierter Kinder allgemein anerkannt und verschiedene kognitive Leistungsgrade von Idiotie über Normal- bis zu Hochbegabung unterschieden. Mit dem Ausbau von Schulen für geistig zurückgebliebene Kinder entstand ein verstärktes Bedürfnis nach objektiven Zuweisungsroutinen. Dies führte in Folge zum Auftrag des französischen Erziehungsministeriums an Binet und Simon, ein zeitsparendes und objektives Verfahren zur Auswahl Minderbegabter zu entwickeln (Groffmann, 1983; Rennen-Allhoff & Allhoff, 1987, Kap. 1).

Zudem bildete sich im Verlauf des 19. Jahrhunderts eine Kinderpsychologie heraus. Zum einen hielten wissenschaftlich interessierte Eltern, unter ihnen auch Forscher wie Charles Darwin (1809-1882) und Thierry William Preyer (1841-1897), ihre Beobachtungen in Tagebüchern fest. Diese Arbeiten boten später umfassendes Material für die Aufgabenformulierung entwicklungsdiagnostischer Verfahren. Zum anderen setzte Granville Stanley Hall (1844-1924) erstmals einen Fragebogen ein, um herauszufinden, welches Wissen Kinder in jeweils welchem Alter besitzen. Infolgedessen war am Ende des 19. Jahrhunderts deutlich geworden, dass Entwicklung zumeist nach ähnlichen Mustern verläuft, die Entwicklungsgeschwindigkeit jedoch unterschiedlich sein kann. Weiters wurde angenommen, dass Verhaltensunterschiede prinzipiell messbar sind. Unklar blieb jedoch, welche Messmethode besonders geeignet wäre (Rennen-Allhoff & Allhoff, 1987, Kap. 1).

Die Anfänge der psychometrischen Entwicklungsdiagnostik lassen sich auf das Jahr 1905 zurückführen, in dem Alfred Binet (1857-1911) und Théodore Simon (1872-1961) ihre Stufenleiter der Intelligenz publizierten. Nach erfolgter Revision im Jahr 1908 umfasste die Stufenleiter 49 Aufgaben und war für das Alter von 3 bis 13 Jahren vorgesehen. Die Aufgaben waren jeweils der Altersstufe zugeordnet, in der sie üblicherweise zuerst bewältigt werden können. Dadurch konnte für jedes Kind ein Intelligenzalter bestimmt werden, welches – mit dem Lebensalter in Relation gesetzt – Aufschluss über altersgemäße, retardierte oder akzelerierte Entwicklung geben sollte (Deimann & Kastner-Koller, 2007; Rennen-Allhoff & Allhoff, 1987, Kap. 1). Für Binet und Simon war die Erfassung der Intelligenz somit gleichbedeutend mit einer Entwicklungsdiagnose (Filipp & Doenges, 1983).

Das Verfahren fand bald weite Verbreitung, insbesondere im angloamerikanischen und deutschen Sprachraum, und wurde mehrmals revidiert und adaptiert. So wurde die

Stufenleiter nach unten bis ins Säuglingsalter und/oder nach oben bis ins Erwachsenenalter ergänzt, auf die Verhältnisse anderer Länder zugeschnitten und hinsichtlich Instruktion, Dokumentation und Interpretation verbessert. Die 1916 von Lewis Madison Terman (1877-1956) vorgelegte Stanford-Revision ist die vermutlich bekannteste Überarbeitung, welche über mehr als 20 Jahre eine führende Rolle in der Intelligenztestung behauptete. Terman realisierte in seinem Verfahren erstmals einen Intelligenzquotienten. Dieser wurde als Verhältnis des Intelligenzalters zum Lebensalter, multipliziert mit 100, errechnet (Rennen-Allhoff & Allhoff, 1987, Kap. 1).

Kurz nach Erscheinen der Stufenleiter von Binet und Simon wurden die ersten Rufe laut, dass die alleinige Überprüfung intellektueller Fähigkeiten für die Einschätzung des Entwicklungsstandes nicht ausreicht (z. B. Kuhlmann, 1912, zitiert nach Filipp & Doenges, 1983). Arnold Gesell (1880-1961) veröffentlichte 1925 erstmals Entwicklungsnormen („developmental schedules“), die das charakteristische Verhalten für Kinder von der Geburt bis zum Schulalter darlegen sollten. Diese auf Basis von Längsschnittuntersuchungen gewonnen Normen umfassten die Entwicklungsbereiche Motorik, Anpassungsfähigkeit, Sprache und personal-soziales Verhalten. Anfang der 40er Jahre publizierte Gesell seine Normen als Entwicklungstest mit der Möglichkeit zur Berechnung eines Entwicklungsquotienten (GDS, Gesell & Amatruda, 1941).

Angeregt durch die frühen Arbeiten Gesells konstruierte der Wiener Arbeitskreis um Charlotte Bühler (1893-1974) ab 1928 unterschiedliche Entwicklungstestreihen für einzelne Altersgruppen. 24-stündige Dauerbeobachtungen, die die Abbildung charakteristischer, den jeweiligen Reifungsstadien entsprechenden Verhaltensweisen zum Ziel hatten, bildeten die Grundlage für die Testentwicklung. Für jedes angenommene Reifungsstadium wurde in Folge eine eigene Entwicklungstestreihe aufgestellt, deren Aufgaben verschiedenen Dimensionen (Sinnliche Rezeption, Körperbeherrschung, Soziale Kontaktaufnahme, Lernen, Betätigung an handgreiflichem Material und Geistige Produktion) zugeordnet wurden. Neben der Quantifizierung der Testleistung durch Berechnung von Entwicklungsalter und -quotient, sahen die Wiener Entwicklungstestreihen die Erstellung eines Entwicklungsprofils vor. Die Darstellung der individuellen Entwicklungsstruktur wurde als besonders wichtig für die Erziehung und praktische Behandlung eines Kindes angesehen (Hetzer, 1950). Die 17 Testreihen, welche für den Zeitraum vom 1. Lebensmonat bis zum vollendeten 6. Lebensjahr vorgesehen waren, wurden als Bühler-Hetzer-Kleinkindertest (BHKT, Bühler & Hetzer, 1932) publiziert,

welcher das bekannteste entwicklungsdiagnostische Verfahren aus dieser Periode ist (Deimann & Kastner-Koller, 2007).

Als weiterer Meilenstein in der Geschichte der Entwicklungsdiagnostik gelten die von Nancy Bayley (1899-1994) im Rahmen der Berkeley Growth Study (vgl. Jones & Bayley, 1941) entwickelten Skalen zur Beurteilung der mentalen und motorischen Entwicklung von Säuglingen und Kleinstkindern. Kontrastierend zu Gesell und Bühler sah Bayley keine Testserien für einzelne Altersstufen vor, sondern ordnete die Items innerhalb des motorischen und kognitiven Bereichs nach ansteigender Schwierigkeit (Rennen-Allhoff & Allhoff, 1987, Kap. 1). Bayley interessierte sich insbesondere für die Frage nach Stabilität bzw. Instabilität individueller Entwicklung und beschäftigte sich somit erstmals mit der Prognose von Entwicklungstestergebnissen.

Eine zweite Hochblüte erreichte die Entwicklungsdiagnostik in den 60er und 70er Jahren, als die Erfolge des amerikanischen Head-Start-Programms die Bedeutung von Frühförderung sozial benachteiligter Kinder im kognitiven und sprachlichen Bereich verdeutlichten. Da die bestehenden Entwicklungstestverfahren sowohl zur objektiven Zuweisung der Kinder zu speziellen Förderprogrammen als auch zur Evaluation der Fördermaßnahmen ungeeignet erschienen, boomte die Entwicklungsdiagnostik (Deimann & Kastner-Koller, 1994, 2007).

Der Schwerpunkt der Testentwicklung lag nun auf der Erfassung umschriebener Verhaltensbereiche, sodass zahlreiche spezifische Verfahren zur Erfassung des Entwicklungsstandes von Kognition, Sprache, Wahrnehmung, Lernen, Aufmerksamkeit, Motorik und sozial-emotionaler Entwicklung konzipiert wurden (vgl. Filipp & Doenges, 1983; Rennen-Allhoff & Allhoff, 1987, Kap.1).

In Hinblick auf die Diagnostik des allgemeinen Entwicklungsstandes wurden testmethodische Verbesserungen angestrebt. So publizierte Bayley Ende der 60er Jahre die Bayley Scales of Infant Development (BSID, Bayley, 1969), welche auf einer Zusammenfassung der für die Berkeley Growth Study entwickelten Skalen basierten, und in die die Erkenntnisse der weiterführenden Forschung Bayleys und ihrer Mitarbeiter einfließen. Das Verfahren war zur Bestimmung des allgemeinen Entwicklungsstandes in den ersten zweieinhalb Jahren bestimmt und umfasste neben einer Mentalen und Motorischen Skala eine standardisierte Verhaltensbeurteilung. Das Manual enthielt neben Durchführungs- und Auswertungsanweisungen umfassende Angaben zur Testgüte des

Verfahrens. Die BSID galten in Folge für lange Zeit als *via regia* der Erfassung des Entwicklungsstandes von Säuglingen und Kleinkindern (Aylward, 2004). Drei Jahre nach Veröffentlichung der BSID erschienen die McCarthy Scales of Children's Abilities (MSCA, McCarthy, 1972), eine der wenigen Neuentwicklungen zur Erfassung der globalen Entwicklung, die in dieser Periode publiziert wurden. Das Verfahren war für Kindern im Alter von 2;6 bis 8;6 Jahren vorgesehen und zielte auf die Erfassung des kognitiven und motorischen Entwicklungsstandes sowie die Darstellung individueller Stärken und Schwächen ab. Das Manual enthielt neben außergewöhnlich umfassenden Anweisungen hinsichtlich Vorgabe, Auswertung und Interpretation Angaben zur Reliabilität und Validität des Verfahrens. Knobloch, Stevens und Malone veröffentlichten 1980 eine revidierte und neu normierte Version der Gesell Developmental Schedules (GDS), welche sich im Gegensatz zum Originalverfahren ebenfalls durch erhöhte Standardisierung auszeichnete und Angaben zur Testgüte enthielt (Rennen-Allhoff & Allhoff, 1987, Kap.1 & 2).

Darüber hinaus entstanden in dieser Zeit erstmals Entwicklungsscreenings, welche in ökonomischer Weise Risikokinder erfassen sollten. Als bedeutendstes dieser Verfahren gilt der Denver Developmental Screening Test (DDST, vgl. Frankenburg & Dodds, 1967), welcher sich insbesondere im pädiatrischen Bereich verbreitete. Zudem wurden Kurzformen allgemeiner Entwicklungstests vorgelegt und Elternfragebögen konzipiert (Rennen-Allhoff & Allhoff, 1987, Kap.1).

Weiters kamen Tests speziell für Säuglinge auf den Markt. Brazelton (1973, zitiert nach Rennen-Allhoff & Allhoff, 1987, Kap. 1) entwickelte ein Verfahren, welches neben Reflexen auch das Verhalten des Neugeborenen und seine Reaktionen auf die Umgebung erfasste. In Deutschland wurde 1978 die Münchner Funktionelle Entwicklungsdiagnostik (MFED-1, Hellbrügge, Lajosi, Menara, Schamberger & Rautenstrauch) publiziert. Dieses Verfahren war ethologisch orientiert und zielte auf die Früherkennung von Entwicklungsstörungen im ersten Lebensjahr ab.

Die folgenden Jahre erbrachten keine nennenswerten Neuerungen in der Allgemeinen Entwicklungsdiagnostik. Allerdings stellten vielfältige Untersuchungen die mess- und testtheoretischen Grundlagen der Verfahren in Frage (z. B. Kastner-Koller & Deimann, 1991). Auch wurden in den 80er und 90er Jahren zunehmend Rufe laut, dass die bestehenden, auf Reifungsmodellen basierenden Stufenleiterverfahren dem mittlerweile vorherrschenden Entwicklungsverständnis nicht gerecht werden (z. B. Stein &

Petermann, 1997). Filipp und Doenges wiesen in diesem Zusammenhang darauf hin, dass „Entwicklungstests immer [nur] so aussagekräftig sind, wie der zugeordnete Entwicklungsbegriff die fraglichen Phänomene angemessen abzubilden in der Lage ist“ (1983, S. 209). Aufgrund der Sichtweise von Entwicklung als lebenslanger Prozess wurde zudem die Forderung erhoben, Entwicklungsdiagnostik auf weitere Altersbereiche auszudehnen. Der erfolgte Paradigmenwechsel zu interaktionistischen Entwicklungstheorien brachte außerdem die Überzeugung mit sich, dass Entwicklung durch die Veränderung von Entwicklungsbedingungen beeinflusst werden kann. Damit wurden neben der Bestimmung des Entwicklungsstandes im Vergleich mit der Altersnorm neue Aufgaben an die Entwicklungsdiagnostik herangetragen (Deimann & Kastner-Koller, 2007). In den Vordergrund rückte nun die Früherkennung individueller Stärken und Schwächen, die in geeigneten Interventionen münden sollten.

Somit begann gegen Ende des 20. Jahrhunderts eine neue Ära der Entwicklungsdiagnostik, die im deutschsprachigen Raum zur Entwicklung von Verfahren wie dem Wiener Entwicklungstest (WET, Kastner-Koller & Deimann, 2002), dem Allgemeinen Entwicklungstest für Kinder von 6 Monaten bis 6 Jahren (ET 6-6, Petermann et al., 2004) und dem Dortmunder Entwicklungsscreening für den Kindergarten (DESK 3-6, Tröster et al., 2004) führte. Mit der bevorstehenden Publikation der Intelligence and Development Scales (IDS, Grob, Meyer & Hagmann-von Arx, zitiert nach Grob et al., 2009) wird zukünftig außerdem ein allgemein orientiertes Entwicklungstestverfahren für fünf- bis zehnjährige Kinder zur Verfügung stehen.

Dennoch sind in der Praxis auch weiterhin reifungsorientierte Entwicklungstests verbreitet. So wurde beispielsweise die überholte zweite Auflage der BSID im Jahr 2007 erstmals deutschsprachig publiziert (BSID-II, Reuner et al., 2007), obwohl mit Bayley-III (Bayley, 2006) bereits eine wesentlich veränderte und verbesserte dritte Auflage vorliegt, die kein Stufenleiterverfahren, sondern eine Testbatterie darstellt. Überdies weisen einige Instrumente, die den Entwicklungsdiagnostika zugeordnet werden (wollen), eine mangelnde testtheoretische Fundierung auf (z. B. DES, Barth, 2008).

3 Methoden der Entwicklungsdiagnostik

Die vorliegende Arbeit fokussiert auf die allgemein orientierte psychometrische Entwicklungsdiagnostik, mit deren Hilfe individuelle Kompetenzen in mehreren Entwicklungsbereichen quantitativ erfasst und mit der Merkmalsverteilung in der Population in Beziehung gesetzt werden. Zu diesem Zweck stehen sowohl Entwicklungsscreenings als auch Entwicklungstests zur Verfügung. Wird in Folge von Entwicklungstestverfahren gesprochen, sind sowohl Screenings als auch Tests gemeint.

Im Zuge des entwicklungspsychologischen Diagnostizierens ist es jedoch nicht ausreichend, lediglich quantitative Daten zu erheben. Es gilt, die Bedürfnisse der Betroffenen und die Entwicklungsbedingungen des in Betracht stehenden Individuums zu erfassen, wofür anamnestische Erhebungen vonnöten sind. Die Beobachtung und Beschreibung des Verhaltens liefert zudem wichtige Informationen, insbesondere hinsichtlich der sozialen und emotionalen Entwicklung. Dies legt nahe, dass erst die Integration verschiedener, multimethodal gewonnener Informationen eine ganzheitliche Betrachtung ermöglicht, welche sodann den Grundstein für eine präzise Diagnose und die Ableitung wertvoller Entwicklungsinterventionen legt.

3.1 Allgemein orientierte Entwicklungsscreenings

Entwicklungsscreenings sind Kurztestverfahren, die mitunter auch von Kinderärzten und Kinderärztinnen sowie Kindergartenpädagogen und Kindergartenpädagoginnen durchgeführt werden. Screenings bezwecken, prognostisch relevante Teilaspekte eines umfassenden Phänomens einfach, schnell und ökonomisch zu erfassen (Melchers et al., 2003).

Screeningverfahren sind defizitorientiert, was impliziert, dass ihre Differenzierungsstärke primär im unteren Bereich liegt: Entwicklungsbezogene Screenings sollen jene zehn Prozent der Klein- und Vorschulkinder erfassen, bei welchen mit Entwicklungsverzögerungen zu rechnen ist (Ettrich, 2000). Ihr vorrangiges Ziel ist die Feststellung von Entwicklungsrisiken zu einem Zeitpunkt, zu welchem noch keine Symptome erkennbar sind. Sie haben somit eine Filterfunktion im Vorfeld einer umfassenderen entwicklungspsychologischen Diagnostik (Deimann & Kastner-Koller, 2007) und unterstützen insbesondere die Früherkennung (Tröster et al., 2004).

Ein Screening zeigt jedoch nur an, ob der Entwicklungsstand hinsichtlich eines zu beurteilenden Merkmals auffällig oder unauffällig ist. Ein festgelegter Leistungsgrenzwert (Cut-off-Wert) beschreibt hierbei die Leistungsgrenze, unter welcher ein Befund als auffällig bewertet wird (Petermann & Macha, 2005).

Die Güte eines Screeningverfahrens wird insbesondere durch dessen Sensitivität und Spezifität ausgedrückt. Während einerseits die Sensitivität ein Maß für die Wahrscheinlichkeit ist, mit der entwicklungsauffällige Kinder als auffällig eingeschätzt werden, ist andererseits aus der Spezifität die Wahrscheinlichkeit abzulesen, mit der altersgemäß entwickelte Kinder als unauffällig eingeschätzt werden (Tröster et al., 2004).

Mit Entwicklungsscreenings soll kostengünstig und effizient die gesamte Risikopopulation erreicht werden und eine Prognose der zukünftigen Entwicklung von Individuen ermöglicht werden. In Populationen mit hohem Risiko einer Entwicklungsverzögerung, beispielsweise bei extremer Frühgeburtlichkeit, spielen Screenings eine untergeordnete Rolle. Zur Beurteilung von Kindern mit bekanntem Risiko sind Entwicklungstests, die Verzögerungen präziser quantifizieren können, von vornherein angebracht (Johnson & Marlow, 2006).

3.2 Allgemein orientierte Entwicklungstests

Deimann und Kastner-Koller (2007) bezeichnen Entwicklungstests als den Königsweg zur Erfassung von Entwicklung: Sofern Verfahren theoretisch und testtheoretisch fundiert sind, stellen sie die bestmögliche diagnostische Annäherung an die kindliche Entwicklung dar.

Breitbandentwicklungstests ermöglichen präzise quantitative Aussagen zu verschiedenen Entwicklungsbereichen und resultieren in einem Entwicklungsprofil. Sie erlauben im Gegensatz zu Screenings neben der Einschätzung von Entwicklungsverzögerungen auch die Darstellung von Entwicklungsvorsprüngen. Hierzu ist jedoch ein erhöhter zeitlicher Testaufwand erforderlich (Petermann & Macha, 2008).

Prinzipiell können nach Petermann und Macha (2003, 2008) drei verschiedene Testformen von Entwicklungstests unterschieden werden:

- Stufenleiterverfahren
- Testbatterien
- Inventare

Stufenleiterverfahren (z. B. Bayley II; Reuner et al., 2007) stellen die älteste Testform dar. Sie orientieren sich an endogenistischen Entwicklungstheorien und sehen Entwicklung demnach als universell an. Infolgedessen werden die Testitems innerhalb umschriebener Entwicklungsbereiche dem Alter nach gereiht, in welchem Kinder einer Normierungsstichprobe die Aufgaben im Mittel lösen konnten. Den einzelnen Aufgaben wird somit ein Entwicklungsalter zugeordnet. Ausgehend von einem leicht unterfordernden Einstiegsniveau werden der Reihe nach die immer schwieriger werdenden Items vorgegeben. Schließlich erfolgt der Testabbruch, wenn die untersuchte Person eine festgelegte Anzahl von Aufgaben nicht mehr lösen kann. Aufgrund des zugrundeliegenden Entwicklungsmodells und der Konstruktionsprinzipien sind Stufenleiterverfahren nur bedingt geeignet, die Variabilität normaler Entwicklung darzustellen und bergen die Gefahr einer Unterschätzung des tatsächlichen Entwicklungsniveaus in sich.

Durch *Testbatterien* (z. B. WET; Kastner-Koller & Deimann, 2002) werden in einer Zusammenstellung homogener Subtests mehrere Entwicklungsdimensionen überprüft. Im Rahmen eines Untertests wird ein eng umschriebener Merkmalsbereich mit gleichartigen Testaufgaben, die sequentiell nach altersspezifischer Schwierigkeit geordnet sind, überprüft. Die Anzahl der gelösten Items ergibt einen Rohwert, welcher in einen Skalenwert transformiert wird und eine kategorielle normorientierte Interpretation (z. B. weit unterdurchschnittlich bis weit überdurchschnittlich) ermöglicht. Testbatterien weisen eine hohe Messgenauigkeit auf und entsprechen den Standards für psychologische Testverfahren.

Inventare (z. B. ET 6-6; Petermann et al., 2004) stellen im Gegensatz zu Testbatterien eine Zusammenstellung von Testaufgaben dar und zielen darauf ab, je Entwicklungsbereich vielfältige qualitative Leistungsaspekte abzubilden. Die Skalen sind somit zwangsläufig heterogen, was zu Einbußen der Reliabilität führt. Streng genommen genügen Inventare deshalb den Standards für psychologische Testverfahren nicht.

3.3 Verhaltensbeobachtung

Die Beobachtung des kindlichen Verhaltens hat einen hohen Stellenwert in der Entwicklungsdiagnostik. Insbesondere bei Kindern, deren Sprachverständnis die Befolgung von Testinstruktionen (noch) nicht erlaubt, stellt sie oft die einzige Möglichkeit dar, einen Eindruck von den kindlichen Kompetenzen zu gewinnen (Kastner-Koller & Deimann, 2009). Zudem gilt sie für eine Reihe diagnostischer Fragestellungen als bester Zugang zu relevanten Informationen, zum Beispiel wenn Interaktionsmuster oder soziales Problemverhalten analysiert werden sollen (La Greca, Kuttler & Stone, 2001). In diesen Fällen entspricht die Verhaltensbeobachtung einem selbständigen Verfahren und wird vorwiegend systematisch und strukturiert durchgeführt.

Allerdings stehen für entwicklungsdiagnostische Fragestellungen kaum standardisierte Beobachtungsinventare zur Verfügung, sodass zumeist individualisierte Wege der Verhaltensbeobachtung beschritten oder Ratingskalen herangezogen werden (Deimann & Kastner-Koller, 2007). Als positives Beispiel kann das DESK 3-6 (Tröster et al., 2004) angeführt werden, welches standardisierte Beobachtungsaufgaben enthält.

Innerhalb des diagnostischen Prozesses wird Verhaltensbeobachtung häufig als begleitendes Verfahren eingesetzt und stellt eine wichtige Ergänzung zur hochstandardisierten testbasierten Diagnostik dar: Ein Testverfahren liefert per se nur die Information ob eine Aufgabe gelöst werden konnte oder nicht, die Art und Weise der Aufgabenbearbeitung wird dadurch nicht abgedeckt (Petermann & Macha, 2008). Durch die Beobachtung in der Testsituation erhält der Untersucher Gelegenheit Arbeitshaltung, Konzentrationsfähigkeit, Frustrationstoleranz und andere nicht kognitive Aspekte des Leistungsverhaltens zu beurteilen (Kastner-Koller & Deimann, 2009).

Auch zur Präzisierung der im diagnostischen Gespräch gewonnen Informationen ist die Verhaltensbeobachtung dienlich, da reales anstatt retrospektiv betrachtetes und verbal bekundetes Verhalten erfasst wird. Somit können Verzerrungen korrigiert werden (Döpfner, Lehmkuhl, Heubrock & Petermann, 2000).

Die Verhaltensbeobachtung ist aber gekennzeichnet durch vielfältige methodische Probleme und birgt nicht wenige Fehlerquellen in sich (zur Problematik von Verhaltensbeobachtung siehe Fassnacht, 1995). Um eine Beobachtung effizient durchzuführen und Beobachtungsdaten korrekt zu interpretieren ist es unerlässlich, sich mit diesen auseinanderzusetzen. Nichtsdestotrotz ist die Verhaltensbeobachtung im Rahmen der

Erstellung einer Entwicklungsdiagnose ein wertvolles Instrument, das insbesondere für die Präzisierung der diagnostischen Beurteilung wertvoll ist (Döpfner et al., 2000).

3.4 Anamnese und Exploration

Kubinger & Deegener (2001) sehen die Anamnese innerhalb des diagnostischen Prozesses als obligatorisches diagnostisches Verfahren an. Nach ihrer Auffassung sind Begriffe wie Exploration, Interview und Biographie dem Anamnesebegriff gleichzusetzen bzw. unterzuordnen und Anamnese somit sehr breit zu verstehen. Die Autoren definieren den Begriff Anamnese als „Sammlung der typischerweise mit dem gegebenen Sachverhalt in Verbindung stehenden Informationen“ (S. 12). Wird in Folge einer der oben genannten Begriffe angesprochen oder der Terminus Diagnostisches Gespräch verwendet, so ist stets die Sammlung der typischerweise mit dem gegebenen Sachverhalt in Verbindung stehenden Informationen gemeint.

Ungleich zur Diagnostik bei Erwachsenen erfolgt die Anamneseerhebung im Rahmen der Entwicklungsdiagnostik üblicherweise sowohl mit dem Klienten – also dem Kind – als auch mit dessen Hauptbezugspersonen.

3.4.1 Befragung von Bezugspersonen

Im Rahmen der entwicklungsdiagnostischen Begutachtung ist die Befragung von relevanten Bezugspersonen gleichermaßen bedeutend wie die Beobachtung des kindlichen Verhaltens und zählt ebenso zum Standard in der entwicklungsdiagnostischen Praxis (Deimann & Kastner-Koller, 2007).

Die wichtigsten Informanten sind die Eltern des zu begutachtenden Kindes (vorausgesetzt das Kind wächst bei diesen auf). Eltern liefern die Informationen zu den Fragen und/oder Problemen, die den Anlass der Begutachtung bilden und können auch Aussagen zur bisherigen Entwicklung und zur familiären Situation treffen, deren Erfassung durch psychometrische Verfahren nicht möglich ist.

Zudem können bedeutsame Informationen von weiteren Verwandten des Kindes (z. B. Großeltern, ältere Geschwister) sowie von Bezugspersonen in anderen sozialen Kontexten (z. B. Krippe, Kindergarten) oder durch Personen bei denen sich das Kind in Behandlung findet (z. B. Kinderarzt bzw. Kinderärztin, Logopäde bzw. Logopädin) erhalten werden (Deegener, 2009).

Einschätzungen des pädagogischen Personals sind besonders dann gefragt, wenn die Probleme des Kindes auch im Kindergarten auftreten oder durch institutionelle Bedingungen möglicherweise beeinflusst werden (Döpfner et al., 2000).

Die anamnestischen Informationen können hierbei in Form von Interviews oder mittels Fragebögen gewonnen werden. Letztere sollten das persönliche Gespräch jedoch nur ergänzen, nicht ersetzen. Sofern der Fragebogen von den Eltern ausgefüllt wurde, ist eine vertiefende nachträgliche Besprechung der so erhaltenen Informationen unumgänglich (Deegener, 2009), denn Fragebögen können biographische Eckdaten zwar zuverlässig abfragen, komplexe Zusammenhänge jedoch häufig nicht abbilden.

Hinsichtlich der erhaltenen Informationen ist zu beachten, dass die Zuverlässigkeit der Beurteilung des Kindes durch dessen Bezugspersonen mitunter fraglich ist. Insbesondere bei Kindern mit Entwicklungsauffälligkeiten ist eine elterliche Überschätzung des tatsächlichen Entwicklungsniveaus gut belegt (Deimann, Kastner-Koller, Benka, Kainz & Schmidt, 2005). Werden beide Elternteile befragt, zeigt sich, dass diese das Gesamtverhalten ihres Kindes ähnlich beurteilen, im Detail jedoch auf unterschiedliche Problembereiche fokussieren (Hay et al., 1999). Im Allgemeinen gelingen Einschätzungen gut, wenn beobachtbares Verhalten beurteilt werden soll oder wenn objektive Vergleichmaßstäbe vorliegen (Deimann & Kastner-Koller, 2007). Dem pädagogischen Kindergartenpersonal scheint es darüber hinaus besser zu gelingen, das Verhalten eines Kindes zutreffend als auffällig oder unauffällig zu klassifizieren (Deimann et al., 2005; Tröster & Reineke, 2006). Zur Vertiefung der Problematik der Zuverlässigkeit bei der Befragung von Bezugspersonen und zu Methoden der Verbesserung derselben wird auf die einschlägige Literatur verwiesen (z. B. Glascoe, 2000; Glascoe & Dworkin, 1995; Rennen-Allhoff, 1991; Rennen-Allhoff, Allhoff, Bowi & Laaser, 1993).

Weitere mögliche Fehlerquellen, welche die Qualität der Anamnese beeinträchtigen, sind in der Sphäre des Interviewers bzw. der Interviewerin begründet: Dessen Bedürfnisse, Interessen, Erfahrungen, Überzeugungen, Vorurteile, impliziten Persönlichkeitstheorien, Erwartungen, gegenwärtigen Befindlichkeiten usw. können dazu führen, die erhaltenen Informationen selektiv wahrzunehmen und verzerrt zu interpretieren (Deegener, 2009). Beobachtungs- und Beurteilungsfehler vermindern die Objektivität der Anamneseerhebung, insbesondere bei sehr intuitiv geführten Gesprächen. Da somit davon ausgegangen werden muss, dass unterschiedliche Personen zu unterschiedlichen

Erkenntnissen gelangen können, wird als Konsequenz auch die Validität reduziert (Kubinger & Deegener, 2001).

3.4.2 Befragung des Kindes

Im Sinne eines multimethodalen Ansatzes kommt der Befragung des zu begutachtenden Kindes hohe Bedeutung zu. Die kindliche Einschätzung hinsichtlich Erziehungsumwelt und Interaktion mit Erziehungspersonen unterscheidet sich nämlich häufig von jener der Erziehungspersonen. Zudem bietet der diagnostische Zugang über die Kinderperspektive einen stärkeren und unmittelbaren Bezug zu den Entwicklungs-, Leistungs- und Persönlichkeitsmerkmalen des Kindes (Grundmann, Welskopf und Sturzbecher, 2001).

Insbesondere bei Kleinkindern sind Interviews jedoch nur eingeschränkt möglich, da sowohl Sprachverständnis und Verbalisierungsfähigkeit nicht vollständig ausgeprägt sind als auch das autobiographische Gedächtnis beschränkt ist. Wenn spezifische Fragen gestellt werden, sind Kinder im Allgemeinen ab einem Alter von drei Jahren einer Befragung zugänglich. Sie sind jedoch noch nicht in der Lage, Erfahrungen und Ereignisse strukturiert und detailliert wiederzugeben (Roebbers & Elischberger, 2002; Tang, 2006). Zudem erinnern sich Klein- und Vorschulkinder schlechter an Einzelereignisse und neigen auch mehr zu Verallgemeinerungen in Skriptform als ältere Kinder. Weiters fällt es ihnen noch schwer, sich von einem typischen Handlungsverlauf abzuwenden und das spezifische einer Situation wiederzugeben (Sturzbecher & Grundmann, 2001).

Überdies sind Klein- und Vorschulkinder besonders anfällig für Suggestionen. Volbert und Pieters (1996) begründen dies mit entwicklungstypischen Besonderheiten dieser Altersgruppe: Neben Defiziten in der Informationsverarbeitung zeigen sie eine höhere Vergessensrate sowie eine noch nicht vollständig ausgeprägte Fähigkeit zur Perspektivenübernahme. Auch reflektieren jüngere Kinder in geringerem Maße die Plausibilität dargebotener Informationen und sind somit eher bereit, suggerierte Informationen zu akzeptieren.

Für die Befragung von Kindern ist umfangreiches entwicklungspsychologisches Wissen vonnöten. Dieses ist zum einen für eine adäquate Gesprächsführung unabdingbar, zum anderen sind die kindlichen Äußerungen auch vor diesem Hintergrund zu beurteilen. Der Interviewer hat jedenfalls seine Sprache dem kindlichen Niveau anzupassen und

suggestive Fragen zu vermeiden. Leitlinien und Tipps für das diagnostische Gespräch mit Klein- und Vorschulkindern können den Ausführungen von Kastner-Koller & Deimann (2009), Roebbers und Elischberger (2002), Sturzbecher und Grundmann (2001) sowie Tang (2006) entnommen werden.

Während die Bezugspersonen des Kindes differenzierte Informationen zum Begutachtungsanlass und zur bisherigen Entwicklung liefern, konzentriert sich das Gespräch mit dem Kind auf seine unmittelbare Umwelt. Kastner-Koller und Deimann (2009) nennen folgende Themen, die für das diagnostische Gespräch mit Klein- und Vorschulkindern geeignet sind:

- Familie (Familienmitglieder, Wohnsituation, Alltag, Beziehung zu Familienmitgliedern)
- Kindergarten (Subjektives Wohlbefinden, Beziehung zum Kindergartenpersonal, Freundschaften, Aktivitäten)
- Interessen (Bevorzugte Spiel- und Freizeitaktivitäten, Medienkonsum)
- Emotionale Befindlichkeit (Wünsche, Sorgen, Ängste, Körperliche Beschwerden)

Auf die psychometrischen Methoden der Allgemeinen Entwicklungsdiagnostik wird im Weiteren noch detaillierter eingegangen. So werden im folgenden Kapitel die Anforderungen, denen Entwicklungstestverfahren zu genügen haben, umfassend betrachtet. Im Anschluss daran werden jene allgemein orientierte Entwicklungsscreenings und -tests, die im deutschen Sprachraum verbreitet Anwendung finden, einer kritischen Würdigung unterzogen.

Es sei an dieser Stelle jedoch nochmals darauf hingewiesen, dass erst die Integration von Informationen, die aus verschiedenen Quellen stammen, eine valide Standortbestimmung der Entwicklung eines Menschen ermöglicht.

4 Qualitätsanforderungen an Entwicklungstestverfahren

Wie bereits oben ausgeführt ist die Entwicklungsdiagnostik Teilgebiet der psychologischen Diagnostik. Folglich haben entwicklungsdiagnostische Verfahren den Anforderungen an psychologische Testverfahren zu genügen. Somit werden in diesem Kapitel die Qualitätsanforderungen der psychologischen Diagnostik beschrieben und in Hinblick auf Entwicklungstestverfahren kommentiert.

Ein psychologischer Test gilt dann als wissenschaftlich fundiert, wenn er hinsichtlich der Erfüllung der Testgütekriterien empirisch überprüft wurde (Moosbrugger & Kelava, 2007).

Die Testgütekriterien dienen der Qualitätsbeurteilung psychologischer Tests. Üblicherweise werden folgende zehn Kriterien unterschieden (Kubinger, 2006; Moosbrugger & Kelava, 2007), wobei die ersten drei unter dem Begriff Hauptgütekriterien subsumiert werden (Lienert & Raatz, 1998):

1. Objektivität
2. Reliabilität
3. Validität
4. Normierung
5. Skalierung
6. Ökonomie
7. Nützlichkeit
8. Zumutbarkeit
9. Unverfälschbarkeit
10. Fairness

Im Manual eines psychologisch-diagnostischen Testverfahrens ist darüber zu informieren, welche Testgütekriterien in welcher Weise erfüllt sind (Moosbrugger & Kelava, 2007).

4.1 Objektivität

Ein Test ist objektiv wenn die durch ihn gewonnen Ergebnisse vom Testleiter, vom Testauswerter und von der Ergebnisinterpretation unabhängig sind, den Testdurchführenden also kein Verhaltensspielraum bei der Durchführung, Auswertung und Interpretation eingeräumt wird (Moosbrugger & Kelava, 2007). Ein hoher Grad an *Durchführungsobjektivität* (Testleiterunabhängigkeit) ist zu erwarten wenn standardisiertes Testmaterial vorhanden ist und das Testmanual Angaben zur Standardisierung der Untersuchungssituation enthält. Eine hohe *Auswertungsobjektivität* wird erreicht wenn standardisierte Auswertungsanleitungen exakt beschreiben, wie Testleistungen zu numerischen Testwerten zu verrechnen sind. Das Ausmaß an Verrechnungssicherheit kann quantifiziert werden indem der Grad an Auswerterübereinstimmung berechnet wird. *Interpretationsobjektivität* ist gegeben, wenn verschiedene Testanwender bei Personen mit demselben Testwert zu denselben Schlussfolgerungen gelangen (Kubinger, 2006; Moosbrugger & Kelava, 2007).

Petermann und Macha (2005) merken hinsichtlich der Interpretationsobjektivität von Entwicklungstestergebnissen an, dass der Leistungsstatus, den ein Kind erbringt, zwar objektiv einzustufen ist, die entwicklungsbezogene Interpretation jedoch häufig nur durch Integration von Informationen aus Anamnese und Verhaltensbeobachtung möglich wird. Dies führt mitunter dazu, dass dasselbe Ergebnis bei verschiedenen Kindern gänzlich anders beurteilt werden muss.

4.2 Reliabilität

Ein Test ist reliabel (zuverlässig), wenn er ein bestimmtes Merkmal exakt, das heißt ohne Messfehler misst. Formal ist die Reliabilität als Anteil der Varianz der wahren Werte an der Gesamtvarianz der beobachteten Testwerte definiert. Das Ausmaß an Messgenauigkeit wird durch einen Reliabilitätskoeffizienten bestimmt, welcher angibt, in welchem Maße ein Testergebnis bei Testwiederholung reproduzierbar ist. Der Reliabilitätskoeffizient kann Werte zwischen Null und Eins annehmen, sollte jedoch einen Wert von 0.7 nicht unterschreiten (Moosbrugger & Kelava, 2007). Ein Wert von 0.9 oder darüber weist auf hohe Messgenauigkeit hin.

Die Reliabilität eines psychologischen Tests kann mittels verschiedener Methoden überprüft werden:

- Bei Anwendung der Paralleltest-Methode werden einer Stichprobe von Testpersonen zwei äquivalente Parallelförmungen eines Tests vorgegeben und deren Ergebnisse korreliert. Diese Methode wird häufig als *via regia* der Reliabilitätsbestimmung bezeichnet (Moosbrugger & Kelava 2007). Allerdings scheitert die Anwendung dieser Methode häufig an der Möglichkeit, tatsächlich indikationsidentische Items zu konstruieren (Ettrich, 2000), weshalb für Entwicklungstests im Allgemeinen keine Parallelversionen vorliegen (Petermann & Macha, 2008). Kubinger (2006) merkt in Hinblick auf die *Paralleltest-Reliabilität* zudem kritisch an, dass die Korrelation zwischen Test und Paralleltest mehr ein Kriterium dafür ist, wie gut die Konstruktion des Paralleltests gelungen ist und weniger die Zuverlässigkeit der Messungen widerspiegelt.
- Im Zuge der Testwiederholungsmethode (*Retest-Reliabilität*) wird derselbe Test derselben Stichprobe wiederholt vorgegeben und die Korrelation der Rohwertpaare ermittelt. Hierbei gilt, dass zwischen den beiden Testzeitpunkten ein hinreichend großer zeitlicher Abstand bestehen muss, um Übungs- und Erinnerungseffekte, die aus der ersten Testung resultieren, möglichst gering zu halten. Die Anwendung dieser Methode auf Entwicklungstests erscheint wenig sinnvoll, da sie mit den Grundannahmen von Entwicklungstheorien nicht in Übereinstimmung zu bringen ist, welche eine grundsätzliche Veränderung über die Zeit postulieren (Ettrich, 2000). So kann auch nicht überprüft werden, inwieweit Leistungsverbesserungen auf Entwicklungsfortschritten oder Lerneffekten beruhen (Petermann & Macha, 2003). Der resultierende Reliabilitätskoeffizient spiegelt vermutlich eher tatsächliche Veränderungen als Zuverlässigkeitsmängel wieder.
- Bei der Testhalbierungsmethode wird ein Test in zwei gleichwertige Hälften geteilt, (bei jeder getesteten Person) das Testergebnis jeder der beiden Testhälften jeweils separat ermittelt und hierauf die Ergebnisse der beiden Hälften korreliert. Die resultierende Korrelation entspricht der Paralleltest-Reliabilität für einen Test von halber Länge und ist somit rechnerisch aufzuwerten. Üblicherweise geschieht dies durch Anwendung der Spearman-Brown-Formel (Kubinger, 2006). Voraussetzung zur Bestimmung der *Split-Half-Reliabilität* sind homogene

Niveautests (Lienert & Raatz, 1998). Da allgemeine Entwicklungstests mehrere Entwicklungsbereiche erfassen, erfordert die Anwendung dieser Methode Verfahren mit homogenen Skalen.

- Die Konsistenzanalyse gründet auf der Annahme, dass man einen Test in ebenso viele Teile zerlegen kann wie Items vorhanden sind. Die Standardmethode zur Schätzung der inneren Konsistenz ist die Berechnung des Chronbach- α -Koeffizienten (Bühner, 2006). Die Konsistenzanalyse ist eine Verallgemeinerung der Split-Half-Methode und unterliegt somit denselben Voraussetzungen.

4.3 Validität

Ein Test ist valide (gültig), wenn er das zu messende Merkmal auch tatsächlich misst (Moosbrugger & Kelava, 2007). Die Validität ist das wichtigste Gütekriterium und der Objektivität und Reliabilität übergeordnet. Schließlich ist es belanglos ob ein Test objektiv und reliabel ist wenn er sich als „ungültig“ erweisen sollte. Gleichzeitig ist die Validität auch das am schwierigsten zu prüfende Gütekriterium (Hartig, Frey & Jude, 2007). Sofern ein Test hohe Validität aufweist, erlauben die hiermit erzielten Ergebnisse die Generalisierung des in der Testsituation beobachteten Verhaltens auf das korrespondierende Verhalten außerhalb der Testsituation (Moosbrugger & Kelava, 2007).

Um ein differenziertes Bild der Gültigkeit eines Tests zu erhalten gilt es verschiedene Aspekte der Validität zu untersuchen: Inhaltsvalidität, Kriteriumsvalidität und Konstruktvalidität.

4.3.1 Inhaltsvalidität

Inhaltliche Validität ist gegeben wenn ein Test bzw. Item das interessierende Merkmal auch tatsächlich repräsentativ erfasst, dieser selbst also das optimale Kriterium für das Merkmal darstellt (Lienert & Raatz, 1998). Sie wird in der Regel nicht numerisch anhand eines Kennwertes bestimmt, sondern theoretisch-argumentativ abgeleitet und durch Expertenurteile abgesichert (Hartig, Frey & Jude, 2007).

Der Begriff der Inhaltsvalidität wird kurioserweise oft mit dem der „Augenscheinvalidität“ verwechselt. Unter letzterem Begriff ist jedoch zu verstehen, dass selbst einem Laien erkennbar ist, was die Messintention eines Verfahrens ist (Kubinger, 2006).

4.3.2 Kriteriumsvalidität

Dieser Validitätsaspekt bezieht sich auf die praktische Anwendbarkeit eines Tests zur Vorhersage von Verhalten und Erleben. Die Überprüfung erfolgt durch Bestimmung der Korrelation zwischen einem Testkennwert und einer als geeignet angesehenen Kriteriumsvariablen. Abhängig davon ob das Außenkriterium bereits in der Gegenwart vorliegt oder erst in der Zukunft vorliegen wird, unterscheidet man zwischen *Übereinstimmungsvalidität* (konkurrenter Validität) und *prognostischer Validität* (Moosbrugger & Kelava, 2007).

Die Übereinstimmungsvalidität kann grundsätzlich für alle Entwicklungstests überprüft werden. Als Außenkriterien bieten sich insbesondere Fremdeinschätzung (beispielsweise durch Eltern und/oder pädagogisches Personal) sowie konstruktnahe oder konstruktferne Testverfahren an (Macha, Proske & Petermann, 2005).

Wie bereits erwähnt zielen Entwicklungstests darauf ab, Entwicklungsprognosen zu erstellen (Johnson & Marlow, 2006). Allerdings ist die prognostische Validität von Entwicklungstests für gewöhnlich eher gering einzuschätzen. Lediglich für klinische Gruppen oder bei deutlichen Entwicklungsabweichungen können genauere Vorhersagen getroffen werden. Zur Verbesserung prognostischer Aussagen empfehlen Johnson und Marlow den Miteinbezug von Umgebungs- und Sozialfaktoren. Dies verdeutlicht die Wichtigkeit von Anamnese und Verhaltensbeobachtung im Zuge des diagnostischen Prozesses, durch welche jene Informationen überhaupt erst gewonnen werden können.

4.3.3 Konstruktvalidität

„Konstruktvalidität umfasst die empirischen Befunde und Argumente, mit denen die Zuverlässigkeit der Interpretation von Testergebnissen im Sinne erklärender Konzepte, die sowohl die Testergebnisse selbst, als auch die Zusammenhänge der Testwerte mit anderen Variablen erklären, gestützt wird“ (Messick, 1995, zitiert nach Hartig, Frey & Jude, 2007, S. 145). Ein Test weist Konstruktvalidität auf, wenn er das theoretische Konstrukt², das er zu erfassen beansprucht, auch tatsächlich erfasst, das Testergebnis also vor dem Hintergrund eines theoretisch definierten Konstrukts interpretiert werden kann. Dieser Validitätsaspekt gilt heute als Kern der psychologischen Validitätstheorie und

² Ein Konstrukt ist ein nicht direkt beobachtbares (latentes) Merkmal, welches nicht vollständig operational definierbar ist, wie z. B. Intelligenz oder Angst.

schließt als übergeordneter Zugang zur Bestimmung der Gültigkeit eines Verfahrens auch die anderen Validitätsarten mit ein (Hartig, Frey & Jude, 2007).

Im Zuge der Bestimmung der Konstruktvalidität werden theoriegeleitete Annahmen über Zusammenhangsstrukturen theoretischer Konstrukte empirisch überprüft. Ausgehend vom Konstrukt und der entsprechenden Theorie werden Hypothesen abgeleitet, die unter Anwendung des zu validierenden Tests überprüft werden. Hierfür stehen sowohl empirisch-korrelative, experimentelle als auch logisch-analytische Methoden zur Verfügung (Lienert & Raatz, 1998):

- Korrelation mit mehreren Außenkriterien (Kriteriumsvalidität)
- Korrelation des Tests mit konstruktnahen Tests (konvergente Validität)
- Korrelation des Tests mit konstruktfernen Tests (divergente bzw. diskriminante Validität)
- Faktorenanalyse des zu validierenden Tests gemeinsam mit Außenkriterien, konstruktnahen und/oder konstruktfernen Tests (faktorielle Validität)
- Nachweis von Testwertunterschieden zwischen Gruppen, deren Bildung aufgrund theoretischer Annahmen zu erfolgen hat (Differenzierungsfähigkeit)
- Analyse intraindividuelle Unterschiede bei wiederholter Vorgabe des Tests, wobei die Durchführungsbedingungen systematisch variiert werden können
- Inhaltlich-logische Analyse der Testelemente (Bestimmung von Interkorrelationen, Itemschwierigkeiten und innerer Konsistenz)

Die vorgenannten Methoden versuchen das Konstrukt, das dem zu validierenden Test zugrunde liegt, von verschiedenen Seiten zu erfassen und ergänzen einander somit. Die Gesamtheit der erhaltenen Resultate aus verschiedenen Analysen ergibt die Konstruktvalidität eines Tests, wobei allerdings kein exaktes Maß hierfür bestimmt werden kann (Lienert & Raatz, 1998). Bühner (2006) weist in diesem Zusammenhang darauf hin, dass jeder Testanwender die Aussage, dass Konstruktvalidität vorliegt, selbst durch einen kritischen Blick auf die ermittelten Kennzahlen prüfen sollte.

Die Konstruktvalidität ist aufgrund der unterschiedlichen Konstruktionsmerkmale (Testform, Skaleneigenschaften) nicht für alle Entwicklungstestverfahren gleichermaßen überprüfbar:

Die theoretische Vorannahme steigender Testleistungen mit zunehmendem Alter wird von allen gängigen Entwicklungstestverfahren erfüllt. Dies zeigt sich in der Regel schon in den publizierten Normen. Auch postulieren Entwicklungstestverfahren altersabhängige Aufgabenschwierigkeiten, die in den Testmanualen üblicherweise auch nachgewiesen werden (Petermann & Macha, 2008). Der überwiegende Teil der im deutschen Sprachraum verbreiteten Verfahren liefert zudem Angaben zu konvergenter und diskriminanter Validität (z. B. Kastner-Koller & Deimann, 2002; Melchers et al., 2003; Petermann et al., 2004; Tröster et al., 2004). Da Entwicklungstestverfahren mehrere Entwicklungsbereiche erfassen, kann die angenommene Dimensionsstruktur prinzipiell durch Berechnung einer Faktorenanalyse überprüft werden. Von den im deutschen Sprachraum verbreiteten allgemeinen Entwicklungstests wurde lediglich für den WET (Kastner-Koller & Deimann) eine Faktorenanalyse berechnet, die die festgelegten Funktionsbereiche im Wesentlichen reproduzieren konnte. Die Berechnung von Interkorrelationen wird in den Manualen der meisten Verfahren angeführt (z. B. Kastner-Koller & Deimann, 2002; Melchers et al., 2003; Petermann et al., 2004; Reuner et al., 2007).

Ein wichtiger Aspekt für die Einschätzung der Validität eines Entwicklungstestverfahrens ist dessen Differenzierungsfähigkeit. Bei Screeningverfahren, die einer ersten Auslese von Risikokindern dienen, ist die Unterscheidung zwischen entwicklungsgefährdeten und unauffälligen Kindern ein zentrales Ziel. Zur Beurteilung der Differenzierungsfähigkeit von Screenings haben sich die Güteindizes Sensitivität und Spezifität durchgesetzt:

- Die *Sensitivität* eines Screenings gilt als Maß für die Fähigkeit des Verfahrens, Kinder mit Entwicklungsgefährdung zu identifizieren. Sie gibt innerhalb der Gesamtheit aller Risikokinder die Höhe des Prozentsatzes jener Kinder an, die richtig-positiv diagnostiziert wurden bzw. einen auffälligen Befund aufweisen.
- *Spezifität* wiederum bezeichnet innerhalb der Gesamtheit aller altersmäßig entwickelten Kinder die Höhe des Anteils der im Screening unauffälligen Kinder und ist somit ein Maß dafür, ob unauffällige Kinder auch richtig-negativ befundet wurden.

Bei der Einschätzung der Differenzierungsfähigkeit eines Screeninginstruments ist zu bedenken, dass zwischen beiden Gütekriterien eine Wechselwirkung besteht, hohe Sensitivität also zu Lasten der Spezifität geht und umgekehrt. Nach Aylward (2004) sind

in der Entwicklungsdiagnostik Sensitivitätswerte von 70% und Spezifitätsraten von 70% bis 80% durchaus realistisch. Die Überprüfung dieser Güteindizes kann durch Gruppenvergleiche, die signifikante Unterschiede aufzeigen sollen, oder durch die Identifikation von Risikokindern, die anhand von zuverlässigen Außenkriterien identifiziert wurden, erzielt werden (Macha et al., 2005). Von den verbreitet Anwendung findenden deutschsprachigen Entwicklungsscreenings, welche im folgenden Kapitel näher betrachtet werden, legen allerdings lediglich die EVU (Melchers et al., 2003) und das DESK 3-6 (Tröster et al., 2004) Angaben zur Differenzierungsfähigkeit vor.

Auch für Entwicklungstests gilt, dass sie zwischen entwicklungsverzögerten und unauffälligen Kindern zu unterscheiden haben. Sie müssen jedoch darüber hinaus Differenzierungsleistungen über den gesamten möglichen Leistungsbereich, beispielsweise anhand der Klassifikation weit unterdurchschnittlich bis weit überdurchschnittlich, erbringen. Angaben zu Testwertunterschieden von unauffälligen Kindern im Vergleich zu Kindern aus Risikogruppen (z. B. Frühgeburtlichkeit, Autismus, Down-Syndrom) finden sich somit in den Manualen der meisten Entwicklungstests (Macha et al., 2005).

4.4 Normierung

Unter Normierung (Eichung) eines Tests versteht man das Erstellen eines Bezugssystems, mit dessen Hilfe individuelle Testergebnisse mit den Merkmalsausprägungen von Personen der Eichstichprobe verglichen werden können (Moosbrugger & Kelava, 2007). Das Gütekriterium der Normierung ist erfüllt, wenn die Population, für die die Normen gelten, definiert ist, die Eichstichprobe für diese Zielpopulation repräsentativ ist und die Normen nicht veraltet sind (Kubinger, 2006). Darüber hinaus ist die Qualität von Normen abhängig von angemessenen Stichprobenumfängen, wobei nach Bühner (2006) eine Zahl von 300 Personen als Daumenregel für die Mindestgröße von Normstichproben gelten kann.

In Zusammenhang mit der Forderung nach aktuellen Normen ist auf den sogenannten Flynn-Effekt zu verweisen, der auf einer Studie von Flynn (1987) basiert: Diese Studie dokumentiert einen durchschnittlichen Anstieg der Intelligenzleistungen um etwa fünf IQ-Punkte während eines Jahrzehnts. Die EPICure Studie (Marlow et al., 2005, zitiert nach Johnson & Marlow, 2006) belegt den Flynn-Effekt fast zwei Jahrzehnte später. Dies

impliziert, dass veraltete Testnormen das kognitive Entwicklungsniveau überschätzen, wodurch Kinder mit Defiziten mitunter als unauffällig beurteilt werden. Hinsichtlich der Entwicklung von Sprache und Motorik werden gegenläufige Kohorteneffekte vermutet, sodass auch in diesen Bereichen veraltete Testnormen zu spürbaren Verzerrungen in der Beurteilung von Testleistungen führen (Petermann & Macha, 2005).

Der überwiegende Teil einschlägiger Entwicklungstestverfahren liefert Normtabellen als Bezugssystem für die individuellen Testergebnisse. Von jenen Verfahren, die im folgenden Kapitel einer kritischen Würdigung unterzogen werden, wurde nur für die DES (Barth, 2005) keine Normierung vorgenommen. Es ist auch anzumerken, dass die deutsche Fassung von Bayley II (Reuner et al., 2007) nur auf veralteten US-Normen beruht. Das Vorhandensein von Normen sowie deren Aktualität ist jedoch Voraussetzung für die korrekte Interpretation von Testergebnissen.

4.5 Skalierung

Ein Test erfüllt dieses Gütekriterium, wenn die aufgrund der jeweiligen Verrechnungsvorschrift ermittelten Testwerte die empirischen Merkmalsrelationen adäquat abbilden, die Relation der Merkmalsausprägung sich also auch in den Testwerten widerspiegelt (Moosbrugger & Kelava, 2007). Das Gütekriterium der Skalierung betrifft somit die Angemessenheit der in einem Testmanual festgesetzten Regeln, die festlegen, wie die einzelnen Testergebnisse zu einem numerischen Testwert zu verrechnen sind (Kubinger, 2006).

Die Umsetzbarkeit dieses Gütekriteriums ist abhängig vom Skalenniveau des Messinstruments und setzt mindestens Ordinalskalenniveau voraus. Intervallskalenniveau erlaubt neben der Bildung einer Rangreihe eine Beurteilung der Größe inter- und intraindividuelle Differenzen (Moosbrugger & Kelava, 2007).

Nach Kubinger (2006) sind die Methoden der Klassischen Testtheorie ungeeignet zur Prüfung des Gütekriteriums der Skalierung: Konsistenzanalysen, Faktorenanalysen und Interkorrelation erlauben lediglich eine Annäherung an die Erfüllung dieses Gütekriteriums. Ob die Verrechnungsregeln eines Tests zu Testwerten führen, die verhaltensadäquate Relationen wiedergeben, kann lediglich im Rahmen der probabilistischen Testtheorie empirisch überprüft werden.

Die gängigen Entwicklungstestverfahren beruhen überwiegend auf der klassischen Testtheorie. Von den im deutschen Sprachraum verbreiteten allgemeinen Entwicklungstests wurde lediglich der Wiener Entwicklungstest (Kastner-Koller & Deimann, 2002) auf Grundlage des probabilistischen Testmodells entwickelt, darüber hinaus auch noch klassisch testanalytisch.

4.6 Ökonomie

Ein Test ist ökonomisch, wenn er, gemessen am diagnostischen Erkenntnisgewinn, relativ wenig Ressourcen beansprucht. Die Ökonomie bezieht sich somit auf die Wirtschaftlichkeit des Verfahrens und wird durch den finanziellen und zeitlichen Aufwand bestimmt, der für eine Testung vonnöten ist. Die Aufwandsminimierung darf jedoch keinesfalls zu Lasten der anderen Gütekriterien gehen (Moosbrugger & Kelava, 2007). Für Screeningverfahren ist die Erfüllung dieses Gütekriteriums eine entscheidende Voraussetzung für deren Akzeptanz (Tröster et al., 2004).

4.7 Nützlichkeit

Ein Test erfüllt das Gütekriterium der Nützlichkeit, wenn für das von ihm gemessene Merkmal praktische Relevanz besteht und die auf seiner Grundlage getroffenen psychologischen Entscheidungen und/oder Maßnahmen mehr Nutzen als Schaden erwarten lassen. Praktisch bedeutsam ist ein Test, wenn er erstens ein Merkmal misst, das im Sinne der Kriteriumsvalidität nützliche Anwendungsmöglichkeiten aufweist und zweitens kein anderes Verfahren vorliegt, das denselben Merkmalsanspruch hat und die übrigen Gütekriterien mindestens genauso gut erfüllt (Moosbrugger & Kelava, 2007).

4.8 Zumutbarkeit

Ein Test ist zumutbar, wenn er die zu testende Person absolut und relativ zu dem aus seiner Anwendung resultierenden Nutzen in zeitlicher, psychischer sowie physischer Hinsicht nicht über Gebühr belastet (Moosbrugger & Kelava, 2007).

Das Gütekriterium der Zumutbarkeit ist von besonderer Bedeutung für die Entwicklungsdiagnostik. Klein- und Vorschulkinder stoßen schneller an ihre physischen und psychischen Grenzen als Erwachsene. Diesem Umstand ist durch die kindgerechte

Gestaltung der Untersuchungssituation und entsprechendes Testleiterverhalten Rechnung zu tragen.

4.9 Unverfälschbarkeit

Ein Test gilt als unverfälschbar, wenn die Testperson ihre Testwerte nicht oder lediglich unwesentlich nach eigenem Belieben beeinflussen bzw. verzerren kann (Kubinger, 2006).

Das Gütekriterium der Unverfälschbarkeit spielt in der Entwicklungsdiagnostik keine Rolle, da Klein- und Vorschulkinder nicht in der Lage sind, die konkreten Ausprägungen ihrer Testwerte bewusst zu steuern.

4.10 Fairness

Ein Test erfüllt das Gütekriterium der Fairness, wenn das Testergebnis einer Person zu keiner systematischen Benachteiligung aufgrund ihrer ethnischen, kulturellen, sozioökonomischen oder geschlechtlichen Gruppenzugehörigkeit führt. Fairness bezieht sich dabei insbesondere auf Aspekte, die unmittelbar mit den Testinhalten und der Testdurchführung verbunden sind. Unfairness in Form eines Itembias liegt vor, wenn Items bestimmte Personen aufgrund ihrer Gruppenzugehörigkeit benachteiligen. Ein Ergebnisbias ist zu erwarten, wenn Sprachprobleme oder Schwierigkeiten bei der technischen Handhabung zur Benachteiligung bestimmter Personen führen. Die Frage der Fairness betrifft zudem auch die Normierung eines Tests und die Interpretation des Testergebnisses (Moosbrugger & Kelava, 2007).

4.11 Zur Auswahl und Durchführung eines Entwicklungstestverfahrens

Bei der Auswahl eines Verfahrens für eine entwicklungsdiagnostische Fragestellung sollte dessen theoretische und testtheoretische Basis stets einer kritischen Bewertung unterzogen werden. Hierbei sind folgende Punkte zu überprüfen:

- Entwicklungstheorie im Hintergrund
- Testmodell sowie Skalen- und Itemkonstruktion
- Durchführungs-, Auswertungs- und Interpretationsobjektivität

- Angaben zur Reliabilität
- Dokumentation von Validierungsstudien
- Darstellung der Normierung, insbesondere in Bezug auf deren Aktualität
- Ausprägung der übrigen relevanten Testgütekriterien

Einer derartigen Überprüfung haben, unabhängig von den unterschiedlichen Zielstellungen, sowohl Entwicklungstests als auch Screenings standzuhalten. Screenings haben zum Ziel, die Gesamtpopulation zu erreichen und klassifizieren den Entwicklungsstand in auffällig oder unauffällig. Allgemeine Entwicklungstests werden auf spezifische Fragestellungen hin angewandt und liefern detailliertere Informationen. Einem auffälligen Screeningbefund hat prinzipiell eine umfassende entwicklungsdiagnostische Untersuchung zu folgen, die auch stets Anamnese und Verhaltensbeobachtung beinhalten sollte.

Die Aussagekraft einer entwicklungspsychologischen Begutachtung ist jedoch nicht nur abhängig von den angewandten Methoden, deren korrekter Handhabung und Güte. Unabhängig von der methodischen Herangehensweise ist bei der Diagnostik kleiner Kinder auch noch zu bedenken, dass deren Mitmachbereitschaft und Ausdauer sehr stark durch Außenreize beeinflusst werden. Diesem Umstand kann durch die Gestaltung der Untersuchungssituation Rechnung getragen werden. Dementsprechend werden abschließend einige Grundregeln formuliert:

- Die Untersuchung hat am Vormittag zu erfolgen, da kleine Kinder dann für gewöhnlich ausgeruht und somit am leistungstärksten sind.
- Der Untersuchungsraum ist kindgerecht zu gestalten.
- Der eigentlichen Untersuchung ist eine Aufwärmphase vorzuschalten, die einen Kontaktaufbau ermöglicht.
- Die Untersuchung ist spielerisch zu gestalten.
- Die Durchführungszeit ist so kurz wie möglich zu halten, Pausen sind jedoch nach Bedarf zu gewähren.

Tipps für eine kindgerechte Diagnostik lassen sich auch den Manualen von Entwicklungstestverfahren entnehmen (z.B. Kastner-Koller & Deimann, 2002; Petermann et al., 2004; Reuner et al., 2007).

5 Ausgewählte deutschsprachige Entwicklungstestverfahren: Darstellung und kritische Würdigung

Dieses Kapitel ist der Beschreibung und anschließender kritischer Würdigung aktueller deutschsprachiger Entwicklungsscreenings und -tests gewidmet. Alle Verfahren sind für den Einsatz in der frühen Kindheit vorgesehen, unterscheiden sich jedoch hinsichtlich der erfassten Altersspanne. Aufgrund ihrer theoretischen und testtheoretischen Fundierung sind die Verfahren zudem höchst unterschiedlich zu bewerten, wodurch die Bandbreite der Qualität jener Entwicklungstestverfahren, die verbreitet Anwendung finden, verdeutlicht wird.

5.1 Entwicklungsscreenings

5.1.1 Dortmunder Entwicklungsscreening für den Kindergarten (DESK 3-6)

Das Dortmunder Entwicklungsscreening für den Kindergarten (DESK 3-6; Tröster et al., 2004) ist ein standardisiertes Verfahren zur Früherkennung von Entwicklungsgefährdungen bei Kindern im Kindergartenalter. Dabei ist nicht die Ermittlung des Entwicklungsstandes Ziel des DESK 3-6, sondern das Finden nach ersten Anhaltspunkten, die einen Verdacht auf Entwicklungsgefährdung begründen, sodass betroffene Kinder einer umfassenden Entwicklungsdiagnostik zugeführt werden können und gegebenenfalls eine notwendige Entwicklungsintervention erhalten.

Das Verfahren entstand aufgrund der Unzufriedenheit der Autoren und der Autorin mit den in Deutschland gesetzlich verankerten Früherkennungsuntersuchungen (U1 - U9). Studien belegen, dass im Rahmen dieser Untersuchungen Entwicklungs- und Verhaltensauffälligkeiten nur unzureichend erkannt werden, wofür vor allem das Fehlen standardisierter Verfahren verantwortlich sein dürfte. Zudem sinkt im Vergleich zu den ersten Vorsorgeuntersuchungen die Inanspruchnahme der letzten beiden vorgesehenen Untersuchungstermine, insbesondere bei Familien aus niedrigen sozialen Schichten. Entwicklungsverzögerungen werden somit häufig erst bei den Einschulungsuntersuchungen erkannt, so dass der optimale Zeitpunkt für eine Frühförderung bereits verpasst ist. Tröster et al. gehen davon aus, dass der Kindergarten bessere Voraussetzungen zur zuverlässigen Früherkennung von Entwicklungsverzögerungen bietet als die pädiatrische Praxis. Zum einen besuchen fast alle Kinder einen

Kindergarten, sodass gute Erreichbarkeit gegeben ist. Zum anderen können die Pädagogen und Pädagoginnen das Kind langfristig in verschiedenen Anforderungssituationen sowie in Interaktion mit Peers beobachten.

Die Zusammenstellung der Entwicklungsaufgaben erfolgte in mehreren Schritten. Auf Grundlage der Sichtung gängiger Entwicklungstestverfahren wurden Entwicklungsaufgaben für drei Altersgruppen (Dreijährige, Vierjährige, Fünf- bis Sechsjährige) zusammengestellt, die in der jeweiligen Altersstufe eine Lösungsfrequenz von 70 bis 90 Prozent aufwiesen. Anschließend wurden die Durchführungshinweise und Beurteilungskriterien der ausgewählten Aufgaben an die Bedingungen im Kindergarten angepasst. Der Itempool wurde dem pädagogischen Kindergartenpersonal zur Beurteilung der Verständlichkeit und Praktikabilität vorgelegt. Die als geeignet angesehenen Aufgaben wurden sodann in Durchführungsaufgaben und Beobachtungsaufgaben klassifiziert. Die Testanalyse erfolgte mit Methoden der klassischen Testtheorie.

Das DESK 3-6 enthält für jede Altersgruppe Entwicklungsaufgaben, die vier Entwicklungsbereichen zugeordnet werden: *Feinmotorik*, *Grobmotorik*, *Sprache* und *Kognition* sowie *Soziale Entwicklung*. Die Lösung der Aufgaben kann entweder als natürliches Vorkommnis des Kindergartenalltages direkt beobachtet und beurteilt werden (Beobachtungsaufgaben) oder aber die Aufgaben werden gezielt vorgegeben und überprüft (Durchführungsaufgaben). Letztere werden in ein Rollenspiel (*Zirkusspiel*) eingebettet, damit die Aufgaben sowohl weitestgehend standardisiert als auch spielerisch mehreren Kindern gleichzeitig vorgegeben werden können.

Die Erzieher bzw. Erzieherinnen haben zu beurteilen ob eine Aufgabe gelöst, unsicher/unvollständig gelöst oder nicht gelöst wurde oder ob ein Verhalten oft/sehr oft, manchmal oder selten/nie gezeigt wurde. Zur Auswertung werden alle vollständig gelösten Aufgaben (Beurteilung mit ja bzw. oft/sehr oft) zusammengezählt. Anhand von Normtabellen können die so erhaltenen Screening-Punkte für jeden Entwicklungsbereich sowie für die Gesamtentwicklung in Normwerte (Stanine-, Prozentrang-Werte) umgewandelt werden, welche die Ermittlung des Screening-Befundes ermöglichen. Kinder mit einem Stanine-Wert von 1 werden als *auffällig* eingestuft, Kinder mit einem Stanine-Wert von 2 erhalten den Screening-Befund *fraglich*, der durch weitere Beobachtung erhärtet werden muss.

Zur Testhandhabung stehen eine Durchführungs- und Auswertungsanleitung sowie Aufgabenhefte und Protokollbögen zur Verfügung. Zur Unterstützung der Interpretation des Screening-Befundes werden Hinweise formuliert. Das für die Durchführungsaufgaben benötigte Material ist nicht enthalten, sondern muss von der beurteilenden Person selbst zusammengestellt werden. Es handelt sich dabei um Material, welches üblicherweise im Kindergarten vorhanden ist (z. B. Knetmasse, Bälle, Malstifte). Zur Überprüfung der Auswertungsobjektivität wurde eine 71 Kinder umfassende Teilstichprobe von zwei Kindergartenpädagoginnen beurteilt und für jede Entwicklungsaufgabe die prozentuelle Übereinstimmung berechnet. Bei der überwiegenden Mehrheit der Aufgaben lag der Prozentsatz an Übereinstimmung über 80 Prozent.

Die innere Konsistenz (Chronbachs Alpha) des DESK 3-6 ist mit Werten zwischen $\alpha = .91$ (Itemset für Dreijährige und Vierjährige) und $\alpha = .93$ (Itemset für die Fünf- und Sechsjährigen) insgesamt als hoch einzuschätzen. Die ermittelte innere Konsistenz der je Altersgruppe vorgesehenen bereichsspezifischen Aufgabenzusammenstellungen (Entwicklungsskalen) liegt überwiegend im guten ($\alpha > .80$) bis befriedigenden ($\alpha > .70$) Bereich. Lediglich die Skala *Soziale Entwicklung* weist bei den Dreijährigen eine geringe innere Konsistenz ($\alpha = .58$) auf.

In Hinblick auf die Konstruktvalidität wird berichtet, dass das Verfahren zwischen verschiedenen Altersgruppen, zwischen entwicklungsauffälligen versus –unauffälligen Kindern, zwischen Kindern mit versus ohne diagnostizierte Entwicklungsbeeinträchtigungen sowie zwischen Kindern mit versus ohne zusätzlichen Förderbedarf differenziert. Prognostische Validität zeigt sich durch Zusammenhänge mit Ergebnissen der pädiatrischen Einschulungsempfehlung. Die Sensitivität des DESK 3-6 liegt je nach Kriterium zwischen 69.6 Prozent bis 84.2 Prozent, die Spezifität zwischen 78.3 Prozent und 93.0 Prozent. Konkurrente Validität wird durch substantielle Zusammenhänge mit dem Gesamtentwicklungsscore des Wiener Entwicklungstest nachgewiesen.

Die Normierung erfolgte an 1492 Kindergartenkindern aus 42 Kindergärten in Nordrhein-Westfalen und Niedersachsen. Die Normtabellen sind für die Drei- und Vierjährigen in Halbjahresschritte und für die Fünf- und Sechsjährigen in Jahresschritte gegliedert und zusätzlich nach Geschlecht getrennt angegeben. Darüber hinaus liegen für Kinder mit nicht-deutscher Muttersprache Normen für die Gesamtentwicklung vor, wobei diese lediglich auf Grundlage der Entwicklungsbereiche Feinmotorik, Grobmotorik und Soziale Entwicklung erstellt wurden.

5.1.2 Die diagnostischen Einschätzskalen (DES) zur Beurteilung des Entwicklungsstandes und der Schulfähigkeit

Die diagnostischen Einschätzskalen zur Beurteilung des Entwicklungsstandes und der Schulfähigkeit (DES; 5. Aufl., Barth, 2008) richten sich vorwiegend an pädagogisches Personal und bieten eine Hilfestellung zur Beurteilung der Lernausgangslage von Kindern im letzten Kindergartenjahr bzw. beim Übergang in die Volksschule. Sie sollen also vorrangig zur Klärung der Schulfähigkeit beitragen. Hierbei stehen jedoch nicht die Erfolgsaussichten im Vordergrund, sondern das Erkennen hinderlicher Prozesse, denen nötigenfalls wirkungsvoll entgegenzusteuern ist. Es wird somit ein förderorientierter Ansatz vertreten.

Die DES basieren auf einem integrativen Entwicklungsansatz, in welchen neuropsychologische Aspekte und Erkenntnisse zur frühen Vorhersage von Lese-/Rechtschreibstörungen sowie Rechenstörungen eingegangen sind. Die Entwicklung des Verfahrens erfolgte insofern theoriegeleitet, indem lediglich Entwicklungsbereiche und Aufgaben bzw. Beobachtungssituationen aufgenommen wurden, die in der Fachliteratur als relevant bezeichnet werden.

Die DES setzen sich aus einer Handanweisung, einem Aufgabenteil und einem Auswertungs- und Einschätzbogen zusammen.

Das Verfahren ist als Inventar konzipiert und erfasst insgesamt 28 Entwicklungsaspekte (19 *primäre* Entwicklungsaspekte und neun Aspekte zur sozial-emotionalen Entwicklung), die jeweils anhand mehrerer Aufgaben und/oder Beobachtungssituationen überprüft werden. In der Handanweisung wird auf die Möglichkeit einer individuellen Kurzfassung hingewiesen. Der durchführenden Person bleibt es dadurch überlassen, ob sie einzelne Aufgaben oder auch ganze Bereiche nicht durchführt. Auch spezifische Erweiterungen sind möglich, die im Handbuch kurz umrissen werden. Manche Aufgabenbereiche können zur Erhöhung der Ökonomie im Gruppensetting vorgegeben werden.

Zur Auswertung ist zunächst zu beurteilen ob die einzelnen Beobachtungs- und Testaufgaben richtig oder nicht gelöst wurden oder ob die Lösung unauffällig, auffällig oder uneindeutig war. Anschließend sind die Leistungen des Kindes je Entwicklungsaspekt auf einer fünfstufigen Skala (sehr ausgeprägt bis stark beeinträchtigt) subjektiv einzustufen. Darüber hinaus sollen weitere Beobachtungen, insbesondere zu

Auffälligkeiten, im Bogen festgehalten werden. Hierfür werden Beobachtungshinweise gegeben. Sämtliche Einschätzungen sind zuletzt in einen Entwicklungsprofilbogen zu übertragen.

Da der Autor für die DES nicht den Anspruch erhebt, ein Test im engeren Sinne zu sein, wurde das Verfahren weder empirisch überprüft noch einer Normierung unterzogen.

5.1.3 Erweiterte Vorsorgeuntersuchung (EVU)

Die Erweiterte Voruntersuchung (EVU; Melchers et al., 2003) ist ein allgemeines Entwicklungsscreening, das im Rahmen der kinderärztlichen Vorsorgeuntersuchungen (U4 bis U9) eingesetzt werden kann. Ziel der EVU ist es, die pädiatrische Vorsorgepraxis um entwicklungspsychologische und -neurologische Aspekte zu erweitern. Entwicklungsauffällige Kinder sollen durch das ergänzende Screening möglichst früh identifiziert und hierauf einer differenzierten Diagnostik und gegebenenfalls einer Förderung bzw. Behandlung zugeführt werden.

Die Entwicklung der EVU gründet auf der Unzufriedenheit mit den in Deutschland gesetzlich verankerten kindlichen Vorsorgeuntersuchungen (KVU). Obwohl diese sich zur Früherkennung und Sekundärprävention somatischer Erkrankungen bewährt haben, erscheinen sie zur Früherkennung von kognitiven, sprachlichen, sozial-emotionalen und neurologisch-motorischen Entwicklungsauffälligkeiten wenig geeignet. Die im Zuge der Vorsorgeuntersuchungen zu erfassenden Entwicklungsgefährdungen sind im sogenannten „Gelben Heft“ nur unzureichend operationalisiert, so dass die Beurteilung in der Praxis häufig intuitiv aufgrund subjektiver Erfahrungen erfolgt oder nach eigenem Ermessen zusammengestellte Testbatterien verwendet werden. Ein derartiges Vorgehen kann jedoch nicht zu validen und reliablen Ergebnissen führen und birgt die Gefahr falsch positiver bzw. falsch negativer Befunde in sich. Vor diesem Hintergrund erschien es notwendig, der pädiatrischen Praxis ein standardisiertes und normiertes Screeningverfahren zur Verfügung zu stellen, welches ökonomisch und objektiv die spezifische und sensitive Detektion von Entwicklungsauffälligkeiten ermöglicht.

Im Zuge der Testkonstruktion sollten zunächst für sechs Untersuchungstermine (U4 bis U9) Aufgaben zusammengestellt werden, die in Hinsicht auf die Dimensionen *Motorik*, *Sprache*, *Kognitive Entwicklung* und *Verhaltensauffälligkeiten* theoretisch fundierte Entwicklungsmeilensteine erfassen und zudem der Variabilität von Entwicklungs-

prozessen Rechnung tragen. Die Aufgaben sollten möglichst einfach und reliabel durchzuführen sein sowie von 90 Prozent der altersgemäß entwickelten Kinder bewältigt werden können. Bei der Aufgabensammlung wurde sowohl auf Items aus bewährten Entwicklungstests zurückgegriffen als auch neue Aufgaben entwickelt. Der so entstandene erste Itempool wurde an einer Stichprobe von 300 Kindern in Hinblick auf Trennschärfe und Aufgabenschwierigkeiten evaluiert. Auf Grundlage der gewonnenen Ergebnisse wurde die Aufgabenzusammenstellung überarbeitet und anschließend normiert.

Die EVU orientiert sich aus praktischen Gründen an den Messzeitpunkten für die routinemäßig durchgeführten Vorsorgeuntersuchungen, die im 3. bis 4. Lebensmonat (U4), im 6. bis 7. Lebensmonat (U5), im 10. bis 12. Lebensmonat (U6), im 21. bis 24. Lebensmonat (U7), im 43. bis 48. Lebensmonat (U8) und im 60. bis 64. Lebensmonat (U9) stattfinden. Für jeden Untersuchungstermin sind Prüfaufgaben aus den Dimensionen *Motorische Entwicklung*, *Sprachentwicklung* und *Kognitive Entwicklung* sowie Fragen zu *Verhaltensauffälligkeiten* vorgesehen. Die Prüfaufgaben werden vom Kinderarzt bzw. der Kinderärztin vorgegeben und dieser/diese hat zu beurteilen ob die einzelnen Aufgaben „regelrecht“ bzw. „auffällig“ ausgeführt wurden. Die „Fragen zum Verhalten“ des Kindes werden mittels eines Elternfragebogens erfasst. Die Bezugsperson beurteilt dabei vor Untersuchungsbeginn ob das Kind das betreffende Verhalten in den vergangenen zwei Monaten „häufig/meistens“ oder „selten/nie“ gezeigt hat.

Zur Auswertung, die noch während der Konsultation zu erfolgen hat, ist zunächst je Entwicklungsbereich (Motorik, Sprache, Kognition) die Anzahl der regelrechten Reaktionen zu ermitteln. Die resultierenden Rohwertsummen werden anschließend zu einem Wert für die Gesamtentwicklung („Summe Entwicklung“) addiert. Die „Fragen zum Verhalten“ werden zu den einzelnen Testzeitpunkten unterschiedlich behandelt, da für diese Dimension aus inhaltlichen und psychometrischen Überlegungen kein homogenes Aufgabenmaterial vorliegt. Bei der EVU 4 und der EVU 5 werden die Verhaltensfragen nicht quantitativ ausgewertet. Hier sind die als „auffällig“ bezeichneten Verhaltensweisen lediglich Grundlage einer eingehenden Nachbesprechung. Ab der EVU 6 ist eine normgestützte Auswertung möglich. Hierfür wird aus den als regelrecht aufzufassenden Antworten auf eine gewisse Teilmenge der Fragen (EVU 6 und EVU 7) respektive auf alle Fragen (EVU 8 und EVU 9) eine „Summe Verhaltensfragen“ errechnet.

Auf Grundlage der Rohwertsummen wird die bereichsspezifische Entwicklung bzw. die Gesamtentwicklung als unauffällig, grenzwertig gefährdet oder auffällig/gefährdet im Vergleich zur Altersgruppennorm beurteilt. Als auffällig gelten Leistungen unterhalb der 10. Perzentile. Ein „Grenzbefund“ resultiert aus Leistungen mit einem Prozentrang < 20 . Die Zuordnung der Summenwerte zu den Referenzwerten der drei Befundgruppen ist unmittelbar auf dem Protokoll- und Auswertungsbogen abzulesen. Zur genaueren Einschätzung des individuellen Entwicklungsstandes stehen zudem Prozentrangtabellen zur Verfügung.

Das Testmaterial umfasst ein Handbuch, einen Durchführungsordner, Durchführungszubehör (z. B. Bilderbuch, Glöckchen, Spielzeughund, Ball), Vorlagen für die grafomotorischen Aufgaben bei der EVU 8 und 9, Fragebögen zum Verhalten (bei EVU 6 bis EVU 9 als Durchschreibsätze) und Protokollbögen (als Durchschreibsätze). Durchführung und Auswertung sind standardisiert. Der Protokollbogen enthält eine Ergebnistafel, die eine schnelle Zuordnung der Ergebniswerte zu den Befundgruppen ermöglicht. Überdies werden im Handbuch klar formulierte Empfehlungen zum Umgang mit „auffälligen“ und „grenzwertig gefährdeten“ Befunden gegeben.

Zur Schätzung der Reliabilität des Verfahrens wurden Konsistenzanalysen (Cronbachs Alpha) für die Entwicklungsbereiche und für die als Gesamtmaß der Entwicklung zu verstehende „Summe Entwicklung“ gerechnet. Die Analysen ergeben für die Entwicklungsbereiche überwiegend Konsistenzkoeffizienten, die prinzipiell als gering einzuschätzen sind ($< .80$). Für das zusammenfassende Entwicklungsmaß werden Koeffizienten von .63 bis .77 berichtet. Melchers et al. betonen im Hinblick auf diese Ergebnisse, dass die Werte in Anbetracht der Kürze der Skalen hinreichender Beleg für die Messgenauigkeit des Verfahrens sind. Für die "Fragen zum Verhalten" wurde die innere Konsistenz nicht ermittelt, da diese von ihrer Konzeption her keine lineare Skala bilden. Die ermittelten Schwierigkeitsindices belegen zudem, dass für den Großteil der Aufgaben die angestrebte Lösungsfrequenz von 90 Prozent besteht.

Zum Nachweis der Konstruktvalidität der EVU wurden verschiedene Methoden herangezogen, über deren Ergebnisse im Folgenden berichtet wird: Erwartungsgemäß niedrige Interkorrelationen der drei Entwicklungsbereiche belegen die Eigenständigkeit der drei untersuchten Bereiche. Die mittleren bis hohen Korrelationen zwischen den Entwicklungsbereichen und der „Summe Entwicklung“ sowie die größtenteils nicht signifikanten Zusammenhänge zwischen den Entwicklungsbereichen und den „Fragen

zum Verhalten“ entsprechen ebenfalls den theoretischen Vorannahmen. Eine Teilstichprobe von 31 bzw. 27 Kindern wurde zu den Zeitpunkten der U 6 und U 7 sowohl mit der EVU als auch mit den Griffiths Entwicklungsskalen (Brandt, 1983) und den BSID II (Bayley, 1995) untersucht. Es zeigten sich Korrelationen in mittlerer Höhe (.63 bis .69) zwischen der „Summe Entwicklung“ und den Entwicklungsmaßen dieser Instrumente (Entwicklungsquotient der Griffiths Entwicklungsskalen, Mental Developmental Index und Psychomotor Development Index der Bayley II). Unter Berücksichtigung der Unterschiede zwischen diesen Verfahren und der EVU hinsichtlich Durchführungsdauer und Aufgabenanzahl ist dies aus Sicht von Melchers et al. ein hinreichender Beleg für die konvergente Validität der EVU. Interkorrelationen der EVU-Entwicklungsbereiche mit Teilbereichen der Griffiths Entwicklungsskalen sowie der BSID II zeigen hohe Zusammenhänge der inhaltlich korrespondierenden Skalen, insbesondere im Bereich der Motorik. Die beiden Studien wurden zudem zur Überprüfung der Differenzierungsfähigkeit herangezogen. Hieraus resultierte eine Sensitivität von 100 Prozent und eine Spezifität von 92.6 Prozent.

Die Normierung der EVU erfolgte an einer Stichprobe von 1743 Kindern, die im Rahmen ihrer regulären Vorsorgeuntersuchungen zusätzlich mit der EVU getestet wurden. Die Datenerhebung erfolgte in mehr als 25 Kinderarztpraxen im Umkreis der Städte Köln/Düren, Braunschweig, Celle und Tegernsee. Nach Melchers et al. kann die regionale Verteilung als für das gesamte Bundesgebiet annähernd repräsentativ angesehen werden. Die Stichprobe wird im Manual in Bezug auf die Wohnortgröße und den Beruf der Eltern charakterisiert und wird vom Autor und den Co-Autoren als begrenzt repräsentativ eingeschätzt. Der Anteil an Migrantenkindern betrug knapp neun Prozent und ist damit geringer als jener in der Grundgesamtheit.

5.1.4 Vergleichende Bewertung der dargestellten Entwicklungsscreenings

Alle drei Screenings sind für den Einsatz in der frühen Kindheit vorgesehen. Die EVU kann außerdem bereits im Säuglings- und Kleinkindalter angewendet werden. Die DES wiederum fokussieren auf Kinder im letzten Kindergartenjahr und decken somit den kleinsten Altersbereich ab.

Durchgeführt wird das DESK 3-6 und die DES von pädagogischem, die EVU hingegen von pädiatrischem Personal.

Sowohl das DESK 3-6 als auch die DES sind gut im Kindergartenalltag einsetzbar. Beide Verfahren beinhalten Beobachtungs- sowie Testaufgaben. Letztere sind beim DESK 3-6 in ein Rollenspiel eingebettet und werden mehreren Kindern gleichzeitig vorgegeben. Bei den DES steht es der untersuchenden Person frei, manche Aufgaben in der Gruppe vorzugeben. Die EVU kann im Zuge der routinemäßigen Vorsorgeuntersuchungen durchgeführt werden.

Bezüglich ihrer entwicklungstheoretischen Fundierung sind sowohl das DESK 3-6 als auch die DES zu kritisieren. Allerdings sind die Angaben, die zum theoretischen Ansatz im Manual des DESK 3-6 gegeben werden, wesentlich umfangreicher als die knappen Hinweise, die die zum konzeptionellen Hintergrund der DES gegeben werden. Die Skalen- und Aufgabenauswahl der EVU dagegen wurden plausibel entwicklungstheoretisch begründet.

Der wesentlichste Unterschied der Verfahren liegt in deren testtheoretischer Fundierung:

- Die DES wurden weder einer empirischen Überprüfung noch einer Normierung unterzogen. Da sowohl bei der Durchführung der beurteilenden Person viel Spielraum einräumt wird als auch die Bewertung nach subjektiven Maßstäben erfolgt, ist dem Verfahren insgesamt geringe Objektivität zu bescheinigen. Unter Berücksichtigung der Aussage des Autors, dass solche Entwicklungsbereiche und Aufgaben in das Inventar aufgenommen wurden, die in der Fachliteratur als relevant bezeichnet werden, kann den DES immerhin in Maßen Inhaltsvalidität zugesprochen werden.
- Das DESK 3-6 und die EVU hingegen wurden sorgfältig empirisch überprüft und überzeugen beide durch hohe Sensitivität und Spezifität. Die Erfüllung dieser Güteindices ist ein Muss, um als gültiges Screeningverfahren bestehen zu können. Zudem sind beide Verfahren ausreichend objektiv und messgenau. Die Normierungsstichproben sind jeweils umfangreich, regional jedoch eingeschränkt (Erhebung in zwei bzw. vier deutschen Bundesländern). Dies schränkt den Geltungsbereich der Verfahren ein.

Sowohl das DESK 3-6 als auch die EVU können zum Screening nach Entwicklungsauffälligkeiten in der frühen Kindheit empfohlen werden. Eine Ausweitung der Normen auf weitere deutsche Bundesländer, Österreich und die Schweiz sind für nachfolgende Auflagen zu wünschen.

Aufgrund der fehlenden testtheoretischen Absicherung kann der Einsatz der DES zur Identifikation von Risikokindern nicht empfohlen werden. Es bleibt zu hoffen, dass eine nachfolgende sechste Auflage nicht lediglich „durchgesehen“ publiziert wird, sondern eine empirische Überprüfung und Normierung enthält.

5.2 Entwicklungstests

5.2.1 Bayley Scales of Infant Development – Second Edition (BSID II, Bayley II)

Die zweite Auflage der Bayley Scales of Infant Development (BSID II; Bayley, 1993), kurz Bayley II, liegt seit 2007 in einer deutschsprachigen Bearbeitung vor (BSID II; Reuner et al., 2007). Mit Bayley II wird der Entwicklungsstand von Kindern im Alter von 1 bis 42 Monaten erhoben.

Die Bayley Scales of Infant Development (BSID) blicken auf eine lange Tradition in der Diagnostik frühkindlicher Entwicklung zurück. Die erste Auflage wurde 1969 veröffentlicht und basierte auf einer Zusammenfassung verschiedener Verfahren, die im Rahmen der Berkeley Growth Study (vgl. Jones & Bayley, 1941) entwickelt worden waren. Bayley ging bei der Konstruktion des Verfahrens davon aus, dass die Fähigkeiten, die sich in den ersten zwei Lebensjahren entwickeln, nicht verschiedenen Faktoren zugeordnet werden können. Demzufolge umfassten die BSID nur eine kognitive (*Mental Scale*) und eine motorische Skala (*Motor Scale*) sowie eine standardisierte Verhaltensbeurteilung (*Infant Behavior Rating*). Dieser Aufbau wurde auch für die 1993 publizierte Originalfassung von Bayley II beibehalten. Die deutschsprachige Fassung beschränkt sich allerdings nur auf die kognitive und die motorische Skala.

Die kognitive Skala erfasst frühe Gedächtnisleistungen, Habituation, Problemlösefähigkeiten, frühe Zahlkonzepte, Klassifikation und Kategorisierungsfähigkeit, Vokalisation und sprachliche Kompetenzen sowie frühe sozial-kommunikative Fähigkeiten. Die Aufgaben der motorischen Skala hingegen dokumentieren die zunehmende Handlungskontrolle bis hin zur Aufrichtung sowie die grob- und feinmotorische Koordination.

Den einzelnen Entwicklungsaufgaben liegen verschiedenste Ergebnisse der Säuglings- und Kleinkindforschung zugrunde, die Basis der Bayley II ist also eklektisch. Das Testmaterial wurde gemäß aktuellerer Erkenntnisse der Kleinkindforschung überarbeitet

und angesichts der Erweiterung des Altersbereichs entsprechend ergänzt. Die neu entwickelten Items wurden im Zuge von drei Pilotphasen hinsichtlich Itemschwierigkeit, Aufforderungscharakter und Problemen bei der Durchführung analysiert und adaptiert oder ausgeschieden. In Folge haben 60 trainierte Beurteiler an einer für die USA repräsentativen und ethnisch gleich verteilten Stichprobe von 643 Kindern die aus alten und neuen Items bestehende vorläufige Version erprobt. Anschließend wurden die Aufgabenschwierigkeiten geschätzt und die Aufgaben entsprechend in den Skalen positioniert. Einige Items wurden entfernt. Hierauf wurde die vorläufige Version mittels Partialkorrelationen und Rasch-Analysen auf geschlechts- bzw. herkunftsspezifische Abweichungen (Bias) hin analysiert und Items wenn nötig wiederum ausgeschieden. Die verbleibenden Aufgaben bildeten die Standardisierungsversion von Bayley II.

Die Items sind in jeder Skala nach steigender Schwierigkeit geordnet. Es sind altersbezogene Aufgabenbereiche (Item-Sets) vorgesehen, die grundsätzlich in Abhängigkeit vom chronologischen bzw. korrigierten Alter des Kindes zu bearbeiten sind. Darüber hinaus werden Einstiegs- und Ausstiegsregeln formuliert. Die Einstiegsregel wird erfüllt, wenn mindestens vier (Motor Scale) bzw. fünf (Mental Scale) Items des altersbezogenen Aufgabenbereichs gelöst werden. Wenn notwendig wird zu einer leichteren Aufgabengruppe gewechselt, in der die Einstiegsregel wiederum anzuwenden ist. Löst ein Kind bei der Bearbeitung der Mental Scale alle Aufgaben eines Item-Sets mit Ausnahme von einer oder zwei, so muss der nächst höhere Aufgabenbereich vorgegeben werden. Bei der Motor Scale ist der ist der nächst höhere Aufgabenbereich vorzugeben, wenn alle bis auf eine Aufgabe richtig gelöst wurden. Anschließend wird solange „aufwärts“ getestet, bis das Kind zwei (Motor Scale) bzw. drei (Mental Scale) Aufgaben eines Bereichs nicht mehr lösen kann. Dem Testleiter wird freigestellt, in welcher Reihenfolge er die Items vorgibt obgleich Durchführungsvorschläge im Manual angeführt sind. Auch der Testeinstieg kann individuell bestimmt werden, wenn aufgrund von Vorinformationen oder Verhaltensbeobachtung vermutet wird, dass das Kind nicht in der Lage ist altersentsprechende Leistungen zu erbringen. Lediglich die Einstiegs- und Ausstiegsregeln sind strikt zu beachten.

Zur Auswertung werden die richtig gelösten Items je Skala zu einem Rohwert addiert und anhand der Normtabellen in einen *Mental Developmental Index (MDI)* bzw. einen *Psychomotor Developmental Index (PDI)* umgewandelt. Diese Indizes entsprechen Standardwerten, für die zusätzlich Konfidenzintervalle angegeben werden. Zudem besteht

die Möglichkeit ein Entwicklungsalter für die beiden Skalen oder für Facetten derselben abzuleiten.

Eine von einem während der Untersuchung anwesenden zweiten Beurteiler durchgeführte Überprüfung der Interrater-Übereinstimmung an 51 Kindern erbrachte allerdings eine Übereinstimmung von .96 für die Kognitive Skala und von .75 für die Motorische Skala.

Konsistenzanalysen erbrachten im Mittel für die Mental Scale ein α von .88 (Werte je nach Alter zwischen .78 und .92) und für die Motor Scale ein α von .84 (Werte von .75 bis .91). Im Manual wird über mehrere Studien zur Retest-Reliabilität berichtet, deren Ergebnisse allerdings unterschiedlich ausfallen und schwache bis befriedigende Zusammenhänge darstellen. Die Stabilität der Testwerte ist für die Kognitive Skala höher als für die Motorische.

Zur Überprüfung der Konstruktvalidität wurden Korrelationen der einzelnen Items mit dem Gesamtwert für jedes relevante Untersuchungsalter berechnet. Die Autoren berichten, dass sich für alle Items höhere Korrelationen mit der jeweils zugewiesenen Skala als mit der alternativen Skala ergaben, ohne jedoch die exakte Höhe der Korrelationen anzugeben. Diskriminante und konvergente Validität konnte durch Vergleichstudien mit anderen Verfahren (z.B. Mc Carthy Scales, Denver II) dargestellt werden. Die Ergebnisse, die Kinder mit Risiken für Entwicklungsstörungen (z.B. HIV, Frühgeborene, Down Syndrom) erbrachten, fielen im Mittel erwartungsgemäß niedriger aus als bei der Normierungsstichprobe.

Die Normierung von Bayley II basiert auf einer stratifizierten, randomisierten US-Stichprobe, die anhand des US-Mikrozensus (1988) festgelegt wurde. Die Normierungsstichprobe umfasste 1700 Kinder, je 100 in jeder Altersgruppe. Um die raschere Entwicklung im ersten Lebensjahr abzubilden, wurden vom ersten bis zum zwölften Monat mehr Altersgruppen gebildet als vom dreizehnten bis zum zweiundvierzigsten Monat.

5.2.1.1 Bewertung des Verfahrens

Die deutsche Bearbeitung von Bayley II wurde 2007 publiziert, obwohl bereits ein Jahr zuvor eine wesentlich veränderte und neu normierte Fassung in den USA auf den Markt kam (BSID III; Bayley, 2006).

Das Verfahren ist als Stufenleiter konzipiert. Das zugrundeliegende Entwicklungsmodell dieser Testform und deren Konstruktionsprinzipien gelten heute als veraltet.

Der bei der Durchführung gewährte Spielraum geht zu Lasten der Objektivität des Verfahrens. Die Anweisungen sind zudem recht komplex, was insbesondere darauf zurückzuführen ist, dass zum einen altersbezogene Aufgabenbereiche vorhanden sind und zum anderen Ein- und Ausstiegsregeln formuliert werden.

Bei der genauen Lektüre des Manuals fällt auf, dass Abbildungen und Tabellen, auf die im Text verwiesen wird, wiederholt falsch beziffert werden, was zur zusätzlichen Verwirrung beiträgt (z. B. Abb. 3.2 auf S. 44-45 wird im Text Abb. 3.3 genannt).

Aufgrund der langen Tradition der Bayley Scales und ihres Einsatzes in vielfältigen unabhängigen Untersuchungen gilt das Verfahren als gut validiert. Auch wenn die quantitativen Ergebnisse der Untersuchungen nicht immer überzeugen, finden sich dennoch vielfältige Hinweise auf den Aussagebereich des Verfahrens. Die berichteten Reliabilitätskoeffizienten genügen zudem den Anforderungen der Psychometrik.

Für die deutsche Bearbeitung wurde weder eine Neu-Normierung noch eine Überprüfung der Normdaten durchgeführt. Das Verfahren beruht auf US-Normen, die vor der Publikation der Originalfassung im Jahr 1993 erhoben wurden und somit bereits veraltet sind.

Für Bayley II spricht die internationale Verbreitung des Verfahrens, welche länderübergreifende Studien ermöglicht. In der Einzelfalldiagnostik sollte jedoch einem Verfahren Vorzug gegeben werden, welches entwicklungstheoretisch und testtheoretisch dem gegenwärtigen Stand der Forschung entspricht.

5.2.2 Basisdiagnostik für umschriebene Entwicklungsstörungen im Vorschulalter (BUEVA)

Die Basisdiagnostik für umschriebene Entwicklungsstörungen im Vorschulalter (BUEVA; Esser und Wyschkon, 2002) ist eine aus bereits publizierten Verfahren zusammengestellte Testbatterie zur Erkennung von Teilleistungsstörungen bei vier- und fünfjährigen Kindern. Eventuelle Defizite in einem umschriebenen Leistungsbereich sollen mithilfe der BUEVA noch vor Schuleintritt erkannt werden, um betroffene Kinder frühzeitig einer entsprechenden Förderung zuführen zu können.

Kinder mit Teilleistungsstörungen sind durch gravierende Defizite in einem einzelnen Leistungsbereich bei ansonsten normaler Entwicklung bzw. durchschnittlicher Intelligenz gekennzeichnet. Im Vorschulalter sind nach Esser und Wyschkon vor allem die Artikulationsstörung, die Expressive und die Rezeptive Sprachstörung sowie Umschriebene Entwicklungsstörungen der motorischen Funktionen relevant. Die Wichtigkeit einer frühzeitigen Diagnose wird aufgrund von mehreren im Manual angeführten Längsschnittuntersuchungen verdeutlicht, welche zeigen, dass umschriebene Entwicklungsstörungen eine hohe Stabilität aufweisen und Zusammenhänge mit Schulleistungen und kinderpsychiatrischen Störungen bestehen.

Die BUEVA wurde entwickelt, da aus Sicht des Autors und der Co-Autorin nahezu alle verfügbaren Vorschultests im oberen Leistungsbereich differenzieren und somit für entwicklungsrückständige Kinder zu schwierig sind. In Anbetracht dieser Problematik hat Esser bereits 1980 im Auftrag der deutschen Kassenärztlichen Bundesvereinigung eine erste Testbatterie zusammengestellt, mit deren Hilfe praktisch bedeutsame Leistungsbereiche untersucht werden sollten. Zwei Voruntersuchungen an insgesamt 120 vier- und fünfjährigen Kindern führten zu Modifikationen an den ausgewählten Verfahren hinsichtlich Instruktion, Darbietung, Auswertung und Schwierigkeitsstaffelung der Items.

Die BUEVA enthält eine Skala zur *Allgemeinen Intelligenz* sowie vier Skalen zur Erfassung umschriebener Entwicklungsstörungen. Diese werden sowohl vier- als auch fünfjährigen Kindern vorgegeben. Für Fünfjährige sind zudem zwei Skalen zur Prüfung der Aufmerksamkeit vorgesehen. Insgesamt umfasst die Testbatterie somit folgende sieben Untertests:

1. *Allgemeine Intelligenz*: Columbia Mental Maturity Scale (CMM, Eggert, 1972)
2. *Sensumotorik*: Subtest „Visumotorische Koordination“ aus Frostigs Entwicklungstest der visuellen Wahrnehmung (FEW, Lockowandt, 1976)
3. *Artikulation*: Möhring-Test (Kurzform)
4. *Expressive Sprache*: Grammatiktest (GT) aus dem Psycholinguistischen Entwicklungstest (PET, Angermaier, 1977)
5. *Rezeptive Sprache*: Wörterergänzen (WE) aus dem PET (Angermaier, 1997)
6. *Aufmerksamkeit I*: Untertest „Konzentration“ aus dem Frankfurter Test für Fünfjährige (FTF-K, Ratz & Moehling, 1971)
7. *Aufmerksamkeit II*: Zahlenfolgegedächtnis aus dem PET (Angermaier, 1977)

Die Testbatterie besteht aus einem Handbuch, den benötigten Materialien, einem Testheft, einem Ergebnisbogen und Schablonen zur Auswertung des Subtests *Sensumotorik*.

Das Manual enthält klare und ausführliche Durchführungsanweisungen, die sich zudem in verkürzter Form im Testheft wiederfinden. Ebenso ist die Auswertung bei fast allen Verfahren eindeutig festgelegt. Laut Esser und Wyschkon können lediglich beim Subtest *Artikulation* Auswertungsdifferenzen in Abhängigkeit von der Toleranz der untersuchenden Person auftreten. Die ermittelten Rohwerte werden je Subtests zu einem Summenscore addiert und mittels altersspezifischer Normtabellen in T-Werte transformiert. Weiters steht eine Tabelle zur Umwandlung der T-Werte in Prozentränge zur Verfügung. Die Rohwerte und Normwerte werden in den *Ergebnisbogen* eingetragen, welcher überdies Hinweise zur Interpretation enthält. So wird eine "Umschriebene Entwicklungsstörung" diagnostiziert, wenn die T-Werte der Untertests 2, 3, 4, oder 5 ≤ 35 betragen und gleichzeitig eine Differenz von mindestens 15 T-Wert-Punkten zu Untertest 1 besteht. Zur Erleichterung der Ergebnisbeurteilung sind im Handbuch auch noch 16 Fallbeispiele angeführt.

Da das Differenzierungsoptimum der BUEVA im unteren Bereich liegen soll, sind die Aufgabenschwierigkeiten gering. Dies führt zwangsläufig dazu, dass auch die Trennschärfe nicht ihr Maximum erreicht.

Die inneren Konsistenzen für die sprachlichen Untertests lagen zwischen .67 und .85, die für den Subtest *Sensumotorik* bei .55 (Vierjährige) bzw. .57 (Fünfjährige). Für den Subtest *Allgemeine Intelligenz* und die beiden Subtests zur Erfassung der Aufmerksamkeit werden lediglich die für die ursprünglichen Fassungen berechneten Reliabilitätskoeffizienten (Innere Konsistenz, Splithalf- und Retest-Reliabilität) angegeben. Die aus der Literatur übernommenen Kennwerte attestieren den Ursprungsverfahren – mit Ausnahme des FEW – zufriedenstellende Reliabilitäten. Esser und Wyschkon weisen im Zusammenhang mit der als niedrig einzuschätzenden Reliabilität des Subtests *Sensumotorik* darauf hin, dass dieser mangels eines Alternativverfahrens in der Testbatterie belassen wurde.

Zur Überprüfung der kriterienbezogenen Validität wurde der WET (Kastner-Koller & Deimann, 1998) eingesetzt. Die Korrelation der gemittelten Gesamtwerte (= Summe der Normwerte, dividiert durch die Anzahl der Subtests) ergab einen Validitätskoeffizienten von .79. Als zweites Kriterium wurde der Potsdamer Erzieherfragebogen (Esser &

Wyschkon, ohne Literaturangabe) herangezogen, welcher den allgemeinen Entwicklungsstand aus Sicht des pädagogischen Kindergartenpersonals erfasst. Es ergab sich eine mäßige Korrelation von $-.59$. Diese fällt negativ aus, da in der BUEVA ein hoher Wert mit einer hohen Leistung verknüpft ist, im Potsdamer Erzieherfragebogen hingegen mit einer niedrigeren. Zur Bestimmung der faktoriellen Validität wurden die BUEVA- und WET-Subtests gemeinsam einer Faktorenanalyse unterzogen. Korrespondierende Untertests sollten auf demselben Faktor laden. Es ergab sich eine Sechs-Faktoren-Lösung, die 69 Prozent der Varianz aufklärt und die hypothetischen Konstrukte der beiden Verfahren weitgehend bestätigte.

Esser und Wyschkon sehen die spätere Schulleistung als das der Zielsetzung des Verfahrens entsprechende wichtigste Validitätskriterium an. Folglich wurde der Darstellung der prognostischen Validität mit 18 Seiten viel Platz eingeräumt: Teilnehmern der „Mannheimer Risikokinderstudie“ (siehe u. a. Esser et al., 1990; Laucht, Esser & Schmidt, 1992; Laucht et al., 1996) wurde im Alter von viereinhalb Jahren die BUEVA vorgegeben. Dreieinhalb Jahre später wurde das Funktionsniveau schulischer Leistungen erhoben und den Kindern Lese-, Rechtschreib- und Rechentests vorgegeben. Die Ergebnisse umfangreicher Analysen zeigen, dass sich die BUEVA gut zur Vorhersage von Rechtschreib- und Rechenleistungen eignet. Die Lesefertigkeiten konnten weniger zutreffend prognostiziert werden. Der Subtest *Artikulation* erwies sich als ungeeigneter Prädiktor für die späteren Schulleistungen.

Die Normierung der BUEVA erfolgte bereits 1980 an einer Kindergartenstichprobe aus drei industriellen Mittelstädten eines westfälischen Kreises. Die Stichprobe umfasste 325 Kindern im Alter von 4;6 bis 4;11 Jahren und 332 Kinder im Alter von 5;6-5;11 Jahren. Die Normen für die verbleibenden Altersgruppen 4;0-4;5 und 5;0-5;5 wurden regressionsanalytisch aus den anderen beiden Altersgruppen gewonnen. Eine Überprüfung der Normierung erfolgte zwischen 1990 und 1992 an einer Zufallsstichprobe von 110 viereinhalbjährigen Teilnehmern der „Mannheimer Risikokinderstudie“. Die Abweichungen betrugen maximal zwei T-Wert-Punkte, so dass die Gültigkeit der ursprünglichen Normen angenommen werden kann. Zusätzlich wurden 1999 Normen für Einschulungsuntersuchungen erhoben, allerdings ohne die Subtests *Rezeptive Sprache (WE)* und *Aufmerksamkeit I (FTF-K)*. Die 1930 Kinder dieser Normierungsstichprobe waren zwischen 5;4 und 7;2 Jahre (durchschnittlich 6;3 Jahre) alt und stammten aus Schleswig-Holstein. Zudem wurden die Daten der 286 Kinder, die zum Zeitpunkt der

Erhebung zwischen 5;6 und 5;11 Jahre alt waren zur Überprüfung der Normen von 1980 herangezogen. Die Abweichungen betrugen hierbei maximal vier T-Wert-Punkte, was nach Ansicht von Esser und Wyschkon den Schluss zulässt, dass auch die nicht verwendeten Subtests keinen großen Veränderungen unterlagen.

5.2.2.1 Bewertung des Verfahrens

Die BUEVA ist eine aus etablierten Verfahren zusammengestellte Testbatterie zur gezielten Erfassung umschriebener Entwicklungsstörungen.

Durchführungs- und Auswertungsobjektivität der Testbatterie sind aufgrund der Standardisierung und der ausführlichen Durchführungs- und Auswertungsrichtlinien im Großen und Ganzen gegeben. Der Autor und seine Co-Autorin weisen auf mögliche Auswertungsdifferenzen beim Subtest *Artikulation* hin. Aus Sicht der Autorin dieser Arbeit fehlen zudem Hinweise, wie vorzugehen ist, wenn einer der Begriffe, die das Kind beim Subtest *Rezeptive Sprache* zu ergänzen hat, diesem aufgrund sprachlicher Besonderheiten nicht bekannt ist. Bei einem Einsatz im Osten Österreichs ist beispielsweise anzunehmen, dass Kindern der Begriff „Eisschrank“ mitunter nicht geläufig sein wird, da das Wort „Eiskasten“ verwendet wird. Die Möglichkeit der Verwendung alternativer Begriffe wird im Handbuch jedoch nicht erwähnt.

Die angeführten Reliabilitätskennwerte weisen prinzipiell auf eine ausreichende Messgenauigkeit der meisten Subtests hin. Für den Subtest *Sensumotorik* fallen die aus eigenen und ursprünglichen Berechnungen hervorgegangenen Reliabilitäten gering aus. Aufgrund der vorgenommenen Modifikationen und in Hinblick auf die Verwendung von mittlerweile mehr als 30 Jahre alten Ursprungsverfahren wäre eine Neubestimmung der inneren Konsistenz und Split-Half-Reliabilität für alle Subtests wünschenswert.

Zur Bestimmung der Kriteriumsvalidität wurde der Zusammenhang zwischen der BUEVA und dem WET sowie dem Potsdamer Erzieherfragebogen ermittelt, wobei sich substantielle Korrelationen ergaben. Seltsam mutet an, dass ein Gesamtwert der BUEVA korreliert wurde, obwohl diese einen solchen gar nicht vorsieht. Für den Zusammenhang mit dem WET würde es sich anbieten, korrespondierende Untertests miteinander in Beziehung zu setzen. Ob der Potsdamer Erzieherfragebogen ein geeignetes Kriterium darstellt kann ob fehlender Literaturangabe nicht nachvollzogen werden. Die Ergebnisse der Untersuchungen zur prognostischen Validität hingegen überzeugen.

Die Normierung der BUEVA erscheint insgesamt problematisch. Die Ursprungsnormen sind veraltet und wurden zudem teilweise regressionsanalytisch bestimmt. Die erste Überprüfung erfolgte 1992, welche aus heutiger Sicht somit jedoch ebenso bereits veraltet ist. Die Überprüfung von 1999 wiederum erfolgte lediglich an 5;6 bis 5;11-jährigen Kindern und es wurden auch nicht alle Untertests miteinbezogen. Weiters wurde die Normierungsstichprobe nur lokal rekrutiert und hat somit auch nur eingeschränkte Gültigkeit. Eine Neunormierung ist daher dringend anzuraten.

Die für einen Entwicklungstest überaus kurze Durchführungsdauer von durchschnittlich 25 Minuten trägt neben der testleiterfreundlichen Handhabung zur Ökonomie des Verfahrens bei.

Im gesamten gesehen scheint die BUEVA dennoch zur Eingangdiagnostik möglicher Entwicklungsstörungen geeignet. Eine Neuauflage, in der die testtheoretischen Mängel behoben sind, ist aber wünschenswert.

5.2.3 Entwicklungstest sechs Monate bis sechs Jahre (ET 6-6)

Der Entwicklungstest sechs Monate bis sechs Jahre (ET 6-6; 2. Aufl.; Petermann et al., 2004) ist kein Test im engeren Sinne, sondern ein Inventar und versucht somit eine Vielzahl qualitativer Entwicklungsmerkmale innerhalb eines breiten Bereiches kindlicher Fertigkeiten abzubilden. Das Verfahren wurde – seinem Namen entsprechend – für Kinder von sechs Monaten bis sechs Jahren konzipiert.

Ausgehend von einem adaptiven und ganzheitlichen Entwicklungsverständnis und der Integration unterschiedlicher interaktionistisch geprägter Entwicklungstheorien entwerfen Petermann et al. ein „Ordnungsschema für Entwicklungsprozesse“ (2004, S. 17 ff). Dieses Modell versucht die theoretische Annahme zu veranschaulichen, dass Entwicklung intra- und interindividuell sehr variabel verläuft und von Inkonsistenzen geprägt ist. Unabhängig von der individuellen Variabilität muss jedes Kind dennoch bestimmte „essentielle Grenzsteine“ absolvieren, die für den Erwerb darauf aufbauender Fertigkeiten notwendig sind.

Bei der Testkonstruktion wurde explizit darauf verzichtet Skalen und durchgehende Subtests für den Altersbereich von sechs Monaten bis sechs Jahren zu bilden, da mit dem ET 6-6 altersspezifisch viele verschiedene qualitative Entwicklungsaspekte abgebildet werden sollen. Die Itemauswahl erfolgte im Einklang mit Erkenntnissen zur Entwicklung

in der frühen Kindheit und unter Berücksichtigung des Grenzsteinprinzips. Eine erste Version wurde an einer Konstruktionsstichprobe von 260 Bremer Kindern erprobt. Die Ergebnisse dieser Studie haben jene Aufgabenzusammenstellung je Entwicklungsdimension und jene Reihung der Items je Altersgruppe determiniert, die der Normierungsstichprobe vorgegeben wurde. Hinsichtlich der zweiten Auflage des ET 6-6 wird im Manual berichtet, dass einzelnen Altersgruppen nunmehr nicht mehr alle ursprünglich vorgesehenen Items vorgegeben werden.

Der ET 6-6 (2. Aufl.) erfasst sechs der Tradition der Entwicklungsdiagnostik entsprechende Beschreibungsdimensionen, die durch zehn sogenannte *Entwicklungsdimensionen* abgebildet werden:

1. *Körpermotorik*
2. *Handmotorik*
3. Kognitive Entwicklung: *Gedächtnis, Handlungsstrategien, Kategorienbildung, Körperbewusstsein*
4. Sprachentwicklung: *rezeptive Sprache, expressive Sprache*
5. *Sozialentwicklung*
6. *Emotionale Entwicklung*

Die sechs Beschreibungsdimensionen werden ab der Mitte des vierten Lebensjahrs durch einen *Subtest Nachzeichnen* ergänzt.

Der Itempool setzt sich aus 113 Testitems, die vom Kind zu bearbeiten sind, und 67 Fragebogenitems, die von der Bezugsperson zu beantworten sind, zusammen. Letztere betreffen Verhaltensweisen, die in einer Testsituation schwierig oder gar nicht zu beobachten sind und beziehen sich vor allem auf die soziale und emotionale Entwicklung des Kindes sowie auf seine grobmotorischen Fähigkeiten.

Der gesamte Altersbereich von sechs Monaten bis sechs Jahren wird in 12 Altersgruppen unterteilt, wobei der Altersbereich *6 bis 24 Monate* in Drei-Monats-Schritte, der Altersbereich *24 Monate und 1 Tag bis 48 Monate* in Sechs-Monats-Schritte und der Altersbereich *48 Monate und 1 Tag bis 60 Monate* in Zwölf-Monats-Schritte unterteilt ist. Für jede dieser Altersgruppen besteht eine altersadäquate Aufgaben- und Fragenzusammenstellung. Dabei zeigen sich Überschneidungen in den Items zwischen den

Altersgruppen. Die Entwicklungsdimensionen werden in den 12 Altersgruppen zudem durch eine unterschiedliche Anzahl von Items repräsentiert, was dadurch begründet wird, dass nicht jede Dimension in allen Altersgruppen gleich bedeutsam ist.

Die Items werden nicht je Entwicklungsdimension vorgegeben, sondern als eine Art Itempool. In der Durchführungshilfe sowie am Protokollbogen ist die vorgesehene Aufgabenreihenfolge ersichtlich. Die Vorgabe erfolgt grundsätzlich mittels des entsprechenden Altersgruppenprotokollbogens. Liegen Hinweise auf eine massive Entwicklungsverzögerung vor, kann ein dem Testleiter angemessen erscheinender Altersgruppenbogen gewählt werden. Während der Durchführung ist zu protokollieren ob die Aufgabe erfüllt oder nicht erfüllt wurde. Eine Ausnahme bildet das Item „Zahlen nachsprechen“, für welches die Länge der längsten reproduzierten Zahlenreihe zu notieren ist. Der „Subtest Nachzeichnen“ ist mit Hilfe einer Schablone auszuwerten. Die sich ergebende Punktzahl ist für jedes der sieben Items des Subtests im Protokollbogen einzutragen, die Antworten der Bezugsperson auf die Fragebogenitems sind ebenfalls zu übertragen. Der Protokollbogen enthält ein Feld zur *Abschließenden Beurteilung der Testdurchführung*. Hier können Beobachtungen zu Wahrnehmungs-, Sprach- und Befindlichkeitsproblemen sowie zum Muskeltonus in halbstrukturierter Form erfasst werden. Darüber hinaus soll die Kommunikation, Motivation und Aufmerksamkeit des Kindes anhand einer vierstufigen Skala bewertet werden.

Anschließend werden die gelösten Items je Entwicklungsdimension summiert und in einen Testwert transformiert. Dieser ergibt sich, indem man die Summe der gelösten Items durch die Anzahl aller Items innerhalb der jeweiligen Entwicklungsdimension dividiert und mit 10 multipliziert. Das resultierende *Entwicklungsprofil* des Kindes wird schließlich mit dem Profil der Referenzgruppe verglichen, deren Mittelwerte und Standardabweichungen angegeben sind. Unterschiede von mehr als einer bis hin zu zwei Standardabweichungen vom Gruppenmittelwert spiegeln überdurchschnittliche bzw. unterdurchschnittliche Leistungen wieder. Divergenzen von mehr als zwei Standardabweichungen gelten als weit überdurchschnittlich bzw. weit unterdurchschnittlich. Bei der Interpretation sollen das Alter des Kindes im Sinne seiner Position im Altersbereich der jeweiligen Altersgruppe, die Itemanzahl in der jeweiligen Entwicklungsdimension, das Verhältnis zwischen Testergebnis und Elternurteil sowie die Itemschwierigkeiten berücksichtigt werden.

Der ET 6-6 umfasst ein Handbuch, die Testmaterialien (z. B. Holzkugeln, verschiedene Bildkarten, drei Puzzles in Form von Schildkrötenpanzern), altersgruppenspezifische Protokoll- und Auswertungsbögen sowie Elternfragebögen, und eine *Hilfe zur Testdurchführung*, der die Materialanforderungen und Testaufgaben (jeweils einschließlich Instruktionen und Bewertungskriterien) pro Altersgruppe zu entnehmen sind.

Die Testdurchführung ist weitgehend standardisiert. Das Handbuch enthält zudem Hinweise zur Gestaltung der Untersuchungssituation und zum Testleiterverhalten. Der untersuchenden Person wird jedoch ein gewisser Spielraum in Hinsicht auf die Itemreihenfolge gegeben. Zudem können einige Items (z. B. in der Altersgruppe 30 – 36 Monate alle Testitems der Entwicklungsdimension *expressive Sprache*) auch durch Elternbefragung erhoben werden. Auswertungs- und Interpretationsrichtlinien sind vorhanden.

Im Manual finden sich altersgruppenspezifische Itemschwierigkeiten und geschlechtsspezifische wie geschlechtsunspezifische Kennwerte der Verteilung der Entwicklungsdimensionsscores in den Altersgruppen (Mittelwerte, Standardabweichungen, Minimum, Maximum und Perzentilwerte). Aufgrund der Item- und Skaleneigenschaften wurde bewusst auf die Berechnung von Itemtrennschärfen verzichtet. Durch die Konstruktion als Inventar wurde eine gewisse Messungenauigkeit bewusst in Kauf genommen.

Hinsichtlich der Differenzierungsfähigkeit des ET 6-6 werden Studien angeführt, deren Ergebnisse darauf hinweisen, dass das Verfahren zwischen Kindern mit niedrigem und normalen Geburtsgewicht, zwischen Frühgeborenen und Termingeborenen sowie zwischen Kindern mit und ohne externalisierenden Verhaltensstörungen und Kindern unterscheidet. Eine Untersuchung von Zusammenhängen zwischen den Entwicklungsdimensionen des ET 6-6 und den Subtests und Summenindices der K-ABC (Melchers & Preuß, 2001) an einer Stichprobe von unauffälligen Vierjährigen erbrachte keine prägnanten statistischen Ergebnisse. Nach Petermann et al. lässt sich dieses Resultat mit den theoretischen Vorannahmen des ET 6-6 (hohe inter- und intraindividuelle Variabilität „normaler“ Entwicklung) in Einklang bringen. Im Zuge einer Studie zur prognostischen Validität wurden 41 Kinder zu Beginn des dritten Lebensjahres und ein Jahr darauf noch einmal getestet. Obwohl sich bei den unauffälligen Zweijährigen prinzipiell stabile Entwicklungsverläufe zeigten, wurden dennoch Leistungsfluktuationen innerhalb des unauffälligen Bereichs (+/- eine Standardabweichung vom Mittelwert)

beobachtet, was als weiterer Beleg für die Gültigkeit der theoretischen Vorannahmen gewertet wurde. In einer vergleichenden Studie mit dem WET (Kastner-Koller & Deimann, 2002), schätzte die Testleiterin im Anschluss an die Testdurchführung mittels des jeweiligen Verfahrens die während der Testung bestehende Motivation der Kinder ein. Sie kam zu dem Ergebnis, dass der ET 6-6 sich besonders günstig auf die Motivation der Kinder auswirkt und somit eine gültige Leistungsprüfung erlaubt. Als weiterer Beleg für die Validität des Inventars wird im Manual angeführt, dass für den *Subtest Nachzeichnen* ein deutlicher Alterstrend nachgewiesen werden konnte.

Die Normierung des ET 6-6 erfolgte vor der Publikation der ersten Auflage. Die Normen beruhen auf einer Stichprobe von 950 Kindern, die im Raum Bremen, Dortmund und Rostock wohnhaft waren und als überwiegend unauffällig galten.

5.2.3.1 Bewertung des Verfahrens

Der ET 6-6 ist kein Testverfahren im herkömmlichen Sinne, sondern ein Inventar. Die Bezeichnung ET (Entwicklungstest) ist somit irreführend und mag dazu verleiten, dem ET 6-6 Testgüte zu bescheinigen, die ihm aber nicht innewohnen kann.

Das Verfahren ist für Kinder von sechs Monaten bis zu sechs Jahren geeignet. Dies erscheint auf den ersten Blick dann vorteilhaft, wenn Entwicklungsverläufe abgebildet werden sollen. Unter Berücksichtigung des bewussten Verzichts auf die Konstruktion von homogenen Subtests und der Tatsache, dass lediglich qualitative Entwicklungsaspekte erfasst werden, erscheint die Verwendung des ET 6-6 in Längsschnittstudien jedoch in einem anderen Licht. Die Vergleichbarkeit der unterschiedlichen altersangepassten Item-Sets kann bei einer Re-Testung nicht ohne weiters vorausgesetzt werden. Entsprechende Studien stehen aber aus.

Objektivität ist nur bedingt gegeben. Die Durchführungsobjektivität wird gestützt durch die standardisierten Testmaterialien und die in Manual und in der separaten Durchführungshilfe enthaltenen Durchführungsanweisungen. Eingeschränkt wird sie durch den Spielraum, der dem Testleiter bzw. der Testleiterin hinsichtlich Itemreihenfolge und Vorgabemodalität (Beobachten bzw. Befragen) gewährt wird. Die Interpretationsobjektivität wird eingeschränkt durch jene Aspekte, die die untersuchende Person bei der Bewertung des Testergebnisses berücksichtigen soll. Eine empirische Überprüfung der Interraterreliabilität wäre hier angebracht gewesen.

Aufgrund der Konstruktion als Inventar wurde auf die Ermittlung klassischer Reliabilitätskoeffizienten verzichtet. Es werden lediglich Itemschwierigkeiten referiert.

Die im Manual berichteten Validitätsstudien liefern Hinweise auf den Gültigkeitsbereich des Verfahrens. Weitere Untersuchungen und insbesondere Längsschnittstudien zur prognostischen Validität sind hier zu wünschen.

Die Normierung erfolgte an einer regional eingeschränkten Stichprobe und schränkt somit den Geltungsbereich des Verfahrens ein.

Der ET 6-6 kann aufgrund des Inventarcharakters und der Abbildung von rein qualitativen Entwicklungsaspekten nur mit Vorbehalt zur Anwendung auf entwicklungsdiagnostische Fragestellungen empfohlen werden. Zur Generierung diagnostischer Hypothesen erscheint er jedoch durchaus sinnvoll.

5.2.4 Wiener Entwicklungstest (WET)

Der Wiener Entwicklungstest (WET; 2. Aufl., Kastner-Koller & Deimann, 2002) ist eine förderdiagnostisch orientierte Testbatterie zur Beurteilung des allgemeinen Entwicklungsstandes von Kindern zwischen drei und sechs Jahren.

Der WET basiert auf der Integration verschiedener entwicklungstheoretischer Ansätze. Den Metarahmen bilden dabei die ökologischen und kontextualistischen Theorien (vgl. Kapitel 1.2). Optimale Entwicklung resultiert demnach aus der optimalen Passung von Individuum und Lernumwelt. Entsprechend dem Schaufelmodell der Entwicklung (vgl. Kapitel 1.3) wird angenommen, dass Erziehungsbemühungen ab dem dritten Lebensjahr zunehmend wirksam werden. Diese wirken jedoch nur dann als Entwicklungsmotor, wenn sie auf den Entwicklungsstand abgestimmt sind. Förderbemühungen, die hierauf keine Rücksicht nehmen, führen demnach zu Über- oder Unterforderung des Kindes. Weiters gehen die Autorinnen davon aus, dass die kindliche Entwicklung über den Altersbereich der Drei- bis Sechsjährigen relativ kontinuierlich in Richtung zunehmender Kompetenz verläuft und keine auffälligen qualitativen Sprünge zu erwarten sind.

Als allgemeiner Entwicklungstest folgt der WET einem ganzheitlichen Entwicklungsverständnis, wonach eine kompetente Auseinandersetzung mit der Lebensumwelt nicht nur intellektuelle Fähigkeiten erfordert. Es wurden Funktionsbereiche ausgewählt, die sowohl für die Bewältigung aktueller Anforderungen notwendig sind als auch eine

(theoretisch und empirisch belegte) Vorhersage der Bewährung in der Einschulungsphase erlauben.

Die Konstruktion des Verfahrens erfolgte unter Berücksichtigung der theoretischen Basis. So galt es den ökologischen Kontext der Zielgruppe (drei- bis sechsjährige Kinder mit deutscher Muttersprache aus dem Kulturkreis westlicher Industrienationen) zu beachten und Aufgaben und Testmaterialien zu schaffen, die an wahrscheinlichen Erfahrungen dieser Zielgruppe anknüpfen und eine spielerische Gestaltung der Testsituation ermöglichen. Zudem sollten die Aufgaben der Subskalen nach einem durchgängigen Prinzip konstruiert sein. Diesen Vorgaben entsprechend wurde zunächst ein breites Spektrum an Subskalen konstruiert und anschließend evaluiert. Die Ergebnisse dieser theoretisch-inhaltlichen und empirischen Überprüfung bestimmten schließlich die aktuelle der aktuellen Form des Verfahrens. Dabei wurden die Entwicklungsaufgaben teilweise aus bewährten Entwicklungstests übernommen und weiterentwickelt, teilweise auch neu konstruiert.

Anhand von 13 Subtests, die vom Kind bearbeitet werden und einem Elternfragebogen erfasst der WET den Entwicklungsstand in den Funktionsbereichen *Motorik*, *Visuelle Wahrnehmung/Visumotorik*, *Lernen und Gedächtnis*, *Kognitive Entwicklung*, *Sprache* sowie *Sozial-Emotionale Entwicklung*. Die Subtests werden in Kapitel 7 ausführlich beschrieben, sodass an dieser Stelle auf eine Darstellung verzichtet wird.

Der WET umfasst die vollständigen Testmaterialien (Arbeitsblätter zum Subtest *Nachzeichnen*, *Elternfragebogen* sowie z. B. Puppenfamilie, zwei Bälle, Fotoalbum), das Manual und den Protokollbogen, welcher der Durchführung, Auswertung und Profilerstellung dient.

Testdurchführung und Auswertung sind standardisiert. Das Manual enthält ausführliche Anweisungen zur Durchführung, eindeutige Kriterien zur Bewertung der Testaufgaben, klare Auswertungsinstruktionen sowie Hinweise zum Testleiterverhalten und zur Gestaltung der Untersuchungssituation.

Zur Auswertung werden die Rohwerte der einzelnen Items je Subskala zu Summenscores addiert und anhand der Normtabellen in altersgruppenspezifische C-Werte transformiert. Die ermittelten Normwerte können zur übersichtlichen Darstellung der Stärken und Schwächen des Kindes in ein Entwicklungsprofil eingetragen werden. Aus der Differenz zwischen bestem und schlechtestem Subtestergebnis kann der Range (ausgewiesen in

Prozentrang-Werten) berechnet werden, welcher Auskunft über die Variabilität des Profils gibt. Darüber hinaus besteht die Möglichkeit einen (in C-Werten und Standardwerten ausgewiesenen) Gesamtentwicklungsscore zu ermitteln.

Die Testanalysen der 13 WET-Subtests erfolgten in erster Linie nach probabilistischen Methoden (Rasch-Analysen, Partial Credit Modell, Mokken-Skalenanalysen) mit dem Ziel, möglichst homogene eindimensionale Skalen für den gesamten Altersbereich von drei bis sechs Jahren zu konstruieren. In Hinblick auf die praktische Anwendung und im Sinne einer methodischen Validierung wurden für die homogenen Items zudem Trennschärfen, Itemschwierigkeiten und Reliabilitätsschätzungen berechnet. Der *Elternfragebogen* zur Erfassung der Selbstständigkeitsentwicklung wurde klassisch testanalysiert.

Die Durchführungsobjektivität wird durch standardisiertes Testmaterial, präzise Testanweisungen sowie zusätzliche Hinweise für das Testleiterverhalten gewährleistet. Auch die Auswertungsobjektivität ist als hoch einzuschätzen, da für den überwiegenden Teil der Skalen die Richtigkeit der Aufgaben eindeutig bestimmt werden kann. Bei den verbalen Subtests und den Subtests *Nachzeichnen* und *Fotoalbum*, welche ungeachtet der ausführlich beschriebenen Kategorisierung einen gewissen Beurteilungsspielraum bei der Auswertung zulassen, liegt die Interrater-Übereinstimmung zudem über .80.

Im Manual finden sich Befunde zur inneren Konsistenz (Cronbachs Alpha) und Split-Half-Reliabilität von 13 Skalen, wobei die verwendeten Maße gut übereinstimmende Ergebnisse erbrachten. In der Gesamtstichprobe variiert die innere Konsistenz zwischen .66 und .99, die Testhalbierungsreliabilität liegt zwischen .72 und .90. Die relativ geringen Reliabilitäten für die Subtests *Lernbär* ($\alpha = .66$) und *Schatzkästchen* ($\alpha = .70$) sind vor dem Hintergrund zu sehen, dass beide Skalen nur über sehr wenige Items verfügen. Für den Subtest *Zahlen Merken* ist eine Konsistenzanalyse aufgrund der Vorgabemodalitäten nicht möglich, sodass die Reliabilität dieser Skala auf Basis einer Testwiederholung (österreichische Stichprobe) ermittelt wurde. Der resultierende Wert von .67 stimmt laut Angaben der Autorinnen gut mit Ergebnissen aus anderen Testverfahren überein.

Die inhaltliche Validität wird durch die theorieorientierte und empirisch fundierte Herleitung der Skalen und Items begründet. Alle Subtests zeigen einen deutlichen Alterstrend und bestätigen somit die entwicklungstheoretischen Vorannahmen. Auch die

Ergebnisse einer im Manual angeführten Längsschnittstudie mit 54 Kindern weisen den altersabhängigen Anstieg der Subtestleistungen über vier Testzeitpunkte hinweg (zu 2;7, 3;1, 4;6 und 5;0 Jahren) nach. Des Weiteren werden im Manual einige experimentelle Studien dargelegt, die die Änderungssensitivität der Testbatterie auf kurz- und mittelfristige Förderungsmaßnahmen belegen. Obwohl die Funktionsbereiche des WET entwicklungstheoretisch nicht voneinander unabhängig sind wurde zur Überprüfung der Dimensionalität eine orthogonale Hauptkomponentenanalyse berechnet, der – analog den sechs Funktionsbereichen – sechs Faktoren, zugrunde gelegt wurden. Die Funktionsbereiche ließen sich dabei erwartungsgemäß nicht vollständig, aber doch weitgehend reproduzieren. Doppelladungen, die sich bei drei Skalen zeigten, sind zudem mit entwicklungstheoretischen Annahmen in Einklang zu bringen. Divergente Validität konnte durch erwartungsgemäß niedrige Korrelationen mit einigen Subtests der K-ABC (Melchers & Preuß, 1991) nachgewiesen werden. Im Manual werden zudem Studien zitiert, aufgrund derer die Differenzierungsfähigkeit des WET nachvollziehbar ist. In diesen Untersuchungen zeigten sich den Erwartungen entsprechende deutliche und syndromspezifische Leistungsunterschiede zwischen nicht klinisch auffälligen Kindern und Kindern aus klinischen Stichproben (Frühgeborene, Down-Syndrom, Autismus, Neurofibromatose).

Die Normierung des WET basiert auf repräsentativen Normwerten, die an 274 österreichischen und 971 deutschen Kindern erhoben wurden. Signifikante Unterschiede, welche die Nationalitäts-, Geschlechts- und Schichtzugehörigkeit betreffen und die in einigen Subtests beobachtet wurden, sind laut Angabe der Autorinnen praktisch irrelevant. Statistisch und praktisch bedeutsam sind lediglich die Unterschiede zwischen den Altersgruppen, die einen deutlichen Entwicklungsfortschritt zeigen. Demzufolge wurden lediglich Altersnormen erstellt, wobei je Halbjahresschritt eine Normtabelle zur Verfügung steht.

Drei unabhängige empirische Studien, deren Ergebnisse von Krampen, Becker, Becker und Thiel (2008) publiziert wurden, liefern weitere Hinweise zur Reliabilität, Validität und Zumutbarkeit des WET:

- Die Befunde aller drei Studien bestätigen durchgängig die im Manual des WET angeführten Split-Half-Reliabilitäten. Erweiternd führten Krampen et al. Berechnungen zu den Reliabilitäten der sechs WET-Funktionsbereiche durch. Die resultierenden Testhalbierungsreliabilitäten weisen zwischen den Studien eine

gute Stabilität auf und genügen unter Berücksichtigung der Tatsache, dass sie nur auf zwei bzw. vier Subskalen mit strukturellen und inhaltlichen Aufgabenheterogenitäten beruhen, zudem den psychometrischen Anforderungen. Der Autor und die Mitautorinnen schlagen deshalb vor, die Testauswertung und -interpretation auf die entwicklungstheoretisch begründeten Funktionsbereiche zu erweitern.

- Vorrangig zur empirischen Prüfung des oben genannten Vorschlags wurde auf Grundlage der Daten aus allen drei Studien eine Hauptkomponentenanalyse mit Varimax-Rotation berechnet, der sechs Faktoren zugrunde gelegt wurden. Die Reproduktion der WET-Funktionsbereiche gelang auch hier weitgehend. Die Ergebnisse der Studien replizieren überdies die Differenzierungsfähigkeit des WET im Hinblick auf die Altersgruppen sowie die weitgehende Unabhängigkeit der WET-Skalenwerte vom Geschlecht. Eine regressionsanalytische Überprüfung der Zusammenhänge der (auf den WET-Funktionsbereichen beruhenden sogenannten) WET-Entwicklungsskalen zu Indikatoren der Konzentrationsfähigkeit und des divergenten Handlungs- und Imaginationspotentials zeigte, dass sowohl die Konzentrationsleistungen als auch die Ideenflüssigkeit und Ideenflexibilität in hohem Maße durch die WET-Entwicklungsskalen determiniert werden. Auf der Stichprobenebene ergibt sich deshalb nach Krampen et al. kein Bedarf, den WET um bereichsspezifische Skalen zur Erfassung der Entwicklung divergenter Leistungen und der selektiven Aufmerksamkeitsleistung zu erweitern.
- Der Vergleich von unmittelbar vor und nach der WET-Testung erhobenen Konzentrationstestleistungen zeigte eine moderate Leistungsverschlechterung zum zweiten Testzeitpunkt. Vor dem Hintergrund der Tatsache, dass die WET-Testung über eine Stunde dauert, spricht dies nach Krampen et al. gegen eine zu starke Belastung der Kinder durch die Testung mit dem WET.

5.2.4.1 Bewertung des Verfahrens

Der WET ist eine entwicklungstheoretisch begründete und testtheoretisch solide entwickelte Testbatterie, die die etablierten Standards für psychologische Testverfahren gut erfüllt und als zeitgemäß zu werten ist. Die Testkonstruktion erfolgte vorrangig auf Basis der probabilistischen Testtheorie. Darüber hinaus wurden klassische Testkennwerte

berechnet, zum einen zur Erleichterung der praktischen Anwendung, zum anderen zur methodischen Validierung.

Durchführungs-, Auswertungs-, und Interpretationsobjektivität sind weitestgehend sichergestellt. Die aus eigenen und unabhängigen Untersuchungen resultierenden Reliabilitätsschätzungen sprechen für eine zufriedenstellende Messgenauigkeit. Die inhaltliche sowie die differentielle Validität des WET können als sehr gut eingeschätzt werden. Unter Berücksichtigung der Tatsache, dass die Funktionsbereiche nicht voneinander unabhängig sind, belegen die durchgeführten Faktorenanalysen die Dimensionalität des Verfahrens erfreulich gut. Zudem enthält das Handbuch weitere Validitätsstudien, die erste Hinweise auf die kriterienbezogene und prognostische Validität des Verfahrens geben. Die Ergebnisse weiterführender Untersuchungen werden mit der Publikation der Neuauflage erwartet.

Einem ganzheitlichen Verständnis von Entwicklung folgend erhebt der WET den Anspruch, alle entwicklungsrelevanten Funktionsbereiche zu überprüfen. Demgemäß werden sechs theoretisch begründete Funktionsbereiche mittels dreizehn homogener eindimensionaler Subtests und einem Elternfragebogen erfasst. Die Ergebnisse von Krampen et al. (2008) legen nahe, dass die sechs Funktionsbereiche zur Abbildung der Gesamtentwicklung gut geeignet sind. Petermann und Macha (2003) heben jedoch kritisch hervor, dass die kognitive Entwicklung nur unzureichend inhaltlich repräsentiert wird. Diese wird bei Drei- und Fünfjährigen anhand von drei, bei Vierjährigen anhand von vier Subtests erfasst. Dieser Kritik wird in der geplanten Neuauflage des WET durch Aufnahme des neuen Subtests *Rechnen* und der Ausweitung des Subtests *Muster Legen* auf Fünfjährige begegnet, sodass zukünftig fünf Subtests die kognitive Entwicklung vier- und fünfjähriger Kinder erfassen.

Die Durchführungsdauer des WET ist mit circa 75 bis 90 Minuten im Vergleich zu anderen Entwicklungstests eher lang. Renziehausen (2003) befürchtet deshalb Motivationseinbußen seitens der Kinder, welche die Aussagekraft der Ergebnisse einschränken könnten. Eigene Erfahrungen sprechen allerdings gegen eine Verschlechterung der Motivation. Der WET erlaubt aufgrund seines kindgerechten Testmaterials und seines Abwechslungsreichtums eine spielerische Vorgabe, sodass nahezu alle getesteten Kinder anscheinend Spaß bei der Bearbeitung hatten und während der gesamten Durchführung hohes Engagement zeigten. Zudem weisen Kastner-Koller und Deimann (2002) darauf hin, dass die Vorgabe bei Anzeichen von Müdigkeit

abzubrechen und ein zusätzlicher Testtermin zu vereinbaren ist. Auch Krampen et al. (2008) kommen zu dem Schluss, dass die Belastung der Kinder durch die WET-Testung nicht allzu hoch ist.

Die Normierung erfolgte an einer Stichprobe von insgesamt 1245 Kindern und ist repräsentativ für Österreich und Deutschland. Im Zuge der geplanten Neuauflage wurden neue Normdaten in Österreich, Deutschland und der Schweiz erhoben, sodass die Aktualität der Normen auch weiterhin sichergestellt sein wird.

Zur Diagnostik der Gesamtentwicklung deutscher und österreichischer Vorschulkinder kann der WET schon derzeit uneingeschränkt empfohlen werden und ist gegenüber den anderen vorgestellten Entwicklungstests zu favorisieren. Es ist zu hoffen, dass die seit der Publikation der zweiten Auflage durchgeführten Untersuchungen und Reanalysen die Testgüte der Neuauflage weiterhin belegen und der Geltungsbereich des Verfahrens auf die Schweiz ausgedehnt werden kann.

6 Resümee aus den theoretischen Ausführungen

Das Ziel des theoretischen Teils bestand in der Darstellung der Grundlagen, Zwecke und Methoden der allgemeinen Entwicklungsdiagnostik sowie der kritischen Würdigung von solchen allgemein orientierten psychometrischen Entwicklungstestverfahren, die im deutschen Sprachraum vermehrt Anwendung finden. Der WET ist eines dieser Verfahren.

Im Folgenden werden die wichtigsten Inhalte, Aussagen und Schlussfolgerungen des theoretischen Teils zusammengefasst:

Im ersten Kapitel wird der Entwicklungsbegriff einer näheren Betrachtung unterzogen und hierauf die verschiedenen Theorien der Entwicklung skizziert. Weiters wird der Entwicklungsverlauf in der frühen Kindheit beschrieben:

- Zunächst wird gezeigt, dass weder eine klar abgrenzbare Definition des Begriffs Entwicklung noch eine allgemein gültige Entwicklungstheorie existiert. Mittlerweile besteht jedoch Übereinstimmung darin, dass Entwicklung durch das Zusammenwirken endogener, exogener und autogener Faktoren determiniert ist und diese sich wechselseitig bedingen. Entwicklung wird heute vorwiegend aus kontextueller Perspektive betrachtet und dem Individuum über die gesamte Lebensspanne hinweg aktiv gestaltender Einfluss zugesprochen.

Entwicklungsdiagnostische Verfahren sind entwicklungstheoretisch zu begründen. Da jedoch keine allgemein gültige Entwicklungstheorie besteht, basieren die Verfahren auf dem Entwicklungsverständnis der jeweiligen Autoren bzw. Autorinnen, wobei zumeist einzelne theoretische Aussagen und Erkenntnisse zu einer eklektizistisch zusammengebauten Theorie verschmolzen werden. Aufgrund der unterschiedlichen theoretischen Basis weisen die Verfahren auch einen unterschiedlichen Geltungsbereich auf.

- Die Entwicklung verläuft in der frühen Kindheit in allen Entwicklungsbereichen vergleichsweise kontinuierlich in Richtung zunehmender Kompetenz. Die in den ersten beiden Lebensjahren erworbenen Fähigkeiten werden nun ausdifferenziert und integriert, sodass hieraus neue Fähigkeiten erwachsen können. Ausgeprägte qualitative Veränderungen sind in diesem Altersabschnitt nicht zu erwarten (Kastner-Koller & Deimann, 2002). Während in den Jahren zuvor biologische und genetische Einflüsse dominieren und die Entwicklung deshalb interindividuell

eher einheitlich verläuft, gewinnen in der frühen Kindheit Umwelteinflüsse zunehmend an Bedeutung. Diese bewirkten eine größere interindividuelle Varianz der Entwicklung, was die Aussagekraft von Entwicklungsprognosen erhöht (McCall, 1987). Die Wirkung gezielter Einflussnahme auf das Entwicklungsgeschehen wird hierdurch ebenfalls verbessert.

Unter Berücksichtigung einer förderdiagnostischen Perspektive erscheint es sinnvoll, den Fokus entwicklungsdiagnostischer Begutachtungen auf die frühe Kindheit zu legen.

Das zweite Kapitel ist der Definition des Begriffs Entwicklungsdiagnostik, der Abgrenzung der allgemeinen von der speziellen Entwicklungsdiagnostik und den Zielsetzungen und Aufgaben der Entwicklungsdiagnostik gewidmet. Zudem wird näher auf die Geschichte der Entwicklungsdiagnostik eingegangen:

- Allgemein orientierte Entwicklungsdiagnostik erfasst mittels qualitativer und quantitativer Methoden entwicklungsbedingte Kompetenzen in mehreren grundlegenden Entwicklungsbereichen und zielt insbesondere auf die Beschreibung und Erklärung des aktuellen Kompetenzniveaus, auf die Ableitung von Entwicklungsinterventionen und auf die Prognose zukünftiger Entwicklung ab.
- Die Anfänge der Entwicklungsdiagnostik können auf zahlreiche Ergebnisse verschiedener Forschungsrichtungen des 19. Jahrhunderts zurückgeführt werden. Als Beginn der psychometrischen Entwicklungsdiagnostik wird die Stufenleiter der Intelligenz der französischen Autoren Binet und Simon (1905) angesehen.

Das dritte Kapitel enthält die Vorstellung der Methoden der Entwicklungsdiagnostik, nämlich Entwicklungsscreenings, Entwicklungstests, Verhaltensbeobachtung sowie die Sammlung der typischerweise mit dem Sachverhalt in Verbindung stehenden Informationen:

- Entwicklungsscreenings haben eine Filterfunktion im Vorfeld einer umfassenden entwicklungspsychologischen Diagnostik. Derartige Verfahren haben den Anspruch, die gesamte mögliche Risikopopulation zu erreichen und zielen darauf ab, all jene Kinder zu identifizieren, bei denen mit Entwicklungsverzögerungen zu rechnen ist.
- Die via regia der Entwicklungsbeurteilung sehen Deimann und Kastner-Koller (2007) in Entwicklungstests. Dennoch reicht es nicht aus, die kindlichen

Kompetenzen ausschließlich psychometrisch zu erfassen. Erst die Integration verschiedener multimethodal gewonnener Informationen ermöglicht eine ganzheitliche Betrachtung, welche sodann den Grundstein für eine präzise Diagnose legt und hieraus die Formulierung von Entwicklungsprognosen und die Ableitung von Entwicklungsinterventionen erlaubt.

- Petermann und Macha (2008) unterscheiden drei verschiedene Formen von Entwicklungstests: Stufenleiterverfahren, Inventare und Testbatterien. Stufenleitern basieren auf einer endogenistischen Sichtweise von Entwicklung und sind demnach als nicht zeitgemäß zu werten. Inventare genügen aufgrund der ihnen zugrunde liegenden Konstruktionsprinzipien nicht den Qualitätskriterien, die an psychologische Testverfahren zu stellen sind. Im Bereich der Entwicklungsdiagnostik ist deshalb Testbatterien der Vorzug zu geben.

Im vierten Kapitel werden jene Anforderungen erläutert, die psychometrische Verfahren und somit auch Entwicklungstestverfahren zu erfüllen haben. Hier werden die Testgütekriterien beschrieben und in Hinblick auf ihre Bedeutung für Entwicklungstestverfahren diskutiert. Das Kapitel endet mit Erläuterungen zur Auswahl und Durchführung von Entwicklungstestverfahren.

Das fünfte Kapitel ist der Beschreibung solcher allgemein orientierter Entwicklungs-screenings und –tests gewidmet, die im deutschen Sprachraum vermehrt Anwendung finden. Die Verfahren werden zudem kritisch beleuchtet:

- Von den hier präsentierten Screeningverfahren können sowohl das DESK 3-6 (Tröster et al., 2004) als auch die EVU (Melchers et. al, 2003) empfohlen werden, da beide Verfahren den Standards entsprechen, die psychologische Testverfahren zu erfüllen haben. Die DES (Barth, 2008) sind wegen ihrer mangelnden testtheoretischen und empirischen Grundlage zum Screening nach Entwicklungsauffälligkeiten ungeeignet.
- Die vorgestellten Entwicklungstests sind ebenfalls unterschiedlich zu bewerten: Die BUEVA (Esser und Wyschkon, 2002) und der WET (Kastner-Koller & Deimann (2002) sind als Testbatterie, der ET 6-6 (Petermann et al., 2004) als Inventar und die BSID II (Reuner et al., 2007) als Stufenleiterverfahren konzipiert. Dies spricht dafür, die BUEVA und den WET gegenüber dem ET 6-6 und den BSID II (ungeachtet des für die Verfahren vorgesehenen Altersbereichs)

zu bevorzugen. Vergleicht man den WET und die BUEVA, so ist der WET der BUEVA testtheoretisch eindeutig überlegen. Erstens beruht der WET sowohl auf der probabilistischen als auch auf der klassischen Testtheorie, die BUEVA hingegen wurde – wie auch alle anderen hier präsentierten Verfahren – nur klassisch testanalysiert. Zweitens überzeugt der WET hinsichtlich seiner Testgüte weitaus mehr als die BUEVA. Während zum WET überzeugende Befunde zur Objektivität, Reliabilität und Validität und zur Normierung des Verfahrens vorliegen, sind bei der BUEVA insbesondere die Reliabilität und die Normierung als problematisch zu werten. Daraus folgt, dass zur entwicklungs-psychologischen Begutachtung drei- bis sechsjähriger Kinder (für welche Altersspanne der WET konzipiert wurde) vorrangig der WET empfehlenswert ist.

- Wenn nun weiters davon ausgegangen wird, dass sowohl der ET 6-6 als auch die BSID II aufgrund ihrer Testform nicht zur Entwicklungsdiagnose empfohlen werden können, so muss man in Hinblick auf die Allgemeine Entwicklungsdiagnostik in der frühen Kindheit zu dem Schluss kommen, dass derzeit kein geeignetes Instrument zur allgemein orientierten Entwicklungsbeurteilung Zweijähriger existiert.

Die ausführlichen theoretischen Erläuterungen sollen vor allem jenes Wissen vermitteln, das für die Einschätzung der Qualität eines Entwicklungstestverfahrens notwendig ist. Weiters soll damit auch ein Blick auf jene Gebiete ermöglicht werden, deren Kenntnis der Durchführende einer in der frühen Kindheit angesetzten entwicklungsdiagnostischen Untersuchung bedarf. Dieses Wissen soll zudem dazu verhelfen, den WET innerhalb des komplexen Feldes der Allgemeinen Entwicklungsdiagnostik (qualitativ) einzuordnen.

EMPIRISCHER TEIL

7 Zielsetzung und Fragestellung der Untersuchung

Die vorliegende Arbeit entstand im Zusammenhang mit der geplanten Neuauflage des Wiener Entwicklungstests (WET; Kastner-Koller & Deimann, 1998, 2002, in Vorbereitung).

Vorrangiges Ziel war die Sammlung von Normierungsdaten für die dritte Auflage des WET. Eine Neunormierung ist angezeigt, da die geplante dritte Auflage zwei zusätzliche Subtests für vier- und fünfjährige Kinder vorsieht.

Die weitere Zielsetzung ergibt sich aufgrund der im theoretischen Teil geäußerten Kritik an den im deutschen Sprachraum verbreiteten Entwicklungstests. Kritisch wurde vor allem die Testgüte der Verfahren gesehen. Das Ziel ist demgemäß eine testanalytische Untersuchung des WET.

Die gegenständliche Untersuchung an einer lokalen Stichprobe von 164 Wiener Kindern im Alter von 4;0 bis 5;11 Jahren beinhaltet die Überprüfung der Konstruktvalidität anhand von zwei Aspekten sowie die Beurteilung der Reliabilität der neu entwickelten Subtests im Sinne einer methodischen Validierung.

- Hinsichtlich der Überprüfung der Konstruktvalidität interessiert die Frage, ob die theoretische Annahme steigender Testleistungen mit zunehmendem Alter bestätigt werden kann.
- Als zweiter Aspekt der Konstruktvalidität wird die Dimensionalität überprüft. Es stellt sich die Frage, inwieweit sich die Subtests den sechs Funktionsbereichen des WET faktoriell zuordnen lassen.
- Zudem ist die Reliabilität der neu entwickelten Subtests von Interesse. Demgemäß gilt es, die innere Konsistenz der Skalen zu überprüfen und Itemschwierigkeiten und Trennschärfen zu berechnen.

Zunächst wird im folgenden Kapitel der WET beschrieben, anschließend auf den Untersuchungsplan eingegangen. Nach Beschreibung der Stichprobe folgt die Darstellung der Ergebnisse.

8 Untersuchungsinstrument

Der Wiener Entwicklungstest (WET, Kastner-Koller & Deimann, 1998, 2002, in Vorbereitung) ist ein förderdiagnostisch orientierter Entwicklungstest zur Erfassung der allgemeinen Entwicklung von Kindern zwischen drei und sechs Jahren.

Die zweite Auflage des WET erfasst anhand von 14 Subtests den Entwicklungsstand in sechs relevanten Funktionsbereichen der kindlichen Entwicklung.

1 - Funktionsbereich Motorik

- Subtest *Lernbär*

Zur Erfassung der Feinmotorik wird das Kind ermuntert, einem Teddybären beim Anziehen zu helfen. Dazu sind ein Druckknopf und eine Gürtelschnalle zu schließen sowie ein Knoten und eine Masche zu binden.

- Subtest *Turnen*

Die Skala Turnen erfasst anhand von 10 zunehmend differenzierten Bewegungsabfolgen die grobmotorischen Fähigkeiten. So ist beispielsweise beim ersten Item ein Ball zu werfen und beim letzten ein Hampelmannsprung auszuführen.

2 – Funktionsbereich Visumotorik / Visuelle Wahrnehmung

- Subtest *Nachzeichnen*

Zur Feststellung der visumotorischen Fähigkeiten sind zunehmend komplexe geometrische Figuren von einer Vorlage abzuzeichnen.

- Subtest *Bilderlotto*

Dieser Subtest überprüft die differenzierte Raumlage-Wahrnehmung unabhängig von fein- und visumotorischer Geschicklichkeit. Das Material umfasst vier Bildtafeln mit jeweils sechs Feldern, auf denen jedes Mal dasselbe Grundmotiv dargestellt wird, jede der sechs Abbildungen sich jedoch in einem Detail von den anderen unterscheidet. Das Kind erhält entsprechende Kärtchen und hat diese den Feldern zuzuordnen.

3 - Funktionsbereich Lernen und Gedächtnis

- Subtest *Zahlen Merken*

Diese Skala erfasst das phonologische Gedächtnis. Dabei hat das Kind Zahlenfolgen zunehmender Länge zu reproduzieren. Die Vorgabe wird abgebrochen, wenn zwei Zahlenfolgen einer Länge nicht reproduziert werden können. Erfasst werden die Gedächtnisspanne sowie die Anzahl der vorgegebenen Zahlenreihen.

- Subtest *Schatzkästchen*

Zur Überprüfung des visuell-räumlichen Speichers dient ein Holzkästchen mit 20 (5x4) buntbemalten Schubladen, in welchem sechs Gegenstände versteckt sind. Jeder Gegenstand ist zudem in zweifacher Ausführung vorhanden. Dem Kind werden die versteckten Gegenstände in einer festgelegten Reihenfolge präsentiert. Anschließend wird ihm ein Gegenstand gezeigt, dessen Zwilling es im Kästchen wiederzufinden hat. Erfasst werden die unmittelbare Gedächtnisleistung, die Anzahl der Lerndurchgänge – bis alle Gegenstände in der richtigen Reihenfolge gefunden werden – und die Behaltensleistung nach 20 Minuten.

4 – Funktionsbereich Kognitive Entwicklung

Die kognitive Entwicklung wird durch zwei sprachliche und zwei nonverbale Skalen erfasst.

- Subtest *Muster Legen*

Dieser Subtest, welcher in der zweiten Auflage des WET für drei- und vierjährige Kinder vorgesehen ist, erfasst das räumliche Denken. Dabei sind neun vorgebaute und ein vorgelegtes Muster mit Mosaiksteinen nachzubauen.

- Subtest *Bunte Formen*

Anhand von Matrizenaufgaben wird das logisch schlussfolgernde Denken überprüft. Als Material dienen 10 Tafeln mit jeweils neun Feldern (3x3), auf denen Kombinationen der Grundelemente Kreis, Dreieck und Quadrat in drei Größen und drei Farben abgebildet sind. Ein Feld ist jeweils leer. Aufgabe des Kindes ist es, aus fünf vorgelegten Kärtchen jenes auszuwählen, welches die 3x3 Matrix richtig ergänzt. Diese Skala ist lediglich für Vier- und Fünfjährige vorgesehen.

- Subtest *Gegensätze*

Das analoge Denken wird mittels unvollständiger sprachlicher Analogien überprüft. Hierbei soll das Kind 15 vorgelesene Sätze sinnvoll beenden.

- Subtest *Quiz*

Diese Skala überprüft das Alltagswissen und die Orientierung des Kindes in seiner Lebenswelt. Hierzu hat es Fragen zu beantworten wie zum Beispiel „Warum soll man nicht auf einer Baustelle spielen?“. Drei-, Vier- und Fünfjährige erhalten Fragen mit altersangepasster Schwierigkeit.

5 - Funktionsbereich Sprache

Zum Funktionsbereich Sprache zählen neben den oben angeführten sprachlich-kognitiven Subtests *Gegensätze* und *Quiz* zwei weitere Untertests.

- Subtest *Wörter Erklären*

Zur Überprüfung der sprachlichen Begriffsbildung wird das Kind aufgefordert, 10 Wörter möglichst differenziert zu erklären.

- Subtest *Puppenspiel*

Diese Skala erfasst das Verständnis für grammatikalische Strukturen. Das Kind soll den Inhalt zunehmend komplexer Sätze mit einer Puppenfamilie nachspielen.

6 - Funktionsbereich Sozial-Emotionale Entwicklung

- Subtest *Fotoalbum*

Anhand von neun Fotos von Erwachsenen und Kindern, die unterschiedliche Emotionen zeigen, soll der mimische Gesichtsausdruck erkannt und das Gefühl entsprechend benannt werden.

- *Elternfragebogen*

Dieser Fragebogen umfasst 22 Items zur Selbständigkeitsentwicklung des Kindes, welche von der Bezugsperson des Kindes auf einer fünf-stufigen Skala zu bewerten sind. Zudem werden im Elternfragebogen demographischen Daten abgefragt.

Für die Neuauflage des WET wurde die oben beschriebene Skala *Muster Legen* für Vierjährige adaptiert und um Items für Fünfjährige erweitert. Die Muster werden Kindern beider Altersgruppen stets vorgebaut und sind zunehmend komplex. Dreijährige erhalten nach wie vor die ursprünglichen 10 Aufgaben. Die Itemsets sind entsprechend der jeweiligen Jahresgruppe unterschiedlich schwierig.

Weiters wurde zur Beurteilung erster mathematischer Kompetenzen vier- und fünfjähriger Kinder ein neuer Subtest entwickelt: *Rechnen* erfasst Zählfertigkeiten, das Verständnis von Ordinalität und erste arithmetische Fertigkeiten. Vier- und Fünfjährige erhalten Aufgaben mit altersangepasster Schwierigkeit.

Da die gegenständliche Untersuchung im Zuge der Normierung für die geplante dritte Auflage des WET lediglich an Kindern im Alter von 4;0 bis 5;11 erfolgte, wurde der in der zweiten Auflage des WET enthaltene Subtest *Muster Legen* durch den adaptierten Untertest ersetzt und zusätzlich die neue Skala *Rechnen* vorgegeben. Somit fand die Erhebung anhand von 13 Subtests der zweiten Auflage des WET und der zwei neu entwickelten Skalen statt.

Die Auswertung der ursprünglichen Subtests erfolgte entsprechend der Auswertungsanweisungen der zweiten Auflage des WET. Zunächst wurden je Subtest die Rohwerte zu Summenscores addiert und anhand der Normtabellen in standardisierte C-Werte transformiert. Zusätzlich wurde der Range, welcher die Variabilität des Entwicklungsprofils veranschaulicht, ermittelt und in Prozentrang-Werte umgewandelt. Schließlich wurde der Gesamtentwicklungsscore berechnet und in C-Werte transformiert.

Für die neu entwickelten Subtests *Muster Legen* und *Rechnen* wurden die richtigen und falschen Lösungen mit 1 bzw. 0 kodiert und jeweils zu Summenscores addiert.

Kapitel 5.2.4 enthält eine umfassende Beschreibung der testtheoretischen Gütekriterien des WET, sodass an dieser Stelle nicht mehr darauf eingegangen wird.

9 Durchführung der Untersuchung

Die empirische Untersuchung gliedert sich in folgende Abschnitte:

- Untersuchungsplanung,
- Datenerhebung,
- Rückmeldung der Ergebnisse an die Eltern der teilnehmenden Kinder,
- Auswertung und statistische Analyse.

Für die Normierung der dritten Auflage des Wiener Entwicklungstests (WET, Kastner-Koller & Deimann, 1998, 2002, in Vorbereitung) sollten in Wien 160 Kinder zwischen 4;0 und 5;11 Jahren getestet werden, welche zu gleichen Teilen auf die beiden Geschlechter und auf vier (in Halbjahresschritte differenzierte) Altersgruppen aufgeteilt werden sollten. Eine weitere Prämisse war, eine sozial geschichtete Stichprobe zu erhalten. Dies sollte durch Testungen in Kindergärten, lokalisiert in verschiedenen Wiener Gemeindebezirken, gewährleistet werden. Auch war zu gewährleisten, dass die Kinder über ausreichende Kenntnisse der deutschen Sprache verfügten.

Zur Durchführung der Datenerhebung in den Kindergärten der Stadt Wien war eine Bewilligung der Magistratsabteilung 10, Wiener Kindergärten (MA 10), vonnöten, sodass deren Leiterin Frau Mag.^a Minich im Juli 2008 schriftlich kontaktiert und um ihr Einverständnis zur geplanten Untersuchung gebeten wurde. Nach der Bewilligung der MA 10, einschließlich der Zuteilung von Kindergärten, wurden die Leiterinnen der ausgewählten Kindergärten der Stadt Wien telefonisch kontaktiert und die Untersuchung avisiert. Zudem wurde Kontakt zu einem katholischen Pfarrkindergarten sowie zu einem Vereinskinderarten der School Sisters of Notre Dame aufgenommen und das Einverständnis der Leiterinnen zur Durchführung der Untersuchung eingeholt.

Im persönlichen Gespräch mit den Kindergartenleiterinnen erhielten diese Informationen zur Untersuchung, die hierauf um Bereitstellung eines geeigneten Untersuchungsraums und um Unterstützung bei der Identifikation entsprechender Kinder gebeten wurden. In Folge wurden die Eltern der geeigneten Kinder schriftlich informiert und deren schriftliches Einverständnis eingeholt. Dies geschah in enger Zusammenarbeit mit dem Kindergartenpersonal.

Die Datenerhebung begann nach Erhalt des Testmaterials im Oktober 2008 und dauerte bis Jänner 2009. Vier Testleiterinnen waren an der Datenerhebung beteiligt: Katerina Lebo, Alexandra Propst, Katherina Rab und Petra Bircsak. Die Daten wurden in insgesamt 11 Kindergärten in 11 verschiedenen Wiener Gemeindebezirken erhoben, wobei die Testungen jeweils vormittags stattfanden, da dies laut Manual des Wiener Entwicklungstests (WET, Kastner-Koller & Deimann, 2002) die günstigste Tageszeit für die Vorgabe des WET ist. Im Anschluss der Testung eines jeden Kindes wurde der Elternfragebogen an dessen Eltern zur Komplettierung übermittelt.

Nach Abschluss der Datenerhebung und Auswertung der Testergebnisse erhielten die Eltern der getesteten Kinder eine schriftliche Rückmeldung über den Entwicklungsstand ihres Kindes. Bei ausgeprägten Defiziten wurde den Eltern empfohlen, eine kostenlose Testwiederholung inklusive Beratungsgespräch am Zentrum für kinder-, jugend- und familienpsychologische Intervention der Fakultät für Psychologie (Universität Wien) in Anspruch zu nehmen.

Schließlich wurden die Testergebnisse in das Statistikprogramm *SPSS 15.0 for Windows* eingegeben, mit welchem die statistische Auswertung vorgenommen wurde.

10 Stichprobenbeschreibung

Insgesamt wurden 167 Kinder getestet. Die Daten von drei Kindern wurden von der Datenanalyse ausgeschlossen, da das Untersuchungsinstrument in diesen Fällen nur zu einem geringen Teil vorgegeben werden konnte. In einem Fall bestand der Grund hierfür in einem Testabbruch seitens der Testleiterin, da das betreffende Kind stark auffälliges Verhalten zeigte und somit als nicht repräsentativ für die Erhebung eingestuft wurde. In den beiden anderen Fällen erfolgte ein Testabbruch seitens der Kinder, die eine weitere Testung verweigerten. Somit gingen die Daten von 164 Kindern in die statistische Analyse ein.

10.1 Die Kindergärten der Kinder

Die Kinder, die an der Erhebung teilnahmen, stammen von 11 Kindergärten. Die angestrebte regionale Streuung konnte aufgrund der Lokalisation der Kindergärten in 11 Wiener Gemeindebezirken erreicht werden. In Tabelle 1 werden diese Kindergärten und die Anzahl der jeweils getesteten Kinder angeführt.

Tabelle 1: Kindergärten und Anzahl der getesteten Kinder

	Häufigkeit	Prozent
Fruchtgasse, 1020 Kindergarten der Stadt Wien	4	2.4
Lustgasse, 1030 Kindergarten der Stadt Wien	14	8.5
Margaretengürtel, 1050 Kindergarten der Stadt Wien	23	14.0
Faßziehergasse, 1070 Kindergarten der Stadt Wien	3	1.8
Währinger Gürtel, 1090 Kindergarten der Stadt Wien	28	17.1
Gudrunstraße, 1100 Kindergarten der Stadt Wien	17	10.4
Albrechtsbergergasse, 1120 Kindergarten der Pfarre Meidling	2	1.2
Friesgasse, 1150 Vereinskindergarten der SSND	18	11.0
Arnethgasse, 1160 Kindergarten der Stadt Wien	16	9.8
Gersthoferstraße, 1180 Kindergarten der Stadt Wien	26	15.9
Radelmayergasse, 1190 Kindergarten der Stadt Wien	13	7.9

52 (31.7 %) Kinder besuchten den Kindergarten halbtags, 86 (52.4 %) ganztags und 15 (9.1 %) Kinder in Teilzeit, d. h. bis circa 14:00 Uhr. Bei 11 (6.7 %) der 164 Kinder fehlten die Angaben zur täglichen Länge des Kindergartenbesuchs.

10.2 Alter und Geschlecht der Kinder

Die Kinder waren zum Testzeitpunkt zwischen 4;0 und 5;11 Jahre alt und wurden in vier Halbjahresgruppen eingeteilt.

83 (50.6 %) der Kinder waren männlichen, 81 (49.6 %) weiblichen Geschlechts.

Die Verteilung der Geschlechter auf die Altersgruppen ist in Tabelle 2 festgehalten.

Tabelle 2: Aufteilung der Kinder nach Altersgruppe und Geschlecht

		Altersgruppe				Total
		4;0 – 4;5	4;6 – 4;11	5;0 – 5;5	5;6 – 5;11	
Geschlecht	männlich	20	21	21	21	83
	weiblich	20	20	20	21	81
Total		40	41	41	42	164

10.3 Familiärer Hintergrund der Kinder

Angaben zum familiären Umfeld wurden zum überwiegenden Teil mittels des Elternfragebogens erhoben. Von 164 ausgegebenen Fragebögen wurden 153 retourniert, wobei einige soziodemographische Angaben zudem nicht vollständig ausgefüllt waren. Der Rest der Informationen wurde durch Befragung des Kindergartenpersonals gewonnen.

Von 141 Müttern und 137 Vätern konnte das Alter in Erfahrung gebracht werden. Die jüngste Mutter war 23, die älteste 47 Jahre alt, der Mittelwert lag bei 35 Jahren. Das Alter der Väter lag zwischen 24 und 60 Jahren, der Mittelwert betrug 38 Jahre.

Alle Mütter, bei denen entsprechende Angaben vorhanden waren (das waren 153), wohnten im selben Haushalt wie das untersuchte Kind. 24 (14.6 %) von den 149 Vätern, über welche die Information gewonnen werden konnte, lebten nicht mit ihrem Kind zusammen.

Der überwiegende Teil der Mütter (43.9 %) ging einer mittelhoch qualifizierten Tätigkeit als Angestellte, Beamtin oder Facharbeiterin nach. 14.6 % übten einen hochqualifizierten Beruf aus oder waren in leitender Position tätig. 13.4 % waren in einfachen Beschäftigungsverhältnissen. 6.7 % waren selbständig oder freiberuflich tätig. 16.5 % waren nicht berufstätig (Hausfrauen, in Ausbildung, Pensionistinnen oder Arbeitslose). Bei 4.9 % fehlten die Angaben zur Berufstätigkeit.

Auch die meisten Väter (39 %) übten eine mittelhoch qualifizierte Tätigkeit aus. Mehr Väter als Mütter (18.3 %) gingen einer einfachen Beschäftigung nach. 16.5 % übten eine hochqualifizierte oder leitende Tätigkeit aus. 14.6 % waren Selbständige oder Freiberufler. Im Gegensatz zu den Müttern waren lediglich 2.4 % nicht berufstätig. Bei 9.1 % fehlten wiederum entsprechende Angaben.

78 Kinder (47.6 %) wohnten mit einem Geschwister, 24 (14.6 %) mit zwei Geschwistern, vier (2.4 %) mit drei Geschwistern und ein Kind (0.6 %) mit vier Geschwistern in einem Haushalt. 47 Kinder (28.7 %) waren Einzelkinder. Für 10 Kinder (6.1 %) fehlten entsprechende Informationen.

Informationen zur Muttersprache der Kinder stammen von den Kindergartenpädagoginnen. 92 der Kinder (56.1 %) haben Deutsch als Muttersprache. In Tabelle 3 sind die unterschiedlichen Muttersprachen der Kinder angeführt.

Tabelle 3: Muttersprache der Kinder

	Häufigkeit	Prozent
deutsch	92	56.1
bosnisch, kroatisch, serbisch	35	21.3
türkisch	13	7.9
arabisch	10	6.1
polnisch	3	1.8
asiatisch	3	1.8
russisch	3	1.8
französisch	1	0.6
sonstige	4	2.4

10.4 Der sozioökonomische Status der Kinder

Wien hat insgesamt 23 Gemeindebezirke. Dem Statistischen Jahrbuch der Stadt Wien 2008 ist zu entnehmen, dass 2006 der höchste durchschnittliche Jahresnettoeinkommen im 1. Bezirk (EUR 29 234) erzielt wurde, gefolgt vom 19. Bezirk mit durchschnittlich EUR 23 716. Das niedrigste Durchschnitts-Jahresnettoeinkommen fand sich im 15. Bezirk (EUR 15 527), das zweitniedrigste im 20. Bezirk mit EUR 16 339. Testungen fanden in jenen Bezirken statt, die in der absteigenden bzw. ansteigenden Rangreihung Platz 2 belegten. Darüber hinaus fanden Testungen in neun weiteren Bezirken statt, in denen das durchschnittliche Jahresnettoeinkommen zwischen EUR 17 170 (16. Bezirk) und EUR 22 139 (18. Bezirk) lag. Betrachtet man das durchschnittliche Einkommen als relevanten Indikator des sozialen Status, kann angenommen werden, dass alle Schichten erreicht wurden.

Ein weiteres Indiz für die soziale Schichtzugehörigkeit eines Bürgers ist die Berufstätigkeit. In Anlehnung an Kastner-Koller und Deimann (2002) ist zur Beschreibung der Schichtzugehörigkeit einer Familie insbesondere die Berufstätigkeit des Vaters relevant, bei Alleinerziehenden die der Mutter. Des Weiteren werden hier berufstätige Mütter als Hauptverdienende klassifiziert, wenn der im selben Haushalt lebende Vater arbeitslos war. Tabelle 4 ist zu entnehmen, welchen Tätigkeiten die Hauptverdienenden nachgingen.

Tabelle 4: Berufstätigkeit der Hauptverdienenden

	Häufigkeit	Prozent
Selbständige, freie Berufe	22	13.4
Angestellte/Beamte: hochqualifizierte oder leitende Tätigkeit	29	17.7
Angestellte/Beamte: mittlere Tätigkeit oder Facharbeiter	70	42.7
Angestellte/Beamte: einfache Tätigkeit, Hilfsarbeit, angelernte Arbeit	31	18.9
in Ausbildung, Pension, Arbeitslos	4	2.4
keine Angabe	8	4.9

Die Verteilung der Berufstätigkeit der Hauptverdienenden gibt einen weiteren Hinweis darauf, dass alle Schichten erfasst wurden. Da insbesondere die Kategorie der mittleren Tätigkeit, welche Berufe der oberen Unterschicht (Facharbeiter) und der unteren bis

mittleren Mittelschicht (mittlere Beamte und Angestellte) umfasst, am häufigsten besetzt ist, kann davon ausgegangen werden, dass die Stichprobe die Grundgesamtheit weitgehend repräsentiert.

10.5 Die allgemeine Entwicklung der Kinder der Gesamtstichprobe

Um die Stichprobe weiter zu beschreiben, sind in Tabelle 5 die Ergebnisse der Kinder im WET anhand wichtiger statistischer Kennwerte angeführt. Die neu entwickelten Skalen wurden zur Beschreibung des Entwicklungsstandes nicht herangezogen, da hierfür keine Normwerte zur Verfügung stehen.

C-Werte von 4 bis 6 weisen auf eine normale Entwicklung in der entsprechenden Fähigkeitsdimension hin. Im Durchschnitt sind die Kinder der Gesamtstichprobe demgemäß in allen Bereichen als normal entwickelt anzusehen. Der geringste Mittelwert ergab sich beim Subtest *Gegensätze* ($M = 4.21$; $SD = 2.03$), der höchste beim Untertest *Wörter Erklären* ($M = 5.89$; $SD = 1.90$). Auch die Variabilität des Entwicklungsprofils, ausgedrückt durch den Range, war im Mittel ausgewogen ($M = 51.69$; $SD = 26.71$).

Minimum und Maximum verdeutlichen, dass die Kinder der Gesamtstichprobe sowohl weit unterdurchschnittliche als auch weit überdurchschnittliche Leistungen erbrachten.

Tabelle 5: Durchschnittliche Leistungen der Gesamtstichprobe im WET

	N	M	SD	Min.	Max.	Md
C-Wert Lernbär	164	5.48	1.195	0	9	5
C-Wert Quiz	164	4.59	2.242	0	10	5
C-Wert Bilderlotto	164	5.01	1.750	0	10	5
C-Wert Puppenspiel	164	4.72	2.047	0	10	4
C-Wert Schatzkästchen	164	5.49	1.974	1	10	5
C-Wert Bunte Formen	164	5.18	2.368	1	9	5
C-Wert Zahlen Merken	164	4.26	2.033	0	10	4
C-Wert Wörter Erklären	164	5.89	1.903	1	10	6
C-Wert Turnen	163 ^a	4.66	1.389	0	8	5
C-Wert Nachzeichnen	164	4.73	1.673	0	9	5
C-Wert Gegensätze	164	4.21	2.498	0	10	4
C-Wert Fotoalbum	164	5.18	1.874	0	10	5
C-Wert Elternfragebogen	144 ^b	5.22	1.973	0	10	5
C-Wert Gesamtentwicklungsscore	164	4.86	1.631	1	9	5
Prozentrang Range	164	51.69	26.714	4	98	59

^a N ist reduziert auf 163, da ein Kind diesen Subtest verweigerte.

^b N ist reduziert auf 144, da nicht alle Elternfragebögen retourniert wurden.

Die Gesamtstichprobe entspricht somit weitgehend der Normierungsstichprobe der zweiten Auflage des WET, welche entsprechend einer C-Skala einen Mittelwert von 5 und eine Standardabweichung von 2 aufwies.

10.6 Entwicklungsunterschiede zwischen Subgruppen der Stichprobe

Zur Überprüfung der Repräsentativität der Stichprobe wurden verschiedene Subgruppen der Stichprobe einer genaueren Analyse unterzogen.

10.6.1 Kinder unterschiedlicher Muttersprache

Wie bereits oben ausgeführt, waren 92 Kinder deutsche Muttersprachler, 72 Kinder hatten eine andere Muttersprache. Es stellte sich demnach die Frage, ob Leistungsunterschiede zwischen Kindern mit deutscher und nicht-deutscher Muttersprache bestehen.

Für die Unterschiedsberechnungen wurde der Mann-Whitney-U-Test herangezogen, da die Voraussetzungen für einen t-Test nur suboptimal gegeben sind (siehe Tabelle A-1 und A-2 im Anhang): Varianzhomogenität ist mit Ausnahme des Subtests *Lernbär* zwar hinsichtlich aller abhängigen Variablen gegeben, der Kolmogorov-Smirnov-Test zeigt jedoch für keinen Subtest Normalverteilung pro Gruppe.

Als Signifikanzniveau wurde ein Alpha von .05 gewählt.

Abhängige Variablen sind die im WET erzielten Leistungen, ausgedrückt in C-Werten bzw. Prozentrangwerten. Als unabhängige Variable dienen die beiden Sprachgruppen.

In den vier sprachlichen Subtests und bezüglich des Gesamtentwicklungsscores zeigen sich signifikante Unterschiede zwischen Kindern mit deutscher und nichtdeutscher Muttersprache. Der Vergleich der mittleren Ränge und Mediane verdeutlicht, dass Kinder mit deutscher Muttersprache bessere Leistungen erbrachten als Kinder mit anderer Muttersprache (siehe Tabelle 6).

Die Unterschiede in den Testleistungen zwischen Kindern mit deutscher und anderer Muttersprache sind vermutlich darauf zurückzuführen, dass der WET für Kinder mit deutscher Muttersprache konzipiert wurde (Kastner-Koller & Deimann, 2002, S. 11). Kinder mit nicht-deutscher Muttersprache scheinen bei ansonsten vergleichbarer motorischer, kognitiver und sozial-emotionaler Entwicklung einen Nachteil bei der

Bearbeitung der sprachlichen Subtests zu haben. Die geringeren Leistungen in den sprachlichen Subtests wirken sich auf den Gesamtentwicklungsscore aus.

Tabelle 6: Signifikante Unterschiede bei den WET-Ergebnissen zwischen Kindern mit deutscher und anderer Muttersprache

	Mutter-sprache	n	Mittlerer Rang	Md	U-Statistik	p*
Quiz	deutsch	92	100.33	5.0	U = 1671.500	< .001
	andere	72	59.72	4.0	Z = -5.492	
Puppenspiel	deutsch	92	96.37	5.0	U = 2036.000	< .001
	andere	72	64.78	4.0	Z = -4.302	
Wörter erklären	deutsch	92	99.33	7.0	U = 1764.000	< .001
	andere	72	61.00	5.0	Z = -5.193	
Gegensätze	deutsch	92	105.80	5.5	U = 1168.500	< .001
	andere	72	52.73	2.0	Z = -7.158	
Gesamtentwicklungsscore	deutsch	92	99.02	5.0	U = 1792.500	< .001
	andere	72	61.40	4.0	Z = -4.5133	

*Exakte Signifikanz: zweiseitig

10.6.2 Geschlechtsunterschiede

Im Zuge der Normierung der zweiten Auflage des WET zeigten sich keine praktisch bedeutsamen Entwicklungsunterschiede zwischen Mädchen und Jungen (Kastner-Koller & Deimann, 2002, S. 22). Demnach stellte sich die Frage, ob Leistungsunterschiede zwischen den Geschlechtern innerhalb der Stichprobe der vorliegenden Untersuchung bestehen.

Die Auswertung der Fragestellung erfolgte aufgrund der suboptimalen Erfüllung der Voraussetzungen für ein parametrisches Verfahren mittels Mann-Whitney-U-Test: Normalverteilte Daten sind hinsichtlich keiner abhängigen Variable gegeben (siehe Tabelle A-3 im Anhang), die Voraussetzung der Varianzhomogenität wurde beim Subtest *Bunte Formen* nicht erfüllt (vgl. Tabelle A-4 im Anhang). Das Signifikanzniveau wurde auf $p = .05$ festgelegt.

Abhängige Variablen sind die WET-Leistungen (C-Werte bzw. Prozentrangwerte), unabhängige Variablen die beiden Geschlechter.

Es zeigte sich lediglich in Bezug auf den Subtest *Lernbär* ein signifikanter Unterschied zwischen den Geschlechtern ($U = 2347.500$; $Z = -3.560$; $p = < .001$). Die Mädchen ($Md = 5$) erzielten in diesem Subtest signifikant bessere Leistungen als die Jungen ($Md = 6$).

10.6.3 Leistungsunterschiede zwischen Kindern in Abhängigkeit vom Testleiter

Da die Testungen von vier verschiedenen Testleiterinnen durchgeführt wurden, ergab sich die Frage nach möglichen Testleitereffekten. Um festzustellen, ob bestehende Unterschiede in den Testleistungen zwischen Kindern, die von unterschiedlichen Testleiterinnen getestet wurden, überhaupt Aussagekraft besitzen, wurden verschiedene Testbedingungen vorab einer näheren Betrachtung unterzogen.

Wie in Kapitel 4.6.1 beschrieben, zeigen sich Leistungsunterschiede in den sprachlichen Subtests zwischen Kindern mit deutscher und anderer Muttersprache. Demnach erschien interessant, wie viele Kinder der jeweiligen Sprachgruppe von den einzelnen Testleiterinnen getestet wurden und ob die beiden Sprachgruppen hinsichtlich der gezeigten Leistungen gleich auf die Testleiter verteilt sind. Hierfür wurde eine Kreuztabelle erstellt (Tabelle 7).

Tabelle 7: Verteilung der Kinder nach Muttersprache und Testleiterin

		Testleiterin WET				Total
		Bircsak	Lebo	Propst	Rab	
Mutter-sprache	deutsch	31 75.6 %	20 48.8 %	31 70.5 %	10 26.3 %	92 56.1 %
	andere	10 24.4 %	21 51.2 %	13 29.5 %	28 73.7 %	72 43.9 %
Total		41 100 %	41 100 %	44 100 %	38 100 %	164 100 %

Frau Lebo testete annähernd gleich viele Kinder beider Sprachgruppen. Frau Bircsak und Frau Propst testeten jeweils 31 Kinder mit deutscher Muttersprache und zu einem wesentlich geringeren Anteil Kinder anderer Muttersprache. Bei Frau Rab zeigt sich ein gegenteiliges Bild: Sie testete überwiegend Kinder mit nicht-deutscher Muttersprache. Um diese scheinbare Ungleichmäßigkeit statistisch zu überprüfen, wurde ein Chi-Quadrat-Unabhängigkeitstest berechnet. Als Zusammenhangsmaß wurde Cramer's V gewählt. Der Test fällt signifikant aus ($V = .387$; $p < .001$) und verdeutlicht, dass die Sprachgruppen auf die Testleiter ungleich verteilt sind.

Um eine sozial geschichtete Stichprobe zu erhalten, fanden die Testungen in 11 Bezirken Wiens statt. In Tabelle 8 ist zu sehen, in welchen Bezirken die Testleiterinnen die Tests durchgeführt haben.

Tabelle 8: Verteilung der Testleiterinnen auf die Kindergärten

	Häufigkeit	Testleiter
Fruchtgasse, 1020	4	Katharina Rab
Lustgasse, 1030	14	Katharina Rab
Margaretengürtel, 1050	23	Katarina Lebo
Faßziehergasse, 1070	3	Katharina Rab
Währinger Gürtel, 1090	28	Alexandra Propst
Gudrunstraße, 1100	17	Katharina Rab
Albrechtsbergergasse, 1120	2	Petra Bircsak
Friesgasse, 1150	18	Katarina Lebo
Arnethgasse, 1160	16	Alexandra Propst
Gersthoferstraße, 1180	26	Petra Bircsak
Radelmayergasse, 1190	13	Petra Bircsak

Zudem wurde der kategorisierte Beruf des hauptverdienenden Elternteils mit den verschiedenen Testleiterinnengruppen in Beziehung gesetzt und wiederum eine Kreuztabelle (Tabelle 9) erstellt.

Tabelle 9: Verteilung der Kinder nach Berufstätigkeit der Hauptverdiener und Testleitergruppen

		Testleiter WET				Total
		Bircsak	Lebo	Propst	Rab	
Beruf des Hauptverdieners	Selbständige, freie Berufe	6 14.6 %	6 14.6 %	6 13.6 %	4 10.5 %	22 13,4 %
	Angestellte/Beamte: hochqualifizierte oder leitende Tätigkeit	11 26.8 %	2 4.9 %	13 29.5 %	3 7.9 %	29 17,7 %
	Angestellte/Beamte: mittlere Tätigkeit oder Facharbeiter	19 46.3 %	18 43.9 %	21 47.7 %	12 31.6 %	70 42,7 %
	Angestellte/Beamte: einfache Tätigkeit, Hilfsarbeit, angelernte Arbeit	3 7.3 %	9 22.0 %	4 9.1 %	15 39.5 %	31 18,9 %
	in Ausbildung, Pension, Arbeitslos	1 2.4 %	3 7.3 %	0 0 %	0 0 %	4 2,4 %
	keine Angabe	1 2.4 %	3 7.3 %	0 0 %	4 10.5 %	8 4,9 %
Total		41 100 %	41 100 %	44 100 %	38 100 %	164 100 %

Bei zeilenweiser Betrachtung der Tabelle fällt insbesondere in den Kategorien „hochqualifizierte oder leitende Tätigkeit“ und „einfache Tätigkeit“ ein Ungleichgewicht zwischen den Testleiterinnen auf. Infolgedessen wurde wiederum ein Chi-Quadrat-Unabhängigkeitstest durchgeführt. Dieser zeigt einen signifikanten Zusammenhang

zwischen den beiden Variablen ($V = .280$; $p = .001$): Die Verteilung der Berufsgruppen der Hauptverdienenden auf die Testleiterinnen unterscheidet sich signifikant von der erwarteten Verteilung.

Zusammenfassend lässt sich feststellen, dass die Kinder, die von den einzelnen Testleiterinnen getestet wurden, sehr heterogen sind. Dies deutet darauf hin, dass eine Unterschiedsanalyse (Varianzanalyse bzw. Kruskal-Wallis-Test) bezüglich erbrachter Testleistungen zwischen Kindern, die von verschiedenen Testleiterinnen getestet wurden, zu einem verzerrten Ergebnis führen würde. Demnach wird auf eine derartige Analyse verzichtet. Auch die Berechnung einer dreifaktoriellen Varianzanalyse mit den unabhängigen Variablen „Muttersprache deutsch/nicht deutsch“, „besuchter Kindergarten“ und „Berufskategorie der Hauptverdienenden“ erscheint unzweckmäßig. Zwar könnten hiermit die vermuteten Interaktionen überprüft werden, der Stichprobenumfang ist hierfür jedoch zu gering. Testleitereffekte können somit weder ausgeschlossen noch verifiziert werden.

10.7 Zusammenfassende Beurteilung der Stichprobe

Die Stichprobe entspricht dem Untersuchungsplan in Hinblick auf den angestrebten Stichprobenumfang, die Alters- und Geschlechtsverteilung und die regionale Streuung der durchgeführten Testungen. Die dargestellten soziodemographischen Daten deuten darauf hin, dass die Stichprobe mit der Grundgesamtheit der Wiener Bevölkerung gut übereinstimmt. Zudem entsprechen die durchschnittlichen Leistungen der Kinder der Gesamtstichprobe den Erwartungen an eine im Mittel normal entwickelte Stichprobe. Ebenso wie bei der Normierungsstichprobe der zweiten Auflage des Wiener Entwicklungstests (WET, Kastner-Koller & Deimann, 2002) zeigen sich keine praktisch bedeutsamen Geschlechtsunterschiede. Auffällig sind jedoch die signifikanten Leistungsunterschiede zwischen Kindern mit deutscher und nicht-deutscher Muttersprache in den sprachlichen Subtests. Ob Unterschiede zwischen Kindern auf Testleitereffekte zurückzuführen sind, lässt sich aufgrund der Heterogenität der Kinder, die von den einzelnen Testleiterinnen getestet wurden, nicht überprüfen.

11 Konstruktvalidität

Konstruktvalidität besteht, wenn ein Test die inhaltlichen Vorannahmen in Bezug auf ein theoretisches Konstrukt auch tatsächlich erfüllt. In einem eng gefassten Sinn zählen hierzu lediglich konvergente, diskriminante und faktorielle Validität. Einem weiter gefassten Begriffsverständnis folgend sind zur Bestimmung der Konstruktvalidität auch nicht-korrelative Methoden anwendbar, die zusätzliche Aussagen über den Erfüllungsgrad der theoretischen Annahmen ermöglichen (vgl. zu Konstruktvalidität auch Kapitel 4.3.3).

Eine theoretische Annahme, die allen Entwicklungstestverfahren zugrunde liegt, sind stetig zunehmende Testleistungen mit fortschreitendem Alter (Petermann & Macha, 2005). Demgemäß gilt es zu überprüfen, ob sich beim Wiener Entwicklungstest (WET; Kastner-Koller & Deimann, 1998, 2002, in Vorbereitung) Leistungsunterschiede zwischen verschiedenen Altersgruppen in postulierter Richtung zeigen.

Allgemeine Entwicklungstests zielen auf die Erfassung mehrerer Entwicklungsbereiche ab. Aufgrund theoretischer Überlegungen wurden sechs Funktionsbereiche für den WET festgelegt und 15 Subtests wie auch ein Elternfragebogen entwickelt. Zur Überprüfung der Dimensionalität des Verfahrens werden die Subtests des WET einer Faktorenanalyse unterzogen.

11.1 Steigende Alterstrends

Kastner-Koller & Deimann (2002) berichten im Handbuch der zweiten Auflage des WET einen deutlichen Alterstrend, wobei signifikante Unterschiede sowohl zwischen den Jahres- als auch den Halbjahresgruppen beobachtet wurden.

Es stellte sich daher die Frage, ob sich diese Alterstrends weiterhin nachweisen lassen und weiters ob Unterschiede in den Subtestleistungen zwischen Vier- und Fünfjährigen wie auch zwischen den Halbjahresgruppen bestehen.

11.1.1 Methode

Der Überprüfung der Altersunterschiede wurden die Summenscores der einzelnen Subtests zugrunde gelegt. Für den Subtests *Schätzkästchen*, der eine dreistufige Auswertung vorsieht, wurde für jede Subtestkomponente eine eigene Analyse durchgeführt.

Zunächst wurden die Voraussetzungen zur Anwendung parametrischer Verfahren überprüft: Die Varianzhomogenität wurde mit dem Levene-Test geprüft. Die Überprüfung der Normalverteilung erfolgte je nach Gruppengröße mittels Kolmogorov-Smirnov-Test oder Shapiro-Wilk-Test. Letzterer liefert vergleichsweise präzisere Ergebnisse, sofern die Gruppen mit weniger als 50 Fällen besetzt sind (Field, 2005; Janssen & Laatz, 2007). Die Ergebnisse der Überprüfung der Voraussetzungen finden sich im Anhang (Tabellen A-5 bis A-10).

Zur Feststellung von Gruppenunterschieden bei Berechnung einer Varianzanalyse wurden a-priori Kontraste festgelegt und anschließend berechnet. Auch wurden die Daten mittels polynomialer Kontraste hinsichtlich eines linearen Trends überprüft. Kontrasten wurde gegenüber Post-Hoc-Tests der Vorzug gegeben, da eine gerichtete Fragestellung vorliegt und Kontraste vergleichsweise mehr Power aufweisen als Post-Hoc-Prozeduren (Field, 2005).

Zur Überprüfung von Gruppenunterschieden bei Berechnung eines Kruskal-Wallis-Tests konnte aufgrund der einseitigen Fragestellung ein Jonckheere-Terpstra-Test durchgeführt werden. Diesem Test wurde gegenüber multiplen Mann-Whitney-U-Tests Vorzug gegeben, da hierfür keine Alpha-Adjustierung vonnöten ist (Field, 2005). Derartige Korrekturen haben zur Folge, dass mit kleiner werdendem Alpha der Beta-Fehler zunimmt.

Das Signifikanzniveau wurde mit $p = .05$ festgelegt.

Ergebnisse, die den Erwartungen aufgrund der gerichteten Fragestellung widersprechen, werden in allen Tabellen dieses Kapitels fett gedruckt.

11.1.2 Unterschiede in den Testleistungen zwischen Vier- und Fünfjährigen

Die Analyse bezüglich der Unterschiede zwischen Vier- und Fünfjährigen erfolgte nur für Subtests, die dieselben Aufgaben für alle Altersgruppen vorsehen.

Zur Auswertung wurde der parameterfreie Mann-Whitney-U-Test herangezogen, da die Voraussetzungen für einen t-Test nicht bzw. nur suboptimal gegeben sind (siehe Tabelle A-5 und A-6 im Anhang): Der Kolmogorov-Smirnov-Test ergibt für keinen Subtest eine Normalverteilung in beiden Altersgruppen, zudem sind die Varianzen teilweise nicht homogen.

Abhängige Variablen sind die erzielten Rohwerte im WET, unabhängige die zwei Altersgruppen.

Tabelle 10: Subtestleistungen Vierjähriger und Fünfjähriger: Ergebnis der U-Tests

Subtest	Alter	n	Mittlerer Rang	Md	U-Statistik	p*
Lernbär	4;0 – 4;11	81	69.55	3	U = 2312.500	<.001
	5;0 – 5;11	83	95.14	3	Z = -4.549	
Bilderlotto: <i>Langform</i>	4;0 – 4;11	79	61.72	13	U = 1715.500	<.001
	5;0 – 5;11	83	100.33	17	Z = -5.248	
Puppenspiel	4;0 – 4;11	81	65.41	8	U = 1977.000	<.001
	5;0 – 5;11	83	99.18	10	Z = -4.587	
Schatzkästchen: <i>Unmittelbares Reproduzieren</i>	4;0 – 4;11	81	68.04	3	U = 2190.500	<.001
	5;0 – 5;11	83	96.61	4	Z = -3.957	
Schatzkästchen: <i>Behaltensleistung nach 20 Minuten</i>	4;0 – 4;11	81	81.66	6	U = 3293.500	.399
	5;0 – 5;11	83	83.32	6	Z = -.257	
Schatzkästchen: <i>Anzahl der Lerndurchgänge</i>	4;0 – 4;11	81	99.98	5	U = 1946.000	<.001
	5;0 – 5;11	83	65.45	2	Z = -4.727	
Bunte Formen	4;0 – 4;11	81	71.36	4	U = 2459.000	.001
	5;0 – 5;11	83	93.37	7	Z = -2.986	
Zahlen merken: <i>Gedächtnisspanne</i>	4;0 – 4;11	81	73.60	3	U = 2640.500	.006
	5;0 – 5;11	83	91.19	4	Z = -2.523	
Wörter erklären	4;0 – 4;11	81	68.16	12	U = 2200.000	<.001
	5;0 – 5;11	83	96.49	14	Z = -3.836	
Turnen	4;0 – 4;11	80	56.36	6	U = 1268.500	<.001
	5;0 – 5;11	83	106.72	8	Z = -6.912	
Nachzeichnen	4;0 – 4;11	81	63.38	5	U = 1812.500	<.001
	5;0 – 5;11	83	101.16	7	Z = -5.192	
Gegensätze	4;0 – 4;11	81	66.40	10	U = 2057.000	<.001
	5;0 – 5;11	83	98.22	12	Z = -4.313	
Fotoalbum	4;0 – 4;11	81	66.80	5	U = 2089.500	<.001
	5;0 – 5;11	83	97.83	7	Z = -4.263	
Elternfragebogen	4;0 – 4;11	71	53.58	83	U = 1248.500	<.001
	5;0 – 5;11	73	90.90	95	Z = -5.370	

* Exakte Signifikanz: einseitig

Wie in Tabelle 10 zu sehen ist, zeigen sich überwiegend signifikante Unterschiede in den Subtestleistungen zwischen den beiden Altersgruppen. Erwartungsgemäß erzielten die Fünfjährigen bessere Leistungen als die Vierjährigen. Hinsichtlich der Subtestkomponente *Schatzkästchen: Behaltensleistung nach 20 Minuten* ($Z = -.257$, $p = .399$) fällt der Leistungsunterschied nicht signifikant aus. Der Vergleich der mittleren Ränge zeigt dennoch einen Alterstrend in erwarteter Richtung: Die Vierjährigen erbringen tendenziell geringere Leistungen als die Fünfjährigen.

11.1.3 Unterschiede in den Testleistungen zwischen den Halbjahresgruppen

Zur anschaulichen Darstellung der unterschiedlichen Leistungen der Halbjahresgruppen in den Subtests wurden Box-Whisker-Plots erstellt. Diese sind im Anhang (Abbildung C-1 bis C-16) zu finden.

11.1.3.1 Leistungsunterschiede bei Subtests mit gleichen Aufgaben für Vier- und Fünfjährige

Varianzhomogenität und Normalverteilung pro Gruppe sind bei keinem Subtest bzw. keiner Subtestkomponente gegeben (siehe Tabellen A-7 und A-8 im Anhang). Da für die Subtests *Bilderlotto*, *Puppenspiel* und *Wörter Erklären* die Homogenität der Varianzen vorausgesetzt werden kann ($p > .05$), die Normalverteilung in drei von vier Gruppen per Konvention gegeben ist ($p > .05$) und ein Signifikanzniveau von .01 bei der Normalverteilungsüberprüfung der vierten Gruppe nicht unterschritten wird, wurde für die Auswertung in diesen Fällen eine Varianzanalyse herangezogen. Dieses Vorgehen gründet in der vielfach beschriebenen Robustheit der Varianzanalyse bei Verletzung der Voraussetzung der Normalverteilung (vgl. Bortz, 2005; Field, 2005). Für alle weiteren Unterschiedsanalysen wurde aufgrund der nichtgegebenen bzw. suboptimalen Erfüllungen der Voraussetzungen für eine Varianzanalyse ein Kruskal-Wallis-Test herangezogen.

Abhängige Variablen sind die erzielten Rohwerte im WET, unabhängige die vier Altersgruppen.

Die Ergebnisse der Varianzanalysen und Kruskal-Wallis-Tests sind in Tabelle B-1 im Anhang zu finden. Alle Analysen führen zu signifikanten Ergebnissen ($p < .05$), so dass weitere Berechnungen nötig waren. Es galt, jene Gruppen zu ermitteln, zwischen denen Unterschiede bestehen sowie Trends zur Leistungserhöhung darzustellen.

Für jene Subtests, die mittels Varianzanalyse ausgewertet wurden, waren a-priori Kontraste geplant worden. Die Kontrastkoeffizienten sind in Tabelle 11 abgebildet.

Tabelle 11: Geplante Kontraste

Kontrast	4,0 - 4,5	4,6 - 4,11	5,0 - 5,5	5,6 - 5,11
1	1	-1	0	0
2	1	0	-1	0
3	1	0	0	-1
4	0	1	-1	0
5	0	1	0	-1
6	0	0	1	-1

Die Ergebnisse der korrespondierenden Kontrast-Tests sind in Tabelle 12 aufgeführt, die entsprechenden deskriptiven Ergebnisse sind in Anhang B (Tabelle B-2) zu sehen: Es zeigen sich signifikante Leistungsunterschiede in allen drei Subtests zwischen der Altersgruppe 4;0 – 4;5 und den beiden Halbjahresgruppen Fünfjähriger. Auch unterscheiden sich die Leistungen 4;6 – 4;11-jähriger signifikant von fünfjährigen Kindern. Keine signifikanten Unterschiede bestehen zwischen der jüngeren und älteren Altersgruppe der Vierjährigen, wie auch zwischen Fünfjährigen.

Tabelle 12: Leistungsunterschiede in den Testleistungen zwischen den vier Altersgruppen: Ergebnis der Kontrast-Tests

	Kontrast	Kontrastwert	SE	T	df	Sig. (1-seitig)
Bilderlotto: <i>Langform</i>	1	-1.29	1.010	-1.277	158	.102
	2	-3.87	1.010	-3.837	158	<.001
	3	-5.41	1.004	-5.386	158	<.001
	4	-2.59	.991	-2.610	158	.005
	5	-4.12	.985	-4.183	158	<.001
	6	-1.53	.985	-1.558	158	.061
Puppenspiel	1	-.67	.513	-1.301	160	.098
	2	-1.74	.513	-3.392	160	<.001
	3	-2.38	.510	-4.655	160	<.001
	4	-1.07	.510	-2.104	160	.018
	5	-1.71	.507	-3.367	160	<.001
	6	-.63	.507	-1.251	160	.106
Wörter Erklären	1	-.42	.707	-.590	160	.278
	2	-2.00	.707	-2.831	160	.003
	3	-2.55	.703	-3.630	160	<.001
	4	-1.59	.703	-2.256	160	.013
	5	-2.14	.699	-3.056	160	.001
	6	-.55	.699	-.787	160	.216

Obwohl sich nicht alle aneinander grenzenden Halbjahresgruppen in ihren Testleistungen signifikant voneinander unterscheiden, zeigt die Berechnung polynomialer Kontraste, dass der Verlauf der Mittelwerte in allen drei Subtests linear signifikant ist (siehe Tabelle B-3 im Anhang). Dies bedeutet, dass sich die Leistungen über die Altersgruppen hinweg signifikant verbessert haben.

In Anschluss an einen signifikanten Kruskal-Wallis-Test kam der Jonckheere-Terpstra-Test zum Einsatz, welcher die Evaluation eines Trends steigender Testleistungen mit zunehmendem Alter ermöglicht.

Tabelle 13 sind die Mittleren Ränge und Mediane der vier Altersgruppen sowie die JT-Statistik der einzelnen Berechnungen zu entnehmen. Es kann demnach überwiegend ein signifikanter Trend zur Leistungserhöhung in erwarteter Richtung festgestellt werden. Lediglich bezüglich der Subtestkomponente *Schatzkästchen: Behaltensleistung nach 20 Minuten* zeigt sich ein unerwartetes Ergebnis: Ein Vergleich der mittleren Ränge verdeutlicht, dass die 4;6 – 4;11-jährigen im Vergleich zu allen anderen Altersgruppen die beste Leistung erzielten.

11.1.3.2 Leistungsunterschiede bei Subtests mit altersangepassten Aufgaben

Die Subtests *Quiz*, *Muster Legen* und *Rechnen* enthalten Aufgaben mit altersangepasster Schwierigkeit für Vier- und Fünfjährige. Infolgedessen lassen sich lediglich die Altersgruppen 4;0 – 4;5 und 4;6 – 4;11 sowie die Altersgruppen 5;0 – 5;5 und 5;6 – 5;11 direkt in ihren Leistungen vergleichen.

Normalverteilung pro Gruppe war lediglich beim Subtest *Quiz* bei den Halbjahresgruppen der Vierjährigen gegeben (siehe Tabelle A-9). Die Varianzen erwiesen sich in allen Fällen als homogen (siehe Tabelle A-10).

Die Auswertung bezüglich Unterschieden in den beim Subtest *Quiz* gezeigten Leistungen von Kindern im Alter von 4;0 bis 4;5 und Kindern im Alter von 4;6 bis 4;11 erfolgte mittels t-Test und Mann-Whitney-U-Test. Für alle anderen Unterschiedsberechnungen wurde aufgrund der suboptimalen Voraussetzungen für einen t-Test ein Mann-Whitney-U-Test berechnet.

Abhängige Variablen sind die erzielten Rohwerte im WET. Als unabhängige Variable wurden die Halbjahresgruppen der Vier- und der Fünfjährigen herangezogen.

Tabelle 13: Leistungsunterschiede in den Testleistungen zwischen den vier Altersgruppen: Ergebnis der Jonckheere-Terpstra-Tests

Subtest	Alter	n	Mittlerer Rang	Md	stand. JT-Statistik	p*
Lernbär	4;0 – 4;5	40	64.90	3.0	4.712	<.001
	4;6 – 4;11	41	74.09	3.0		
	5;0 – 5;5	41	92.33	3.0		
	5;6 – 5;11	42	97.88	3.0		
Schatzkästchen: <i>Unmittelbares Reproduzieren</i>	4;0 – 4;5	40	59.04	3.0	4.417	<.001
	4;6 – 4;11	41	76.83	3.0		
	5;0 – 5;5	41	91.94	4.0		
	5;6 – 5;11	42	101.17	4.0		
Schatzkästchen: <i>Behaltensleistung nach 20 Minuten</i>	4;0 – 4;5	40	69.80	5.0	1.381	.082
	4;6 – 4;11	41	93.23	6.0		
	5;0 – 5;5	41	78.17	6.0		
	5;6 – 5;11	42	88.35	6.0		
Schatzkästchen: <i>Anzahl der Lerndurchgänge</i>	4;0 – 4;5	40	108.64	5.0	-5.690	<.001
	4;6 – 4;11	41	91.52	4.0		
	5;0 – 5;5	41	79.46	3.0		
	5;6 – 5;11	42	51.76	2.0		
Bunte Formen	4;0 – 4;5	40	63.48	4.0	3.823	<.001
	4;6 – 4;11	41	79.05	5.0		
	5;0 – 5;5	41	83.77	5.0		
	5;6 – 5;11	42	102.75	8.0		
Zahlen merken: <i>Gedächtnisspanne</i>	4;0 – 4;5	40	63.63	3.0	2.250	.006
	4;6 – 4;11	41	83.33	3.0		
	5;0 – 5;5	41	96.27	4.0		
	5;6 – 5;11	42	86.23	4.0		
Turnen	4;0 – 4;5	39	36.21	5.0	7.692	<.001
	4;6 – 4;11	41	75.52	7.0		
	5;0 – 5;5	41	103.21	8.0		
	5;6 – 5;11	42	110.14	8.0		
Nachzeichnen	4;0 – 4;5	40	47.54	5.0	5.998	<.001
	4;6 – 4;11	41	78.83	6.0		
	5;0 – 5;5	41	97.27	7.0		
	5;6 – 5;11	42	104.96	7.0		
Gegensätze	4;0 – 4;5	40	64.50	10.0	4.481	<.001
	4;6 – 4;11	41	68.24	10.0		
	5;0 – 5;5	41	90.16	11.0		
	5;6 – 5;11	42	106.08	12.5		
Fotoalbum	4;0 – 4;5	40	56.85	5.0	4.588	<.001
	4;6 – 4;11	41	76.50	6.0		
	5;0 – 5;5	41	95.32	7.0		
	5;6 – 5;11	42	100.27	7.0		
Elternfragebogen	4;0 – 4;5	36	48.13	82.0	6.025	<.001
	4;6 – 4;11	35	59.20	84.0		
	5;0 – 5;5	36	79.82	92.5		
	5;6 – 5;11	37	101.68	97.0		

* Monte Carlo Signifikanz: einseitig

Tabelle 14 sind die Mittelwerte, Mittleren Ränge und Mediane der beiden Altersgruppen sowie die U-Statistik der einzelnen Berechnungen zu entnehmen. Signifikante Leistungsunterschiede zeigen sich sowohl zwischen 4;0 bis 4;5-jährigen und 4;6 - 4;11-jährigen ($Z = -2.987$; $p = .001$) als auch zwischen 5;0 bis 5;5-jährigen und 5;6 bis 5;11-jährigen ($Z = -3.521$; $p = < .001$) im Subtest *Muster Legen*. Ebenfalls den Erwartungen entsprechend sind die signifikanten Leistungsunterschiede zwischen den Altersgruppen 4;0 – 4;5 und 4;6 – 4;11 im Subtest *Rechnen* ($Z = -2.370$; $p = .009$). Die Leistungsunterschiede zwischen 5;0 – 5;5-jährigen im selben Subtest waren hingegen nicht signifikant ($Z = -1.390$; $p = .083$). Ein Vergleich der Mittleren Ränge zeigt jedoch einen Trend in die erwartete Richtung: Die ältere Halbjahresgruppe erzielt bessere Leistungen als die jüngere. Hinsichtlich der im Subtest *Quiz* erzielten Leistungen zeigen sich weder signifikante Unterschiede zwischen 4;0 – 4;5-jährigen und 4;6 – 4;11-jährigen ($T = .374$; $df = 79$; $p = .355$ bzw. $Z = -.390$; $p = .350$) noch zwischen 5;0 – 5;5-jährigen und 5;6 – 5;11-jährigen ($Z = -.831$; $p = .205$). Bei den Fünfjährigen ist der erwartete Unterschied zwischen den Halbjahresgruppen tendenziell gegeben: Der Mittlere Rang der 5;6 – 5;11-jährigen ist höher als der der 5;0 – 5;5-jährigen. Der Vergleich der Mittelwerte bzw. mittleren Ränge der Altersgruppen 4;0 – 4;5 und 4;6 – 4;11 weist ein unerwartetes Ergebnis auf: Die jüngere Halbjahresgruppe erbringt geringfügig bessere Leistungen als die ältere.

Tabelle 14: Unterschiede zwischen den Halbjahresgruppen bei Subtests mit altersangepasster Schwierigkeit

Subtest	Alter	n	M	Mittlerer Rang	Md	U-Statistik	p*
Quiz	4;0 – 4;5	40	6.03	42.03	6.0	U = 1640.000	.350
	4;6 – 4;11	41	5.80	40.00	6.0	Z = -.390	
Quiz	5;0 – 5;5	41	6.90	39.79	8.0	U = 1631.500	.205
	5;6 – 5;11	42	7.36	44.15	8.0	Z = -.831	
Muster Legen	4;0 – 4;5	40	4.40	33.16	4.0	U = 1337.500	.001
	4;6 – 4;11	41	6.07	48.65	6.0	Z = -3.521	
Muster Legen	5;0 – 5;5	41	4.02	32.62	3.0	U = 1570.500	<.001
	5;6 – 5;11	42	6.52	51.15	7.0	Z = -3.957	
Rechnen	4;0 – 4;5	40	6.50	34.79	6.0	U = 1570.500	.009
	4;6 – 4;11	41	8.24	47.06	9.0	Z = -1.390	
Rechnen	5;0 – 5;5	41	5.66	38.30	6.0	U = 1946.000	.083
	5;6 – 5;11	42	6.52	45.61	6.5	Z = -4.727	

* Exakte Signifikanz: einseitig

11.1.4 Zusammenfassende Beurteilung der Annahme steigender Alterstrends

Bezüglich aller Subtests mit gleichen Aufgaben für Vier- und Fünfjährige bestehen signifikante Leistungsunterschiede zwischen den Jahresgruppen. Lediglich hinsichtlich der Subtestkomponente *Schatzkästchen: Behaltensleistung nach 20 Minuten* sind die Unterschiede zwischen vier- und fünfjährigen Kindern nur auf Trendniveau. Die Ergebnisse der polynomialen Kontraste und der Jonckheere-Terpstra-Tests zeigen bei Subtests ohne altersangepasste Aufgaben überwiegend signifikante lineare Trends zur Leistungserhöhung mit zunehmendem Alter. Lediglich bezüglich der Behaltensleistung nach 20 Minuten beim Subtest *Schatzkästchen* ist kein Trend feststellbar. Die 4;6 – 4;11-jährigen Kinder erbringen hier die besten visuell-räumlichen Gedächtnisleistungen, wie ein Vergleich der mittleren Ränge zeigt. Bei den mittels Varianzanalyse untersuchten Subtests *Bilderlotto*, *Puppenspiel* und *Wörter Erklären* zeigen Kontrasttests zudem signifikante Unterschiede zwischen 4;6 – 4;11-Jährigen und 5;0 – 5;5-Jährigen.

Bei jenen Subtests, die altersangepasste Aufgaben für die beiden Jahresgruppen vorsehen, wurden nur die Halbjahresgruppen vierjähriger und fünfjähriger Kinder einem Vergleich unterzogen. Im Subtest *Muster Legen* zeigen sich signifikante Unterschiede sowohl zwischen Vierjährigen als auch Fünfjährigen. Hinsichtlich der im Subtest *Rechnen* erzielten Leistungen bestehen signifikante Unterschiede nur zwischen den Vierjährigen, bei den Fünfjährigen erlangen die Unterschiede lediglich Trendniveau. Im Subtest *Quiz* zeigen sich weder zwischen den Vierjährigen noch zwischen den Fünfjährigen signifikante Unterschiede. Bei den Fünfjährigen lässt sich zwischen den Altersgruppen wiederum ein Trend in erwarteter Richtung feststellen. Eine Ausnahme bildet der Leistungsunterschied zwischen den Altersgruppen 4;0 – 4;5 und 4;6 – 4;11: Die jüngere Altersgruppe der Vierjährigen ($M = 6,03$) erzielt geringfügig bessere Leistungen als die ältere ($M = 5,80$).

Die Annahme steigender Testleistungen mit zunehmendem Alter kann aufgrund der vorliegenden Stichprobendaten durchaus bestätigt werden. Dem unerwarteten Ergebnis, dass die 4;6 – 4;11-jährigen hinsichtlich der Subtestkomponente *Schatzkästchen: Behaltensleistung nach 20 Minuten* die besten Leistungen erzielen, ist wenig Gewicht beizumessen, da in Hinblick auf die anderen beiden Subtestkomponenten erwartungskonforme Ergebnisse beobachtet werden konnten. Auch der nicht erwartungsgemäße Unterschied zwischen 4;0 – 4;5-Jährigen und 4;6 – 4;11-Jährigen im

Subtest Quiz ist praktisch unbedeutend. Diese überraschenden Ergebnisse sind vermutlich auf Besonderheiten der Stichprobe zurückzuführen.

11.2 Faktorielle Validität

Zur Überprüfung der Dimensionalität der geplanten Neuauflage des Wiener Entwicklungstests (WET; Kastner-Koller & Deimann, 1998, 2002, in Vorbereitung) wurde eine Faktorenanalyse berechnet. Entsprechend den sechs Funktionsbereichen des WET wurden sechs Faktoren festgelegt und eine Hauptkomponentenanalyse mit anschließender orthogonaler Varimax-Rotation durchgeführt. Ladungen unter .40 wurden unterdrückt.

In die Berechnung gingen die C-Werte der ursprünglichen Subtests der zweiten Auflage des WET (mit Ausnahme des Elternfragebogens) und die je Altersgruppe z-transformierten Rohwerte der neuen Subtests *Rechnen* und *Muster Legen* ein. Von der Analyse wurden die Daten eines Kindes ausgeschlossen, da aufgrund von Verweigerung kein C-Wert für den Subtest *Turnen* zur Verfügung steht. Die Stichprobe umfasste somit 163 Personen.

Der KMO-Test (Kaiser-Meyer-Olkin Measure of Sampling Adequacy) erreicht einen Wert von $> .80$ und weist darauf hin, dass der Stichprobenumfang zur Durchführung einer Faktorenanalyse gut geeignet ist. Die Kommunalitäten aller Variablen sind zudem größer .05, was diese Annahme bestätigt. Da auch der Bartlett-Test zeigt, dass die Korrelationsmatrix mit einer Sicherheit von $p < .001$ eine Einheitsmatrix darstellt, kann von der Brauchbarkeit der Daten für eine Faktorenanalyse ausgegangen werden.

In Tabelle 15 ist der Eigenwertverlauf ersichtlich. Unter Berücksichtigung des Kaiser-Kriteriums (Eigenwert > 1) wäre eine Drei-Faktoren-Lösung angemessen, die 50 % der Gesamtvarianz erklärt. Diese Wahl wäre jedoch theoretisch nicht sinnvoll. Nach erfolgter Rotation weisen allerdings alle aufgrund der Funktionsbereiche des WET festgelegten sechs Faktoren einen Eigenwert größer als 1 auf und erklären insgesamt 70.02 % der Gesamtvarianz.

Tabelle 15: Eigenwertverlauf und erklärte Varianz vor und nach der Rotation

Faktor	Anfängliche Eigenwerte			Rotierte Summe der quadrierten Ladungen		
	Eigenwert	% der Varianz	Kumulierte %	Eigenwert	% der Varianz	Kumulierte %
1	4.189	29.920	29.920	2.994	21.385	21.385
2	1.768	12.626	42.546	1.508	10.772	32.157
3	1.133	8.094	50.640	1.457	10.405	42.562
4	.966	6.899	57.539	1.401	10.004	52.566
5	.951	6.793	64.332	1.238	8.841	61.408
6	.796	5.686	70.018	1.205	8.610	70.018

Die Ergebnisse der Hauptkomponentenanalyse nach erfolgter Varimax-Rotation sind in Tabelle 16 dargestellt und werden im Folgenden interpretiert:

Faktor 1 entspricht dem Funktionsbereich Sprache, auf welchem die sprachlichen Subtests *Quiz*, *Puppenspiel*, *Wörter Erklären* und *Gegensätze* mit Ladungen zwischen .75 und .82 laden. Daneben zeigen auch *Fotoalbum* und *Rechnen* bedeutsame Ladungen. Dies impliziert, dass sowohl das Erkennen und Benennen des mimischen Gefühlsausdrucks als auch erste mathematische Grundfertigkeiten einen Zusammenhang mit sprachlicher Kompetenz aufweisen.

Faktor 2 entspricht dem Funktionsbereich Visumotorik/Visuelle Wahrnehmung. Hier laden sowohl *Nachzeichnen* mit .87 als auch *Bilderlotto* mit .50. Zudem weist der Subtest *Muster Legen* eine Nebenladung in relevanter Höhe auf. Dies ist nicht überraschend, da dieser Subtest aufgrund der Aufgabengestaltung stark an Wahrnehmungsprozesse angelehnt ist (Kastner-Koller & Deimann, 2002).

Auf Faktor 3 lädt *Zahlen Merken* mit .89 sehr hoch. Darüber hinaus zeigt sich eine substantielle Ladung beim Subtest *Rechnen* (.62). Dies deutet auf einen Zusammenhang zwischen dem phonologischen Arbeitsgedächtnis und rechnerischem Denken hin, welcher auch in der Literatur diskutiert wird (z. B. D'Amico & Guarnera, 2005; Schuchardt et al., 2006).

Faktor 4 kann als kognitiver Faktor interpretiert werden, auf welchem die nonverbal-kognitiven Subtests *Bunte Formen* (.90) und *Muster Legen* (.47) lokalisiert sind. Zudem weist der Subtest *Bilderlotto*, der die differenzierte Raum-Lage-Wahrnehmung erfasst, eine bedeutsame Nebenladung auf diesem Faktor auf.

Auf Faktor 5 lädt der Subtest *Lernbär* am höchsten (.78). Darüber hinaus ist hier *Schatzkästchen* enthalten (.47). *Fotoalbum* korreliert negativ mit dem Faktor (-.58). Scheinbar besteht ein negativer Zusammenhang zwischen dem Benennen des mimischen Gefühlsausdrucks und feinmotorischen Fähigkeiten als auch visuellen Gedächtnisleistungen.

Faktor 6 enthält einerseits den Subtest *Turnen*, der hier mit .90 sehr hoch lädt, als auch *Schatzkästchen* (.50).

Tabelle 16: Ergebnisse der Hauptkomponentenanalyse nach Varimax-Rotation

	Faktoren					
	1	2	3	4	5	6
C-Wert Lernbär					.781	
C-Wert Quiz	.755					
C-Wert Bilderlotto		.497		.477		
C-Wert Puppenspiel	.754					
C-Wert Schatzkästchen					.472	.504
C-Wert Bunte Formen				.901		
C-Wert Zahlen Merken			.887			
C-Wert Wörter Erklären	.822					
C-Wert Turnen						.904
C-Wert Nachzeichnen		.870				
C-Wert Gegensätze	.807					
C-Wert Fotoalbum	.449				-.584	
Z-Wert Muster Legen		.449		.471		
Z-Wert Rechnen	.425		.622			

Insgesamt zeigt die Faktorenanalyse, dass die sechs Funktionsbereiche mittels der vorliegenden Stichprobendaten zum Teil reproduziert werden können: Die Faktoren 1, 2 und 4 sind inhaltlich gut interpretierbar und entsprechen den Funktionsbereichen Sprache, Visumotorik/Visuelle Wahrnehmung und Kognitive Entwicklung. Die Faktoren 3, 5 und 6 können nicht eindeutig interpretiert werden. *Schatzkästchen* lädt auf den Faktoren 5 und 6 in ähnlicher Höhe, *Zahlen Merken* auf Faktor 3. Es findet sich demnach kein gemeinsamer Gedächtnisfaktor. Auch die Subtests des Funktionsbereichs Motorik *Lernbär* und *Turnen* sind auf zwei unterschiedlichen Faktoren lokalisiert. Auch findet sich kein eigener Faktor für den Funktionsbereich Sozial-Emotionale Entwicklung. *Fotoalbum* zeigt eine positive Ladung auf dem sprachlichen Faktor und eine negative auf Faktor 5. Zudem lässt sich der Subtest *Rechnen* nicht wie erwartet dem kognitiven Bereich zuordnen. Er lädt gemeinsam mit *Zahlen Merken* auf einem Faktor und zeigt eine Nebenladung auf dem sprachlichen Faktor.

12 Reliabilität der neu entwickelten Subtests

Die für die geplante dritte Auflage des Wiener Entwicklungstests (WET, Kastner-Koller & Deimann, 1998, 2002, in Vorbereitung) neu entwickelten Subtests *Muster Legen* und *Rechnen* wurden aufgrund der probabilistischen Testtheorie entwickelt. Die beiden Gesamtskalen sehen zwei Kompetenzniveaus vor, die in unterschiedlichen Itemsets für Vierjährige und Fünfjährige münden. Entsprechend den Annahmen des Rasch-Modells sollten die Aufgaben für beide Jahresgruppen homogen sein und somit dieselbe Fähigkeit über den gesamten Altersbereich von 4;0 bis 5;11 erfassen.

Die folgenden Reliabilitätsberechnungen sind demgemäß im Sinne einer methodischen Validierung zu verstehen, die neben Aussagen zur Reliabilität einen Hinweis auf die Konstruktvalidität des WET geben.

Die Reliabilitätsanalysen erfolgen jeweils getrennt für die Aufgabengruppen der Vierjährigen und Fünfjährigen.

12.1 Reliabilität der Skala Muster Legen

Die neu entwickelte Skala *Muster Legen* umfasst für jede Altersgruppe 10 Items. Vierjährige Kinder erhalten die Items 4 bis 13, Fünfjährige die Items 8 bis 17.

Die Überprüfung der inneren Konsistenz des Itemsets für Vierjährige ($N = 81$) ergibt ein Cronbach- α von .757. Im Mittel konnten die Kinder 5.26 Items lösen. Die Schwierigkeitsindizes und Trennschärfekoeffizienten sowie Cronbach's Alpha, wenn das jeweilige Item weggelassen wird, sind Tabelle 17 zu entnehmen.

Tabelle 17: Itemanalyse der Skala Muster Legen (Vierjährige)

	Itemschwierigkeit	Trennschärfe	Cronbach- α , wenn Item weggelassen
Item 4	.85	.057	.777
Item 5	.88	.378	.743
Item 6	.62	.566	.713
Item 7	.81	.379	.742
Item 8	.63	.372	.744
Item 9	.36	.452	.732
Item 10	.40	.597	.708
Item 11	.32	.514	.722
Item 12	.22	.407	.738
Item 13	.17	.437	.735

Unter Berücksichtigung aller 10 Items ist die innere Konsistenz der Aufgabenzusammenstellung für die Vierjährigen als akzeptabel einzuschätzen. Die Schwierigkeitsindizes verdeutlichen, dass Items aller Schwierigkeitsstufen enthalten sind (.17 bis .88). Die mittlere Itemschwierigkeit beträgt .53, die mittlere Trennschärfe .42. Neun Items weisen zufriedenstellende bis gute Trennschärfen auf (.372 bis .597). Lediglich das Item 4 erzielt mit einer korrigierten Item-Skala-Korrelation von .057 eine sehr geringe Trennschärfe. Die Aussonderung dieses Items hätte eine Erhöhung der Reliabilität auf .777 zur Folge.

Für die Skala der Fünfjährigen (N = 83) zeigt sich ein Cronbach- α von .874. Im Mittel lösten die Fünfjährigen 5.30 Items. Die Ergebnisse der Itemanalysen sind in Tabelle 18 aufgeführt.

Tabelle 18: Itemanalyse der Skala *Muster Legen* (Fünfjährige)

	Itemschwierigkeit	Trennschärfe	Cronbach- α , wenn Item weggelassen
Item 8	.84	.412	.874
Item 9	.65	.579	.863
Item 10	.69	.576	.863
Item 11	.57	.572	.864
Item 12	.59	.630	.859
Item 13	.52	.640	.858
Item 14	.42	.752	.849
Item 15	.46	.667	.856
Item 16	.27	.527	.867
Item 17	.30	.573	.864

Die innere Konsistenz der Skala *Muster Legen* für die Fünfjährigen ist als gut einzuschätzen ($\alpha > .80$). Die Schwierigkeitsindizes erreichen Werte zwischen .84 und .27, wobei Items mittlerer Schwierigkeit überwiegen. Die mittlere Itemschwierigkeit beträgt .53. Alle 10 Items weisen gute Trennschärfen auf (.412 bis .752), die mittlere Trennschärfe liegt bei .59.

Zusammenfassend lässt sich feststellen, dass das Itemset für Fünfjährige in dieser Stichprobe eine höhere Messgenauigkeit erzielt als die Aufgabenzusammenstellung für Vierjährige. Die mittlere Itemschwierigkeit ist mit .053 in beiden Skalen ident. Die Skala

für Vierjährige enthält jedoch drei Aufgaben, die von mehr als 80 % der Kinder gelöst werden konnten, die Skala für Fünfjährige hingegen nur eine. Item 13 erwies sich für die Vierjährigen als schwierigste Aufgabe und wurde von 17 % gelöst. Item 16 war das schwierigste Item für Fünfjährige, konnte aber von immerhin 27 % gelöst werden. Items 8 bis 13, die beiden Altersgruppen vorgegeben werden, weisen für die Fünfjährigen erwartungsgemäß geringere Schwierigkeitsindizes auf als für die Vierjährigen. Die mittlere Trennschärfe der Aufgaben für Vierjährige liegt bei .42, die der Fünfjährigen bei .59. Item 4 der Skala für vierjährige Kinder weist in dieser Stichprobe eine sehr geringe Trennschärfe auf (.057) und würde bei einer Eliminierung zu einer nicht unbedeutenden Erhöhung der inneren Konsistenz führen.

Es ist demnach zu schlussfolgern, dass die Aufgabenzusammenstellungen für die beiden Altersgruppen die zu messende Dimension in dieser Stichprobe unterschiedlich genau erfassen.

12.2 Reliabilität der Skala Rechnen

Der Subtest Rechnen sieht für jede Altersgruppe 11 Items vor. Vierjährige Kinder erhalten die Items 10 bis 20, Fünfjährige die Items 18 bis 28.

Die Überprüfung der inneren Konsistenz für das Itemset der Vierjährigen (N=81) ergibt ein Cronbach- α von .849. Im Mittel lösten die Vierjährigen 7.43 Items. Die Schwierigkeitsindizes und Trennschärfekoeffizienten sowie Cronbach's Alpha, wenn das jeweilige Item weggelassen wird, sind in Tabelle 19 aufgeführt.

Tabelle 19: Itemanalyse der Skala *Rechnen* (Vierjährige)

	Itemschwierigkeit	Trennschärfe	Cronbach- α , wenn Item weggelassen
Item 10	.80	.325	.851
Item 11	.83	.324	.850
Item 12	.86	.430	.844
Item 13	.64	.606	.830
Item 14	.60	.650	.826
Item 15	.84	.407	.845
Item 16	.62	.581	.832
Item 17	.56	.621	.828
Item 18	.64	.596	.831
Item 19	.62	.672	.824
Item 20	.42	.586	.832

Die Messgenauigkeit der Aufgabenzusammenstellung für Vierjährige ist insgesamt als gut zu bewerten ($\alpha > .80$). Die Itemschwierigkeiten sind mit Werten zwischen .86 und .42 als relativ leicht bis mittelschwer einzuschätzen. Die mittlere Itemschwierigkeit liegt bei .68. Alle Items weisen zufriedenstellende bis gute Trennschärfen auf (.372 bis .597), im Mittel eine Trennschärfe von .53. Das Cronbach- α würde sich minimal erhöhen, wenn Item 10 und 11 weggelassen würden.

Für die Skala der Fünfjährigen (N=83) ergibt sich ein Cronbach- α von .746. Im Mittel lösten die Fünfjährigen 6.10 Aufgaben. Die Ergebnisse der Items sind in Tabelle 20 dargestellt.

Tabelle 20: Itemanalyse der Skala *Rechnen* (Fünfjährige)

	Itemschwierigkeit	Trennschärfe	Cronbach- α , wenn Item weggelassen
Item 18	.89	.156	.752
Item 19	.95	.173	.749
Item 20	.39	.513	.709
Item 21	.63	.306	.741
Item 22	.33	.335	.736
Item 23	.64	.526	.708
Item 24	.70	.338	.735
Item 25	.75	.439	.721
Item 26	.41	.525	.707
Item 27	.18	.493	.716
Item 28	.24	.453	.719

Die innere Konsistenz der Skala *Rechnen* für Fünfjährige ist mit einem Wert von .746 als akzeptabel einzuschätzen. Die Aufgabenschwierigkeiten reichen von sehr leicht (.95) bis eher schwer (.18), im Mittel .56. Die mittlere Trennschärfe beträgt .39. Die Trennschärfen der Items 20 bis 28 sind als durchwegs befriedigend zu beurteilen. Item 18 und 19, welche die geringste Aufgabenschwierigkeit zeigen, weisen niedrige Trennschärfen ($< .30$) auf. Bei Eliminierung dieser Items würde sich das Cronbach- α marginal erhöhen, aber dennoch keinen Wert von größer .80 erreichen.

Zusammenfassend kann gesagt werden, dass die Aufgabenzusammenstellung für die Vierjährigen in dieser Stichprobe eine höhere Reliabilität erzielt als jene für Fünfjährigen. Die Aufgaben für Vierjährige zeigten sich im Mittel einfacher (.68) als die Aufgaben für Fünfjährige (.53). Die Skala für Vierjährige enthält vier Aufgaben, die von mehr als 80 %

der Kinder gelöst werden konnten, die Skala für Fünfjährige hingegen nur zwei. Item 20 erwies sich für die Vierjährigen als die schwierigste Aufgabe, konnte aber dennoch von 42 % gelöst werden. Item 27 war das schwierigste Item für Fünfjährige, konnte aber lediglich von 18 % gelöst werden. Die Items 19 bis 20, die beiden Altersgruppen vorgegeben werden, weisen für die Fünfjährigen erwartungsgemäß geringere Schwierigkeitsindizes auf als für die Vierjährigen. Die mittlere Trennschärfe der Aufgaben für die Vierjährigen liegt bei .53, die der Fünfjährigen bei .39. Die Skala für fünfjährige Kinder enthält zwei Aufgaben (Item 18 und 19) mit einer Trennschärfe unter .30, die bei einer Eliminierung allerdings nur zu einer praktisch unbedeutenden Erhöhung der inneren Konsistenz führen würden. Die Trennschärfen aller Items für Vierjährige sind dagegen durchwegs zufriedenstellend.

Insgesamt zeigt sich, dass die zugrundeliegende Dimension anhand der vorliegenden Daten in den beiden Altersgruppen unterschiedlich genau erfasst wird.

13 Diskussion

Ziel des empirischen Teils war es, einen Beitrag zur Testanalyse der geplanten Neuauflage des WET zu liefern: Die Daten der Wiener Normierungsstichprobe wurden deshalb zur Überprüfung mehrerer Aspekte der Konstruktvalidität herangezogen. Drei Fragestellungen, die im Folgenden nacheinander behandelt werden, waren hierbei von Interesse:

Erstens interessierte die Frage, ob die theoretische Annahme steigender Testleistungen mit fortschreitendem Alter bestätigt werden kann.

Alle Subtests mit gleichen Aufgaben für Vier- und Fünfjährige sowie der Elternfragebogen zeigen einen deutlichen Alterstrend. Signifikante Unterschiede sind hier sowohl zwischen den Jahresgruppen als auch zwischen den Halbjahresgruppen zu beobachten. Auch bei den Subtests *Quiz*, *Muster Legen* und *Rechnen*, die altersangepasste Aufgaben für vier- und fünfjährige Kinder vorsehen, zeigt sich überwiegend ein Alterstrend, der allerdings nur in Hinblick auf den Subtest *Muster Legen* sowohl für Vier- als auch für Fünfjährige signifikant ausfällt. Beim Subtest *Rechnen* sind die Leistungsunterschiede zwischen Vierjährigen auf Signifikanzniveau, bei den Fünfjährigen zeichnet sich ein erwartungsgemäßer Trend ab. Ebenfalls auf Trendniveau sind die Unterschiede zwischen den Halbjahresgruppen Vierjähriger bzw. Fünfjähriger. Der Trend verläuft bei den Fünfjährigen in entsprechender Richtung, bei den Vierjährigen erzielt die jüngere Halbjahresgruppe allerdings geringfügig bessere Leistungen als die ältere. Dieses unerwartete Ergebnis ist jedoch weder statistisch noch praktisch bedeutsam.

Die Ergebnisse der gegenständliche Untersuchung belegen die Differenzierungsfähigkeit des WET in Hinblick auf verschiedene Altersgruppen eindeutig und bestätigen außerdem sowohl die Befunde von Kastner-Koller und Deimann (2002) als auch die der unabhängigen Studie von Krampen et al. (2008).

Zweitens war von Interesse, inwieweit sich die 14 Subtests (der Elternfragebogen wurde nicht berücksichtigt) der Neuauflage den sechs Funktionsbereichen des WET zuordnen lassen. Es wurde also die Dimensionalität des Verfahrens überprüft.

Durch die durchgeführte Hauptkomponentenanalyse mit anschließender orthogonaler Varimax-Rotation lassen sich die Funktionsbereiche teilweise reproduzieren. Die Funktionsbereiche *Sprache*, *Visumotorik/Visuelle Wahrnehmung* und *Kognitive*

Entwicklung lassen sich faktoriell deutlich anhand der ihnen zugeordneten Subtests identifizieren. Drei Faktoren sind weniger eindeutig interpretierbar, spiegeln jedoch Zusammenhänge wider, die in der Literatur beschrieben werden:

- *Zahlen Merken* lädt mit *Rechen* auf einem Faktor, was auf einen Zusammenhang zwischen dem phonologischen Arbeitsgedächtnis und rechnerischem Denken hindeutet und durchaus dem aktuellen Forschungsstand entspricht (z. B. D'Amico & Guarnera, 2005; Hitch & McAuley, 1991; Lee & Kang, 2002; Kurdek & Sinclair, 2001; Schuchhardt et al., 2006). Diese Beziehung findet sich zudem im Index *Arbeitsgedächtnis* der aktuellen Auflage des Hamburger-Wechsler-Intelligenztest für Kinder (HAWIK IV; Petermann & Petermann, 2007) wieder, in den die Ergebnisse der Subtests *Zahlennachsprechen*, *Buchstaben-Zahlen-Folgen* und *Rechnerisches Denken* einfließen. Auch in der dritten Auflage (HAWIK III; Tewes, Rossmann & Schallberger, 2002) war dieser Zusammenhang bereits durch einen Index *Unablenkbarkeit* repräsentiert.
- Der Subtest *Schatzkästchen* ist mit ähnlich hohen Ladungen auf zwei Faktoren lokalisiert. Zum einen lädt er mit *Lernbär*, zum anderen mit *Turnen* auf einem Faktor. Die beiden Faktoren können als perzeptiv-motorische Faktoren interpretiert werden und weisen auf die Bedeutung des visuell-räumlichen Speichers für motorische Leistungen hin. Dieser Zusammenhang ist in der Literatur ebenfalls gut belegt (Wilkening & Krist, 2008).

Die faktorielle Validierung der Neuauflage des WET ist aufgrund der vorliegenden Stichprobendaten zum Teil gelungen. Die erzielte Faktorenlösung ist jedoch weniger zufriedenstellend als jene, die Kastner-Koller und Deimann (2002) sowie Krampen et al. (2008) für die zweite Auflage referieren.

Drittens sollte die Reliabilität der – für Vier- und Fünfjährige – neu entwickelten bzw. adaptierten Subtests Rechnen und Muster Legen überprüft werden. Zweck der Reliabilitätsanalysen der nach der probabilistischen Testtheorie konstruierten Subtests war eine methodische Validierung derselben. Hierzu wurden die inneren Konsistenzen der jeweiligen dem Alter entsprechenden Itemsets sowie die entsprechenden Itemschwierigkeiten und Trennschärfen berechnet.

In Hinblick auf den Subtest *Muster Legen* zeigt sich, dass die Aufgabenzusammenstellung für Fünfjährige ($\alpha = .87$) eine höhere Messgenauigkeit aufweist als die

für Vierjährige ($\alpha = .76$). Beim Subtest *Rechnen* hingegen erzielt das Itemset für die Vierjährigen ($\alpha = .85$) eine höhere Reliabilität als das für Fünfjährige ($\alpha = .75$).

Es zeigt sich also, dass die den Subtests zugrundeliegenden Dimensionen anhand der vorliegenden Daten in den beiden Altersgruppen unterschiedlich genau erfasst werden. Da die Reliabilitäten der einzelnen Aufgabenzusammenstellungen dennoch stets als hinreichend zu bewerten sind, sollte diesem Ergebnis keine allzu hohe Bedeutung beigemessen werden. Insgesamt scheint die methodische Validierung in Maßen gelungen zu sein.

Die im Rahmen dieser Arbeit durchgeführte Untersuchung zielt darauf ab, zur Testanalyse der Neuauflage des WET (Kastner-Koller & Deimann, in Vorbereitung) beizutragen. Die hierbei gewonnenen Ergebnisse liefern Hinweise auf die Konstruktvalidität des WET. Zur weiteren Einschätzung der Güte des Verfahrens werden im Folgenden verschiedene Qualitätsaspekte, die in dieser Arbeit Erwähnung gefunden haben, genauer betrachtet:

Im Theorieteil dieser Arbeit wurde mehrfach darauf hingewiesen, dass die einem Entwicklungstest zugrunde liegende Entwicklungstheorie ein Maß für dessen Güte ist.

Kastner-Koller und Deimann (2002) legen dem WET eine eklektizistische Theorie zugrunde, wobei die ökologischen und kontextualistischen Ansätze den Rahmen ihrer Theorie bilden. Die Autorinnen betonen, dass Entwicklung durch die „Gestaltung von Umwelten“ veränderbar ist. In diesem Zusammenhang referieren sie das Schaufelmodell von Kopp und McCall (1982) und gehen diesem entsprechend davon aus, dass Förderbemühungen ab dem dritten Lebensjahr zunehmend wirksam werden, diese jedoch nur dann greifen, wenn sie entwicklungskontingent erfolgen. In Hinblick auf den Entwicklungsverlauf sind die Autorinnen der Ansicht, dass die kindliche Entwicklung über den Altersbereich der Drei- bis Sechsjährigen relativ kontinuierlich in Richtung zunehmender Kompetenz verläuft und keine auffälligen qualitativen Sprünge zu erwarten sind.

Der WET wurde gewissenhaft entwicklungstheoretisch belegt. Die Erfüllung des dem Verfahren innewohnenden Anspruchs, ein förderdiagnostisches Verfahren zu sein, kann schlüssig nachvollzogen werden.

Als allgemein orientierter Entwicklungstest folgt der WET einem ganzheitlichen Entwicklungsverständnis, wonach eine kompetente Auseinandersetzung mit der Umwelt nicht nur intellektuelle Fähigkeiten erfordert. Das Verfahren zielt somit darauf ab, die Gesamtentwicklung drei- bis sechsjähriger Kinder umfassend und adäquat abzubilden.

Die Ergebnisse einer Studie von Krampen et al. (2008) legen nahe, dass die sechs Funktionsbereiche des WET hierfür auch gut geeignet sind und es jedenfalls keiner Erweiterung um Skalen zur Entwicklung von divergenten Leistungen und selektiven Aufmerksamkeitsleistungen bedarf.

In Kapitel 4.4 wird auf das Gütekriterium der Normierung eingegangen. Dieses gilt als erfüllt, wenn die Population, für die die Normen gelten, definiert ist, die Eichstichprobe repräsentativ ist und die Normen nicht veraltet sind (Kubinger, 2006).

Die Normierungsstichprobe der aktuellen Auflage des WET (Kastner-Koller & Deimann, 2002) kann als repräsentativ für Österreich und Deutschland angesehen werden. Die Erhebung der österreichischen Stichprobe ($n = 274$) erfolgte im Jahr 1996, die der deutschen Stichprobe im Jahr 2001 ($n = 971$). Bezogen auf den Flynn-Effekt (vgl. Flynn, 1987) sind die österreichischen Normdaten bereits veraltet.

Für die Normierung der geplanten dritten Auflage wurde in den Jahren 2008 und 2009 eine Datenerhebung an vier- und fünfjährigen Kindern aus Österreich, Deutschland und der Schweiz durchgeführt. Die Aktualität der Normen für Vier- und Fünfjährige wird somit wieder sichergestellt sein, wobei hier jedoch nichts über die Repräsentativität der Gesamtnormierungsstichprobe ausgesagt werden kann.

Die im Rahmen dieser Arbeit gesammelten Normdaten von 164 Wiener Kindern sind Teil der Normierung. Die Wiener Stichprobe stimmt mit der Grundgesamtheit der Wiener Bevölkerung gut überein und ist als repräsentativ zu werten.

Das Gütekriterium der Zumutbarkeit (vgl. Kapitel 4.8) ist für Entwicklungstests von besonderer Bedeutung, da Kinder schneller an ihre physischen und psychischen Grenzen stoßen als Erwachsene. Insbesondere jüngere Kinder sind nur eingeschränkt in der Lage, ihre Aufmerksamkeit über eine längere Zeit hinweg auf eine Testdurchführung zu fokussieren und aufrecht zu erhalten (Petermann & Macha, 2008). Durch die kindgerechte Gestaltung der Untersuchungssituation und entsprechendes Testleiterverhalten kann diesem Umstand begegnet werden. Kastner-Koller und Deimann (2002) unterstützen den Testanwender in dieser Hinsicht mit wertvollen Hinweisen.

Die im Vergleich zu anderen Entwicklungstests vergleichsweise lange Testdauer von mehr als 60 Minuten schränkt die Zumutbarkeit des WET auf den ersten Blick ein. In einer Studie von Krampen et al. (2008) konnte jedoch gezeigt werden, dass die WET-Testung lediglich geringe Auswirkungen auf die kurzzeitige Konzentrationsleistung hat. Dies spricht gegen eine zu starke Belastung der Kinder durch die Testung.

Der WET kann somit durchaus als zumutbar bezeichnet werden.

Die in dieser Arbeit gewonnen Erkenntnisse resultieren abschließend in der Ableitung einiger Empfehlungen für die zukünftige Qualitätssicherung des WET:

- Die im Rahmen dieser Arbeit durchgeführte testanalytische Untersuchung weist einen eingeschränkten Geltungsbereich auf, da deren Stichprobe lokal auf Wien beschränkt ist. Sie liefert jedoch erste Hinweise auf die Konstruktvalidität der Neuauflage. Dennoch ist es notwendig, die erzielten Ergebnisse anhand der Daten der Gesamtstichprobe zu überprüfen.
- Die Studie von Krampen et al. (2008) zur Aufmerksamkeitsbelastung der Kinder durch die WET-Testung kam zu dem Ergebnis, dass die Belastung der Kinder durch die Testung nicht allzu hoch ist. Die Ergebnisse betreffen jedoch die zweite Auflage des WET. Da die dritte Auflage zwei zusätzliche Subtests für Vier- und Fünfjährige vorsieht, und die Testdauer dadurch verlängert wird, erscheint eine Replikation der Studie empfehlenswert.
- Die Neunormierung erfolgte lediglich an vier- und fünfjährigen Kindern, also in jenem Altersbereich, für den die neuen Subtests vorgesehen sind. Für dreijährige Kinder werden weiterhin die Normen der zweiten Auflage verwendet. Die vor 13 Jahren erhobenen österreichischen Normdaten sind unter Bezugnahme auf den Flynn-Effekt jedoch als veraltet anzusehen. In Hinblick auf eine verantwortungsvolle Testpflege sollte somit in den nächsten Jahren eine Neunormierung auch an Dreijährigen erfolgen.
- Da Entwicklung in der frühen Kindheit relativ kontinuierlich in Richtung zunehmender Kompetenz verläuft (vgl. Kapitel 1) und diese Altersspanne zudem als optimale Phase für Entwicklungsinterventionen angesehen wird (McCall, 1987), sollte überlegt werden, ob die Skalen des WET auf den Altersbereich der

Zweijährigen ausgedehnt werden können oder eine sonstige Erweiterung des Verfahrens möglich und sinnvoll ist. Das ist vor allem auch insofern wünschenswert als für Zweijährige derzeit kein empfehlenswerter Breitbandentwicklungstest vorhanden ist.

Zusammenfassung

Die Arbeit entstand im Zusammenhang mit der geplanten dritten Auflage des Wiener Entwicklungstests (WET; Kastner-Koller & Deimann, 1998, 2002, in Vorbereitung), wobei die darin enthaltene Sammlung von Normierungsdaten Bestandteil der notwendigen Vorarbeiten für diese überarbeitete und erweiterte Auflage ist. Ziel der empirischen Untersuchung ist es, zur Testanalyse der Neuauflage beizutragen.

Die umfangreichen Ausführungen des theoretischen Teils bezwecken, den Leser bzw. die Leserin in die Allgemeine Entwicklungsdiagnostik in der frühen Kindheit einzuführen und sollen vor allem jenes Wissen vermitteln, das für die Einschätzung der Qualität von Entwicklungstests notwendig ist. Es werden sowohl die Grundlagen, Ziele und Methoden der Allgemeinen Entwicklungsdiagnostik dargelegt als auch aktuelle deutschsprachige Entwicklungstests und -screenings vorgestellt und anschließend kritisch gewürdigt.

Die testanalytische Untersuchung beruht auf den an einer Stichprobe von 164 Wiener Kindern im Alter zwischen 4;0 und 5;11 Jahren gewonnenen Normierungsdaten. Ihre Ergebnisse stützen die theoretische Vorannahme steigender Testleistungen mit zunehmendem Alter und belegen somit die Differenzierungsfähigkeit des WET in Hinblick auf verschiedene Altersgruppen. Sie deuten zudem auf eine hinreichende faktorielle Validität des Verfahrens hin. Die methodische Validierung der neu entwickelten Subtests *Rechnen* und *Muster Legen* hingegen ist lediglich in Maßen gelungen.

Insgesamt weisen die Ergebnisse darauf hin, dass die dritte Auflage des WET den Anforderungen an die Konstruktvalidität psychologischer Tests genügt. Die erzielten Ergebnisse sind jedoch anhand der Daten der Gesamtnormierungsstichprobe zu überprüfen.

In Hinblick auf eine verantwortungsvolle Testpflege wird empfohlen, in den nächsten Jahren eine Neunormierung auch an dreijährigen Kindern vorzunehmen.

Abstract

The thesis is linked to the forthcoming publication of the third edition of the Wiener Entwicklungstest (WET; Kastner-Koller & Deimann, 1998, 2002, in press). The included collection of normative data is an inherent part of the preliminaries, which have been necessary for this revised and extended edition. The goal of the empirical study is to make a contribution to the test analysis of the new edition.

The intention of the detailed explanations of the theoretical part is to introduce the reader to the field of general developmental diagnostics during early childhood and impart knowledge that is necessary for estimating the quality of developmental tests. Therefore, the fundamentals as well as the purposes and the methods of general developmental diagnostics are described. Additionally, current German developmental tests and screenings will be reviewed.

The test analysis is based on normative data that was collected from a sample of 164 Viennese children (ages 4;0 to 5;11). Results corroborate the theoretical assumption that performance increases with age and thereby prove the age-related differentiation ability of the WET. Furthermore, the results indicate that the factorial validity of the WET is sufficient. The methodical validation of the new subtests *Rechnen* and *Muster Legen* was moderately successful.

In conclusion, the results indicate that the WET (third edition) meets the criteria for construct validity of psychological tests. The results should however be reassessed on the basis of the data of the total standardization sample.

A standardization of the WET in the coming years also with three-year-old children is recommended to maintain the quality of the instrument.

Literaturverzeichnis

- Anglin, J. (1993). Vocabulary development: A morphological analysis. *Monographs of the Society for Research in Child Development*, 58 (10), 1-165.
- Atkinson, R. C. & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 2, pp. 89-195). New York: Academic Press.
- Aylward, G. P. (2004). Measures of infant and early childhood development. In G. Goldstein & S. R. Beers (Eds.), *Comprehensive handbook of psychological assessment: Vol. 1 Intellectual and neuropsychological assessment* (pp. 87-97). New York: Wiley.
- Baddeley, A. (2009). Working memory. In A. Baddeley, M. W. Eysenck & M. C. Anderson (Eds.), *Memory* (pp. 41-68). East Sussex: Psychology Press.
- Baddeley, A. D. & Lewis, V. J. (1981). Inner active processes in reading: The inner voice and the inner eye. In A. M. Lesgold & C. A. Perfetti (Eds.), *Interactive processes in reading* (pp. 107-129). Hillsdale: Erlbaum.
- Bayley, N. (1969). *Bayley Scales of Infant Development*. San Antonio: Psychological Corporation.
- Bayley, N. (2006). *Bayley Scales of Infant and Toddler Development* (3rd ed.). San Antonio: Psychological Corporation.
- Barth, K. (2008). *Die diagnostischen Einschätzskalen (DES) zur Beurteilung des Entwicklungsstandes und der Schulfähigkeit* (5., durchgesehene Aufl.). München: Reinhardt.
- Bauer, P. J. (2004). Getting explicit memory off the ground: Steps toward construction of a neuro-developmental account of changes in the first two years of life. *Developmental Review*, 24, 347-373.
- Berk, L. E. (2005). *Entwicklungspsychologie* (3., aktualisierte Aufl.) [E. Aralikatti, Übers.]. München: Pearson. (Original erschienen 2004: *Development through the lifespan*)
- Bortz, J. (2005). *Statistik für Human- und Sozialwissenschaftler* (6., vollständig überarbeitete und aktualisierte Aufl.). Heidelberg: Springer.
- Brannon, E. M. (2002). The development of ordinal numerical knowledge in infancy. *Cognition*, 83, 223-240.
- Bühler, C. & Hetzer, H. (1932). *Kleinkindertests. Entwicklungstests vom 1. bis 6. Lebensjahr*. Leipzig: Barth.
- Bühner, M. (2006). *Einführung in die Test- und Fragebogenkonstruktion* (2., aktualisierte Aufl.). München: Pearson.
- Cain, K., Oakhill, J. & Bryant, P. (2004). Children's reading comprehension ability: Concurrent prediction by working memory, verbal ability, and component skills. *Journal of Educational Psychology*, 96, 31-42

- Case, R. (1999). *Die geistige Entwicklung des Menschen: von der Geburt bis zum Erwachsenenalter* (H. A. Mund, Übers.). Heidelberg: Winter. (Original erschienen 1985: Intellectual development – Birth to adulthood)
- D'Amico, A. & Guarnera, M. (2005). Exploring working memory in children with low arithmetical achievement. *Learning and Individual Differences*, 15, 189-202.
- Deegener, G. (2009). Eltern und Erzieher als Informationsquellen. In D. Irblich & G. Renner (Hrsg.), *Diagnostik in der Klinischen Kinderpsychologie. Die ersten sieben Lebensjahre* (S. 86-96). Göttingen Hogrefe.
- Deimann, P. & Kastner-Koller, U. (1994). Wie gut ist mein Kind entwickelt? Entwicklungsdiagnostik im Vorschulalter. In G. Gittler, M. Jirasko, U. Kastner-Koller, C. Korunka & A. Al-Roubaie (Hrsg.). *Die Seele ist ein weites Land: Aktuelle Forschung am Wiener Institut für Psychologie* (S. 232-244). Wien: WUV-Universitätsverlag.
- Deimann, P. & Kastner-Koller, U. (2007). Entwicklungsdiagnostik. In M. Hasselhorn & W. Schneider (Hrsg.), *Handbuch der Entwicklungspsychologie* (S. 558-569). Göttingen: Hogrefe.
- Deimann, P., Kastner-Koller, U., Benka, M., Kainz, S. & Schmidt, H. (2005). Mütter als Entwicklungsdiagnostikerinnen: Der Entwicklungsstand von Kindergartenkindern im Urteil ihrer Mütter. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 37 (3), 122–134.
- Döpfner, M., Lehmkuhl, G., Heubrock, D. & Petermann F. (2000). *Leitfaden Kinder- und Jugendpsychotherapie (Bd. 2). Diagnostik psychischer Störungen im Kindes- und Jugendalter*. Göttingen: Hogrefe.
- Dufresne, A. & Kobasigawa, A. (1989). Children's spontaneous allocation of study time: Differential and sufficient aspects. *Journal of Experimental Child Psychology*, 47, 274-296.
- Esser, G. & Wyschkon, A. (2002). *Basisdiagnostik für umschriebene Entwicklungsstörungen im Vorschulalter (BUEVA)*. Göttingen: Beltz Test
- Ettrich, K. U. (2000). *Entwicklungsdiagnostik im Vorschulalter. Grundlagen, Verfahren, Neuentwicklungen, Screenings*. Göttingen: Vandenhoeck & Ruprecht.
- Faßnacht, G. (1995). *Systematische Verhaltensbeobachtung. Eine Einführung in die Methodologie und Praxis* (2., völlig neubearbeitete Aufl.). München: Reinhardt.
- Field, A. (2005). *Discovering statistics using SPSS* (2nd ed.). London: Sage.
- Filipp, S.-H. & Doenges, D. (1983). Entwicklungstests. In K.-J. Groffmann & L. Michel (Hrsg.), *Enzyklopädie der Psychologie: Themenbereich B Methodologie und Methoden, Serie II Psychologische Diagnostik, Bd. 2 Intelligenz- und Leistungsdiagnostik* (S. 202-306). Göttingen: Hogrefe.
- Flammer, A. (2003). *Entwicklungstheorien: Psychologische Theorien der menschlichen Entwicklung* (3., korrigierte Aufl.). Bern: Huber.
- Flavell, J. H. (1971). First discussant's comments: What is memory development the development of? *Human Development*, 14, 272-278.
- Flavell, J. H. (1979). *Kognitive Entwicklung* (H. Steiner, Übers.). Stuttgart: Klett-Cotta. (Original erschienen 1977: Cognitive development)

- Flynn, J. R. (1987). Massive IQ gains in 14 nations: What IQ tests really measure. *Psychological Bulletin*, 101, 171-191.
- Frankenburg, W. K. & Dodds, J. B. (1967). The Denver Developmental Screening Test. *Journal of Pediatrics*, 71, 181-191
- Fuiko, R. & Wurst, E. (2003). Entwicklungsdiagnostik. In K. Kubinger & R. Jäger (Hrsg.), *Schlüsselbegriffe der Psychologischen Diagnostik* (S. 119-123). Weinheim: Psychologie Verlags Union.
- Gathercole, S. E., Tiffany, C., Briscoe, J. & Thorn, A. (2005). Developmental consequences of poor phonological short-term memory function in childhood: a longitudinal study. *Journal of Child Psychology and Psychiatry*, 46, 598-611.
- Gesell, A. & Amatruda, C. (1941). Developmental diagnosis. Normal and abnormal child development. New York: Harper & Row.
- Glascoe, F. P (2000). Evidence-based approach to developmental and behavioural surveillance using parents' concerns. *Child: Care, Health and Development*, 26, 137-149.
- Glascoe, F. P & Dworkin P. H. (1995). The role of parents in the detection of developmental and behavioral problems. *Pediatrics*, 95, 829-836.
- Goswami, U. (2001). *So denken Kinder: Einführung in die Psychologie der kognitiven Entwicklung* (M. Wengenroth, Übers.). Bern: Huber (Original erschienen 1998: Cognition in children)
- Grimm, H. (2003). *Störungen der Sprachentwicklung: Grundlagen - Ursachen - Diagnose - Intervention - Prävention*. (2., überarbeitete Aufl). Göttingen: Hogrefe.
- Grob, A., Hagmann-von Arx, P. & Bodmer, N. M. (2009). Entwicklungsdiagnostik. In S. Schneider & J. Margraf (Hrsg.), *Lehrbuch der Verhaltenstherapie, Bd. 3 Störungen im Kindes- und Jugendalter* (S. 145-158). Heidelberg: Springer.
- Groffmann, K.-J. (1983). Die Entwicklung der Intelligenzmessung. In K.-J. Groffmann & L. Michel (Hrsg.), *Enzyklopädie der Psychologie: Themenbereich B Methodologie und Methoden, Serie II Psychologische Diagnostik, Bd. 2 Intelligenz- und Leistungsdiagnostik* (S. 2-103). Göttingen: Hogrefe.
- Grube, D. & Hasselhorn, M. (2006). Längsschnittliche Analysen zur Lese-, Rechtschreib- und Mathematikleistung im Grundschulalter: zur Rolle von Vorwissen, Intelligenz, phonologischem Arbeitsgedächtnis und phonologischer Bewusstheit. In I. Hosenfeld & F.-W. Schrader, *Schulische Leistung: Grundlagen, Bedingungen, Perspektiven* (S. 87-105). Münster: Waxmann.
- Grundmann, M., Welskopf, R. & Sturzbecher, D. (2001). Wozu braucht man kindliche Einschätzungen von Betreuungskontexten? In D. Sturzbecher (Hrsg.), *Spielbasierte Befragungstechniken. Interaktionsdiagnostische Verfahren für Begutachtung, Beratung und Forschung* (S. 11-23). Göttingen: Hogrefe.
- Habermas, T. & Döll-Hentschker, S. (2007). Psychoanalytische Grundlagen der Entwicklungspsychologie. In M. Hasselhorn & W. Schneider (Hrsg.), *Handbuch der Entwicklungspsychologie* (S. 62-70). Göttingen: Hogrefe.

- Hartig, J., Frey, A. & Jude, N. (2007). Validität. In H. Moosbrugger & A. Kelava (Hrsg.). *Testtheorie und Fragebogenkonstruktion* (S. 135-164). Heidelberg: Springer.
- Hay, D. F., Pawlby, S., Sharp, D., Schmücker, G., Mills, A., Allen, H. & Kumar, R. (1999). Parents' judgements about young children's problems: Why mothers and fathers might disagree yet still predict later outcomes. *Journal of Child Psychology and Psychiatry*, 40, 1249-1258.
- Hellbrügge, T., Lajosi, F., Menara, D., Schamberger, R. & Rautenstrauch, T. (1978). *Münchener Funktionelle Entwicklungsdiagnostik. Erstes Lebensjahr*. München: Urban & Schwarzenberg.
- Hetzer, H. (1950). *Entwicklungstestverfahren. Aufbau, Ausbau und praktische Anwendung*. Lindau: Piorkowski.
- Hitch, J. G. & McAuley, E. (1991). Working memory in children with specific arithmetical learning difficulties. *British Journal of Psychology*, 82, 375-386.
- Hitch, G. J., Halliday, S., Schaafstal, A. M. & Schraagen, J. M. C. (1988). Visual working memory in young children. *Memory & Cognition*, 16, 120-132.
- Janssen, J. & Laatz, W. (2007). *Statistische Datenanalyse mit SPSS für Windows: Eine anwendungsorientierte Einführung in das Basissystem und das Modul Exakte Tests* (6., neu bearbeitete und erweiterte Aufl.). Heidelberg: Springer.
- Johnson, S. & Marlow, N. (2006). Developmental screen or developmental testing? *Early Human Development*, 82, 173-183.
- Jones, H. E. & Bayley, N. (1941). The Berkeley Growth Study. *Child Development*, 12 (2), 167-173.
- Kail, R. (1991). Developmental changes and speed processing during childhood and adolescence. *Psychological Bulletin*, 109, 490-501.
- Kastner-Koller, U. & Deimann, P. (1991). Entwicklungsdiagnostik im Vorschulalter: Raschanalyse der Testreihen III-VI des Bühler-Hetzer-Kleinkindertests nach inhaltlicher Itemselektion. *Schweizerische Zeitschrift für Psychologie*, 50, 111-118.
- Kastner-Koller, U. & Deimann, P. (2002). *Wiener Entwicklungstest (WET). Ein Verfahren zur Erfassung des allgemeinen Entwicklungsstandes bei Kindern von 3 bis 6 Jahren* (2., überarbeitete und neu normierte Aufl.). Göttingen: Hogrefe.
- Kastner-Koller, U. & Deimann, P. (2009). Beobachtung und Befragung von Kindern. In D. Irlich & G. Renner (Hrsg.), *Diagnostik in der Klinischen Kinderpsychologie. Die ersten sieben Lebensjahre* (S. 97-107). Göttingen Hogrefe.
- Knopf, M. & Schneider, W. (1998). Die Entwicklung des kindlichen Denkens und die Verbesserung der Lern- und Gedächtniskompetenzen. In F. E. Weinert (Hrsg.), *Entwicklung im Kindesalter* (S. 75-94). Weinheim: Psychologie Verlags Union.
- Kopp, C. B. & McCall R. B. (1982). Predicting later mental performance für normal, at-risk and handicapped infants. In P. B. Baltes & O. G. Brim, Jr. (Eds.), *Life-span development and behavior*, Vol. 4 (pp. 33-61). New York: Academic Press.
- Krajewski, K. (2008). Vorschulische Förderung bei beeinträchtigter Entwicklung mathematischer Kompetenzen. In J. Borchert, B. Hartke und P. Jogschies (Hrsg.), *Frühe Förderung entwicklungsauffälliger Kinder und Jugendlicher* (S. 122-135). Stuttgart: Kohlhammer.

- Kubinger, K. D. (2006). *Psychologische Diagnostik. Theorie und Praxis psychologischen Diagnostizierens*. Göttingen: Hogrefe.
- Kubinger, K. D. & Deegener, G. (2001). *Psychologische Anamnese bei Kindern und Jugendlichen*. Göttingen: Hogrefe.
- Kurdek, L. A. & Sinclair, R. J. (2001). Predicting reading and mathematics achievement in fourth-grade children from kindergarten readiness scores. *Journal of Educational Psychology*, 93, 451-455.
- Laible, D. J. & Thompson, R. A. (1998). Attachment and emotional understanding in preschool children. *Developmental Psychology*, 34, 1038-1045.
- La Greca, A. M., Kuttler, A. F. & Stone, W. L. (2001). Assessing children through interviews and behavioral observations. In C. E. Walker & M. C. Roberts (Eds.), *Handbook of clinical child psychology* (pp. 90-110). New York: Wiley.
- Lee, K.-M. & Kang, S.-Y. (2002). Arithmetic operation and working memory: differential suppression in dual tasks. *Cognition*, 83, 63-68.
- Lienert, G. A. & Raatz, U. (1998). *Testaufbau und Testanalyse* (6. Aufl.). Weinheim: Psychologie Verlags Union.
- Macha, T., Proske, A. & Petermann, F. (2005). Validität von Entwicklungstests. *Kindheit und Entwicklung*, 14, 150-162.
- Magistrat der Stadt Wien, Magistratsabteilung 5 (Finanzwirtschaft, Haushaltswesen und Statistik (Hrsg.). (2008). *Statistisches Jahrbuch der Stadt Wien 2008*. Wien: Herausgeber.
- Mähler, C. (2007). Kindergarten- und Vorschulalter. In M. Hasselhorn & W. Schneider (Hrsg.), *Handbuch der Psychologie, Bd. 7 Handbuch der Entwicklungspsychologie* (S. 164-174). Göttingen: Hogrefe.
- Mähler, C. (2008). Das Kindergarten- und Vorschulalter: 4. bis 7. Lebensjahr. In M. Hasselhorn & R. K. Silberstein (Hrsg.), *Enzyklopädie der Psychologie: Themenbereich C Theorie und Forschung, Serie V Entwicklungspsychologie, Bd. 4 Entwicklungspsychologie des Säuglings- und Kindesalters* (S. 177-239). Göttingen: Hogrefe.
- McCall, R. B. (1981). Nature-nurture and the two realms of development: A proposed integration with respect to mental development. *Child Development*, 52, 1-12.
- McCall, R. B. (1987). Predictionists and interventionists: A synthesis. *Peabody Journal of Education*, 65, 66-86.
- McCarthy, D. (1972). *McCarthy Scales of Children's Abilities*. New York: The Psychological Corporation.
- Melchers, P., Floß, S., Brandt, I., Eßer, K. J., Lehmkuhl, G., Rauh H. & Sticker, E. (2003). *Erweiterte Vorsorgeuntersuchung (EVU). Entwicklungsneurologische und entwicklungspsychologische Ergänzung zur pädiatrischen Vorsorge bei U4 bis U9*. Leiden: PITS.
- Menyuk, P. (2000). Wichtige Aspekte der lexikalischen und semantischen Entwicklung. In H. Grimm (Hrsg.), *Enzyklopädie der Psychologie: Themenbereich C Theorie und Forschung, Serie III Sprache, Bd. 3 Sprachentwicklung* (S. 171-192). Göttingen: Hogrefe.

- Mietzel, G. (2002). *Wege in die Entwicklungspsychologie: Kindheit und Jugend* (4., vollständig überarbeitete Aufl.). Weinheim: Psychologie Verlags Union.
- Miller, P. H. (1993). *Theorien der Entwicklungspsychologie* (A. Hildebrandt-Essig, Übers.). Heidelberg: Spektrum. (Original erschienen 1993: *Theories of development*, 3rd ed.)
- Montada, L. (1985). Entwicklungsberatung als Angewandte Entwicklungspsychologie. In J. Brandstätter & H. Gräser (Hrsg.), *Entwicklungsberatung unter dem Aspekt der Lebensspanne* (S. 30-43). Göttingen: Hogrefe.
- Moosbrugger, H. & Kelava, A. (2007). Anforderungen an einen psychologischen Test (Testgütekriterien). In H. Moosbrugger & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (S. 7-26). Heidelberg: Springer.
- Müntefering, F. (2008). Frühförderung behinderter Kinder als Auftrag an alle. In C. Leydendecker (Hg.), *Gemeinsam Handeln statt Behandeln: Aufgaben und Perspektiven der Komplexleistung Frühförderung* (S. 15-21). München: Reinhardt.
- Oerter, R. (1977). Zur Rolle der Entwicklungspsychologie. *Psychologische Rundschau*, 28, 22-30.
- Oerter, R. & Montada, L. (Hrsg.). (2008). *Entwicklungspsychologie* (6., vollständig überarbeitete Aufl.). Weinheim: Psychologie Verlags Union.
- Petermann, F. & Macha, T. (2003). Strategien in der testgestützten allgemeinen Entwicklungsdiagnostik. *Monatszeitschrift Kinderheilkunde*, 151, 6-13.
- Petermann, F. & Macha, T. (2005). Entwicklungsdiagnostik. *Kindheit und Entwicklung*, 14, 131-139.
- Petermann, F. & Macha, T. (2008). Entwicklungsdiagnostik. In F. Petermann & W. Schneider (Hrsg.), *Enzyklopädie der Psychologie: Themenbereich C Theorie und Forschung, Serie V Entwicklungspsychologie, Bd. 7 Angewandte Entwicklungspsychologie* (S. 19-59). Göttingen: Hogrefe.
- Petermann, F. & Petermann, U. (Hrsg.). (2007). *Hamburg-Wechsler-Intelligenztest für Kinder - Vierte Auflage (HAWIK IV). Übersetzung und Adaptation der WISC-IV von David Wechsler*. Bern: Huber.
- Petermann, F. & Wiedbusch, S. (2003). *Emotionalhe Kompetenz bei Kindern*. Göttingen: Hogrefe.
- Petermann, F., Stein I. A. & Macha T. (2004). *Entwicklungsdiagnostik mit dem ET 6-6*. (2., veränderte Aufl.). Frankfurt a. M.: Harcourt.
- Piaget, J. (1966). *Psychologie der Intelligenz* (2. Aufl., vollständig überarbeitete Übers.). Zürich: Rascher. (Original erschienen 1947: *La psychologie de l'intelligence*)
- Piaget, J. (1974). *Der Aufbau der Wirklichkeit beim Kinde* (J.-U. Sandberger, C. Thirion & H.-L. Wunberg, Übers.). Stuttgart: Klett. (Original erschienen 1950: *La construction du réel chez l'enfant*)
- Piaget, J. (2003a). *Das Erwachen der Intelligenz beim Kinde* (5. Aufl.) [B. Seiler, Übers.]. Stuttgart: Klett-Cotta. (Original erschienen 1959: *La naissance de l'intelligence chez l'enfant*)
- Piaget, J. (2003b). *Meine Theorie der geistigen Entwicklung* (R. Fatke, Hrsg; H. Kober, Übers.). Weinheim: Beltz. (Original erschienen 1970: *Piaget's theory*)

- Piaget, J. & Inhelder, B. (1972). *Die Psychologie des Kindes* (L. Häfliger, Übers.). Olten: Walter. (Original erschienen 1966: *La psychologie de l'enfant*)
- Rauh, H. (1985). Diagnostik und Beratung in der frühkindlichen Entwicklung. In J. Brandstätter & H. Gräser (Hrsg.), *Entwicklungsberatung unter dem Aspekt der Lebensspanne* (S. 44-64). Göttingen: Hogrefe.
- Rennen-Allhoff B (1991). Wie verlässlich sind Elternangaben? *Praxis der Kinderpsychologie und Kinderpsychiatrie*, 40, 333-338.
- Rennen-Allhoff, B. & Allhoff, P. (1987). *Entwicklungstests für das Säuglings-, Kleinkind- und Vorschulalter*. Berlin: Springer.
- Rennen-Allhoff, B., Allhoff, P., Bowi, U. & Laaser, U. (1993). *Elternbeteiligung bei Entwicklungsdiagnostik und Vorsorge. Verfahren zur Früherkennung von Krankheiten und Entwicklungsstörungen im Säuglings- und Kindesalter durch Eltern*. Weinheim: Juventa.
- Renziehausen, A. (2003). Wiener Entwicklungstest (WET) von Ursula Kastner-Koller und Pia Deimann. Ein Verfahren zur Erfassung des allgemeinen Entwicklungsstandes bei Kindern von 3 bis 6 Jahren (1. Aufl. 1998; 2., überarbeitete und neu normierte Aufl. 2002) [Testinformationen]. *Diagnostica*, 49 (3), 140-145.
- Reuner, G., Rosenkranz, J., Pietz, J. & Horn, R. (Hrsg.) (2007). *Bayley Scales of Infant Development, 2nd ed. (Bayley II). Deutsche Fassung (Bayley II)*. Frankfurt a. M.: Harcourt Test Services.
- Roebbers, C. M. & Elischberger, H. B. (2002). Autobiographische Erinnerung bei jungen Kindern: Möglichkeiten und Grenzen bei der Verbesserung ihrer freien Berichte. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 34, 2-10.
- Rousselle, L., Palmers, E. & Noël, M.-P. (2004). Magnitude comparison in preschoolers: What counts? Influence of perceptual variables. *Journal of Experimental Child Psychology*, 87, 57-84.
- Scheid, V. (1994). Motorische Entwicklung in der frühen Kindheit. In J. Baur, K. Bös & R. Singer (Hrsg.), *Motorische Entwicklung. Ein Handbuch* (S. 260-275). Schorndorf: Hofmann.
- Schenk-Danzinger, L. (2006). *Entwicklungspsychologie* (2. Aufl., völlig neu bearbeitet von K. Rieder). Wien: Österreichischer Bundesverlag & HPT.
- Schneider, W. & Büttner, G. (2008). Entwicklung des Gedächtnisses bei Kindern und Jugendlichen. In R. Oerter & L. Montada (Hrsg.), *Entwicklungspsychologie* (6., vollständig überarbeitete Aufl., S. 480-501). Weinheim: Psychologie Verlags Union.
- Schmidt, H. D. (1970). *Allgemeine Entwicklungspsychologie*. Berlin: VEB Deutscher Verlag der Wissenschaften.
- Schuchardt, K., Kunze, J., Grube, D. & Hasselhorn, M. (2006). Arbeitsgedächtnisdefizite bei Kindern mit schwachen Rechen- und Schriftsprachleistungen. *Zeitschrift für Pädagogische Psychologie*, 20, 261-268.
- Schwarzer, G. & Jovanovic, B. (2007). Kleinkindalter. In M. Hasselhorn & W. Schneider (Hrsg.), *Handbuch der Psychologie, Bd. 7 Handbuch der Entwicklungspsychologie* (S. 153-163). Göttingen: Hogrefe.

- Siegler, R. S. (2001). *Das Denken von Kindern* (3. Aufl.) [J. Schmidt, Übers.]. München: Oldenbourg. (Original erschienen 1998: *Children's thinking*, 3rd ed.)
- Sodian, B. (2008). Entwicklung des Denkens. In R. Oerter & L. Montada (Hrsg.), *Entwicklungspsychologie* (6., vollständig überarbeitete Aufl., S. 436-479). Weinheim: Psychologie Verlags Union.
- Spangler, G. & Schwarzer, G. (2008). Kleinkindalter. In M. Hasselhorn & R. K. Silberreisen (Hrsg.), *Enzyklopädie der Psychologie: Themenbereich C Theorie und Forschung, Serie V Entwicklungspsychologie, Bd. 4 Entwicklungspsychologie des Säuglings- und Kindesalters* (S. 127-176). Göttingen: Hogrefe.
- Starkey, P., Spelke, E. S. & Gelman, R. (1990). Numerical abstraction by human infants. *Cognition*, 36, 97-127.
- Stein, I. A. & Petermann, F. (1997). Konstruktion eines neuen Entwicklungstests für Kinder im Alter von 0 bis 6 Jahren. *Kindheit und Entwicklung*, 6, 183-186.
- Steinebach, C. (2000). *Entwicklungspsychologie*. Stuttgart: Klett-Cotta.
- Stern, E. (1998). Die Entwicklung schulbezogener Kompetenzen. In F. E. Weinert (Hrsg.), *Entwicklung im Kindesalter* (S. 95-113). Weinheim: Psychologie Verlags Union.
- Sturzbecher, D. & Grundmann, M. (2001). Vorschulkinder als unsichere Informationsquelle – alte Klischees und neue Perspektiven. In D. Sturzbecher (Hrsg.), *Spielbasierte Befragungstechniken. Interaktionsdiagnostische Verfahren für Begutachtung, Beratung und Forschung* (S. 37-50). Göttingen: Hogrefe.
- Tang, C. M. (2006). Developmentally sensitive forensic interviewing of preschool children: Some guidelines drawn from basic psychological research. *Criminal Justice Review*, 31, 132-145.
- Tewes, U., Rossmann, P. & Schallberger, U. (Hrsg.). (2002). *Hamburg-Wechsler-Intelligenztest für Kinder - Dritte Auflage (HAWIK III). Übersetzung und Adaptation der WISC-III Wechsler Intelligence Scale for Children - Third Edition von David Wechsler* (3., überarbeitete und ergänzte Aufl.). Bern: Huber.
- Trautner, H. M. (1992). *Lehrbuch der Entwicklungspsychologie: Bd. 1 Grundlagen und Methoden* (2. Aufl.). Göttingen: Hogrefe.
- Trautner, H. M. (2003). *Allgemeine Entwicklungspsychologie* (2. Aufl.). Stuttgart: Kohlhammer.
- Trautner, H. M. (2006). Entwicklungsbegriffe. In W. Schneider & F. Wilkening (Hrsg.), *Enzyklopädie der Psychologie: Themenbereich C Theorie und Forschung, Serie V Entwicklungspsychologie, Bd. 1 Theorien, Modelle und Methoden der Entwicklungspsychologie* (S. 59-89). Göttingen: Hogrefe.
- Tröster, H., Flender, J. & Reineke, D. (2004). *Dortmunder Entwicklungsscreening für den Kindergarten (DESK 3-6)*. Göttingen: Hogrefe.
- Tröster, H. & Reineke, D. (2006). Wie gut können Entwicklungsauffälligkeiten von Schulanfängern bereits im Kindergarten entdeckt werden. *Psychologie in Erziehung und Unterricht*, 53, 22-34.
- Volbert, R. & Pieters, V. (1996). Suggestive Beeinflussung von Kinderaussagen. *Psychologische Rundschau*, 47, 183-198.

- Wendt, D. (1997). *Entwicklungspsychologie. Eine Einführung*. Stuttgart: Kohlhammer.
- Weinert, E. F. (1998). Überblick über die psychische Entwicklung im Kindesalter: Was wir darüber wissen, was wir noch nicht wissen und was wir wissen sollten. In F. E. Weinert (Hrsg.), *Entwicklung im Kindesalter* (S. 1-36). Weinheim: Psychologie Verlags Union.
- Weinert, S. (2007). Spracherwerb. In M. Hasselhorn & W. Schneider (Hrsg.), *Handbuch der Psychologie, Bd. 7 Handbuch der Entwicklungspsychologie* (S. 221-231). Göttingen: Hogrefe.
- Weinert, S. & Grimm, H. (2008). Sprachentwicklung. In R. Oerter & L. Montada (Hrsg.), *Entwicklungspsychologie* (6., vollständig überarbeitete Aufl., S. 502-534). Weinheim: Psychologie Verlags Union.
- Weinert, S. & Weinert, E. F. (2006). Entwicklung der Entwicklungspsychologie: Wurzeln, Meilensteine, Entwicklungslinien. In W. Schneider & F. Wilkening (Hrsg.), *Enzyklopädie der Psychologie: Themenbereich C Theorie und Forschung, Serie V Entwicklungspsychologie, Bd. 1 Theorien, Modelle und Methoden der Entwicklungspsychologie* (S. 3-58). Göttingen: Hogrefe.
- Wellman, H. M., Harris, P. L., Banerjee, M. & Sinclair, A. (1995). Early understanding of emotion: Evidence from natural language. *Cognition and Emotion*, 9, 117-149.
- Wilkening, F. & Krist, H. (2008). Entwicklung der Wahrnehmung und Psychomotorik. In R. Oerter & L. Montada (Hrsg.), *Entwicklungspsychologie* (6., vollständig überarbeitete Aufl., S. 411-435). Weinheim: Psychologie Verlags Union.
- Winter, R. & Roth, K. (1994). Entwicklung motorischer Fertigkeiten. In J. Baur, K. Bös & R. Singer (Hrsg.), *Motorische Entwicklung. Ein Handbuch* (S. 217-237). Schorndorf: Hofmann.
- Wynn, K. (1995). Infants possess a system of numerical knowledge. *Current Directions in Psychological Science*, 4, 172-177.
- Xu, F. (2003). Numerosity discrimination in infants: Evidence for two systems of representations. *Cognition*, 89, B15-B25.

Tabellenverzeichnis

Tabelle 1: Kindergärten und Anzahl der getesteten Kinder	106
Tabelle 2: Aufteilung der Kinder nach Altersgruppe und Geschlecht	107
Tabelle 3: Muttersprache der Kinder.....	108
Tabelle 4: Berufstätigkeit der Hauptverdienenden.....	109
Tabelle 5: Durchschnittliche Leistungen der Gesamtstichprobe im WET.....	110
Tabelle 6: Signifikante Unterschiede bei den WET-Ergebnissen zwischen Kindern mit deutscher und anderer Muttersprache.....	112
Tabelle 7: Verteilung der Kinder nach Muttersprache und Testleiterin.....	113
Tabelle 8: Verteilung der Testleiterinnen auf die Kindergärten.....	114
Tabelle 9: Verteilung der Kinder nach Berufstätigkeit der Hauptverdiener und Testleitergruppen.....	114
Tabelle 10: Subtestleistungen Vierjähriger und Fünfjähriger: Ergebnis der U-Tests	118
Tabelle 11: Geplante Kontraste	120
Tabelle 12: Leistungsunterschiede in den Testleistungen zwischen den vier Altersgruppen: Ergebnis der Kontrast-Tests	120
Tabelle 13: Leistungsunterschiede in den Testleistungen zwischen den vier Altersgruppen: Ergebnis der Jonckheere-Terpstra-Tests	122
Tabelle 14: Unterschiede zwischen den Halbjahresgruppen bei Subtests mit altersangepasster Schwierigkeit	123
Tabelle 15: Eigenwertverlauf und erklärte Varianz vor und nach der Rotation	126
Tabelle 16: Ergebnisse der Hauptkomponentenanalyse nach Varimax-Rotation	127
Tabelle 17: Itemanalyse der Skala <i>Muster Legen</i> (Vierjährige)	128
Tabelle 18: Itemanalyse der Skala <i>Muster Legen</i> (Fünfjährige)	129
Tabelle 19: Itemanalyse der Skala <i>Rechnen</i> (Vierjährige).....	130
Tabelle 20: Itemanalyse der Skala <i>Rechnen</i> (Fünfjährige)	131

Anhang

- A** Überprüfung der Voraussetzungen für parametrische Verfahren
(Tabelle A-1 bis A-10)
- B** Ergebnisse der Varianzanalysen und Kruskal-Wallis-Tests
(Tabelle B-1 bis B-3)
- C** Graphische Darstellung der Leistungsunterschiede der vier Altersgruppen
(Abbildung C-1 bis C-16)
- D** Elternbrief einschließlich Einverständniserklärung

Anhang A: Überprüfung der Voraussetzungen für parametrische Verfahren

Tabelle A-1: Normalverteilung der WET-Ergebnisse (Normwerte) nach Muttersprache

	Sprachgruppe	Kolmogorov-Smirnov ^a		
		Z	df	Sig.
C-Wert Lernbär	deutsch	.237	92	< .001
	andere	.274	72	< .001
C-Wert Quiz	deutsch	.126	92	.001
	andere	.146	72	.001
C-Wert Bilderlotto	deutsch	.145	92	< .001
	andere	.125	72	.007
C-Wert Puppenspiel	deutsch	.198	92	< .001
	andere	.214	72	< .001
C-Wert Schatzkästchen	deutsch	.121	92	.002
	andere	.142	72	.001
C-Wert Bunte Formen	deutsch	.126	92	.001
	andere	.128	72	.005
C-Wert Zahlen Merken	deutsch	.143	92	< .001
	andere	.133	72	.003
C-Wert Wörter Erklären	deutsch	.145	92	< .001
	andere	.132	72	.003
C-Wert Turnen	deutsch	.194	91	< .001
	andere	.187	72	< .001
C-Wert Nachzeichnen	deutsch	.136	92	< .001
	andere	.179	72	< .001
C-Wert Gegensätze	deutsch	.140	92	< .001
	andere	.177	72	< .001
C-Wert Fotoalbum	deutsch	.168	92	< .001
	andere	.163	72	< .001
C-Wert Elternfragebogen	deutsch	.154	.81	< .001
	andere	.120	.63	.024
C-Wert Gesamtentwicklungsscore	deutsch	.138	92	< .001
	andere	.161	72	< .001
Prozentrang Range	deutsch	.191	92	< .001
	andere	.204	72	< .001

a. Signifikanzkorrektur nach Lilliefors.

Tabelle A-2: Homogenität der Varianzen – WET-Ergebnisse (Normwerte) nach Muttersprache

	Levene Test auf Varianzgleichheit			
	F*	df1	df2	Sig.
C-Wert Lernbär	4.528	1	162	.035
C-Wert Quiz	.418	1	162	.519
C-Wert Bilderlotto	2.132	1	162	.146
C-Wert Puppenspiel	.115	1	162	.735
C-Wert Schatzkästchen	.017	1	162	.896
C-Wert Bunte Formen	.001	1	162	.976
C-Wert Zahlen Merken	1.732	1	162	.190
C-Wert Wörter Erklären	.141	1	162	.708
C-Wert Turnen	.430	1	161	.513
C-Wert Nachzeichnen	.114	1	162	.736
C-Wert Gegensätze	2.121	1	162	.147
C-Wert Fotoalbum	.244	1	162	.622
C-Wert Elternfragebogen	1.066	1	142	.304
C-Wert Gesamtentwicklungsscore	.158	1	162	.692
Prozentrang Range	.110	1	162	.741

* Basierend auf dem Mittelwert.

Tabelle A-3: Normalverteilung der WET-Ergebnisse (Normwerte) nach Geschlecht

	Geschlecht	Kolmogorov-Smirnov ^a		
		Z	df	Sig.
C-Wert Lernbär	männlich	.291	83	< .001
	weiblich	.226	81	< .001
C-Wert Quiz	männlich	.110	83	.015
	weiblich	.122	81	.005
C-Wert Bilderlotto	männlich	.131	83	.001
	weiblich	.136	81	.001
C-Wert Puppenspiel	männlich	.172	83	< .001
	weiblich	.178	81	< .001
C-Wert Schatzkästchen	männlich	.136	83	.001
	weiblich	.121	81	.005
C-Wert Bunte Formen	männlich	.151	83	< .001
	weiblich	.135	81	.001
C-Wert Zahlen Merken	männlich	.150	83	< .001
	weiblich	.120	81	.005
C-Wert Wörter Erklären	männlich	.156	83	< .001
	weiblich	.169	81	< .001
C-Wert Turnen	männlich	.154	82	< .001
	weiblich	.209	81	< .001
C-Wert Nachzeichnen	männlich	.163	83	< .001
	weiblich	.149	81	< .001
C-Wert Gegensätze	männlich	.110	83	.014
	weiblich	.123	81	.004
C-Wert Fotoalbum	männlich	.142	83	< .001
	weiblich	.171	81	< .001
C-Wert Elternfragebogen	männlich	.157	70	< .001
	weiblich	.153	74	< .001
C-Wert Gesamtentwicklungsscore	männlich	.172	83	< .001
	weiblich	.115	81	.010
Prozentrang Range	männlich	.175	83	< .001
	weiblich	.218	81	< .001

a. Signifikanzkorrektur nach Lilliefors.

Tabelle A-4: Homogenität der Varianzen – WET-Ergebnisse (Normwerte) nach Geschlecht

	Levene Test auf Varianzgleichheit			
	F*	df1	df2	Sig.
C-Wert Lernbär	1.074	1	162	.302
C-Wert Quiz	.021	1	162	.884
C-Wert Bilderlotto	.200	1	162	.655
C-Wert Puppenspiel	.013	1	162	.908
C-Wert Schatzkästchen	.186	1	162	.667
C-Wert Bunte Formen	11.175	1	162	.001
C-Wert Zahlen Merken	1.995	1	162	.160
C-Wert Wörter Erklären	.348	1	162	.556
C-Wert Turnen	.014	1	161	.905
C-Wert Nachzeichnen	.165	1	162	.685
C-Wert Gegensätze	.699	1	162	.404
C-Wert Fotoalbum	.473	1	162	.492
C-Wert Elternfragebogen	3.346	1	142	.069
C-Wert Gesamtentwicklungsscore	1.216	1	162	.272
Prozentrang Range	.041	1	162	.840

*Basierend auf dem Mittelwert.

Tabelle A-5: Normalverteilung der WET-Ergebnisse (Rohwerte) bei Subtests mit gleichen Aufgaben für Vierjährige und Fünfjährige pro Jahresgruppe

	Jahresgruppe	Kolmogorov-Smirnov ^a		
		Z	df	Sig.
Lernbär	4;0 – 4;11	.417	81	< .001
	5;0 – 5;11	.401	83	< .001
Bilderlotto: <i>Langform</i>	4;0 – 4;11	.098	79	.057
	5;0 – 5;11	.108	83	.018
Puppenspiel	4;0 – 4;11	.099	81	.049
	5;0 – 5;11	.150	83	< .001
Schatzkästchen: <i>Unmittelbares Reproduzieren</i>	4;0 – 4;11	.208	81	< .001
	5;0 – 5;11	.151	83	< .001
Schatzkästchen: <i>Behaltensleistung nach 20 Minuten</i>	4;0 – 4;11	.355	81	< .001
	5;0 – 5;11	.335	83	< .001
Schatzkästchen: <i>Anzahl der Lerndurchgänge</i>	4;0 – 4;11	.206	81	< .001
	5;0 – 5;11	.231	83	< .001
Bunte Formen	4;0 – 4;11	.105	81	.027
	5;0 – 5;11	.175	83	< .001
Zahlen Merken: <i>Gedächtnisspanne</i>	4;0 – 4;11	.270	81	< .001
	5;0 – 5;11	.310	83	< .001
Wörter Erklären	4;0 – 4;11	.079	81	.200*
	5;0 – 5;11	.113	83	.011
Turnen	4;0 – 4;11	.147	80	< .001
	5;0 – 5;11	.190	83	< .001
Nachzeichnen	4;0 – 4;11	.188	81	< .001
	5;0 – 5;11	.155	83	< .001
Gegensätze	4;0 – 4;11	.123	81	.004
	5;0 – 5;11	.140	83	< .001
Fotoalbum	4;0 – 4;11	.133	81	.001
	5;0 – 5;11	.177	83	< .001
Elternfragebogen	4;0 – 4;11	.082	71	.200*
	5;0 – 5;11	.152	73	< .001

*, Dies ist eine untere Grenze der echten Signifikanz.

a. Signifikanzkorrektur nach Lilliefors.

Tabelle A-6: Homogenität der Varianzen – WET-Ergebnisse (Rohwerte) bei Subtests mit gleichen Aufgaben für Vierjährige und Fünfjährige pro Jahresgruppe

	Levene Test auf Varianzgleichheit			
	F*	df1	df2	Sig.
Lernbär	12.513	1	162	.001
Bilderlotto: <i>Langform</i>	.330	1	160	.566
Puppenspiel	.424	1	162	.516
Schatzkästchen: <i>Unmittelbares Reproduzieren</i>	2.007	1	162	.158
Schatzkästchen: <i>Behaltensleistung nach 20 Minuten</i>	.008	1	162	.929
Schatzkästchen: <i>Anzahl der Lerndurchgänge</i>	1.597	1	162	.208
Bunte Formen	.553	1	162	.458
Zahlen Merken: <i>Gedächtnisspanne</i>	3.052	1	162	.083
Wörter Erklären	.174	1	162	.677
Turnen	1.230	1	161	.269
Nachzeichnen	1.513	1	162	.220
Gegensätze	6.397	1	162	.012
Fotoalbum	17.338	1	162	< .001
Elternfragebogen	1.381	1	142	.242

* Basierend auf dem Mittelwert.

Tabelle A-7: Normalverteilung der WET-Ergebnisse (Rohwerte) bei Subtests mit gleichen Aufgaben für Vierjährige und Fünfjährige pro Halbjahresgruppe

	Altersgruppe	Shapiro-Wilk		
		W	df	Sig.
Lernbär	4;0 – 4;5	.743	40	< .001
	4;6 – 4;11	.511	41	< .001
	5;0 – 5;5	.580	41	< .001
	5;6 – 5;11	.614	42	< .001
Bilderlotto: <i>Langform</i>	4;0 – 4;5	.955	38	.133
	4;6 – 4;11	.983	41	.801
	5;0 – 5;5	.965	41	.236
	5;6 – 5;11	.930	42	.013
Puppenspiel	4;0 – 4;5	.974	40	.464
	4;6 – 4;11	.965	41	.234
	5;0 – 5;5	.949	41	.065
	5;6 – 5;11	.936	42	.021
Schatzkästchen: <i>Unmittelbares Reproduzieren</i>	4;0 – 4;5	.879	40	< .001
	4;6 – 4;11	.938	41	.026
	5;0 – 5;5	.929	41	.014
	5;6 – 5;11	.916	42	.004
Schatzkästchen: <i>Behaltensleistung nach 20 Minuten</i>	4;0 – 4;5	.790	40	< .001
	4;6 – 4;11	.594	41	< .001
	5;0 – 5;5	.646	41	< .001
	5;6 – 5;11	.603	42	< .001
Schatzkästchen: <i>Anzahl der Lerndurchgänge</i>	4;0 – 4;5	.874	40	< .001
	4;6 – 4;11	.914	41	.005
	5;0 – 5;5	.896	41	.001
	5;6 – 5;11	.751	42	< .001
Bunte Formen	4;0 – 4;5	.913	40	.005
	4;6 – 4;11	.946	41	.051
	5;0 – 5;5	.920	41	.007
	5;6 – 5;11	.844	42	< .001
Zahlen Merken: <i>Gedächtnisspanne</i>	4;0 – 4;5	.853	40	< .001
	4;6 – 4;11	.888	41	.001
	5;0 – 5;5	.806	41	< .001
	5;6 – 5;11	.851	42	< .001
Wörter Erklären	4;0 – 4;5	.984	40	.816
	4;6 – 4;11	.975	41	.493
	5;0 – 5;5	.969	41	.324
	5;6 – 5;11	.942	42	.034
Turnen	4;0 – 4;5	.934	39	.025
	4;6 – 4;11	.921	41	.008
	5;0 – 5;5	.908	41	.003
	5;6 – 5;11	.906	42	.002
Nachzeichnen	4;0 – 4;5	.954	40	.101
	4;6 – 4;11	.886	41	.001
	5;0 – 5;5	.939	41	.029
	5;6 – 5;11	.941	42	.031

Tabelle A-7: Fortsetzung

Gegensätze	4;0 – 4;5	.947	40	.061
	4;6 – 4;11	.929	41	.013
	5;0 – 5;5	.930	41	.014
	5;6 – 5;11	.867	42	< .001
Fotoalbum	4;0 – 4;5	.972	40	.416
	4;6 – 4;11	.939	41	.028
	5;0 – 5;5	.934	41	.020
	5;6 – 5;11	.914	42	.004
Elternfragebogen	4;0 – 4;5	.970	36	.429
	4;6 – 4;11	.987	35	.945
	5;0 – 5;5	.949	36	.099
	5;6 – 5;11	.871	37	.001

Tabelle A-8: Homogenität der Varianzen – WET-Ergebnisse (Rohwerte) bei Subtests mit gleichen Aufgaben für Vierjährige und Fünfjährige pro Halbjahresgruppe

	Levene Test auf Varianzgleichheit			
	F*	df1	df2	Sig.
Lernbär	7.555	3	160	< .001
Bilderlotto: <i>Langform</i>	.443	3	158	.722
Puppenspiel	2.010	3	160	.115
Schatzkästchen: <i>Unmittelbares Reproduzieren</i>	.864	3	160	.461
Schatzkästchen: <i>Behaltensleistung nach 20 Minuten</i>	2.727	3	160	.046
Schatzkästchen: <i>Anzahl der Lerndurchgänge</i>	2.804	3	160	.042
Bunte Formen	.168	3	160	.918
Zahlen Merken: <i>Gedächtnisspanne</i>	2.061	3	160	.108
Wörter Erklären	.396	3	160	.756
Turnen	1.691	3	159	.171
Nachzeichnen	1.225	3	160	.302
Gegensätze	2.930	3	160	.035
Fotoalbum	4.626	3	160	.004
Elternfragebogen	1.691	3	140	.172

* Basierend auf dem Mittelwert.

Tabelle A-9: Normalverteilung der WET-Ergebnisse (Rohwerte) bei Subtests mit altersangepassten Aufgaben für Vierjährige und Fünfjährige pro Halbjahresgruppe

	Altersgruppe	Shapiro-Wilk		
		W	df	Sig.
Quiz	4;0 – 4;5	.945	40	.051
	4;6 – 4;11	.952	41	.079
	5;0 – 5;5	.921	41	.007
	5;6 – 5;11	.925	42	.009
Muster Legen	4;0 – 4;5	.937	40	.027
	4;6 – 4;11	.962	41	.179
	5;0 – 5;5	.913	41	.004
	5;6 – 5;11	.905	42	.002
Rechnen	4;0 – 4;5	.933	40	.020
	4;6 – 4;11	.834	41	< .001
	5;0 – 5;5	.973	41	.419
	5;6 – 5;11	.945	42	.041

Tabelle A-10: Homogenität der Varianzen – WET-Ergebnisse (Rohwerte) bei Subtests mit gleichen Aufgaben für Vierjährige und Fünfjährige pro Halbjahresgruppe

		Levene Test auf Varianzgleichheit			
		F*	df1	df2	Sig.
Quiz	4;0 – 4;11	.021	1	79	.884
	5;0 – 5;11	.107	1	81	.745
Muster Legen	4;0 – 4;11	.156	1	79	.694
	5;0 – 5;11	.180	1	81	.672
Rechnen	4;0 – 4;11	.267	1	79	.607
	5;0 – 5;11	2.705	1	81	.104

* Basierend auf dem Mittelwert.

Anhang B: Ergebnisse der Varianzanalysen und Kruskal-Wallis-Tests

Tabelle B-1: Leistungsunterschiede in den Testleistungen zwischen den vier Altersgruppen: Ergebnis von Varianzanalysen und Kruskal-Wallis-Tests

Subtest	Ergebnis der Varianzanalysen und Kruskal-Wallis-Tests
Lernbär	$\chi^2 = 22.505$; df = 3; p = <.001
Bilderlotto: <i>Langform</i>	F = 11.986; df = 3; p = <.001
Puppenspiel	F = 8.698; df = 3; p = <.001
Schatzkästchen: <i>Unmittelbares Reproduzieren</i>	$\chi^2 = 19.485$; df = 3; p = <.001
Schatzkästchen: <i>Behaltensleistung nach 20 Minuten</i>	$\chi^2 = 7.861$; df = 3; p = .049
Schatzkästchen: <i>Anzahl der Lerndurchgänge</i>	$\chi^2 = 32.329$; df = 3; p = <.001
Bunte Formen	$\chi^2 = 14.472$; df = 3; p = .002
Zahlen Merken: <i>Gedächtnisspanne</i>	$\chi^2 = 11.361$; df = 3; p = .010
Wörter Erklären	F = 6.101; df = 3; p = .001
Turnen	$\chi^2 = 62.527$; df = 3; p = <.001
Nachzeichnen	$\chi^2 = 36.656$; df = 3; p = <.001
Gegensätze	$\chi^2 = 21.086$; df = 3; p = <.001
Fotoalbum	$\chi^2 = 22.012$; df = 3; p = <.001
Elternfragebogen	$\chi^2 = 35.100$; df = 3; p = <.001

Tabelle B-2: Leistungsunterschiede in den Testleistungen zwischen den vier Altersgruppen - deskriptive Ergebnisse der Varianzanalysen

		N	M	SD	SE	95%		Min.	Max.
						Konfidenzintervall			
						Unter- grenze	Ober- grenze		
Bilderlotto: <i>Langform</i>	4;0 – 4;5	38	11.71	4.299	.697	10.30	13.12	3	19
	4;6 – 4;11	41	13.00	4.438	.693	11.60	14.40	4	23
	5;0 – 5;5	41	15.59	4.878	.762	14.05	17.13	5	24
	5;6 – 5;11	42	17.12	4.290	.662	15.78	18.46	4	23
	Total	162	14.42	4.923	.387	13.66	15.18	3	24
Puppenspiel	4;0 – 4;5	40	7.63	2.009	.318	6.98	8.27	3	12
	4;6 – 4;11	41	8.29	2.713	.424	7.44	9.15	3	13
	5;0 – 5;5	41	9.37	2.211	.345	8.67	10.06	3	13
	5;6 – 5;11	42	10.00	2.242	.346	9.30	10.70	4	13
	Total	164	8.84	2.468	.193	8.45	9.22	3	13
Wörter Erklären	4;0 – 4;5	40	11.90	3.550	.561	10.76	13.04	4	20
	4;6 – 4;11	41	12.32	3.134	.489	11.33	13.31	6	18
	5;0 – 5;5	41	13.90	2.948	.460	12.97	14.83	7	19
	5;6 – 5;11	42	14.45	3.078	.475	13.49	15.41	8	20
	Total	164	13.16	3.329	.260	12.65	13.67	4	20

Tabelle B-3: Leistungsunterschiede in den Testleistungen zwischen den vier Altersgruppen – Ergebnis polynomialer Kontraste

			SS	df	MS	F	Sig.
Bilderlotto: <i>Langform</i>	Linearer Term	Ungewichtet	707.832	1	707.832	35.189	< .001
		Gewichtet	711.473	1	711.473	35.370	< .001
		Abweichung	11.812	2	5.906	.294	.746
	Quadratischer Term	Ungewichtet	.603	1	.603	.030	.863
		Gewichtet	.545	1	.545	.027	.869
		Abweichung	11.267	1	11.267	.560	.455
Puppenspiel	Linearer Term	Ungewichtet	137.707	1	137.707	25.819	< .001
		Gewichtet	137.705	1	137.705	25.818	< .001
		Abweichung	1.475	2	.737	.138	.871
	Quadratischer Term	Ungewichtet	.012	1	.012	.002	.963
		Gewichtet	.013	1	.013	.002	.961
		Abweichung	1.462	1	1.462	.274	.601
Wörter Erklären	Linearer Term	Ungewichtet	175.025	1	175.025	17.281	< .001
		Gewichtet	175.264	1	175.264	17.305	< .001
		Abweichung	10.122	2	5.061	.500	.608
	Quadratischer Term	Ungewichtet	.181	1	.181	.018	.894
		Gewichtet	.167	1	.167	.016	.898
		Abweichung	9.955	1	9.955	.983	.323

Anhang C: Graphische Darstellung der Leistungsunterschiede der vier Altersgruppen

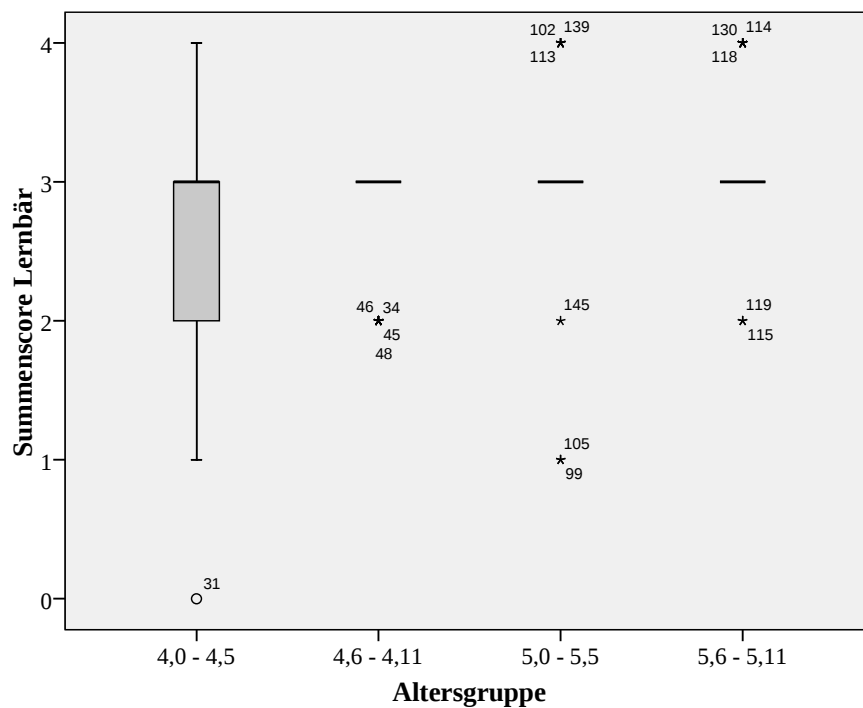


Abbildung C-1: Subtest *Lernbär* - Anstieg der Rohwerte über die Altersgruppen

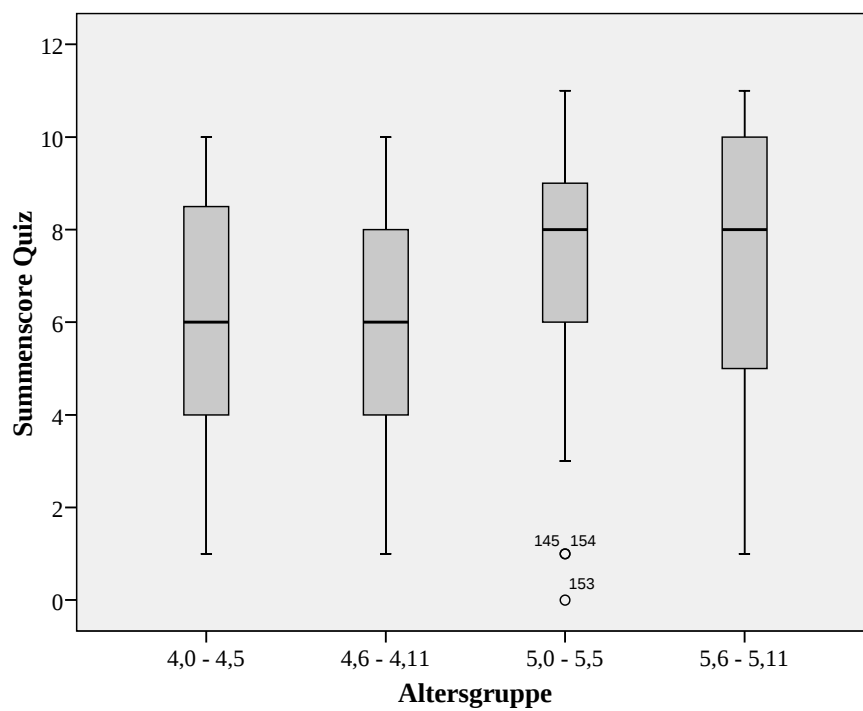


Abbildung C-2: Subtest *Quiz* - Anstieg der Rohwerte über die Altersgruppen

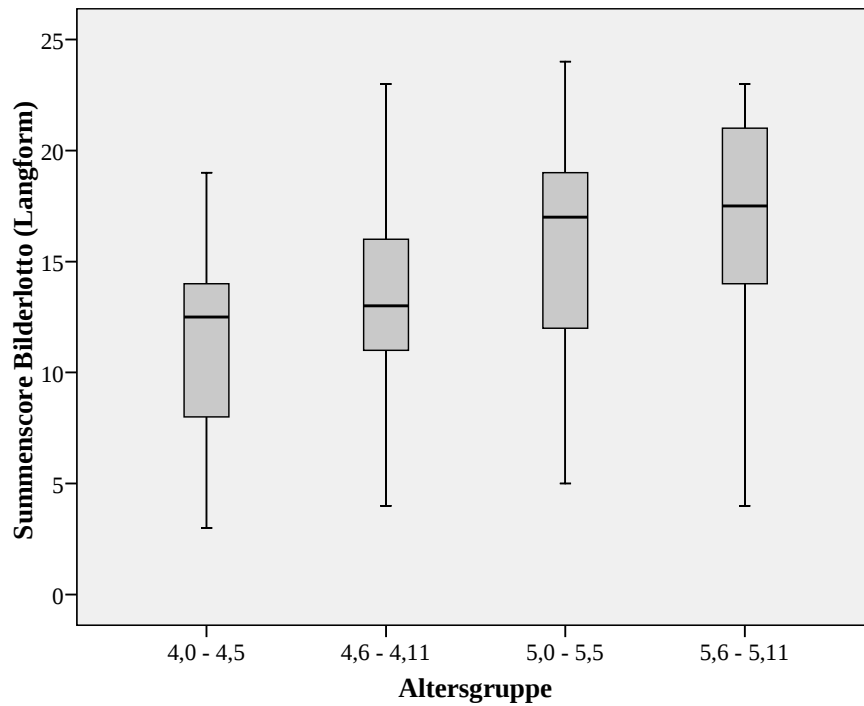


Abbildung C-3: Subtest *Bilderlotto* - Anstieg der Rohwerte über die Altersgruppen

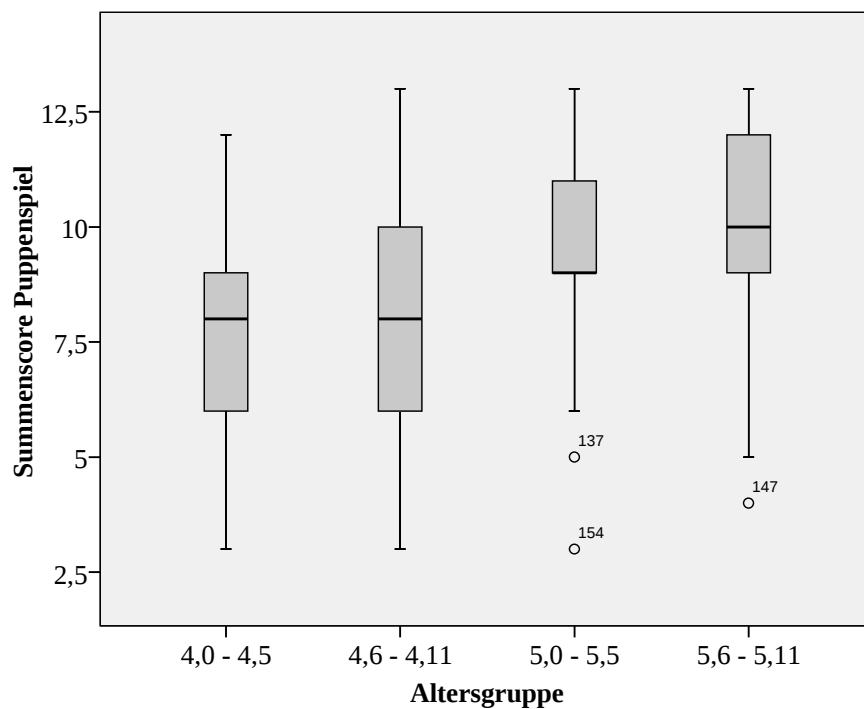


Abbildung C-4: Subtest *Puppenspiel* - Anstieg der Rohwerte über die Altersgruppen

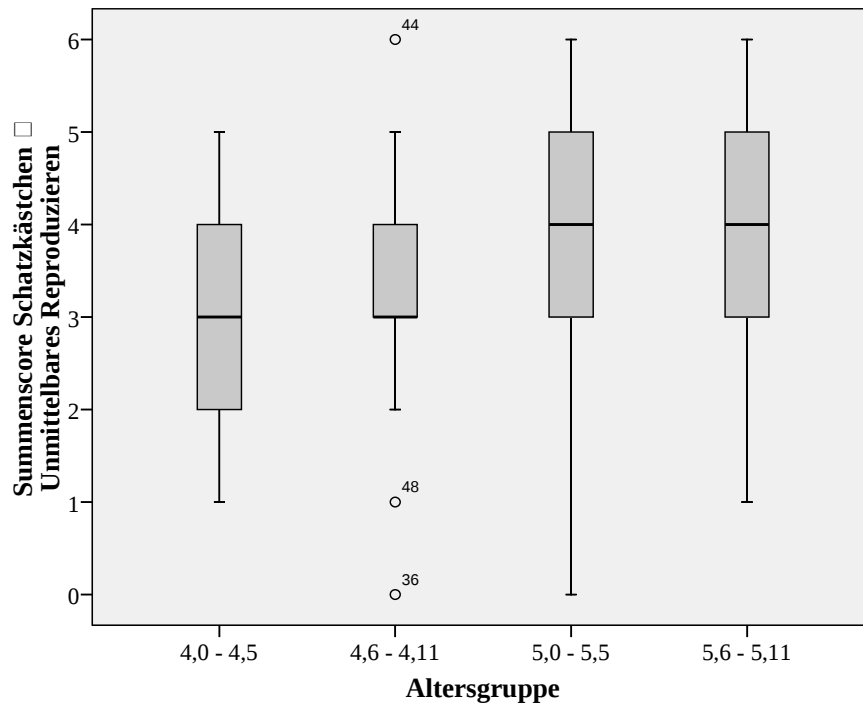


Abbildung C-5: Subtest *Schatzkästchen: Unmittelbares Reproduzieren* - Anstieg der Rohwerte über die Altersgruppen

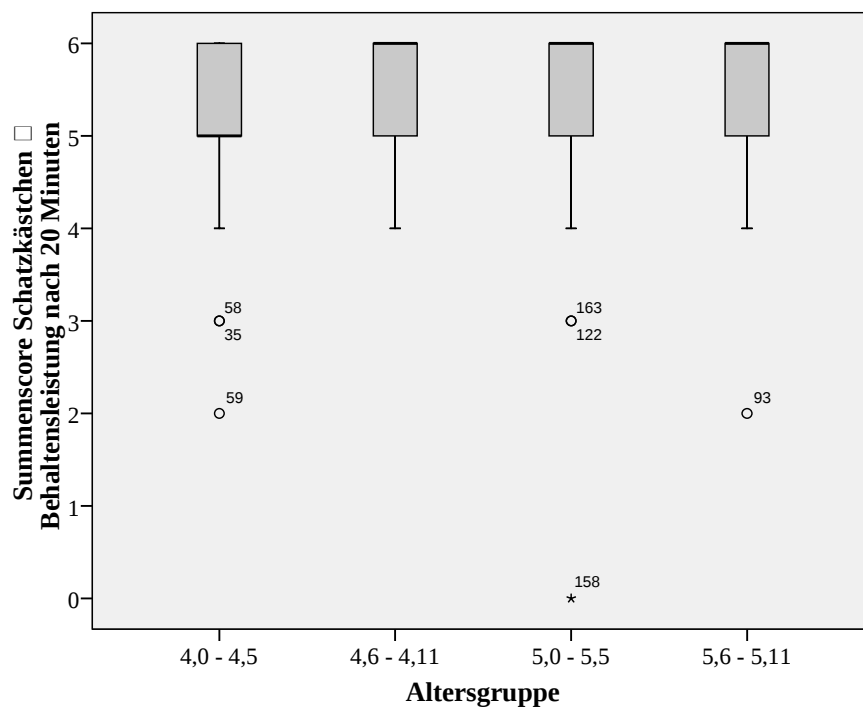


Abbildung C-6: Subtest *Schatzkästchen: Behaltensleistung nach 20 Minuten* - Anstieg der Rohwerte über die Altersgruppen

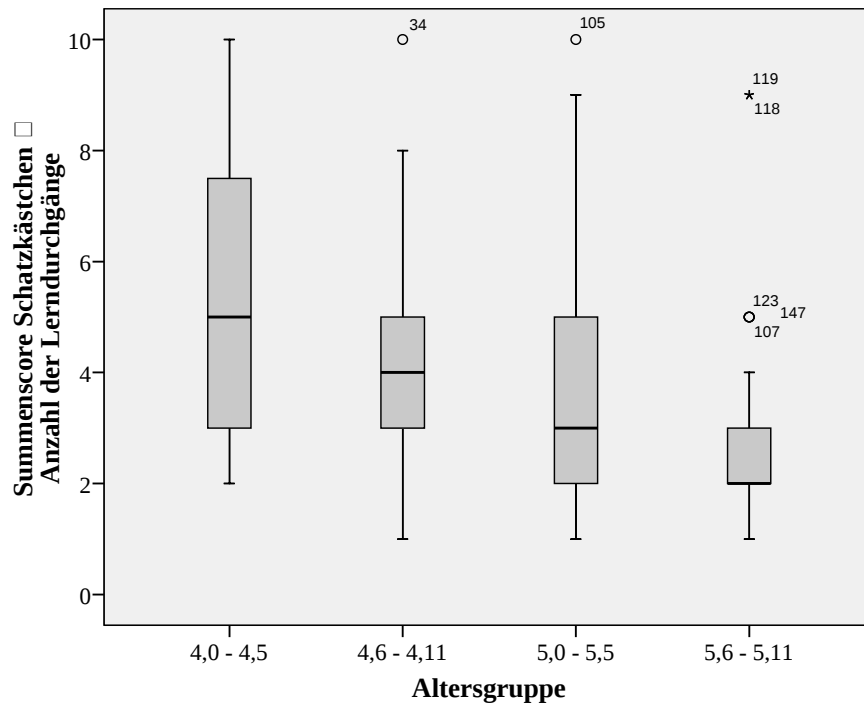


Abbildung C-7: Subtest *Schatzkästchen*: Anzahl der Lerndurchgänge - Anstieg der Rohwerte über die Altersgruppen

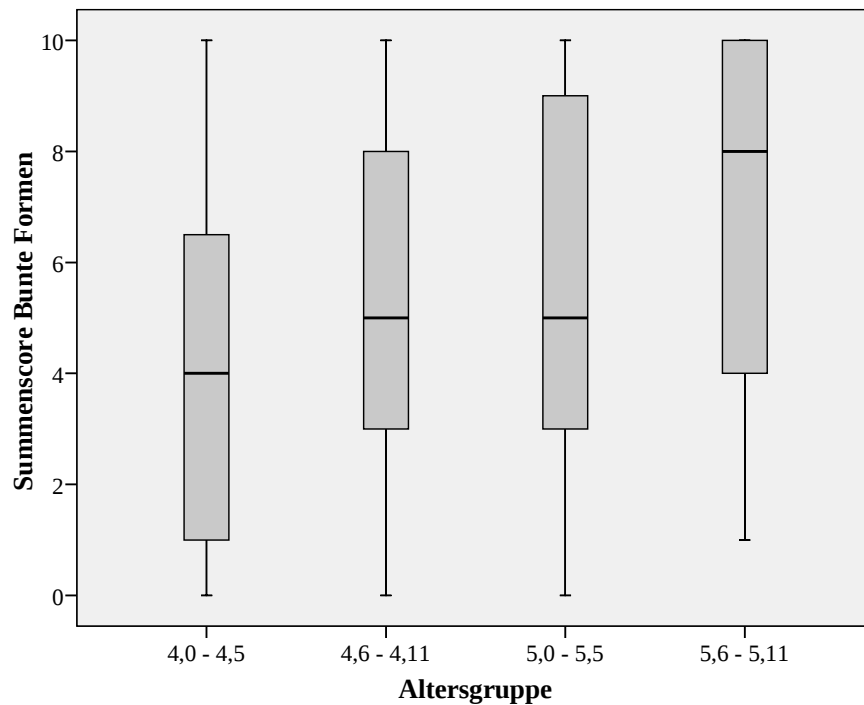


Abbildung C-8: Subtest *Bunte Formen* - Anstieg der Rohwerte über die Altersgruppen

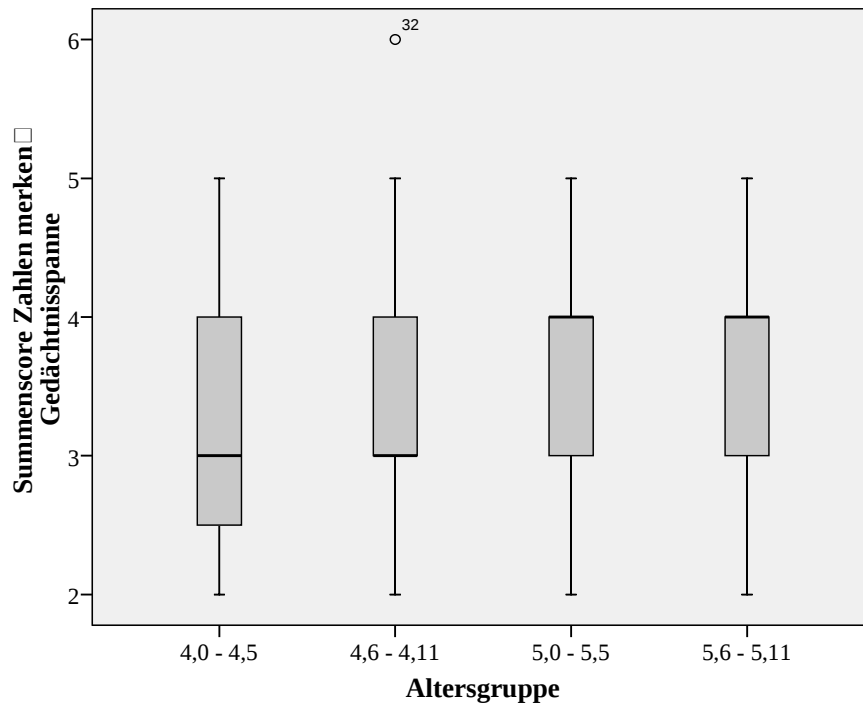


Abbildung C-9: Subtest Zahlen Merken: Gedächtnisspanne - Anstieg der Rohwerte über die Altersgruppen

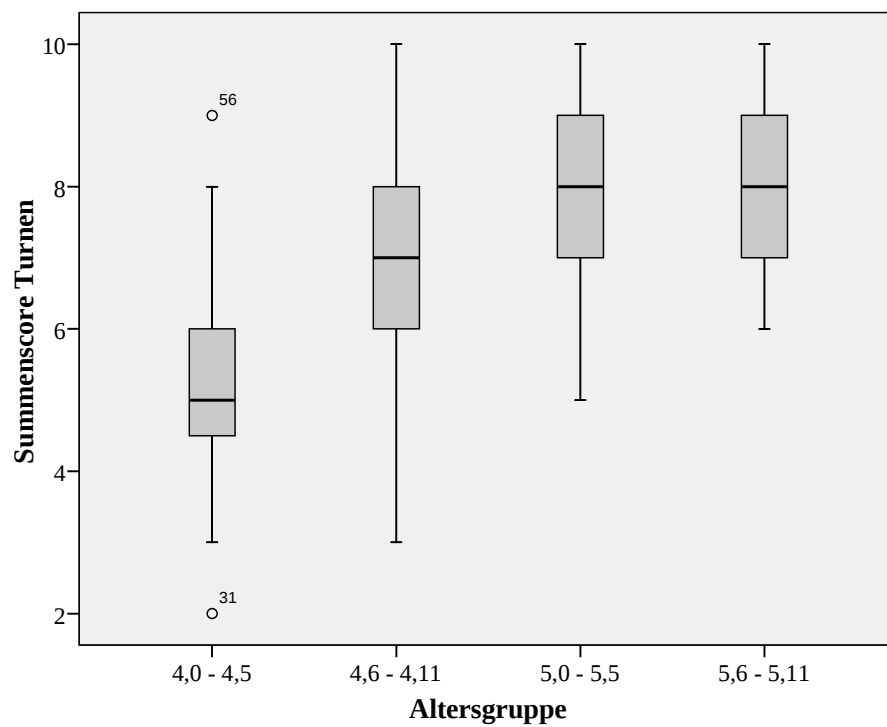


Abbildung C-10: Subtest Turnen - Anstieg der Rohwerte über die Altersgruppen

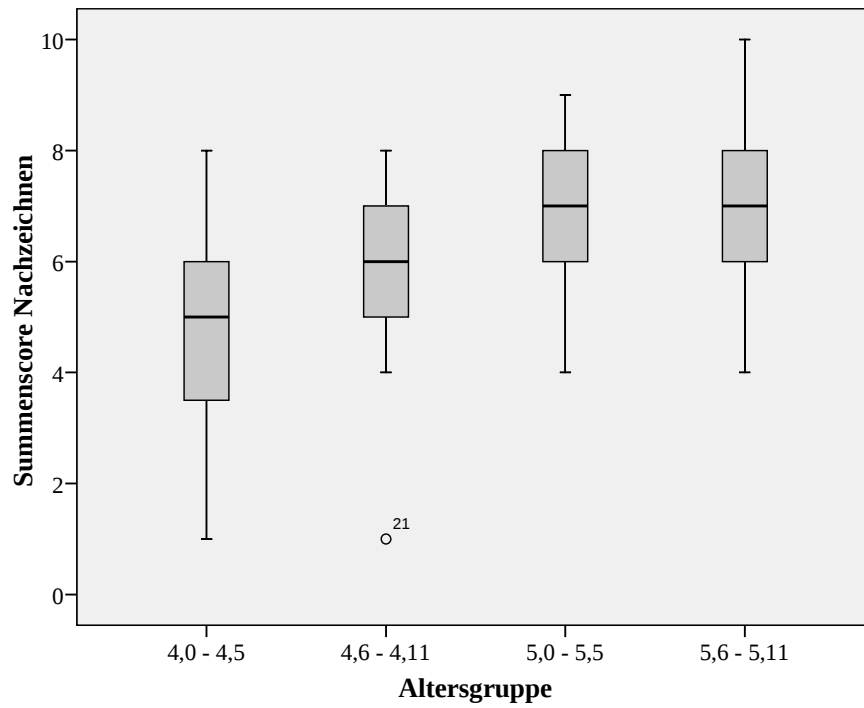


Abbildung C-11: Subtest *Nachzeichnen* - Anstieg der Rohwerte über die Altersgruppen

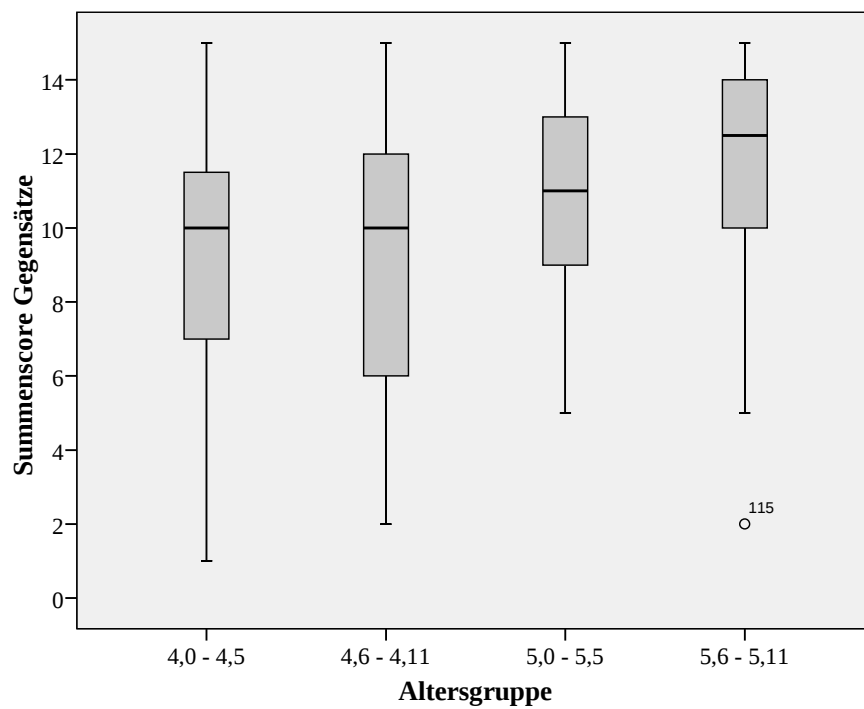


Abbildung C-12: Subtest *Gegensätze* - Anstieg der Rohwerte über die Altersgruppen

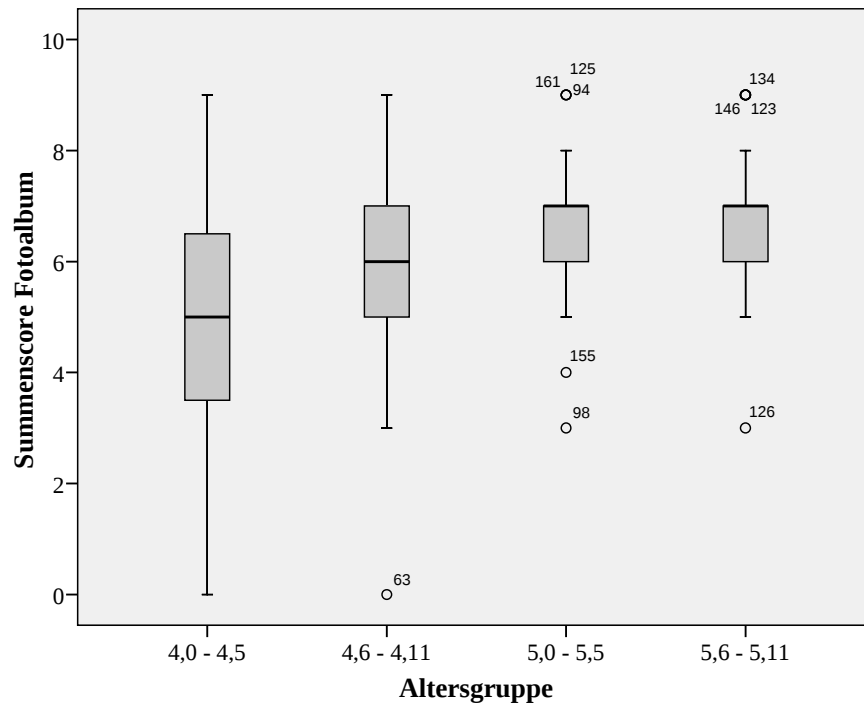


Abbildung C-13: Subtest *Fotoalbum* - Anstieg der Rohwerte über die Altersgruppen

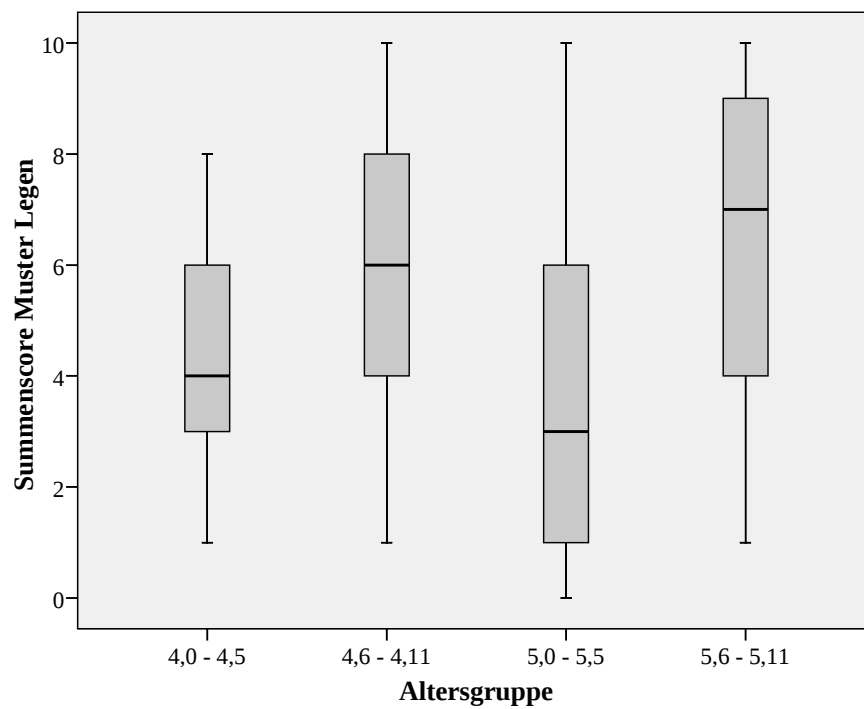


Abbildung C-14: Subtest *Muster Legen* - Anstieg der Rohwerte über die Altersgruppen

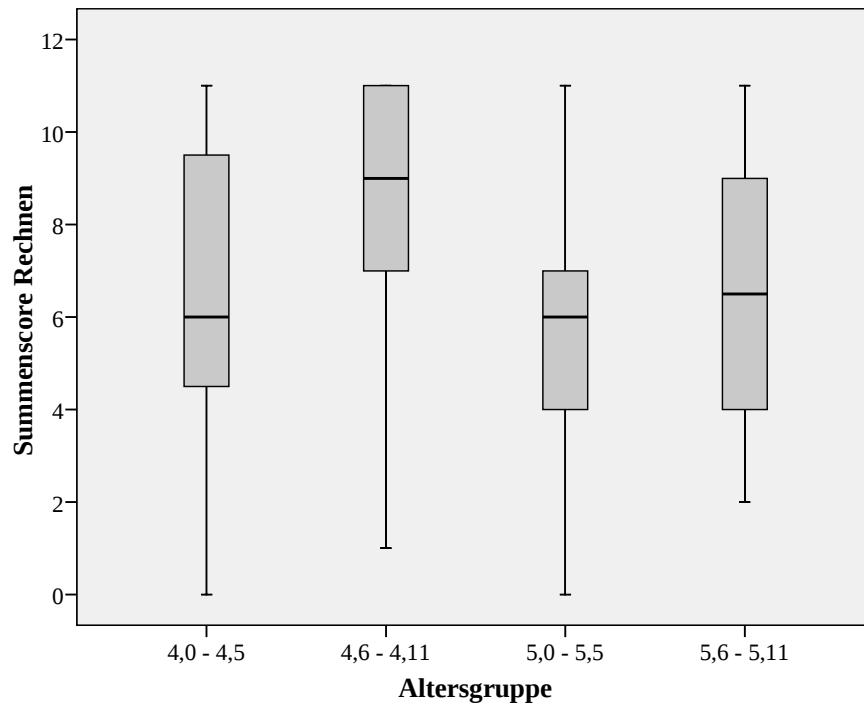


Abbildung C-15: Subtest *Rechnen* - Anstieg der Rohwerte über die Altersgruppen

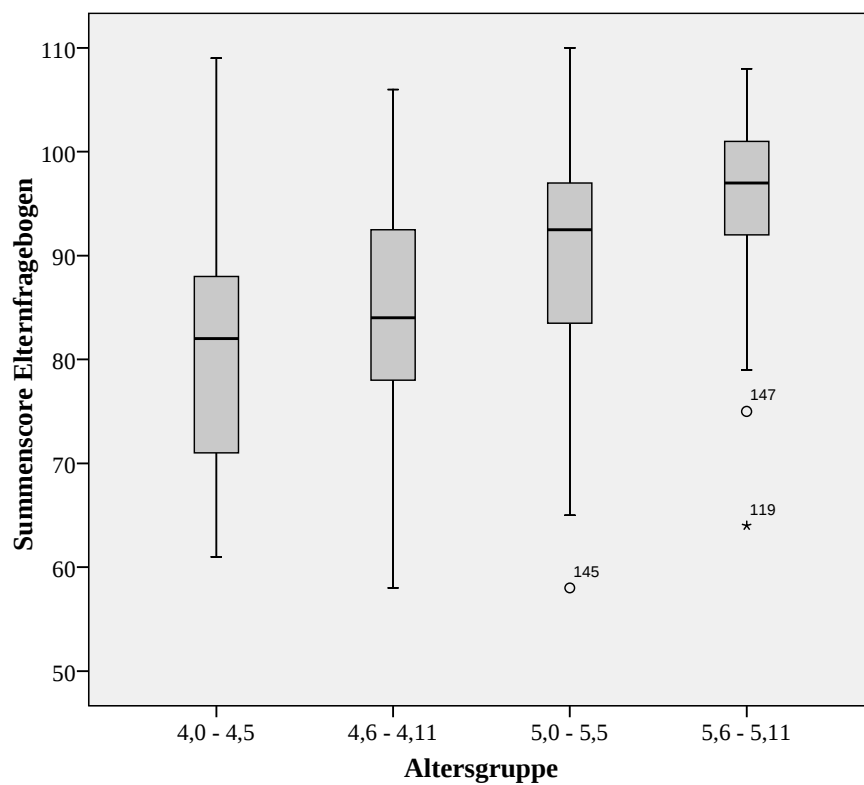


Abbildung C-16: Subtest Elternfragebogen - Anstieg der Rohwerte über die Altersgruppen

Anhang D: Elternbrief

UNIVERSITÄT  WIEN

Fakultät für Psychologie

**Institut für Entwicklungspsychologie
und Psychologische Diagnostik**

Ass. Prof. Dr. Ursula Kastner-Koller Tel +43-1-4277-47861 E-Mail ursula.kastner-koller@univie.ac.at
Ass. Prof. Dr. Pia Deimann Tel +43-1-4277-47867 E-Mail pia.deimann@univie.ac.at

Liebe Eltern!

Um Kindern mit Problemen helfen zu können, muss man das übliche Verhalten normal entwickelter Kinder kennen. Zu diesem Zweck werden in österreichischen Kindergärten Erhebungen über den allgemeinen Entwicklungsstand und die Entwicklung der Lateralität (Händigkeit) von Kindergartenkindern im Alter von 4-6 Jahren durchgeführt.

Dazu möchten wir Kindern den Wiener Entwicklungstest und einen, ebenfalls an der Universität Wien entwickelten, Händigkeitstest vorgeben. Beim Wiener Entwicklungstest handelt es sich um ein Verfahren zur Erfassung des allgemeinen Entwicklungsstandes für Kindergartenkinder, der folgende Entwicklungsbereiche abdeckt: Motorik, Visumotorik/Visuelle Wahrnehmung, Gedächtnis, kognitive Entwicklung, Sprache und emotionale Entwicklung.

Die Erhebung wird im Kindergarten stattfinden und ist selbstverständlich anonym. Alle Aufgaben sind spielerisch gestaltet und machen den Kindern erfahrungsgemäß viel Spaß.

Für Sie bietet die Untersuchung die Möglichkeit, Informationen über den Entwicklungsstand Ihres Kindes zu erhalten. Wir stehen Ihnen diesbezüglich gerne für Rückmeldungen und Gespräche zur Verfügung.

Wir würden uns sehr freuen, wenn Ihr Kind an unserer Untersuchung teilnehmen darf!

Mit freundlichen Grüßen

Petra Bircsak

Ass. Prof. Dr. Ursula Kastner-Koller

Ass. Prof. Dr. Pia Deimann

EINVERSTÄNDNISERKLÄRUNG

Ich bin einverstanden, dass mein Kind.....

geboren am....., an der Untersuchung teilnimmt.

Mein Kind ist ☐ weiblich ☐ männlich (bitte ankreuzen)

Ich denke mein Kind ist ☐ LinkshänderIn ☐ RechtshänderIn

☐ Ich möchte Informationen über den Entwicklungsstand meines Kindes.

Name:.....

Adresse:.....

Telefon/Email:.....

.....
Unterschrift des Erziehungsberechtigten

Danksagung

Diese Arbeit entstand im Zuge der Vorarbeiten für die geplante dritte Auflage des Wiener Entwicklungstests (Kastner-Koller & Deimann, 1998, 2002, in Vorbereitung). Mein Dank gebührt an erster Stelle somit Frau Dr. Kastner-Koller und Frau Dr. Deimann, die mir die Teilnahme an den Vorarbeiten dieses Projekts und damit diese Diplomarbeit erst ermöglicht haben. Sie haben überdies mein Interesse für die Entwicklungsdiagnostik entfacht und mich auch zur zukünftigen Arbeit mit Kindern inspiriert.

Weiters möchte ich mich bei all jenen bedanken, die mich ebenso unterstützt und entscheidend zum Gelingen dieser Arbeit beigetragen haben:

- der Magistratsabteilung 10, Dezernat II unter der Leitung von Frau Mag.^a Minich,
- dem Personal der an der Untersuchung beteiligten Kindergärten,
- allen Kindern, die an der Untersuchung teilgenommen haben sowie auch deren Eltern, die das ermöglicht haben,
- meinen Kolleginnen Katerina Lebo, Alexandra Propst und Katherina Rab,
- meinem lieben Freund Christian (Landschützer) für die kritische Durchsicht des Textes sowie vor allem
- Niki und meiner Familie.

Curriculum Vitae

Petra Bircsak
geb. am 12.11.1972 in Innsbruck

Ausbildung

- 2006 - Studium der Psychologie, Universität Wien**
Schwerpunkt: Angewandte Kinder- und Jugendpsychologie
- 2003 - 2005 Executive Management MBA Program, SMA, Wien**
Titel der Master Thesis: *Die betriebliche Leistungsbeurteilung: Grundlagen, Verfahren, Probleme und Tendenzen*
Abschluss: magna cum laude
- 1993 - 1999 Studium der Psychologie, Universität Innsbruck**
Schwerpunkt: Arbeits- und Organisationspsychologie
- 1992 - 1994 Studium der Medizin, Universität Innsbruck**
- 1986 - 1990 Katholisches Oberstufenrealgymnasium, Innsbruck**
Matura im Juni 1990

Berufliche Tätigkeit

- 11/02 – 10/08 Selbständige Tätigkeit / Unternehmensberatung**
Interimsmanagement; externes HR-Management; Reorganisation von Personalabteilungen; Beratung bei der Einführung von Personalbeurteilung sowie Konzeption entsprechender Systeme
- 09/01 – 09/02 SBB Software GmbH, Wolfsgraben**
Manager Human Resources
Verantwortlich für die HR-Agenden des Unternehmens in Österreich, Deutschland, Schweiz, UK und USA; Entwicklung eines Personalbeurteilungssystems u. eines Provisionsmodells für die Vertriebsmitarbeiter
- 07/00 – 08/01 gameplay GmbH, Berlin**
Personalleiterin
Betreuung der fünf deutschsprachigen Standorte (D & A) der gameplay plc (UK) in HR-Belangen; Umsetzung von Konzernrichtlinien der gameplay plc; Mitarbeit an der Verschmelzung von 3 Unternehmen auf die gameplay GmbH; Reorganisation und Personalabbau

- 05/99 – 07/00** **I-D Media AG, Aalen – Berlin - London**
Head of Human Resources
 Aufbau u. Leitung des HR-Teams; Implementierung diverser HR Systeme; Recruiting und Personalmarketing
- 04/98 – 04/99** **Hill International (Personalberatung), Innsbruck**
Freie Mitarbeiterin
 Telefonische Kundenakquise; Vorgabe und Auswertung der computergestützten Kompetenzanalyse; Erstellung von Gutachten unter Supervision

Ehrenamtliche Tätigkeiten, studentische Tätigkeiten und Praktika

- 03/09 – 11/09** **Pädiatrische Neuro-Onkologie - AKH, Wien**
 Clinic Nanny
- 10/07 – 01/08** **Fakultät für Psychologie, Universität Wien**
 student mentor
- 10/06 – 06/07** **WoMentoring - Club alpha, Wien**
 Mentorin
- 02/99 – 04/99** **Institut für Medizinische Psychologie, Universität Innsbruck**
 Mitarbeit an einer Studie zur Suchprävention an Schulen
- 08/98 – 11/98** **Mummert + Partner (Unternehmensberatung), Hamburg**
Praktikant
 Mitarbeit an der Entwicklung eines Personalbeurteilungssystems sowie eines Gesprächstrainings für Führungskräfte
- 06/98 – 07/98** **Institut für Sozialmedizin, Universität Innsbruck**
 Mitarbeit an der Umweltverträglichkeitsprüfung Eisenbahnachse Brenner, Zulaufstrecke Nord, Unteres Inntal
- 11/97 – 01/98** **Hill International (Personalberatung), Innsbruck**
Praktikant
 Vorgabe der computergestützten Kompetenzanalyse, Verwaltung der Bewerberdaten, Erstellung von Gutachten unter Supervision