



universität
wien

Diplomarbeit

Titel der Diplomarbeit

Die Überarbeitung und testtheoretische Überprüfung
eines spielbasierten Itempools zur Erfassung der
Entwicklung im dritten Lebensjahr
unter besonderer Berücksichtigung der kognitiven Entwicklung

Verfasserin

Sigrid Anna Punz

Angestrebter akademischer Grad

Magistra der Naturwissenschaften (Mag. rer. nat.)

Wien, 2013

Studienkennzahl: 298

Studienrichtung: Psychologie

Betreuerin: Ass.-Prof. Dr. Pia Deimann

Danksagung

Während des Verfassens meiner Diplomarbeit standen mir viele Personen unterstützend zur Seite, denen ich die folgenden Zeilen widme, um meinen Dank auszusprechen.

Zuallererst bedanke ich mich bei Frau *Ass.-Prof. Dr. Deimann* und Frau *Ass.-Prof. Dr. Kastner-Koller* für die außerordentlich hilfsbereite und konstruktive Betreuung meiner Diplomarbeit.

Mein aufrichtiger Dank gilt den Eltern und Kindern, die an der Studie teilgenommen haben – durch ihre Bereitschaft konnte meine Diplomarbeit entstehen.

Meiner Kollegin und Freundin, *Viktoria Kronberger* danke ich herzlich für die ausgesprochen gute Zusammenarbeit, den konstruktiven Austausch und die Möglichkeit freudige Momente zu teilen.

Ein großer Dank richtet sich an meine Eltern, meinen Bruder und meine Freundinnen *Carmen, Hanna, Hannah, Lisa, Martina, Nana* und *Olivia* für den motivierenden Zuspruch sowie die wertvollen Anregungen. Im Besonderen bedanke ich mich bei *Carmen* für die äußerst geduldige und gewissenhafte Korrektur meiner Arbeit.

Danke

Inhaltsverzeichnis

1. Einleitung	1
I. Theoretischer Teil.....	3
2. Grundlagen der Entwicklungsdiagnostik	5
2.1. Entwicklungsbegriff	5
2.2. Definition Entwicklungsdiagnostik.....	7
2.3. Definition Entwicklungstest.....	8
2.4. Klassifikation von Entwicklungstests	8
2.5. Begriffsabgrenzung zu Intelligenzdiagnostik.....	9
3. Besonderheiten der Diagnostik im Kleinkindalter.....	11
3.1. Das Kind als Klient	11
3.2. Anforderungen an die untersuchende Person.....	12
3.3. Anwesenheit der Bezugsperson	13
3.4. Zeitliche Rahmenbedingungen.....	14
3.5. Räumliche Rahmenbedingungen	14
3.6. Untersuchungsablauf.....	15
4. Kritik an allgemeinen Entwicklungstests.....	17
4.1. Gütekriterien	17
4.1.1. Objektivität.....	17
4.1.2. Reliabilität	18
4.1.3. Validität.....	20
4.1.4. Normierung	22
4.1.5. Ökonomie	22
4.1.6. Nützlichkeit	22
4.1.7. Zumutbarkeit	22
4.2. Kritik an der aktuellen allgemeinen Entwicklungsdiagnostik.....	23
4.3. Darstellung und Kritik gängiger allgemeiner Entwicklungstests.....	26
4.3.1. Die Münchener funktionelle Entwicklungsdiagnostik zweites und drittes Lebensjahr (MFED 2-3)	26
4.3.2. Bayley Scales of Infant Development - Second Edition – deutsche Fassung (Bayley II oder BSID-II)	28
4.3.3. Bayley Scales of Infant and Toddler Development – Third Edition	30
4.3.4. Entwicklungstest sechs Monate bis sechs Jahre (ET 6-6).....	31
4.4. Zusammenfassende Bewertung: Relevanz der Entwicklung eines neuen Verfahrens	33

5. Spielbasierte Verfahren.....	34
5.1. Definition spielbasierter Verfahren	34
5.2. Erste spielbasierte Verfahren.....	35
5.3. Transdisciplinary Play-Based Assessment	37
5.4. Anwendungsmöglichkeiten spielbasierter Verfahren.....	38
5.5. Gegenüberstellung spielbasierter Ansätze und allgemeiner Entwicklungstests	39
5.6. Zusammenfassende Bewertung der Bedeutung spielbasierter Verfahren	42
6. Kognitive Entwicklung im dritten Lebensjahr	43
6.1. Definition Kognition	43
6.2. Studienergebnisse zur kognitiven Entwicklung	44
6.2.1. „Theory of Mind“	44
6.2.2. Numerische Fähigkeiten	48
6.2.3. Schlussfolgerndes Denken.....	52
6.3. Relevanz der Erhebung kognitiver Entwicklung.....	56
6.4. Diagnostik kognitiver Entwicklung.....	57
6.4.1. Diagnostik kognitiver Fähigkeiten in allgemeinen Entwicklungstests.....	57
6.4.2. Diagnostik kognitiver Fähigkeiten in Intelligenztests	59
6.5. Geschlechtsspezifische Unterschiede der kognitiven Entwicklung.....	60
6.6. Zusammenfassung kognitive Entwicklung.....	61
II. Empirischer Teil.....	63
7. Zielsetzung und Fragestellung.....	65
8. Methode	66
8.1. Stichprobe.....	66
8.2. Untersuchungsinstrument.....	69
8.2.1. Erhebung der sprachlichen Entwicklung	70
8.2.2. Erhebung der kognitiven Entwicklung	71
8.2.3. Erhebung der Entwicklung des Gedächtnisses	72
8.2.4. Elternfragebogen	73
8.2.5. Änderungen	73
8.3. Die Untersuchung.....	75
8.3.1. Räumliche und zeitliche Rahmenbedingungen	75
8.3.2. Anwesende Personen.....	75
8.3.3. Ablauf der Untersuchung.....	76

9. Auswertung	77
9.1. Auswertung: Objektivität.....	77
9.2. Auswertung: Itemanalyse und Reliabilität.....	78
9.3. Validität.....	80
9.3.1. Auswertung: Alterstrends und Geschlechtsunterschiede.....	80
9.3.2. Auswertung: Zusammenhang zwischen Elternfragebogen und beobachtetem Verhalten.....	81
10. Ergebnisse	82
10.1. Zumutbarkeit.....	82
10.2. Objektivität.....	83
10.3. Itemanalyse und Reliabilität.....	84
10.3.1. Sprachliche Entwicklung: Itemanalysen und Reliabilität.....	85
10.3.2. Kognitive Entwicklung: Itemanalysen und Reliabilität.....	88
10.3.3. Gedächtnis: Itemanalysen und Reliabilität.....	91
10.4. Inhaltsvalidität.....	92
10.5. Konstruktvalidität.....	93
10.5.1. Sprachliche Entwicklung: Alterstrends und Geschlechtsunterschiede.....	93
10.5.2. Kognitive Entwicklung.....	96
10.5.3. Gedächtnis: Alterstrends und Geschlechtsunterschiede.....	99
10.6. Kriteriumsvalidität.....	100
11. Diskussion	101
11.1. Sprachliche Entwicklung.....	102
11.2. Kognitive Entwicklung.....	104
11.2.1. „Theory of Mind“.....	105
11.2.2. Numerisches Wissen.....	107
11.2.3. Deduktives Denken.....	110
11.3. Entwicklung des Gedächtnisses.....	111
11.4. Zusammenfassende Bewertung.....	112
12. Abkürzungsverzeichnis	114
13. Literaturangabe	115
14. Abbildungsverzeichnis	122
15. Tabellenverzeichnis	124
16. Anhang	126

1. Einleitung

Im *theoretischen Teil* der vorliegenden Arbeit soll nach der Definition grundlegender Begriffe aufgezeigt werden, dass für den Altersbereich der zweijährigen Kinder sehr wenige allgemeine Entwicklungstests existieren. Die vorhandenen Verfahren weisen zudem oft wesentliche Kritikpunkte wie eine unzureichende Erfüllung der Gütekriterien, mangelhafte theoretische Fundierung bzw. Probleme in der praktischen Anwendung auf. Es werden ebenfalls spielbasierte Verfahren beschrieben, die einen kindgerechten Ansatz mit dem Versuch die Probleme der klassischen Entwicklungstests zu überwinden, darstellen. Sie versuchen Verhaltensweisen der Kinder in einer möglichst natürlichen Spielsituation zu beobachten. An diesen Verfahren ist jedoch die mangelnde Überprüfung der Gütekriterien zu kritisieren.

Diese Informationen belegen die Notwendigkeit der Entwicklung eines neuen Verfahrens. Im Rahmen eines Projekts von Ass.-Prof. Dr. Pia Deimann und Ass.-Prof. Dr. Ursula Kastner-Koller wurde daher 2011 an der psychologischen Fakultät ein spielbasierter Itempool entworfen, der die Entwicklung von Kindern im dritten Lebensjahr erheben soll. Es sollen die Stärken der klassischen Entwicklungstests und der spielbasierten Ansätze kombiniert werden. Der Itempool soll demnach einerseits den Gütekriterien entsprechen und andererseits eine möglichst spielerische und kindgerechte Testsituation bieten.

Der Itempool wurde im Rahmen dieser Arbeit zum wiederholten Mal überarbeitet und mithilfe der Ergebnisse einer Erhebung an 20 Kindern testtheoretisch überprüft. Die Änderungen und Ergebnisse sollen im *empirischen Teil* dieser Arbeit beschrieben und diskutiert werden, um eine Basis für die erneute Weiterentwicklung des Verfahrens zu bieten. Als langfristiges Ziel gilt es, der Praxis eine Testbatterie zur Verfügung zu stellen, die den Gütekriterien entspricht, eine repräsentative Normierung vorweist und eine spielerische Erhebung der Entwicklung von zweijährigen Kindern ermöglicht.

In dieser Arbeit wird die „kognitive Entwicklung“ besonders berücksichtigt. Im empirischen Teil der Arbeit werden auch die Ergebnisse der Skalen „Sprachliche Entwicklung“ und „Entwicklung des Gedächtnisses“ angeführt und diskutiert. Für die Beschreibung der übrigen Bereiche mit dem Schwerpunkt „visuelle Wahrnehmung“ muss auf die Arbeit von Kronberger (in Vorbereitung) verwiesen werden.

I. Theoretischer Teil

2. Grundlagen der Entwicklungsdiagnostik

In diesem Kapitel werden grundlegende Begriffe dieser Arbeit definiert. Zuerst wird ein Überblick über traditionelle und moderne Entwicklungskonzepte gegeben, gefolgt von einer Definition von Entwicklungsdiagnostik. Des Weiteren sollen die Definition und Klassifikation von Entwicklungstests dargestellt werden. Zuletzt erfolgt eine Abgrenzung von Entwicklungs- und Intelligenzdiagnostik.

2.1. Entwicklungsbegriff

Seit Anbeginn der empirischen Entwicklungspsychologie bestehen unterschiedliche Konzeptionen von Entwicklung nebeneinander und die Definition von Entwicklung unterliegt einem historischen Wandel. Bis heute findet sich in der Literatur kein einheitlicher Entwicklungsbegriff (Esser & Petermann, 2010; Montada, 2008). Allen Konzeptionen ist jedoch die Beschäftigung mit Veränderungen im Lebenslauf gemein (Montada, 2008). Im Folgenden sollen nun verschiedene Konzeptionen von Entwicklung beschrieben werden.

Der *traditionelle bzw. enge Entwicklungsbegriff* basiert auf Reifungsprozessen und Stufen- bzw. Phasenmodellen (Montada, 2008; Petermann & Macha, 2003; Petermann, Stein & Macha, 2008). In traditionellen Phasenmodellen wird angenommen, dass Menschen alterstypische Phasen durchlaufen. In Konzepten zu Entwicklungsstufen wird ebenfalls von solchen Phasen ausgegangen und zusätzlich wird eine bestimmte Abfolge und ein Endstadium angenommen (Montada, 2008).

Ettrich (2000) konnte fünf Kriterien für die Entwicklung von Kindern und Jugendlichen aus verschiedenen traditionellen Stufenmodellen subsumieren: Die Veränderungen vollziehen sich demnach in unumkehrbaren Stufen (*Sequentialität und Irreversibilität*) und sind unidirektional nach einem Endziel ausgerichtet (*Unidirektionalität*). Der Verlauf ist für alle Kinder und Jugendlichen ähnlich, aber die Geschwindigkeit variiert (*Universalität*). Das Kriterium des *Strukturalismus* besagt, dass die Veränderungen qualitativ sind, d.h. jede neue Entwicklungsstufe geht mit neuen Systemeigenschaften einher.

Dieser traditionelle Entwicklungsbegriff stellt die Entwicklung demnach stark vereinfacht dar und reduziert Entwicklung auf biologische Aspekte (Ettrich, 2000).

Montada (2008) kritisiert am engen Entwicklungsbegriff, dass viele Veränderungen nicht schrittweise durchlaufen werden. Ein höheres Niveau eines Merkmals zu besitzen, ist nicht immer als positiv zu erachten (z.B. bei Persönlichkeitsmerkmalen) und ein Endzustand der Entwicklung ist meist nicht zu definieren, da von einer lebenslangen Entwicklung in Abhängigkeit von unterschiedlichsten Umweltfaktoren auszugehen ist.

Viele empirische Studien zeigen auf, dass aufgrund der Variabilität von Entwicklungsverläufen ein *erweitertes Entwicklungskonzept* angemessener ist (Esser & Petermann, 2010; Petermann & Macha, 2003; Petermann, Stein & Macha, 2008). Diese modernen Konzepte gehen von einer lebenslangen Entwicklung aus und berücksichtigen, dass neben der genetischen Komponente viele verschiedene externale und internale Faktoren diese Entwicklung beeinflussen. Sie bieten somit Erklärungen für Entwicklungsunterschiede. Entwicklungspfade werden dieser Ansicht nach durch biopsychosoziale Wechselwirkungen bestimmt, wobei die Umwelt den Menschen beeinflusst und der Mensch die Umwelt. Wichtige Aspekte sind hierbei Reifung, Lernen und Selbststeuerung (Ettrich, 2000; Montada, 2008; Petermann & Macha, 2003). Entwicklung wird als ungerichtet und multidimensional verstanden (Petermann, Stein & Macha, 2008). Die einzelnen Entwicklungsbereiche verändern sich unabhängig voneinander (Reuner & Pietz, 2006). Jede Art der Veränderung wird als Entwicklung verstanden. Entwicklung kann demnach nicht nur Gewinn, sondern auch Verlust bedeuten (Montada, 2008).

Die beiden Entwicklungsbegriffe bestehen nach wie vor nebeneinander und je nach Entwicklungsdimension und Abschnitt ist der Erklärungswert der beiden Begriffe unterschiedlich zu diskutieren (Esser & Petermann, 2010).

2.2. Definition Entwicklungsdiagnostik

Als Definition von Entwicklungsdiagnostik soll jene von Deimann und Kastner-Koller (2007) angegeben und erläutert werden:

„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“ (S. 558).

Dieser Definition liegt ein moderner Entwicklungsansatz zugrunde. Moderne Ansätze beachten den Einfluss der Umwelt und weisen demnach auf die Bedeutung und Möglichkeiten von Entwicklungsdiagnostik hin (Esser & Petermann, 2010). Auch Hagmann-von Arx, Meyer und Grob (2008) betonen in ihrer Definition die Möglichkeiten der Diagnostik. Sie sehen Entwicklungsdiagnostik „als ganzheitliche dem Individuum angepasste Diagnostik, die Anfänge von Störungsbildern erkennt und rechtzeitig Fördermaßnahmen respektive wenn notwendig Therapiemaßnahmen einleitet“ (S. 240).

Das oben stehende Zitat definiert Entwicklungsdiagnostik im Allgemeinen und nicht nur Entwicklungstests. Viele Autoren/Autorinnen setzen jedoch den Begriff Entwicklungsdiagnostik mit dem Begriff Entwicklungstest gleich (Reuner & Pietz, 2006). Rümmele (2011) meint, dass Entwicklungsdiagnostik das Ziel verfolgt, „lebenslaufbezogene Veränderungen psychischer Phänomene anhand unterschiedlicher Entwicklungstests zu identifizieren und beschreiben“ (S. 929). Dem modernen Entwicklungsbegriff folgend inkludiert Entwicklungsdiagnostik jedoch neben Entwicklungstests auch ein Anamnese- bzw. Explorationsgespräch, um Informationen über die Lebensgeschichte und die Entwicklungsumwelt des Kindes zu erhalten sowie die Beobachtung (Ettrich, 2000; Sarimski, 2009).

Diesem modernen Verständnis entsprechend soll Entwicklungsdiagnostik folgende Aufgaben erfüllen: Sie soll normale Entwicklung differenziert erheben und gleichzeitig auch Entwicklungsstörungen, -stillstand bzw. -defizite möglichst früh diagnostizieren. Entwicklungsdiagnostik soll aber auch Auskünfte über Ressourcen und eventuelle beschleunigte Entwicklung geben. Basierend auf diesen durch Test, Gespräch und

Beobachtung gewonnenen Informationen soll man eine Förderung planen können und die Beratung durchführen. Die Informationen aus der Entwicklungsdiagnostik sollen auch helfen Entwicklungsprognosen zu erstellen sowie dem Monitoring und der Evaluation von Präventionen und Interventionen dienen (Ettrich, 2000; Petermann, Stein & Macha, 2008).

2.3. *Definition Entwicklungstest*

Eine verbreitete Definition von Tests geben Lienert und Raatz (1998). Sie verstehen unter einem Test „ein wissenschaftliches Routineverfahren zur Untersuchung eines oder mehrerer empirisch abgrenzbarer Persönlichkeitsmerkmale mit dem Ziel einer möglichst quantitativen Aussage über den relativen Grad der individuellen Merkmalsausprägung“ (S. 1). Diese Definition betont die Standardisierung von Tests und die Quantifizierung bzw. das Messen von Merkmalen. Bei Entwicklungstests handelt es sich beim Untersuchungsgegenstand bzw. bei den zu erhebenden Persönlichkeitsmerkmalen um Merkmale der Entwicklung (Quaiser-Pohl, 2010).

2.4. *Klassifikation von Entwicklungstests*

Bei Entwicklungstests unterscheidet man zwischen den Verfahrenstypen Screenings, allgemeine Entwicklungstests und spezifische Entwicklungstests (Esser & Petermann, 2010).

Screenings sind Kurztestverfahren, die unter Aufwendung geringer zeitlicher und finanzieller Ressourcen einen groben Überblick über den Entwicklungsstand bzw. eine erste Einschätzung über mögliche Auffälligkeiten liefern sollen. Wenn aufgrund dessen ein Verdacht für das Vorliegen einer Entwicklungsstörung vorliegt, müssen weitere entwicklungsdiagnostische Verfahren vorgegeben werden (Esser & Petermann, 2010; Rümmele, 2011).

Spezifische Entwicklungstests beschränken sich auf die Erhebung eines speziellen Entwicklungsbereichs. Es bestehen unter anderem Verfahren zur Erhebung von motorischen oder sprachlichen Leistungen, der Wahrnehmungsentwicklung oder der kognitiven Entwicklung (Esser & Petermann, 2010). Zu den spezifischen Entwicklungstests sind auch Intelligenztests zu zählen (Hagmann-von Arx et al., 2008).

Allgemeine Entwicklungstests sollen möglichst alle Bereiche der Entwicklung erfassen. Es sollen so Aussagen über den allgemeinen Entwicklungsstand eines Kindes und über die Entwicklung in den unterschiedlichen Bereichen getätigt werden können (Esser & Petermann, 2010). Laut Esser und Petermann (2010) beinhalten die meisten allgemeinen Entwicklungstests die Erfassung folgender Bereiche:

- Körpermotorik („Grobmotorik“)
- Handmotorik („Auge-Hand-Koordination“ „Visuomotorik“)
- Wahrnehmung
- kognitive Entwicklung
- Sprachentwicklung
- Sozialentwicklung
- emotionale Entwicklung
- Selbstständigkeit

Der spielbasierte Itempool erhebt ebenfalls diese Bereiche. Da er der letztgenannten Verfahrensgruppe zugehörig ist, beschäftigen sich vor allem das Kapitel 4 (Kritik an allgemeinen Entwicklungstests) genauer mit allgemeinen Entwicklungstests.

2.5. Begriffsabgrenzung zu Intelligenzdiagnostik

Unter Intelligenzdiagnostik versteht man die Erhebung der kognitiven Entwicklung (Renner & Irblich, 2009). Da sich diese Arbeit auf die Erhebung der kognitiven Entwicklung im Rahmen von Entwicklungstests konzentriert, soll der Begriff der Entwicklungsdiagnostik vom Begriff der Intelligenzdiagnostik abgegrenzt werden. Die Anfänge der Intelligenzdiagnostik sowie der allgemeinen Entwicklungsdiagnostik sind auf den Test von Simon und Binet (1905) zurückzuführen (Hagmann-von Arx et al., 2008). Daher soll die historische Entwicklung der Diagnostik skizziert werden.

Ab dem 19. Jahrhundert war man darum bemüht, Unterschiede in Entwicklungsverläufen darzustellen. In Frankreich wurde daher von Binet und Simon 1905 ein Intelligenztest (Binet Test) entwickelt, um geistige Behinderungen bei Kindern ausfindig machen und sie daraufhin Sonderschulen zuweisen und eine passende Förderung anbieten zu können. Bald wurde jedoch klar, dass für die Entwicklungsdiagnostik im Gegensatz zur Intelligenzdiagnostik neben kognitiven

Fähigkeiten auch andere Verhaltensbereiche von Interesse sind. Der Ansatz von Binet und Simon wurde von Gsell auch auf andere Entwicklungsbereiche ausgeweitet. So wurden ab den 1940er Jahren erste Entwicklungstests veröffentlicht (Hagmann-von Arx et al., 2008; Petermann, Stein & Macha, 2008).

Neben diesem großen Unterschied die Breite des Untersuchungsgegenstandes betreffend finden sich noch zahlreiche andere Differenzierungsmerkmale. So wird der Begriff Intelligenztest meist für Testungen kognitiver Fähigkeiten von Kindern ab dem dritten Lebensjahr verwendet. Bei jüngeren Kindern werden im Rahmen von Entwicklungstests Vorläuferfähigkeiten der späteren Intelligenz erhoben (Renner & Irblich, 2009).

Die beiden Verfahrensgruppen wollen das momentane Funktionsniveau von Kindern erheben, jedoch liegen ihnen unterschiedliche Konstrukte zugrunde. Entwicklungstests wurden entwickelt, um beurteilen zu können, ob ein Kind bestimmte Fertigkeiten, die in seiner Altersgruppe beobachtbar sind, schon erworben hat (Renner & Irblich, 2009). Sie basieren auf theoretischen Modellen der Entwicklung und empirischen Daten. Diese Verfahren erheben jene Bereiche, die die größte Varianz in einer bestimmten Altersgruppe darstellen (Bayley, 2006). Während bei Konstruktionen von Entwicklungstests meist berücksichtigt wird, dass die Entwicklung variantenreich verläuft und die Veränderungen qualitativer Natur sind, gehen Intelligenztests davon aus, dass die Fähigkeiten normalverteilt sind und zu verschiedenen Alterszeitpunkten verschiedene Fähigkeitsausprägungen erreicht werden. Bei Entwicklungstests ist es aufgrund qualitativer Veränderungen oft schwierig, Fähigkeit eines späteren Zeitpunkts vorherzusagen. Bei Intelligenztest geht man jedoch von der Vorhersagbarkeit späterer Leistungen aus (Reuner, Rosenkranz, Pietz & Horn, 2007).

3. Besonderheiten der Diagnostik im Kleinkindalter

Im Folgenden wird auf die Besonderheiten eingegangen, die sich bei Testungen mit Kleinkindern im dritten Lebensjahr ergeben. Es werden daraus Empfehlungen für die Testung von Kleinkindern abgeleitet, die später als Grundlage für die Testungen im empirischen Teil dienen sollen.

3.1. Das Kind als Klient

Es bestehen große Unterschiede darin, ob ein Kleinkind ein älteres Kind oder ein Erwachsener Klient einer psychologischen Diagnostik ist. In den meisten Fällen wenden sich die Bezugspersonen von Kleinkindern mit der Bitte um Abklärung der Entwicklung an den Psychologen/die Psychologin. Kleine Kinder können sich jedoch nicht gezielt für oder gegen eine psychologische Diagnostik aussprechen (Esser & Petermann, 2010).

Die Kinder bringen auch andere Verhaltensweisen und Fähigkeiten mit in die Untersuchungssituation als Erwachsene. Sie sind in ihrem Verhalten oft unvorhersehbar. Die Kinder sind impulsiv und reagieren eventuell manchmal mit Gefühlsausbrüchen. Die Aufmerksamkeitspanne der Kinder ist meist gering, die Ablenkbarkeit erhöht und die Aufmerksamkeit unterliegt starken Schwankungen (Esser & Petermann, 2010; Irblich & Renner, 2009; Macha, Proske & Petermann, 2005; Reuner & Pietz, 2006).

Der Motivation von Kleinkindern kommt daher eine große Bedeutung zu. Kleinkinder sind nicht in der Lage, die Überzeugungen der untersuchenden Person bzw. den Grund der Testung zu verstehen (Kapitel 6.2.1. Theory of Mind). Einerseits führt das dazu, dass kleine Kinder nicht versuchen können ein Testergebnis zu verfälschen andererseits kann es zu mangelnder intrinsischer Motivation und Unsicherheit auf der Seite des Kindes führen. Zudem ist die Kooperationsbereitschaft der Kinder unter sechs Jahren oft sehr sprunghaft (Irblich & Renner, 2009).

Kinder werden bei der Testung mit ganz anderen Gefühlen konfrontiert als Erwachsene. Die Testung stellt für Kinder eine unbekannt Situation dar, die oft viele Anforderungen an sie richtet und verschiedene Emotionen wie Angst auslöst. Schon der Kontaktaufbau zu Kindern im dritten Lebensjahr gestaltet sich oft als schwierig, da viele Fremden gegenüber Angst empfinden bzw. scheu sind (Irblich & Renner, 2009).

Die Testsituation muss an die Bedürfnisse des Kleinkindes angepasst sein. Für den Diagnostiker/die Diagnostikerin bedeutet dies, dass er/sie dem Kind vor allem Sicherheit und Motivation bieten muss. Müdigkeit, Hunger und andere Bedürfnisse des Kindes müssen stärker beachtet werden als bei Testungen mit Erwachsenen (Esser & Petermann, 2010, Irblich & Renner, 2009). Wie genau die Gestaltung einer kinderechten Testung aussehen sollte, wird in den folgenden Kapiteln beschrieben.

3.2. Anforderungen an die untersuchende Person

Generell für alle psychologisch-diagnostischen Untersuchungen gilt, dass der Testleiter/die Testleiterin sehr gut mit den Materialien und dem Ablauf vertraut sein muss, um eine möglichst standardisierte und objektive Testsituation zu gewährleisten. Dies erfordert eine sehr gute Vorbereitung der Untersuchung. Bei Testungen mit Kleinkindern muss jedoch ein Kompromiss geschlossen werden: Einerseits sollte die Testung möglichst standardisiert sein, andererseits muss man bei Kleinkindern individuell und einfühlsam vorgehen und die Informationen aus Beobachtung, Testung und Elternbefragung kombinieren (Esser & Petermann, 2010).

Bei Testungen mit Kleinkindern ergeben sich somit höhere Anforderungen an die untersuchende Person als bei älteren Kindern. Ihr wird ein hohes Maß an Kompetenz und psychologischem Geschick abverlangt (Esser & Petermann, 2010, Irblich & Renner, 2009). Viel Erfahrung mit Kindern und die Fähigkeit, flexibel auf die individuellen Bedürfnisse einzugehen, sind notwendige Voraussetzungen für eine kindgerechte Untersuchung. Die untersuchende Person muss sensibel für die Bedürfnisse des Kindes sein, um entsprechend darauf reagieren zu können. Sie sollte auch Blickkontakt zum Kind suchen, um Fragen dann zu stellen, wenn das Kind aufnahmefähig ist (Irlich & Renner, 2009).

Die Untersuchungsperson muss außerdem viel Freude an der Arbeit mit Kindern und ein ehrliches Interesse an deren Ansichten aufweisen, da Kinder sehr sensibel auf Gleichgültigkeit oder Ablehnung reagieren. Dem Kind soll eine wertschätzende Haltung entgegengebracht werden. Dies wird auch erreicht, indem sich der Diagnostiker/die Diagnostikerin um eine kindgerechte Ausdrucksweise und freundliche Sprache bemüht (Irlich & Renner, 2009).

3.3. Anwesenheit der Bezugsperson

Bei Kindern unter vier Jahren ist meist ein Elternteil bei der Untersuchung anwesend. Die Eltern müssen einerseits über den Ablauf, das Ziel, die Methoden und die Dauer informiert werden und ihr Einverständnis geben, andererseits sollen sie dazu beitragen, dass sich das Kind wohl fühlt. Nur wenn sich das Kind wohl fühlt, kann das typische Verhalten des Kindes erhoben werden (Bayley, 2006; Esser & Petermann, 2010).

Die Anwesenheit der Eltern bietet weitere Informationen über die Eltern-Kind-Interaktion. Außerdem dient die Befragung der Eltern als notwendige Informationsquelle bei der Anamnese und Exploration. Sie können aber auch Informationen über mögliche störende Einflüsse geben, wie Erkrankungen, Konflikte vor der Untersuchung, Müdigkeit, Hunger oder Durst des Kindes (Irblich & Renner, 2009).

Reuner et al. (2007) empfehlen im Manual zur Bayley II, dass auch einzelne Aufgaben von der Bezugsperson an das Kind gestellt werden können. Jedoch muss es für ein solches Vorgehen strenge Vorschriften geben, wann und wie die Eltern miteinbezogen werden sollen. Ansonsten sollen sich die Eltern nicht an der Untersuchung beteiligen. Sie müssen sich während der Testung im Hintergrund halten. Die untersuchende Person sollte beobachten, ob Verhaltensweisen der Bezugspersonen eventuell dennoch Einfluss auf das Verhalten des Kindes haben (Irblich & Renner, 2009).

Die Eltern müssen in jedem Fall mit Respekt behandelt werden. Sie erleben die Anwesenheit bei der Testung oft als sehr stressig, da sie meist wünschen, dass ihr Kind gut abschneidet. Wenn ein Kind dann aber keine guten Leistungen zeigt, führt dies zu Nervosität bei manchen Eltern (Kamphaus, 2005). Die Erwartungen, Ängste und Wünsche der Eltern sollen deshalb vorab besprochen werden. Esser und Petermann (2010) empfehlen auch, die Eltern darauf hinzuweisen, dass das Kind nicht alle Aufgaben lösen muss, um ein gutes Ergebnis zu erzielen. So sollen durch Information ein möglicher Druck oder dadurch motivierte Hilfestellungen seitens der Eltern vermieden und eine realistische Erwartungshaltung erzeugt werden.

3.4. Zeitliche Rahmenbedingungen

Der Testzeitpunkt sollte unter Berücksichtigung des Schlaf-Wach-Rhythmus des Kindes und unter Absprache mit den Bezugspersonen gewählt werden. Esser und Petermann (2010) empfehlen des Weiteren Testungen am Vormittag, da die meisten Kinder zu dieser Tageszeit am leistungsfähigsten und interessiertesten seien. Wartezeiten sollten vermieden werden. Die Untersuchungsdauer sollte so kurz wie möglich gehalten werden und eine Zeit von dreißig bis fünfzig Minuten nicht überschreiten. Je nach Durchhaltevermögen und Aufmerksamkeit des Kindes kann der Test in ein oder zwei Sitzungen abgehalten werden (Reuner et al, 2007).

Kinder können den Wunsch nach einer Pause meist nicht selbst äußern. Daher sollte das Verhalten des Kindes hinsichtlich Anzeichen für den Bedarf einer Pause beobachtet werden (Kamphaus, 2005). Anzeichen hierfür können motorische Unruhe, ein abschweifender Blick, längere Reaktionszeiten, ablenkende Fragen und Suche nach Nähe zu den Eltern sein (Irblich & Renner, 2009).

3.5. Räumliche Rahmenbedingungen

Das Zimmer sollte hell und freundlich gestaltet sein, sodass sich alle Anwesenden im Raum wohlfühlen. Ähnlichkeiten zu einer Arztpraxis sollen vermieden werden, um das Kind nicht an schmerzhaftes, angstvolle Erfahrungen zu erinnern. Außerdem sollte auf eine angenehme Raumtemperatur und genügend frische Luft geachtet werden (Irblich & Renner, 2009; Kamphaus, 2005). Für die Sicherheit des Kindes muss stets gesorgt sein, so sollten sich z.B. Kabel und Stromquellen außerhalb der Reichweite des Kindes befinden (Esser & Petermann, 2010).

Im Testraum werden häufig unterschiedliche Bereiche für Gespräch, Spiel und Untersuchung abgegrenzt. Ein Spielteppich mit einigen Spielmaterialien soll das Kind zum Spielen einladen. Im Untersuchungsbereich sind Sitzmöglichkeiten und Möbel in kindgerechter Größe wichtig. Es wird empfohlen, dass der Testleiter/die Testleiterin und das Kind vis-a-vis oder übereck sitzen (Irblich & Renner, 2009). Der Untersuchungsbereich sollte zum Wohlfühlen einladen, gleichzeitig aber möglichst wenig Ablenkung bieten. Ratsam ist daher, Spiel- oder Untersuchungsmaterial, das gerade nicht verwendet wird, aus dem Blickfeld des Kindes zu räumen. Andere

Störquellen wie ablenkende Geräusche oder blendendes Licht sollten ebenfalls vermieden werden (Esser & Petermann, 2010).

3.6. Untersuchungsablauf

Im Folgenden soll ein Überblick über konkrete Empfehlungen für den Ablauf einer Untersuchung im Kleinkindalter gegeben werden:

Kontaktaufnahme: Die Begrüßung stellt einen sehr wichtigen Aspekt für den Beziehungsaufbau dar. Der Testleiter/die Testleiterin soll das Kind auf Augenhöhe begrüßen und es beim Namen nennen. Ist das Kind schüchtern, sollte sich der Psychologe/die Psychologin zunächst der Bezugsperson zuwenden und dem Kind ermöglichen, sich als passiver Beobachter einen Eindruck von dem Testleiter/der Testleiterin zu machen (Irblich & Renner, 2009; Petermann, Stein & Macha, 2008). In jedem Fall ist es ratsam viel Zeit in den Aufbau einer Beziehung zum Kind zu investieren. Bei sehr lebhaften Kindern soll der Diagnostiker/die Diagnostikerin zwar freundlich sein, aber es ist notwendig, von Anfang an die Grundregeln klarzustellen (Irblich & Renner, 2009; Reuner et al., 2007).

Untersuchung: Bevor mit der Untersuchung mit dem Kind begonnen wird, sollten der Psychologe/die Psychologin den weiteren Ablauf erklären. Für den Beziehungsaufbau mit dem Kind erweist es sich als hilfreich, das Einverständnis der Eltern für diesen Ablauf einzuholen, da die Kinder eher kooperieren, wenn sie merken, dass die Eltern mit dem Geschehen einverstanden sind (Irblich & Renner, 2009).

Dem Kind soll die Gelegenheit geboten werden, sich im Untersuchungsraum umzusehen (Petermann, Stein & Macha, 2008). Das im Raum bereitgestellte Spielzeug erleichtert den Einstieg in die Untersuchungssituation. Der richtige Zeitpunkt für den Beginn der Testung ist dann gegeben, wenn das Kind Bereitschaft für die Teilnahme signalisiert. Anzeichen hierfür sind, dass es sich immer mehr von den Eltern löst, mit den Spielmaterialien beschäftigt und in Kontakt mit dem Untersucher/der Untersucherin tritt (Irblich & Renner, 2009).

Für Kinder unter vier Jahren empfehlen Irblich und Renner (2009) einen spielerischen Zugang zur Testung. Durch interessante Testmaterialien sollen die Neugier und das Interesse der Kleinkinder geweckt werden. Um die Kooperationsbereitschaft hoch zu

halten, soll die Itemreihenfolge flexibel sein und sich an den Interessen und Bedürfnissen des Kindes orientieren. Es empfiehlt sich jedoch, mit einfachen Aufgaben zu beginnen. Wenn ein Kind in der Untersuchungssituation noch wenig bis gar nichts gesprochen hat, sollten zuerst nonverbale Untertests vorgegeben werden. Die Dauer der Bearbeitung der Aufgaben bestimmt das Kind. Jedoch muss der Testleiter/die Testleiterin abwägen, ob das Tempo für die Aufrechterhaltung der Aufmerksamkeit angemessen ist (Reuner et al., 2007).

Reuner et al. (2007) betonen auch, dass es wichtig sei, Aufgaben, die nicht gelöst wurden, zu einem späteren Zeitpunkt nochmals vorzugeben. Oft ist es der Fall, dass Kinder zu Beginn einer Testung noch schüchtern und zurückhaltend sind und sie erst bei einer späteren Vorgabe der Aufgabe die Lösung finden. Durch den Aufbau einer Beziehung während der Untersuchung kann Kindern auch die Bearbeitung von weniger interessanten Aufgaben schmackhaft gemacht werden. Der Testleiter/die Testleiterin sollte stets beobachten, ob das Kind Anzeichen von Müdigkeit zeigt, wann Motivationseinbrüche drohen und wann die Aktivität gewechselt werden sollte.

Beobachtung: Während der gesamten Untersuchungssituation muss der Testleiter/die Testleiterin das Verhalten des Kindes beobachten. So können zum Beispiel Informationen über die Arbeitshaltung, Konzentrationsfähigkeit und Frustrationstoleranz beiläufig beim Bearbeiten von Aufgaben erhoben werden. In die Interpretation der Testwerte sollten Motivation und Äußerungen des Kindes mit einfließen. Auch die Beobachtung der Interaktion zwischen Mutter und Kind kann wichtige Informationen für die Interventionsplanung liefern (Irblich & Renner, 2009, Kastner-Koller & Deimann, 2009).

Testende: Am Ende sollte sich der Testleiter/die Testleiterin beim Kind bedanken und das Kind nach einer Rückmeldung über die Testung fragen. Der Testleiter/die Testleiterin soll eventuell aufgetretene Probleme ansprechen und motivierende Worte für einen eventuellen nächsten Termin finden. Das Kind soll die Testung mit einem angenehmen Gefühl verlassen (Irblich & Renner, 2009).

4. Kritik an allgemeinen Entwicklungstests

Dieses Kapitel beschäftigt sich mit der Kritik an Entwicklungstests. Es sollen die methodischen Anforderungen an Verfahren der psychologischen Diagnostik – die sogenannten Gütekriterien – dargestellt werden. Dann werden gängige allgemeine Entwicklungstests für den deutschsprachigen Raum vorgestellt und unter anderem hinsichtlich der Gütekriterien kritisiert.

4.1. Gütekriterien

Diagnostische Verfahren müssen bestimmte methodische Anforderungen – die sogenannten Gütekriterien – erfüllen. Mithilfe der Gütekriterien kann die Qualität eines Verfahrens beurteilt werden. Zu den Hauptgütekriterien der klassischen Testtheorie zählt man Objektivität, Reliabilität und Validität. Als Nebengütekriterien gelten Normierung bzw. Eichung, Ökonomie und Nützlichkeit. Zusätzlich soll auch die Zumutbarkeit als weiteres Kriterium für psychologisch-diagnostische Verfahren beschrieben werden (Bortz & Döring, 2006; Kubinger, 2009).

Die Gütekriterien werden hinsichtlich ihrer Bedeutung für allgemeine Entwicklungstests beleuchtet und die Probleme von Verfahren im Kleinkindalter, einigen dieser Kriterien zu entsprechen (Esser & Petermann, 2010), werden erläutert.

4.1.1. Objektivität

Ein Verfahren ist objektiv, wenn es unabhängig von der untersuchenden Person ist, d.h. wenn verschiedene Untersucher/Untersucherinnen bei demselben Klienten/derselben Klientin zu dem gleichen Ergebnis gelangen (Lienert & Raatz, 1998). Die Objektivität berechnet man, indem man die interpersonelle Übereinstimmung der untersuchenden Personen ermittelt (Esser & Petermann, 2010). Man unterscheidet verschiedene Aspekte der Objektivität je nachdem, welche Phase des Tests betrachtet wird (Lienert & Raatz, 1998):

Die *Durchführungsobjektivität* bezieht sich auf die Unabhängigkeit des Verfahrens von äußeren Umständen wie räumlichen und zeitlichen Bedingungen, Material, Instruktion und dem Verhalten der untersuchenden Person (Esser & Petermann, 2010; Kubinger, 2009). Kubinger (2009) fasst den Begriff der Durchführungsobjektivität enger und bevorzugt den Begriff *Testleiterunabhängigkeit*. Eine ausführliche schriftliche

Instruktion des Testleiters/der Testleiterin und eine standardisierte Testdurchführung helfen, den Einfluss eines Untersuchers/einer Untersucherin auf die Testperson zu vermeiden (Lienert & Raatz, 1998). Das Manual sollte neben wörtlichen Testanweisungen und der genauen Beschreibung der Vorgabe der Aufgaben auch genaue Informationen geben, wie man mit schwierigen Situationen wie z.B. Rückfragen des Kindes umgehen soll (Renner, 2009).

Bei der Auswertung der Antworten der Testperson soll die *Auswertungsobjektivität* bzw. Verrechnungssicherheit gewährleistet werden. Notwendig hierfür sind klare Regeln, wie welches Testverhalten zu verwerten ist bzw. welchem Verhalten welcher numerischer oder kategorialer Wert zuzuordnen ist. Das Manual sollte außerdem auch Lösungsbeispiele beinhalten (Renner, 2009). Bei geschlossenen Antwortformaten wird diesem Kriterium leichter entsprochen als bei offenen Antwortformaten (Kubinger, 2009; Lienert und Raatz, 1998).

Die *Interpretationsobjektivität* bzw. Interpretationseindeutigkeit ist gegeben, wenn die Schlussfolgerungen aus den Ergebnissen unabhängig vom/von der interpretierenden Untersucher/Untersucherin sind. Je genauer die Anweisungen im Manual sind, wie welche Ergebnisse zu beurteilen sind, desto objektiver ist die Interpretation. Diese Form der Objektivität wird durch den Vergleich des Testwerts einer Testperson mit den Ergebnissen der vergleichbaren Normstichprobe erreicht (Kubinger, 2009; Lienert & Raatz, 1998).

Die Forderung nach standardisierten Testbedingungen stellt gerade bei jungen Kindern ein Problem dar, da man sich oft an den individuellen Bedürfnissen des Kindes orientieren muss. Mithilfe von Informationen aus Befragung, Beobachtung und Tests wird die Entwicklung des Kindes zusammenfassend eingeschätzt (Esser & Petermann, 2010).

4.1.2. Reliabilität

Die Reliabilität eines Tests gibt den Grad der Genauigkeit an, mit dem ein Verfahren ein Merkmal misst. Genau misst der Test dann, wenn die Ergebnisse der Testperson fehlerfrei angegeben werden können. Die Reliabilität ist jedoch kein Maß dafür, ob der Test auch das Merkmal misst, das er zu messen beansprucht (siehe Kapitel 4.1.3. Validität).

Der Reliabilitätskoeffizient gibt an, in welchem Maße Testwerte einer Testperson, die unter gleichen Bedingungen gewonnen wurden, übereinstimmen (Lienert & Raatz, 1998). Ein Koeffizient von null spricht dafür, dass das Ergebnis ausschließlich aus Messfehlern besteht und ein Koeffizient von eins spricht für keinen Messfehler (Esser & Petermann, 2010). Werte ab .7 können als akzeptabel angesehen werden (Kline, 2000, S. 13).

Laut Lienert und Raatz (1998) kann die Messgenauigkeit mithilfe folgender Methoden berechnet werden:

Sowohl die *Retest-Reliabilität* als auch die *Paralleltest-Reliabilität* wird mittels Testwiederholung und anschließender Korrelation der Ergebnisse zu den beiden Zeitpunkten gemessen. Bei der Retest-Methode wird einer Stichprobe zweimal der gleiche Test vorgegeben. Bei der zweiten Variante wird beim zweiten Zeitpunkt ein Paralleltest, also eine gleichwertige Nachahmung des Tests, verwendet (Lienert & Raatz, 1998). Bei Verfahren für Kleinkinder ist es aufgrund der sehr geringen Anzahl an möglichen Items für die noch stark begrenzten Entwicklungsbereiche schwer möglich Paralleltests zu konstruieren (Esser & Petermann, 2010).

Eine Testwiederholung mit nur wenig verstrichener Zeit zwischen den Testzeitpunkten ist aufgrund von Übungs- und Erinnerungseffekten nicht zumutbar. Daher sind längere Abstände zwischen den Testzeitpunkten zu bevorzugen (Kubinger, 2009). Jedoch gilt sowohl für die Paralleltest- als auch für die Retest-Reliabilität, dass sich diese Methoden kaum für Verfahren für das Kleinkindalter eignen, da in diesem Altersbereich rasche Veränderungen individueller Geschwindigkeit und unterschiedlicher Verlaufsformen stattfinden. Unterschiede zwischen den Ergebnissen der beiden Testzeitpunkte resultieren daher nicht nur aus der Ungenauigkeit der Messung, sondern auch aus zwischenzeitlichen Entwicklungsprozessen und eventuellen Übungseffekten. Der Einfluss dieser Prozesse ist umso größer, je länger die Testzeitpunkte auseinander liegen (Esser & Petermann, 2010; Renner, 2009).

Ettrich (2000) empfiehlt daher, die innere Konsistenz eines entwicklungsdiagnostischen Verfahrens zu berechnen. Die *innere Konsistenz* eines Tests sagt aus, wie gut die einzelnen Teile dasselbe messen (Kubinger, 2009; Lienert und Raatz, 1998). Man

berechnet sie mittels der Testhalbierungsmethode oder der Konsistenzanalyse. In beiden Fällen wird der Test einmal einer Stichprobe dargeboten.

Bei der *Split-half-Methode* werden die Aufgaben eines Verfahrens in zwei Hälften aufgeteilt und die Korrelation der Testergebnisse der beiden Hälften wird berechnet. Die *Konsistenzanalyse* ist ein Spezialfall der Testhalbierung. Das Verfahren wird hierbei in so viele Teile gegliedert, wie es Aufgaben gibt (Kubinger, 2009; Lienert & Raatz, 1998). Genaueres zur Berechnung findet sich in „Kapitel 9.1. Auswertung: Ite-manalyse und Reliabilität“. Da Verfahren für Kleinkinder häufig aus nur wenigen Items bestehen und die Skalen oft die Voraussetzung der Konsistenzanalyse für Homogenität verletzen, ist diese Berechnung bei entwicklungsdiagnostischen Analysen oft nicht passend (Esser & Petermann, 2010).

4.1.3. Validität

Unter Validität eines Tests ist zu verstehen, dass er tatsächlich jenes psychische Merkmal misst, welches es zu messen beansprucht (Kubinger, 2009, S.55). Die Validität gilt als das wichtigste, aber am schwierigsten zu prüfende Gütekriterium (Kubinger, 2009; Lienert & Raatz, 1998). Objektivität und Reliabilität stellen die Messgenauigkeit sicher und bieten die Voraussetzungen für eine hohe Validität. Wenn die Reliabilität niedrig ist, kann die Validität keinen hohen Wert erreichen. Nur wenn ein Verfahren eine hohe Validität besitzt, können seine Testergebnisse auch für andere Situationen außerhalb der Testsituation generalisiert werden (Moosbrugger & Kevala, 2012). Die Validität wird formal mittels Korrelation des Testergebnisses mit einer Variablen, die außerhalb der Testsituation liegt, bestimmt (Moosbrugger & Kevala, 2012). Werte zwischen .40 und .90 sind als mittelmäßig anzusehen, Werte unter .40 als niedrig und jene über .90 als hoch (Weise, 1975). Drei Arten von Validität können in Abhängigkeit der verwendeten Kriterien unterschieden werden (Kubinger, 2009):

Der *inhaltlichen Gültigkeit* wird entsprochen, wenn ein Test selbst das optimale Kriterium für das zu erhebende Merkmal darstellt und dieses repräsentiert (z.B. Schreibprobe zur Erfassung der Schreibgeschwindigkeit einer Sekretärin). Um die inhaltliche Gültigkeit zu erlangen und somit herauszufinden, ob die Items messen, was sie laut operationaler Definition messen sollen, müssen sie hinsichtlich ihres Inhalts betrachtet werden. Meist geschieht dies mittels „Expertenrating“. Darunter versteht man

eine Beurteilung der Validität des Verfahrens auf einer Ratingskala durch eine Gruppe kompetenter Fachleute (Kubinger, 2006; Lienert & Raatz, 1998).

Mittels *Konstruktvalidität* wird angegeben, in welchem Maße die Aufgaben eines Verfahrens mit dem theoretischen Konstrukt, das gemessen werden soll, übereinstimmen. Es kann z.B. mithilfe der Berechnung einer Faktorenanalyse, der Skaleninterkorrelationen und der Alterstrends nachgewiesen werden, ob dem Konstrukt entsprochen wird oder nicht. Bei allgemeinen Entwicklungstests kommt der Differenzierungsfähigkeit zwischen auffälligen und unauffälligen Kindern besondere Bedeutung zu (Macha, Proske & Petermann, 2005). Probleme bei allgemeinen Entwicklungstests ergeben sich dadurch, dass der allgemeine Entwicklungsstand eine sehr heterogene und sich ständig verändernde Variable ist und sich daher schwer als ein einheitliches Konstrukt verstehen lässt (Esser & Petermann, 2010).

Mittels Korrelation des Testergebnisses mit einer davon unabhängig erhobenen, relevanten Variable (sog. Außenkriterium) kann man ein Maß für die *kriteriumsbezogene Validität* angeben (Lienert & Raatz, 1998). Zu unterscheiden sind die Übereinstimmungs- und die prognostische Validität. Die *Übereinstimmungsvalidität* wird für ein Verfahren mittels Korrelation mit einem anderen Verfahren, das dasselbe Konstrukt erheben soll, berechnet. Bei der *prognostischen Validität* wird die Korrelation des Testergebnisses mit einer Variablen, die in der Zukunft liegt, berechnet (Kubinger, 2009). Die prognostische Aussagekraft ist bei Entwicklungstests durch die Schwierigkeit, Vorläuferfähigkeiten für komplexe Kompetenzen zu finden, und aufgrund der schnellen individuellen Entwicklungsprozesse oft eingeschränkt. Dennoch gilt die prognostische Validität für Verfahren zur Früherkennung von Entwicklungsabweichungen als sehr wichtig (Sarimski, 2009). Ettrich (2000) betont, dass für eine befriedigende Entwicklungsvorhersage neben Ergebnissen von Entwicklungstests auch Informationen über die Lebensgeschichte und die Umwelt des Kindes herangezogen werden müssen. Spezifische Entwicklungsbereiche sind dem Autor zufolge geeigneter für langfristige Prognosen.

4.1.4. Normierung

Ein Test, der den Hauptgütekriterien entspricht, jedoch nicht normiert wurde, besitzt – wenn überhaupt – nur geringe Brauchbarkeit (Lienert & Raatz, 1998). Ein Verfahren ist normiert, wenn ein Bezugssystem vorliegt, mit dem man das Testergebnis einer Person im Vergleich zu anderen einordnen bzw. relativieren kann (Kubinger, 2009; Lienert & Raatz, 1998).

Kubinger (2009) formuliert als Kriterien für die Erfüllung des Gütekriteriums, dass die Eich Tabellen nicht veraltet sein dürfen, die Population definiert und die herangezogene Stichprobe repräsentativ sein muss. Die Normen dürfen nicht älter als zehn Jahre sein (Macha et al., 2005). Esser und Petermann (2010) betonen die Bedeutung einer für den niedrigen Leistungsbereich repräsentativen Normstichprobe für die Diagnostik von Entwicklungsstörungen. Nur dann kann optimal zwischen unterschiedlich niedrigen Leistungen im unteren Leistungsbereich differenziert werden.

4.1.5. Ökonomie

Ein Test ist ökonomisch, wenn gemessen am diagnostischen Informationsgewinn wenige Ressourcen aufgewendet werden müssen: Er sollte demnach wenig Zeit und Materialien bei der Durchführung, Auswertung und Interpretation benötigen, als Gruppentest durchführbar und einfach zu handhaben sein (Kubinger, 2009).

4.1.6. Nützlichkeit

Nützlich ist ein Test dann, wenn in der Praxis ein Bedarf an der Messung des vom Test erhobenen psychischen Merkmals besteht und der Test durch keinen anderen Verfahren vertreten werden kann (Kubinger, 2009, Lienert & Raatz, 1998). Die Entscheidungen, die basierend auf dem Testergebnis getroffen werden, müssen mehr Nutzen als Schaden erwarten lassen (Kubinger, 2009).

4.1.7. Zumutbarkeit

Ein Test ist zumutbar, „wenn er die Testperson absolut und relativ zu dem aus seiner Anwendung resultierenden Nutzen in zeitlicher, psychischer (insbesondere energetisch-motivationaler und emotionaler) sowie körperlicher Hinsicht schont (Kubinger, 2009; S. 15). Es ist sehr schwierig, zu konkretisieren, was als zumutbar bzw. nützlich anzusehen ist. In der Entwicklungsdiagnostik sind in jedem Fall Überlegungen über die

Zumutbarkeit der Schwierigkeit der Aufgaben, der Testdauer und der Pausengestaltung, sowie des Untersuchungsablaufs, anzustellen (Kubinger, 2009).

4.2. Kritik an der aktuellen allgemeinen Entwicklungsdiagnostik

Für den deutschsprachigen Raum existieren nur sehr wenige allgemeine Entwicklungstests (Quaiser-Pohl & Köhler, 2010). Im Brickenkamp Handbuch psychologischer und pädagogischer Tests wird der Bestand an allgemeinen Entwicklungstests im deutschen Sprachraum mit sieben Verfahren angegeben, wobei davon nur drei Verfahren Aufgaben für das dritte Lebensjahr beinhalten (Brähler et al., 2002). In dieser Auflistung sind die Bayley-Skalen noch nicht enthalten.

Laut Renner (2009) hat sich das Angebot an diagnostischen Verfahren für das Kleinkind- und Vorschulalter in den letzten Jahren zwar erweitert, jedoch erfüllen viele dieser Verfahren die Gütekriterien nicht hinreichend. Neuentwicklungen stammen meist aus dem angloamerikanischen Raum (Petermann, Stein & Macha, 2008). Inwiefern diese Normierungen dann im deutschsprachigen Raum repräsentativ sind, ist zu hinterfragen (Kubinger, 2009). Für deutschsprachige Verfahren liegt selten eine aktuelle deutsche Normierung vor und sie können meist nur eine geringe Altersspanne abdecken (Petermann, Stein & Macha, 2008).

Eine Schwierigkeit zeigt sich auch bezüglich der prognostischen Validität. Die Informationen aus Entwicklungstests über die ersten Lebensjahre zeigen vor allem für Kinder mit unauffälliger Entwicklung schlechte Vorhersagbarkeit für spätere Leistungen (Petermann & Macha, 2008).

Kritisch anzumerken ist des Weiteren, dass den Verfahren kein einheitliches theoretisches Fundament zugrunde liegt, da keine allgemein gültige Entwicklungstheorie existiert. Die Verfahren stellen eine Sammlung von altersbezogenen Aufgaben dar, die sich in der Praxis bewährt haben (Sarimski, 2009). Somit gibt es auch keine einheitliche Regelung, welche Entwicklungsdimensionen wie zusammenspielen müssen, um von einer normalen Entwicklung sprechen zu können (Petermann & Macha, 2003)..

Die Messung und die Interpretation der Entwicklung lassen darauf schließen, ob die Verfahren auf dem traditionellen oder modernen Entwicklungsbegriff beruhen

(Petermann & Macha, 2003). Viele Verfahren basieren auf einem überholten traditionellen Entwicklungsbegriff (Kapitel 2.1. Entwicklungsbegriff) und lassen variable Entwicklungsverläufe unberücksichtigt, da sich invariante Entwicklungsverläufe leichter erheben lassen. Dieses Vorgehen kann zu einer fehlerhaften Einschätzung der Entwicklung führen (Petermann & Macha, 2003; Petermann, Stein & Macha, 2008).

Für die Kritik an der Auswertung verschiedener Verfahren sollen drei Testformen unterschieden werden (Petermann & Macha, 2003), die kurz dargestellt und kritisiert werden sollen:

Viele deutschsprachige Verfahren sind als *Stufenleiterverfahren* konzipiert. Die Aufgaben sind hierfür nach ihrer Schwierigkeit geordnet. Je nach Alter des Kindes werden bezüglich der Schwierigkeit passende Items für den Testeinstieg gewählt. Das Kind bekommt solange immer schwierigere Aufgaben, bis es eine bestimmte Anzahl aufeinander folgender Items nicht mehr lösen kann und die Vorgabe abgebrochen wird. Aus dem letzten gelösten Item wird der Entwicklungsstand berechnet (Petermann & Macha, 2003). Zu kritisieren ist an dieser Vorgangsweise, dass die Kinder mit einem Misserfolgsereignis die Testung beenden (Witzlack, 2001).

Es wird meist ein Entwicklungsalter oder ein Entwicklungsquotient angegeben. Das Entwicklungsalter besagt, ob die Leistung eines Kindes seinem Alter oder den zu erwartenden Leistungen jüngerer oder älterer Kinder entspricht. Der Entwicklungsquotient stellt individuelle Leistungen den durchschnittlichen Leistungen Gleichaltriger gegenüber (Petermann & Macha, 2003).

Diese Verfahren bieten Übersichtlichkeit und eine schnelle Auswertung, jedoch kann bei einer Reihung der Aufgaben nach ihrer Schwierigkeit nicht die Variabilität der Entwicklungsverläufe berücksichtigt werden. So kann es vorkommen, dass manche Kinder eine schwierigere Aufgabe schon lösen können, eine leichtere jedoch nicht. Ein niedrigeres Entwicklungsalter als das tatsächliche Alter kann daher ein Hinweis für eine bedeutsame Entwicklungsabweichung oder aber eine Abweichung innerhalb der normalen Variationsbreite der Entwicklung in der Altersgruppe sein.

Der Quotient oder das Entwicklungsalter lassen außerdem keine Aussagen über die einzelnen Entwicklungsdimensionen zu (Petermann, Stein & Macha, 2008; Sarimski,

2009). Die Angabe eines Entwicklungsalters kann zu falschen Interpretationen führen. Wenn ein fünf- jähriges Kind Leistungen zeigt, die durchschnittlich schon von Dreijährigen gezeigt werden, so erbringt es diese Leistungen wahrscheinlich auf unterschiedliche Weise (Petermann & Macha, 2003). Die Bayley-Skalen (Kapitel 4.3.2. und 4.3.3.) und die „Münchener funktionelle Entwicklungsdiagnostik“ (Kapitel 4.3.1.) sind Beispiele für Stufenleiterverfahren.

Inventare wollen einen möglichst breiten Bereich an Entwicklungsdimensionen und viele Aspekte der kindlichen Entwicklung abbilden. Die einzelnen Dimensionen bestehen aus heterogenen Aufgaben, die viele unterschiedliche Leistungen erheben. Die Messgenauigkeit kann daher für diese Verfahren nicht bestimmt werden, daher werden sie oft nicht als Entwicklungstests im engeren Sinn angesehen (Petermann & Macha, 2003). Der „Entwicklungstest sechs Monate bis sechs Jahre“ (Kapitel 4.3.4.) ist ein Inventar.

Testbatterien bestehen aus mehreren Untertests, die jeweils eng umschriebene Entwicklungsbereiche mit homogenen Aufgaben unterschiedlicher Schwierigkeit erheben. Die Entwicklung wird abgebildet, da sich mit steigendem Alter die Leistungen in den Untertests verbessern. Diese Verfahren geben sehr genaue Informationen über eine begrenzte Anzahl an Leistungen. Für viele Fragestellungen sind diese Informationen über einige wenige eng umschriebene Bereiche nicht umfassend genug. Der Itempool der vorliegenden Arbeit stellt eine solche Testbatterie dar (Petermann & Macha, 2003, Sarimski, 2009).

Neben dieser methodischen Kritik gibt es Kritikpunkte der praktischen Durchführbarkeit an der standardisierten Testung, die eine unnatürliche Situation mit fremden Erwachsenen für die Kinder darstellt. Diese Vorgehensweise entspricht nur unzureichend den Besonderheiten des Kleinkindalters. Junge Kinder wollen nicht still sitzen, Anweisungen befolgen etc. Sie können so in der Testsituation ihre Fähigkeiten, die sie in anderen Situationen aber besitzen, nicht zeigen. Viele Entwicklungstests bieten außerdem zu wenige Informationen für die spätere Interventionsplanung. Noch deutlicher zeigen sich die Unzulänglichkeiten standardisierter Verfahren jedoch bei der Testung von Kindern mit starken Verzögerungen und Behinderungen. Diese Kinder können meist mit Entwicklungstests nicht diagnostiziert werden (Bagnato, 2005).

4.3. Darstellung und Kritik gängiger allgemeiner Entwicklungstests

Nach Sarimski (2009) zählen zu den in Deutschland am häufigsten verwendeten allgemeinen Entwicklungstests für das dritte Lebensjahr die „Münchener Funktionelle Entwicklungsdiagnostik für das zweite bis dritte Lebensjahr“ (MFED 2-3, Hellbrügge, 1994, 2001), der „Entwicklungstest sechs Monate bis sechs Jahre“ (ET 6-6; Petermann, Stein & Macha, 2004) und die „Bayley Scales of Infant Development – Second Edition – deutsche Fassung“ (Bayley II; Reuner, Rosenkranz, Pietz & Horn, 2008). Für die „Bayley Scales of Infant and Toddler Development – Third Edition“ (Bayley III; Bayley, 2006) gibt es noch keine deutschsprachige Übersetzung, sie sollen aber aufgrund der Bemühungen um die Verbesserung der psychometrischen Qualität und der Anwendbarkeit kurz dargestellt werden. Diese Entwicklungstest werden u. a. hinsichtlich der bereits beschriebenen Gütekriterien (Kapitel 4.1. Gütekriterien) kritisiert.

4.3.1. Die Münchener funktionelle Entwicklungsdiagnostik zweites und drittes Lebensjahr (MFED 2-3)

Die „Münchener funktionelle Entwicklungsdiagnostik – zweites und drittes Lebensjahr“ (Hellbrügge, 1994) versteht sich als ethologisches Verfahren. Die ethologische Diagnostik geht davon aus, dass sich der Entwicklungsverlauf in unterschiedliche Einheiten unterteilen lässt. Daher ist die MFED 2-3 ein Stufenleiterverfahren.

Das Verfahren erhebt sieben verschiedene Funktionsbereiche. Mit dem „Laufalter“ soll die Entwicklung der Körperbeherrschung erhoben werden. Das „Handgeschicklichkeitsalter“ dient der Erhebung der Feinmotorik. Das „Perzeptionsalter“ gibt die Entwicklung des Erfassens von Zusammenhängen an. Für die Sprache bestehen die zwei Funktionsbereiche „Sprechalter“ (aktive Sprache) und „Sprachverständnisalter“. Die letzten zwei Bereiche sind das „Sozialalter“ (soziales Verhalten) und das „Selbstständigkeitsalter“ (Esser & Petermann, 2010; Hellbrügge, 1994). Die Durchführungsdauer beträgt 40 bis 60 Minuten.

Der Test gibt keinen globalen Entwicklungsquotienten an, da die Ergebnisse der sieben Einzelbereiche so in einen wenig aussagekräftigen Wert zusammengefasst werden würden. Es sollen die einzelnen Bereiche betrachtet werden und eine Grundlage für die Therapie bieten. Zur Interpretation wird das Ergebnis des Kindes in den einzelnen

Funktionsbereichen mit dem 95. Perzentil als Toleranzgrenze verglichen und das Entwicklungsalter für jeden Bereich angegeben. Überschreitet das Kind die Toleranzgrenze, besteht der Verdacht einer Retardierung in diesem Bereich.

Kritik: Esser und Petermann (2010) beurteilen das Verfahren als übersichtlich und praktikabel. Es ermöglicht die Diagnose von Rückständen in mehreren Bereichen und erleichtert vor allem die darauf basierenden Interventionen. Das Manual beinhaltet hierfür Informationen für Maßnahmen.

Trotz der hohen Beliebtheit des Verfahrens weist es jedoch große Mängel auf. Die Aufgaben des Tests wurden von internationalen Entwicklungstests und Beobachtungen der Autoren abgeleitet (Esser & Petermann, 2010). Die Zuordnung der Aufgaben zu den einzelnen Bereichen basiert nicht auf statistischen Analysen (Reuner & Pietz, 2006). Die theoretische Fundierung und die Methodik sind demnach unzureichend und das verwendete Stufenkonzept ist veraltet (Esser & Petermann, 2010). Für das Verfahren gelten die Kritikpunkte der Stufenleiterverfahren (Kapitel 4.2.).

Auch den Gütekriterien werden nicht hinreichend entsprochen: Die Standardisierung ist unzureichend. Das Manual beinhaltet zu wenige konkrete Anweisungen wie wörtliche Anweisungen an das Kind. Die Reliabilität ist als fragwürdig und die Validierung als nicht ausreichend einzuschätzen. Zur Differenzierungsfähigkeit werden keine Angaben gemacht. (Macha et al, 2005; Reuner & Pietz, 2006). Es werden kaum Informationen zur Normierungsstichprobe gegeben. Angegeben wird, dass Daten von 85 unauffälligen Kindern mittels Elternbefragung erhoben wurden. Die Normen gelten als fehlend, da die Informationen aus den Befragungen keine angemessenen empirischen Daten für eine Normierung darstellen (Esser & Petermann, 2010).

4.3.2. Bayley Scales of Infant Development - Second Edition – deutsche Fassung (Bayley II oder BSID-II)

Die “Bayley Scales of Infant Development” in der zweiten Version wurden von Bayley 1993 veröffentlicht. Für das Verfahren wurde von Reuner et al. (2007) eine deutsche Fassung publiziert. Die Skalen sind als Stufenleiterverfahren konzipiert. Das Verfahren ist ein „Individualtest zur Untersuchung des Entwicklungsniveaus von Kindern zwischen einem und 42 Monaten“ (Reuner, et al., 2007, S.8). Es wird meist zur Diagnose von Entwicklungsverzögerungen und zur Planung einer Frühförderung verwendet. Die Bayley-Skalen basieren auf aktueller Forschung, daher liegt ihnen keine spezifische Theorie zugrunde.

Die drei Teile des Verfahrens sind die motorische Skala, die kognitive Skala und eine Skala zur Verhaltensbeurteilung. Die letztgenannte Skala ist nicht für die deutsche Fassung übersetzt worden. Die Bayley II beinhaltet demnach Aufgaben zur Erhebung des kognitiven, sprachlichen, persönlich-sozialen und des grob- und feinmotorischen Funktionsniveaus. Nach der Auswertung liegt für jedes Kind je ein Standardwert für die kognitive und die motorische Skala vor, der mit den Ergebnissen der Normstichprobe verglichen wird. Es können aber auch die Stärken und Schwächen durch die Analyse der einzelnen Facetten angegeben werden.

Kritik: Für die Bayley-Skalen treffen wiederum die Kritikpunkte der Stufenleiterverfahren zu, jedoch werden Standardwerte berechnet, damit die Leistung des Kindes anhand der Leistungsverteilung der Referenzstichprobe relativiert werden kann. Diese Vorgehensweise berücksichtigt die Variabilität der Entwicklungsverläufe besser als bei anderen Stufenleiterverfahren (Sarimski, 2009). Bei der Bayley II wird zusätzlich auch ein Entwicklungsalter angegeben, um das Gespräch mit den Eltern zu erleichtern (Sarimski, 2009).

Die deutsche Fassung der Bayley II ist ein sehr beliebtes Verfahren, das jedoch vor allem hinsichtlich der amerikanischen Normen eine begrenzte Aussagekraft hat (Esser & Petermann, 2010). Diese sind zudem als veraltet anzusehen. Die Normierung der englischen wie auch der deutschen Fassung der Bayley Scales of Infant Development – Second Edition basiert auf der randomisierten und stratifizierten Stichprobe von 1700 amerikanischen Kindern im Jahr 1993 (Reuner et al., 2007). Die Skalen müssen daher im deutschsprachigen Raum sehr vorsichtig interpretiert werden (Frey, 2012).

Auch die Reliabilitäten wurden basierend auf den amerikanischen Stichproben berechnet. Insgesamt sind alle Reliabilitätswerte als akzeptabel, aber nicht besonders hoch einzuschätzen. Esser und Petermann (2010) weisen jedoch darauf hin, dass die Reliabilitäten überschätzt werden, da bei den Berechnungen alle nicht getesteten Aufgaben aus jüngeren Altersbereichen als gelöst angenommen werden.

Es liegen ebenfalls viele Studien mit guten Werten zur Validitätsbestimmung vor (Frey, 2012). Die Inhaltsvalidität wird angenommen, da das Verfahren basierend auf Expertenbefragungen konstruiert und anschließend beurteilt wurde. Für die Konstruktvalidität spricht, dass die Items mit der ihnen zugeordneten Skala höher korrelieren als mit der anderen Skala. Korrelationen zwischen der kognitiven Skala und anderen Intelligenztests liegen zwischen .73 und .79. Korrelationsstudien mit anderen Entwicklungs- und Leistungstests sprechen ebenfalls für eine gute Validität. Auch die Differenzierungsfähigkeit hinsichtlich verschiedener Risikogruppen bestätigt die Validität. Die genauen Angaben im Manual sprechen für die Durchführungs- und Auswertungsobjektivität.

Das Material ist attraktiv und motiviert die Kinder zum Mitmachen (Sarimski, 2009). Das Verfahren deckt eine große Bandbreite von Entwicklungsbereichen ab, was jedoch zu einer langen Durchführungszeit führen kann. Esser und Petermann (2010) empfehlen, dass das Verfahren nur von erfahrenen Untersuchungspersonen eingesetzt werden sollte, wenn viel Zeit in eine ausführliche standardisierte Beobachtung investiert werden kann.

Sarimski (2009) rät davon ab, die Bayley-Skalen zur Identifikation umschriebener Entwicklungsstörungen zu verwenden. Es liegen keine Informationen vor, wann Unterschiede zwischen den Ergebnissen in sprachlichen und nichtsprachlichen Aufgaben bzw. wann Unterschiede in den Ergebnissen der einzelnen Facetten zufällige Schwankungen oder schwerwiegende Profildiskrepanzen abbilden. Ein weiterer Kritikpunkt besteht darin, dass ca. 30 % der Items der kognitiven Skala sprachliche Aufgaben sind. Dies kann zur Unterschätzung der kognitiven Fähigkeiten von zweijährigen oder älteren Kindern mit sprachlichen Defiziten kommen. Bei Kindern mit sensorischen oder körperlichen Behinderungen ist eine standardisierte Durchführung nicht möglich (Sarimski, 2009).

4.3.3. Bayley Scales of Infant and Toddler Development – Third Edition

Die „Bayley Scales of Infant and Toddler Development - Third Edition“, die von Bayley 2006 zuletzt veröffentlicht wurden, sollen hier nur kurz dargestellt werden, da keine deutsche Version hierfür vorliegt, sie aber dennoch eine wichtige Weiterentwicklung der Bayley Scales darstellen. Mit dem Verfahren soll der klinische Nutzen erhöht, die Administration erleichtert und die psychometrische Qualität der BSID verbessert werden. Das Verfahren besteht in der dritten Version aus fünf Skalen. Neben der sprachlichen und kognitiven wird die motorische Skala erhoben. Mittels Befragung der Eltern werden Informationen zur sozial-emotionalen Skala und der Verhaltensskala erhoben (Bayley, 2006).

Kritik: Die Ziele der Überarbeitung der Bayley-Skalen wurden alle größtenteils erreicht. Alle Gütekriterien liegen über dem kritischen Wert, wobei die meisten als hoch einzuschätzen sind. Die Reliabilität konnte für jüngere Kinder erhöht werden (Albers & Grieve, 2007). Es liegen zum Beispiel Analysen zur inneren Konsistenz des Verfahrens vor, die mit Werten für die verschiedenen Altersgruppen über die Subtests zwischen .71 bis .98 als akzeptabel bis hoch einzuschätzen sind.

Die Inhaltsvalidität wird ebenfalls als gegeben angesehen, da eine große Bandbreite an Fähigkeiten abgedeckt wird und die Items basierend auf aktueller Literatur und mittels Expertenreviews aktualisiert wurden. Mittels Faktorenanalyse konnte das Drei-Faktoren-Modell (kognitive, sprachliche und motorische Skala) bestätigt werden. Die Kriteriumsvalidität wird aufgrund von Korrelationsstudien mit vielen anderen Verfahren wie z.B.: BSID II, WPPSI III, PLS-4 als angemessen erfüllt angesehen. Viele Studien liegen zur Differenzierungsfähigkeit vor (Bayley, 2006). Das Ziel der neuen Normierung wurde ebenfalls erreicht. Diese wurde 2004 in Amerika vorgenommen und beinhaltet stratifizierte Daten von 1700 Kindern aus 17 Altersgruppen. Die Normierungsdaten wurden mit Daten von Kindern mit Entwicklungsdefiziten und besonders guten Vorschulkindern erweitert, um Decken- und Bodeneffekten zu entgehen (Bayley, 2006). Ein Ziel, das nicht erreicht werden konnte, betrifft die Ökonomie: Für das Verfahren sind aufwändige Schulungen notwendig und die Durchführungsdauer beträgt 50 – 90 Minuten (Frey, 2012).

Auf Grund der Neuerungen und Vorteile ist eine deutsche Version mit aktuellen Normen für deutschsprachige Gebiete wünschenswert (Frey, 2012).

4.3.4. Entwicklungstest sechs Monate bis sechs Jahre (ET 6-6)

Der „Entwicklungstest sechs Monate bis sechs Jahre“ wurde 2008 in seiner dritten Auflage von Petermann, Stein und Macha veröffentlicht. Er ist ein „Inventar kriteriumsorientierter Entwicklungsdiagnostik mit Altersnormen für den Altersbereich vom Säuglingsalter bis zum Alter der Schulreife“ (Sarimski, 2001, S.107). Das Inventar distanziert sich von traditionellen Stufenleiterverfahren und berücksichtigt die Variabilität der Entwicklung. Die Autoren gehen von Entwicklungsbahnen aus, die sich zunächst unabhängig voneinander entwickeln. Mehrere dieser Bahnen sind notwendig, um eine Fähigkeit zu erlangen. Diese Bahnen lassen sich nicht direkt erheben. Obwohl die Entwicklung variabel verläuft, müssen Kinder zu bestimmten Zeitpunkten bestimmte Fähigkeiten erreicht haben.

Kindern im Alter zwischen sechs Monaten und sechs Jahren werden ihrer jeweiligen Altersgruppe entsprechende Aufgaben aus sechs Bereichen vorgegeben. Diese sind Körpermotorik, Handmotorik, kognitive Entwicklung, Sprachentwicklung, Sozialentwicklung und emotionale Entwicklung. Zusätzlich wird noch der Subtest Nachzeichen vorgegeben. Die individuelle Reihenfolge der Aufgaben soll gewährleistet werden. Auch werden Informationen mittels Elternfragebogen eingeholt. Die Ergebnisse werden durch ein Entwicklungsprofil dargestellt, wobei die Summenwerte in standardisierte Testwerte umgewandelt werden. Die Ergebnisse des Profils werden mit den durchschnittlichen Testleistungen Gleichaltriger und der Streuung dieser Werte in Bezug gesetzt, um dann eine Einschätzung vornehmen zu können, ob das Kind „unauffällig“ ist, im „Risikobereich“ liegt oder „gravierende Entwicklungsdefizite“ aufweist.

Kritik: Der ET 6-6 ist aufgrund seines innovativen Charakters in den vergangenen Jahren zu einem sehr beliebten Verfahren geworden. Der Entwicklungstest basiert auf einem modernen Entwicklungsbegriff und beinhaltet Aufgaben für einen weiten Altersbereich (Esser & Petermann, 2010). Er eignet sich besonders dafür, den Verlauf der Entwicklung eines Kindes abzubilden, da versucht wird, über das Alter hinweg möglichst einheitliche Dimensionen zu erfassen (Sarimski, 2001). Das Verfahren berücksichtigt außerdem die Besonderheiten, die sich bei der Testung von Kindern in unterschiedlichen Altersgruppen ergeben. Das Material ist besonders attraktiv für die Kinder und der Ablauf ist abwechslungsreich und individuell für jedes Kind zu

gestalten. Dies motiviert auch deutlich beeinträchtigte Kinder zum Mitmachen. Eine Schwäche stellt jedoch die Unübersichtlichkeit des Tests dar (Esser & Petermann, 2010; Petermann et al., 2008).

Bezüglich der Gütekriterien ergibt sich folgendes Bild: Es werden zwar keine Angaben zur Objektivität im Manual gemacht, jedoch ist das Verfahren aufgrund seiner eindeutigen Anweisungen als objektiv zu bewerten. Es ist außerdem gut validiert. Dafür sprechen erste Validierungsversuche bezüglich der Differenzierungsfähigkeit und der Skalenstruktur (Esser & Petermann, 2010). Außerdem konnte durch Längsschnittstudien gezeigt werden, dass der ET 6-6 die Variabilität der Entwicklung im Normalbereich berücksichtigt (Esser & Petermann, 2010). Bezüglich der Stabilität bestätigten Lissmann, Domsch und Lohaus (2006) signifikante Beziehungen zwischen zwei Testzeitpunkten, die nicht mehr als sechs Monate auseinanderliegen. Für längere Zeiträume lassen sich Leistungen von jungen Kindern nicht vorhersagen. Die Normierung wurden 1999/2000 vorgenommen und gilt daher als veraltet. Die Repräsentativität der Normen ist eingeschränkt. Vor allem im jungen Alter sind die Normen nur wenig differenziert (Reuner & Pietz, 2006). Die Erweiterung der Normstichprobe ist daher von Nöten (Lücking & Scheithauer, 2006). Durch das leicht handhabbare Erhebungs- und Auswertungsblatt für jede Altersgruppe und die kurze Testdurchführungsdauer zwischen zehn und sechzig Minuten gestaltet sich der Test als sehr ökonomisch (Petermann & Esser, 2010).

Aufgrund der Berücksichtigung von qualitativen Entwicklungsveränderungen und der Konzeption als Inventar, wurde jede Dimension bezüglich heterogener Aspekte erfasst und interpretiert. Da aufgrund dieser Konzeption keine Angaben über die Messgenauigkeit gemacht werden, kann das Verfahren nicht als Entwicklungstest im engeren Sinne verstanden werden (Kastner-Koller & Deimann, 2011).

Das Verfahren eignet sich vor allem für Fragen der Förderdiagnostik, da es im unteren Leistungsbereich gut differenziert. Im oberen Leistungsbereich zeigen sich jedoch Deckeneffekte (Esser & Petermann, 2010). Es empfiehlt sich, Kinder mit schweren Störungen nicht mit diesem Verfahren zu testen, da nicht eindeutig festgelegt ist, welche Aufgaben ihnen vorgegeben werden sollen, und Informationen über die erreichte Entwicklungsstufe bzw. Anhaltspunkte für die Interventionsplanung fehlen (Sarimski, 2001).

4.4. Zusammenfassende Bewertung: Relevanz der Entwicklung eines neuen Verfahrens

Entwicklungstests müssen –um den qualitativen Anforderungen gerecht zu werden–den Hauptgütekriterien Objektivität, Reliabilität und Validität sowie den Nebengütekriterien Normierung, Ökonomie und Nützlichkeit entsprechen.

Viele Verfahren (Bayley-Scales, MFED) sind als Stufenleiterverfahren konstruiert, die von einem veralteten Stufenmodell der Entwicklung ausgehen und die Variabilität der Entwicklung nicht berücksichtigen. Testbatterien und Inventare bieten Alternativen hierfür.

Wie gezeigt werden konnte, stehen nur wenige Verfahren zur allgemeinen Entwicklungsdiagnostik im dritten Lebensjahr zur Verfügung. Häufig verwendete Verfahren für den deutschen Sprachraum, die auch Aufgaben für das dritte Lebensjahr beinhalten, sind die „Münchener funktionelle Entwicklungsdiagnostik – zweites und drittes Lebensjahr“ (Hellbrügge, 1994), der Bayley Scales of Infant Development – Second Edition – deutsche Fassung (Reuner, et al, 2007) und der „Entwicklungstest sechs Monate bis sechs Jahre“ (ET 6-6; Petermann, Stein & Macha, 2004).

Diese Verfahren erfüllen oft die Gütekriterien nicht oder die Gütekriterien wurden nicht genügend untersucht. So geben die Testautoren des ET 6-6 aufgrund der Konstruktionsweise des ET als Inventar keine Informationen zur Reliabilität an. Alle beschriebenen deutschsprachigen Verfahren haben veraltete Normen (MFED 2-3, Bayley II; ET 6-6). Nur der englischsprachige Bayley III hat noch aktuelle Normen. Der Bayley II weist zudem amerikanische Normen auf, die für den Einsatz in deutschsprachigen Gebieten nicht repräsentativ sind. Der ET 6-6 wurde mittels Erhebungen in Deutschland normiert. Für den Gebrauch in Österreich ist selbst bezüglich deutscher Normierungsstichproben die Repräsentativität fraglich.

Gefordert ist daher die Entwicklung von Verfahren für die Diagnostik von Kindern in den ersten Lebensjahren, die den Gütekriterien entsprechen (Renner 2009). Neue Verfahren soll sich an einem möglichst modernen Entwicklungsbegriff orientieren und sich aufgrund der Kritik an Stufenleiterverfahren von diesen abheben. Des Weiteren ist deutlich geworden, dass die einzelnen Entwicklungsdimensionen auch getrennt voneinander betrachtet werden müssen.

5. Spielbasierte Verfahren

Aufgrund der Unzufriedenheit und der Kritik an standardisierten Tests wurde ein neuer Ansatz entwickelt, der die Entwicklung der Kleinkinder in einer sehr natürlichen Spielsituation erheben möchte. Diese spielbasierten Verfahren sollen nun dargestellt und ihre Stärken und Schwächen den standardisierten Tests gegenübergestellt werden.

5.1. Definition spielbasierter Verfahren

Um die Definition spielbasierter Verfahren zu ermöglichen, muss zuvor der zugrunde liegende Begriff „Spiel“ definiert werden. Unter dem Begriff „Spiel“ werden viele verschiedene Aktivitäten wie das Manipulieren von Objekten durch den Säugling oder Rollen-, Sport- und Regelspiele (z.B. Schach) summiert. Deshalb ist eine einheitliche Definition von Spiel schwierig (Sturzbecher, 2001).

Oerter (1999) gibt daher drei handlungsorientierte Merkmale des Spiels an, um Spiel zu definieren. Das erste Merkmal ist die „*Handlung um der Handlung willen*“. Demnach ist Spiel eine besondere Art der Handlung, die gekennzeichnet ist von Zweckfreiheit und intrinsischer Motivation. Das Spiel wird um seiner selbst willen betrieben. Es werden keine weiteren Folgen damit angestrebt, denn die Tätigkeit an sich ist belohnend. Das zweite Merkmal ist der „*Wechsel des Realitätsbezugs*“. Spielende konstruieren eine völlig neue Realität, die ihren Bedürfnissen und Zielen entspricht. Das dritte Merkmal lautet „*Wiederholung und Ritual*“. Wiederholung beobachtet man nicht nur bei Regelspielen wie Schach und Kartenspielen, sondern auch beim Funktionspiel von Säuglingen, die einzelne Handlungen und Bewegungen ständig wiederholen. Wiederholen kann aber auch als Variation von Handlungsschemen angesehen werden bzw. als das Verarbeiten von Erlebtem im Spiel. Schon ab dem zweiten Lebensjahr zeigen Kinder des Weiteren eine Vorliebe für Rituale, die einer festgelegten Handlungsreihenfolge entsprechen. Beim Zubettgehen muss zuerst das Kuscheltier in die Arme gelegt werden, dann müssen eine Gutenachtgeschichte und der Gutenachtkuss folgen. Daher ist das Spiel einerseits durch Handlungsfreiheit bestimmt, andererseits bestehen bestimmten Regeln, die den Ritualen entspringen.

Während sich diese Konzeption von Spiel auf die Handlung konzentriert, stehen bei anderen Definitionen der soziale oder kognitive Aspekt im Vordergrund (Lifter, Mason & Barton, 2011). Für die spielbasierte Diagnostik ist laut Lifter et al. (2011) die

Verwendung von Objekten der zentrale Aspekt der Definition von Spiel. Spiel besteht daher aus spontanen und natürlich auftretenden Aktivitäten mit Objekten, die das Interesse oder die Aufmerksamkeit erregen, unabhängig davon, wer noch mitspielt und welche Gefühle oder Vorstellungen damit verbunden sind. Es wird also deutlich, dass viele verschiedene Konzepte von Spiel vorhanden sind.

Sturzbecher (2001) versteht nun unter spielbasierten Verfahren, dass Spiele als Handlungsrahmen bzw. Medium dienen, um diagnostische Intentionen zu realisieren. In einer ähnlichen Definition von Kelly-Vance und Ryalls (2005) stellen die Autorinnen den spielbasierten Ansatz als eine Beobachtung der Fähigkeiten eines Kindes im Rahmen eines Spieles dar. Von professionellen Anwendern/Anwenderinnen des Verfahrens werden die Fähigkeiten des Kindes in verschiedenen Bereichen dahingehend beurteilt, ob sie signifikant von der typischen Entwicklung Gleichaltriger abweichen.

5.2. Erste spielbasierte Verfahren

Schon früh bestand die Absicht, das Spiel zu deuten, um Rückschlüsse auf Fähigkeiten, Bedürfnisse, Ängste und Sichtweisen des Kindes zu ziehen. Die Theorie von Freud (1909) galt als Ausgangspunkt der psychoanalytischen Deutung des Spiels und der Spielergebnisse (Witzlack, 2001).

Klein gilt als die Begründerin der Spieldiagnostik. Ihre Spielanalyse wurde 1976 unter dem Namen „Analytische Spieltechnik“ veröffentlicht. Sie stellte dem Kind bestimmte Spielutensilien wie Tiere, Häuser, Fahrzeuge und Papier in einem Spielzimmer zur Verfügung und analysierte die verwendeten Materialien als Symbole der Triebbefriedigung. Das Verfahren war jedoch nicht standardisiert. Ihrer Ansicht nach stellt das Kind im Spiel verschiedene Erfahrungen nach (nach Witzlack, 2001).

Lowenfeld veröffentlichte zeitgleich 1976 die „Welt-Technik“ und beeinflusste damit die Spieldiagnostik nachhaltig. Sie distanzierte sich von der psychoanalytischen Deutung des Spiels. In ihrem Ansatz sollten die Kinder mit Spielsachen wie Häusern, Menschen und Plastilin eine Welt aufbauen. Im anschließenden gemeinsamen Gespräch mit dem Kind sollte diese Welt gedeutet werden. Lowenfeld war der Überzeugung, dass man durch den Weltenbau nicht nur Informationen über Eigenschaften und Entwicklungsdimensionen, sondern vielmehr Einblicke in die Daseinsauffassung des Kindes erhält. Lowenfeld sah ihre Technik als Methode der Therapie und weniger als

diagnostischen Ansatz, dennoch beeinflusste ihre Vorgehensweise die Entwicklung vieler diagnostischer Verfahren. So war vor allem Bühler darum bemüht, das Verfahren zu standardisieren und veröffentlichte eine modifizierte Variante.

Aus diesem Verfahren und unter Einfluss von Methoden, die mit der Verwendung von Handpuppen arbeiteten, wurde wiederum der „Scenotest“ von v. Staabs (1951) entwickelt, der noch heute weltweite Verbreitung findet (Witzlack, 2001). Der Test soll dem Kind ermöglichen sein Erleben in einer Miniaturwelt darzustellen, um der untersuchenden Person, die mit psychoanalytischer Deutung arbeitet, Informationen über eventuelle unterbewusste Probleme und Konflikte zu eröffnen (Dietrich, 2001). Da sich auch dieses Verfahren nicht auf die Erhebung von Entwicklung bezieht, wird es in dieser Arbeit nicht näher beschrieben.

Mit dem Ziel, Entwicklung zu diagnostizieren, entwickelte Witzlack 1988 ein spielbasiertes Verfahren, das als Einschulungstest in der DDR dienen sollte. Er kritisierte, dass die traditionelle Diagnostik Basiseigenschaften erhebt, die situationsinvariant und unabhängig davon sind, welche Anforderungen tatsächlich an das Kind gestellt werden. Somit sind Aussagen über Entwicklungsbeeinträchtigungen und Förderungen nur schwer abzuleiten. Im Rahmen der Reformierung der Einschulungsdiagnostik, die die Analyse des Spiels als wichtigste Methode erachtete, entwickelte Witzlack 1988 das standardisierte Rollenspiel „Kaufmannsladen“. Um die Bedingungen für alle Kinder gleich zu gestalten, standardisierte er das Verfahren in dem er den Rollenspielinhalt (Kaufmannsladen), die Spielrolle des Kindes und des Spielpartners/der Spielpartnerin (Verkäufer/Verkäuferin oder Käufer/Käuferin), sowie die Spielgegenstände festlegte und das Spiel in verschiedene Handlungssequenzen, in denen dem Kind nur bestimmte Materialien vorgegeben wurden, gliederte. Die Instruktionen sind ebenfalls für alle Kinder gleich. Erhoben werden können zum Beispiel numerische Fähigkeiten (Gib mir zwei Bananen.), praktisches Problemlöseverhalten, konstruktive Fähigkeiten und Konfliktverhalten (z.B. beim Aufbau der Regale, beim Öffnen der Verriegelung der Boxen (nach Witzlack, 2001).

Wie bereits angedeutet wurde, entstand die spielbasierte Diagnostik der Entwicklung aus der Unzufriedenheit mit den traditionellen Tests. Die beiden anderen treibenden Kräfte waren Studien, die das Spiel als Zugang zur Erhebung der Entwicklung belegten, und das „Transdisciplinary Play-Based Assessment“ von Linder (1993; nach Farmer-

Dougan & Kaszuba, 1999). Dieses Verfahren wird daher im folgenden Kapitel beschrieben.

5.3. Transdisciplinary Play-Based Assessment

Es existieren nur sehr wenige Verfahren, die der spielbasierten Diagnostik zuzurechnen sind. Als eine der drei gängigsten Methoden soll das „Transdisciplinary Play-Based Assessment“ (TPBA; Linder, 1993) als Beispiel für spielbasierte Verfahren dienen (Kelly-Vance & Ryalls, 2008). Es handelt sich um ein Verfahren für Kinder von der Geburt bis zu sechs Jahren, wobei die Kodierungsrichtlinien nur für Kinder bis 72 Monate vorliegen (nach Kelly-Vance & Ryalls, 2008; Vig & Sanders, 2007). Das Verfahren erhebt die kognitiven, sprachlichen, sowie fein- und grobmotorischen Fähigkeiten und das Sozialverhalten von Kindern durch die Beobachtung von Spiel. Neben der Kodierung von unterschiedlichen Spielarten werden im kognitiven Bereich unter anderem die Fähigkeiten Imitation, Problemlösen, Klassifikation, Eins-zu-Eins Zuordnung usw. erhoben.

Vor den Beobachtungen werden die Eltern über ihre Bedenken bezüglich der Entwicklung ihres Kindes befragt und diese Informationen bestimmen den zentralen Fokus der Beobachtung. Zuerst wird in etwa 30 Minuten lang das freie, unstrukturierte Spiel des Kindes beobachtet. Danach wird das Kind von einem Erwachsenen dazu hingeleitet, noch nicht präsentierte Verhaltensweisen (z.B.: Zeichnen) zu zeigen. Später soll das Kind auch noch mit einem zweiten Kind spielen und mit einem Elternteil interagieren.

Insgesamt dauert die Beobachtung zwischen 60 und 90 Minuten. Beobachtet wird das Spiel von einem interdisziplinären Team, das sich über die Beobachtungen austauscht. Die Ergebnisse in den einzelnen Kategorien und den Unterkategorien beschreiben die Stärken des Kindes und die Bereiche in denen das Kind Intervention benötigt. Welche Interventionen passend sind, wird ebenfalls in den Richtlinien von Linder (1993) beschrieben (nach Kelly-Vance, Ryalls & Glover, 2002; Kelly-Vance & Ryalls, 2008).

5.4. Anwendungsmöglichkeiten spielbasierter Verfahren

Während einige Autoren/Autorinnen spielbasierte Verfahren als dem Entwicklungstest überlegen ansehen und den substituierenden Einsatz dieses alternativen Ansatzes vertreten (z.B. Witzlack, 2001), fordern andere wie Vig und Sanders (2007) eine Kombination der beiden Verfahrensarten.

Die Vorgabe von standardisierten Tests ist im Kleinkindalter mit vielen Schwierigkeiten und Störquellen (z.B. geringe Motivation und Aufmerksamkeitspanne der Kinder, siehe Kapitel 3.1. Das Kind als Klient) verbunden. Laut Vig und Sanders (2007) darf sich daher die Erhebung der kognitiven Entwicklung nie ausschließlich auf standardisierte Entwicklungstests stützen, sondern muss immer durch andere Ansätze ergänzt werden. Einer dieser Ansätze ist der spielbasierte Ansatz.

Die Beobachtungen spielbasierter Verfahren können – wenn sie zusätzlich zu standardisierten Entwicklungstests vorgegeben werden – Informationen über die Validität des Tests oder über beobachtbare Verhaltensweisen wie Aufmerksamkeitsfokussierung und Organisationsfähigkeit zur Zielerreichung geben.

Die Beobachtung des Kindes in einer Situation, die es selbst strukturieren kann, im Vergleich zu einer von außen strukturierten Situation (z.B. Testsituation) kann differenzierte Hinweise für das Vorliegen von Entwicklungsproblemen bzw. Informationen für die Interventionsplanung liefern. Dafür spricht, dass in mehreren Studien unterschiedliche Charakteristiken des Spiels für Kinder mit bestimmten Entwicklungsproblemen festgestellt werden konnten. Viele Kinder sprechen in der Spielsituation viel freier und viel mehr, was für eine möglichst natürliche Einschätzung der verbalen Fähigkeiten von Vorteil ist (Vig & Sanders, 2007).

Sollte ein Kind mit einem standardisierten Entwicklungstest nicht getestet werden können, da es die Bearbeitung sehr vieler Aufgaben verweigert oder aufgrund von starken Entwicklungsverzögerungen oder Behinderungen nicht zur Bearbeitung in der Lage ist (z.B. blinde Kinder, Bagnato & Neisworth, 1994), kann mit spielbasierten Verfahren eine Einschätzung der kognitiven Entwicklung vorgenommen werden (Vig & Sanders, 2007).

5.5. Gegenüberstellung spielbasierter Ansätze und allgemeiner Entwicklungstests

Spielbasierte Ansätze sollen standardisierte Verfahren ersetzen oder ergänzen. Sie weisen hierfür zahlreiche *Vorteile* auf:

Spielbasierte Verfahren distanzieren sich von der prüfungsähnlichen, hochstrukturierten Situation standardisierter Entwicklungstests, in der sich Kleinkinder oft nur schwer konzentrieren können und motivieren lassen (Esser & Petermann, 2010; Kelly-Vance et al., 2002; Witzlack, 2001). Kinder in diesem Alter spielen sehr gerne und verbringen mehrere Stunden pro Tag damit. Spielbasierte Verfahren erheben daher die Entwicklung des Kindes in einer für das Kind natürlichen, spaßigen Situation (Kelly-Vance et al., 2002).

Während bei Entwicklungstests, die nach dem Stufenleiterverfahren ihre Aufgaben nach der Schwierigkeit ordnen, die Vorgabe abgebrochen wird, wenn das Kind mehrere aufeinander folgende Aufgaben nicht lösen kann, schließt das Kind bei spielbasierten Ansätzen nicht mit einem Misserfolgsereignis ab (Witzlack, 2001). Als Vorteil des spielbasierten Verfahrens kann auch das völlig adaptive und flexible Vorgehen genannt werden (Kelly-Vance et al., 2002; Witzlack, 2001). Außerdem können alle Kinder damit getestet werden – unabhängig davon, ob sie starke Entwicklungsverzögerungen oder Behinderungen haben (Bagnato, 2005).

Witzlack (2001) betont, dass man einerseits mit Rollenspielen die gleichen Informationen gewinnen kann wie mit Entwicklungstests, andererseits aufgrund des großen Handlungsspielraums des Kindes sogar mehr Informationen erhält. Dem Kind wird im Gegensatz zu Entwicklungstests, die eine eingeschränkte Auswahl an Aufgaben und Aufgabenschwierigkeiten haben, die Möglichkeit geboten, seine optimalen Leistungen zu demonstrieren. Außerdem liefern spielbasierte Verfahren auch Informationen über Stärken des Kindes und Informationen, die für die Interventionsplanung und -überwachung relevant sind. Diese Informationen fehlen häufig bei standardisierten Verfahren (Kelly-Vance et al., 2002).

Doch neben den zahlreichen Vorteilen zeigen die spielbasierten Ansätze auch einige *Mängel*: Eine deutliche Schwäche ist die unzureichende Untersuchung des Nutzens und

den Gütekriterien der Verfahren (Farmer-Dougan & Kaszuba, 1999; Kelly-Vance et al., 2002; Kelly-Vance & Ryalls, 2008).

Die Standardisierung der meisten Verfahren ist als unzulänglich zu beurteilen, es liegt keine Normierung der Verfahren vor und zu Reliabilität und Validität bestehen für die meisten Verfahren kaum Studien (Farmer-Dougan & Kaszuba, 1999). Der Versuch von Kelly-Vance und Ryalls (2005), ein spielbasiertes Verfahren zu standardisieren, stellte die Beobachtung einer völlig unstrukturierten Spielsituation ohne die Teilnahme eines Erwachsenen dar. Die Autorinnen konnten jedoch feststellen, dass die Kinder in dieser unstrukturierten Spielsituation nicht die gesamte Bandbreite ihrer Fähigkeiten zeigten und die Möglichkeit fehlte, die Kinder zu motivieren, ein bestimmtes, für die Kodierung relevantes Verhalten zu zeigen (z.B. Klassifikationen, numerische Fähigkeiten). Zur Standardisierung der Strukturierung des Spiels durch Erwachsene bestehen noch keine Studien (Kelly-Vance & Ryalls, 2005).

Erste Studien zur Validität und Reliabilität zeigen vielversprechende Ergebnisse: Einige Forschungsbemühungen erheben den Zusammenhang zwischen spielbasierten Verfahren und standardisierten Entwicklungstests. So zeigt eine Studie, dass Vorschulkinder mit einem niedrigeren Niveau im Spiel sowohl geringere Werte bei einem standardisierten Verfahren zur Erhebung der kognitiven Fähigkeiten, als auch bei einem Verfahren der sozialen Fähigkeiten erreichten bzw. von ihren Pädagogen/Pädagoginnen diesbezüglich schlechter eingestuft wurden. Kinder, die komplexes Spiel zeigten, erreichten in den standardisierten Verfahren bzw. in der Bewertung durch Pädagogen/Pädagoginnen dagegen hohe Werte (Farmer-Dougan & Kaszuba, 1999). Die Interpretation dieser Ergebnisse ist durch die geringe, wenig repräsentative Stichprobenanzahl nur eingeschränkt möglich. Die Befragung der Pädagogen/Pädagoginnen ist außerdem eine sehr subjektive Methode. Eine andere Studie spricht für die Kriteriumsvalidität des TPBA, da das Verfahren mit den Bayley-Scales in der zweiten Version signifikant hoch korreliert (Kelly-Vance, Needelman, Troia & Ryalls, 1999; nach Kelly-Vance et al., 2002).

Viele Studien bestätigen zudem, dass sich Kinder mit Entwicklungsstörungen (z.B. Autismus) in ihrem Spielverhalten von Kindern ohne Entwicklungsstörungen unterscheiden (Lifter et al., 2011). Eine erste Studie von Kelly-Vance et al. (2002) spricht ebenso für die Validität der Kodierung von Spielarten. Bei Testungen von 16

zwei- oder dreijährigen Kindern mit einem Codierungssystem, das eine Modifizierung des Kodierungssystems von Linders TPBA (1993) darstellt, stellte sich heraus, dass ältere Kinder während freier Spielsituationen komplexere Spielformen zeigten als jüngere Kinder. Dieselbe Studie zeigte, dass eine hohe Inter-Rater-Reliabilität erreicht wurde. Die Aussagekraft dieser Studie ist durch die kleine und wenig repräsentative Stichprobe jedoch eingeschränkt. In einer Nachfolgerstudie von Kelly-Vance und Ryalls (2005) wurde das unstrukturierte Spiel von Vorschulkindern mithilfe eines ähnlichen Kodierungssystems beobachtet. Auch hier konnte eine hohe Inter-Rater-Reliabilität festgestellt werden. Anhand einer Testwiederholung drei Wochen nach der ursprünglichen Testung konnte eine moderate Stabilität beobachtet werden.

Problematisch ist, dass die meisten Studien unterschiedliche Vorgehensweisen und Kodierungssysteme verwenden. Die zu kodierenden Stufen von Spiel unterscheiden sich in den Studien, die Definitionen der Spielarten variieren und das Ausmaß an unstrukturiertem Spiel bzw. Interaktion ist unterschiedlich. Daher können keine generellen Aussagen über den Nutzen von spielbasierten Verfahren getroffen werden (Kelly-Vance et al., 2005).

Ein weiterer Kritikpunkt an den spielbasierten Verfahren, der im Zusammenhang mit dieser Arbeit steht, betrifft die Tatsache, dass die meisten Verfahren für das Vorschulalter konzipiert wurden bzw. wenige empirische Daten für die Verwendung spielbasierter Ansätze für Kinder im dritten Lebensjahr vorliegen. Außerdem stammen die meisten entwickelten Verfahren aus dem amerikanischen Raum und ob die Verwendung im deutschsprachigen Raum angemessen ist, ist fraglich.

5.6. Zusammenfassende Bewertung der Bedeutung spielbasierter Verfahren

Spielbasierte Verfahren stellen die Beobachtung der Fähigkeiten von Kindern in einem Spiel dar. Während erste Ansätze unter Spieldiagnostik die psychoanalytische Deutung des Spiels hinsichtlich möglicher Konflikte und Probleme verstanden, so liegen nun einige spielbasierte Verfahren zur Diagnostik der Entwicklung vor, die sich aus der Unzufriedenheit mit den vorhandenen Entwicklungstests entwickelten. Diese Ansätze werden entweder als alleinige Erhebungsmethode eingesetzt oder sollen die Erhebung der Entwicklung mit Entwicklungstests ergänzen.

Im Gegensatz zu der prüfungsähnlichen Situation von Tests erheben die spielbasierten Verfahren die Entwicklung des Kindes in einer natürlichen Situation, in der es viel Spaß hat. So können viele zusätzliche Informationen über Fähigkeiten des Kindes gewonnen werden.

Standardisierte Tests sind den spielbasierten Ansätzen hinsichtlich der Prüfung der Gütekriterien überlegen. Es bestehen wenige Studien über den Nutzen und die Einhaltung der Gütekriterien der spielbasierten Ansätze und eine generelle Aussage über die psychometrischen und praktischen Qualitäten kann daher noch nicht getroffen werden. Jedoch zeigen erste Studien, dass es sich um einen vielversprechenden Ansatz für die Diagnostik junger Kinder handelt.

Eine Kombination der Vorteile des spielbasierten Ansatzes und der Entwicklungstests unter gleichzeitiger Vermeidung der Schwachpunkte waren die grundlegende Idee der Entwicklung des spielbasierten Itempools, der im empirischen Teil beschrieben wird. So sollen eine standardisierte Vorgabe von Aufgaben, die in das Spiel eingebettet sind, und die Beobachtung vieler Fähigkeiten während des Spiels eine Erfüllung der Gütekriterien ermöglichen.

6. Kognitive Entwicklung im dritten Lebensjahr

Da der Schwerpunkt dieser Arbeit bei der kognitiven Entwicklung liegt, werden nach einer grundlegenden Definition des Begriffs „Kognition“ Studienergebnisse zur Entwicklung kognitiver Fähigkeiten im dritten Lebensjahr beschrieben. Danach wird die Frage beantwortet, warum die kognitive Entwicklung bei Kleinkindern erhoben wird und wie die kognitiven Fähigkeiten im dritten Lebensjahr in anderen Entwicklungs- oder Intelligenztests erfasst werden.

6.1. Definition Kognition

Die „Kognition“ umfasst die gesamte Bandbreite psychologischer Prozesse, die mentalen Ereignissen unterliegen, wie die Wahrnehmung von Informationen, das Verstehen, das Gedächtnis und die Intelligenz (Solso, 2008).

Obwohl es keine einheitliche Definition des Intelligenzbegriffs gibt, meint Solso (2008), dass alle höheren Formen der Kognition unter dem Begriff der „Intelligenz“ zusammengefasst werden können. Rindermann und Kwiatkowski (2010) definieren Intelligenz als „die Fähigkeit zum denkgestützten Lösen von Aufgaben und Problemen in Situationen, die für die Person neu sind, die Fähigkeit zum induktiv und deduktiv schlussfolgernden Denken, die Fähigkeit zum abstrakten Denken und die Fähigkeit zu Verständnis und Einsicht – zum Erkennen und zur Herstellung von Strukturen, Beziehungen, Sinnzusammenhängen und Bedeutungen“ (S.102 f.).

Da der Begriff Intelligenz für die Diagnostik kognitiver Fähigkeiten im dritten Lebensjahr nicht angemessen ist, wird die Bezeichnung „kognitive Entwicklung“ verwendet (siehe Kapitel 2.5. Begriffsabgrenzung zu Intelligenzdiagnostik).

Der spielbasierte Itempool von Kuchler, Sapper, Deimann und Kastner-Koller (2011) beinhaltet Aufgaben zur „Theory of Mind“, zum „numerischen Wissen“ und zum „deduktiven Denken“. Im folgenden Kapitel werden diese Bereiche mit Studienergebnisse zur Entwicklung im dritten Lebensjahr beschrieben.

6.2. Studienergebnisse zur kognitiven Entwicklung

6.2.1. „Theory of Mind“

Die „Theory of Mind“ (ToM) oder „naive Alltagspsychologie“ ist die Fähigkeit, anderen oder sich selbst mentale Zustände zuzuschreiben. Wünsche; Überzeugungen, Gefühle und Absichten anderer werden dabei unabhängig von den eigenen gesehen. So kann das Verhalten darauf basierend vorhergesagt werden (Goswami, 2008; Sodian, 2008).

Erst ab einem Alter von vier Jahren scheinen Kinder über die eigentliche „Theory of Mind“ zu verfügen, jedoch haben sie schon ab dem ersten Lebensjahr ein psychologisches Handlungsverständnis (Sodian, 2008). Im Zuge der Entwicklung des „Als-ob-Spiels“ lernen sie außerdem, eine reale und eine fiktive Welt zu unterscheiden. In den folgenden Kapiteln sollen daher die Entwicklung des Verstehens von Wünschen und Absichten und des „Als-ob-Spiels“ als die Vorläuferfähigkeiten der „Theory of Mind“ beschrieben werden (Sodian, 2008).

6.2.1.1. Verstehen von Wünschen und Handlungsabsichten

Die Fähigkeit, Wünsche und Absichten anderer zu verstehen, entwickelt sich früher als die Fähigkeit, andere Überzeugungen zu verstehen. In der englischsprachigen Literatur spricht man in diesem Zusammenhang von „Theory-Theories“ (Wellman, 2011) und unterscheidet zwischen „desires“ (Wünsche; z.B. Tom möchte einen Burger.) und „beliefs“ (Überzeugungen; z.B. Er glaubt im Restaurant einen Burger zu bekommen.), aus denen Vorhersagen über ein künftiges Verhalten gemacht werden können (z.B. Er geht ins Restaurant, um sich einen Burger zu kaufen). Kinder können bis zum fünften Lebensjahr demnach verstehen, was andere Menschen wollen, können aber nicht berücksichtigen, was andere Menschen glauben. Viele Autoren/Autorinnen sehen in dieser Fähigkeit, die Motiviertheit und Zielgerichtetheit (Intentionalität) des Verhaltens anderer zu erkennen, den Beginn der kindlichen „Theory of Mind“ (Rauh, 2008).

Studien zeigen, dass schon zweijährige Kinder zwischen eigenem und fremdem mentalem Zustand unterscheiden können und Gefühle, Wünsche und Absichten anderer unabhängig von eigenen Handlungsintentionen sehen. In einer Studie von Carpenter, Call und Tomasello (2002) lernten zweijährige Kinder eine komplexe Aufgabe (das Öffnen einer Box) schneller nach dem Vorbild eines Erwachsenen nachzuahmen, wenn

sie seine Intention kannten (das in der Box versteckte Spielzeug zu erlangen). Wellman und Woolley (1990) fragten in ihrer Studie Kinder nach deren Vorhersage der Reaktion und der Emotion von Erwachsenen in drei verschiedenen Situationen. In der ersten Situation sucht ein Erwachsener ein Objekt. Als er im ersten von zwei möglichen Räumen sucht, findet er das gesuchte Objekt. In der zweiten Situation findet er das gesuchte Objekt im ersten Raum nicht und in der dritten Situation findet er im ersten Raum nicht das gesuchte, sondern ein anderes Objekt. Die Kinder sollen nun beantworten, ob der Erwachsene in den unterschiedlichen Situationen weitersucht oder nicht und wie er sich fühlt.

In einem zweiten Experiment wurden den Kindern verschiedene möglichst einfache Aufgaben zum Verstehen der Überzeugung vorgegeben. Zum Beispiel wurde den Kindern gesagt, dass eine Person wieder nach einem Gegenstand an zwei möglichen Orten sucht und die Person glaubt, den Gegenstand z.B. am ersten Ort zu finden. Bevor die Kinder nach einer Vorhersage des Verhaltens der agierenden Person gefragt wurden, wurden sie gefragt, wo sie das Objekt vermuteten. Konnten sie ihre eigene Überzeugung von der Überzeugung der handelnden Person unterscheiden, spricht dies für die „Theory of Mind“. Die Ergebnisse zeigen jedoch, dass die Zweijährigen Reaktionen und Emotionen in Bezug auf einfache Wünsche hervorsagen konnten, bei Aufgaben zu Überzeugungen anderer jedoch versagten. Wellman und Wooley (1990) unterscheiden daher zwei Formen der „naiven Psychologie“ („Theory of Mind“) nämlich die „simple desire psychology“ von Kindern unter 4 Jahren und die „belief-desire psychology“ älterer Kinder und Erwachsener.

Zu erwähnen bleibt noch, dass einige Autoren/Autorinnen meinen, dass schon zweijährige Kinder die Fähigkeit haben Überzeugungen zu verstehen, diese Fähigkeit aber in den meisten Aufgaben nicht zeigen können (Ritblatt, 2000). Jedoch sprechen die meisten Studien – wie z.B. die Ergebnisse einer Metaanalyse von dreizehn Studien zeigt – für die Theorie von Wellman und Wooley (1990; Wellman & Liu, 2004).

Eine Erklärung für diese Unterschiede zwischen den Anfängen der „Theory of Mind“ bei jungen Kindern und späterer „naiver Alltagspsychologie“ sehen Wellman und Wooley (1990) im Verständnis mentaler Repräsentationen. Einfache Wünsche nach konkreten Objekten machen keine mentalen Repräsentationen notwendig, Überzeugungen jedoch schon. Denkt also ein Kind über einfache Wünsche eines

Anderen nach, so denkt es, der Andere habe ein inneres Verlangen nach einem externen Objekt (z.B. Er will einen Apfel.). Denkt jemand aber über die Überzeugung eines Anderen nach (z.B. Er denkt, dies sei ein Apfel.), so muss er wissen, dass der Andere eine mentale Repräsentation der Welt, also auch des Apfels, besitzt. Die Schwierigkeiten junger Kinder, diese mentalen Repräsentationen zu verstehen, erklären ihre Probleme, Überzeugungen Anderer zu verstehen (Perner, 1991; Walker & Muracher, 2012).

Sogenannte Metarepräsentationen sind wichtige Bedingungen, mentale Zustände Anderer zu verstehen. Metarepräsentation ist die Fähigkeit, den eigenen Wissenstand (bzw. die mentale Repräsentation) oder den eines anderen Menschen zu repräsentieren. Diese Metarepräsentation entwickelt sich aus drei Fähigkeiten: der Imitation, dem „Als-ob-Spiel“ und der Sprache (Goswami, 2008).

6.2.1.2. „Als-ob-Spiel“

Schon wenn man den Begriff „Spiel“ definieren möchte, kommt man nicht ohne das Charakteristikum der Konstruktion einer neuen Realität aus. Oerter (1999) nennt neben den Merkmalen „*Handlung um der Handlung willen*“ und „*Wiederholung und Ritual*“ auch den „*Wechsel des Realitätsbezugs*“ als wichtiges Charakteristikum des Spiels. Spielende konstruieren eine völlig neue Realität, die ihren Bedürfnissen und Zielen entspricht (Kapitel 5.1. Definition spielbasierter Verfahren). Auch Sturzbecher (2001) betont, dass jedes Spiel auf einer erfundenen Situation basiert.

Man unterscheidet verschiedene Formen des Spieles, wobei das „pretend play“ die Form darstellt, die vor allem durch den Als-ob-Charakter des Spiels zu definieren ist (Lillard et al., 2013). Deshalb soll im Folgenden „pretend play“ mit „Als-ob-Spiel“ übersetzt werden. In vielen Studien zeigten schon zweijährige Kinder „Als-ob-Spiel“ (Nielsen & Dissanayake, 2004).

Es entwickelt sich basierend auf Beobachtung und Imitation Anderer. Die Kinder nehmen im Spiel verschiedene Rollen an und tun zum Beispiel so, *als ob* sie kochen würden. Im Spiel können sie all ihre Fähigkeiten und Einsichten zeigen und benötigen dafür viel Vorstellungsvermögen und Kreativität (Sheridan, 1999). Das Kind tut, als ob etwas der Fall sei, obwohl dem nicht so ist (z.B. verwendet es die Banane als Telefon) und weiß dabei auch, dass es sich um eine erfundene Situation handelt (Lillard, 1993).

Die drei Kriterien des „Als ob Spiels“, die die Unterscheidung des „funktionellen Spiels“, das der Verwendung des Spielzeugs in seiner rein konventionellen, funktionellen Weise entspricht, ermöglichen, sind die „Objekt-Substitution“, die Zuschreibung erfundener Eigenschaften und frei erfundene Objekte. Bei der „Objekt-Substitution“ verwendet das Kind einen Gegenstand in einer anderen als seiner eigentlichen Funktion und tut so, als wäre zum Beispiel ein Becher ein Hut. Beim „Als ob Spiel“ schreiben die Kinder Objekten Eigenschaften zu, die sie eigentlich nicht besitzen (z.B. Das Kind tut so, als wäre der Becher heiß,) oder erfinden Objekte vollständig (z.B. Das Kind tut so, als halte es einen Löffel in der Hand, obwohl kein Löffel zur Verfügung steht.; Leslie, 1987).

Dieses „Tun-als-ob“ ist mittels Repräsentationen möglich. Im „Als-ob-Spiel“ bestehen zwei Repräsentationen nebeneinander. Die eine Repräsentation bezieht sich darauf, wie die Situation tatsächlich wahrgenommen wird (primäre Repräsentation: Banane) und die andere beinhaltet die erfundene Situation (Als-ob- bzw. sekundäre Repräsentation: Telefon). Die beiden Repräsentationen widersprechen einander dabei. Das Kind weiß jedoch, dass die primäre Repräsentation der Wahrheit entspricht und die Als-ob-Repräsentation wird von ihr getrennt als eine Repräsentation dieser Repräsentation gesehen.

Aus diesen frühen Repräsentationen scheinen sich die Metarepräsentationen zu entwickeln, die notwendig für eine „Theory of Mind“ sind (Lillard et al., 2011; Perner, 1991; Walker & Muracher, 2012). Leslie (1987) geht sogar davon aus, dass die Repräsentationen des „Als-ob-Spiels“ dieselben Metarepräsentationen erfordert, wie sie bei der ToM notwendig sind und die Fähigkeit zum „Tun-als-ob“ als Beginn der Fähigkeit des Verstehens der Kognition zu sehen ist. Einig ist man sich darüber, dass die Metarepräsentation eine wichtige Bedingung ist, um die mentalen Zustände Anderer zu verstehen (Leslie, 1987; Lillard et al., 2011; Perner, 1991).

Erst kürzlich veröffentlichte Lillard et al. (2013) eine Überblicksarbeit über Studien aus den letzten 50 Jahren, die den Einfluss des „Als-ob-Spiels“ auf die kognitive Entwicklung erheben. In Bezug auf die Beziehung zwischen „Theory of Mind“ und „Als-ob-Spiel“ zeigen sich voneinander abweichende Ergebnisse. Bei möglichen Zusammenhängen zwischen den beiden Variablen betonen die Autoren/Autorinnen, dass sie durch dritte konfundierende Variablen erklärt werden könnten bzw. ein

Hinweis dafür bestünde, dass nicht das „Als-ob-Spiel“ die ToM beeinflusst sondern umgekehrt. Zusammenfassend lässt sich sagen, dass die Forschung ein inkonsistentes Bild zeigt und noch weitere Studien hierzu notwendig sind. Walker und Chopnik (2013) reagieren auf die Arbeit von Lillard et al. (2013) mit einer neuen Interpretationsmöglichkeit: Das „Als-ob-Spiel“ und die kognitiven Fähigkeiten beeinflussen einander wechselseitig. Lillard et al (2013) betonen jedoch auch, dass viele Studien starke methodische Mängel aufweisen. Bergen (2013) meint, dass bei vielen Studien anzuzweifeln ist, ob das „Als-ob-Spiel“ valide erhoben wurde und daher generelle Aussagen ohne weitere Studien zum Zusammenhang zwischen „Als-ob-Spiel und kognitiver Entwicklung bzw. „Theory of Mind“ nicht angemessen sind.

Neben dem „Als-ob-Spiel“ werden auch die Fähigkeit der Kinder, sich selbst im Spiegel zu erkennen, und die synchronische Imitation als Vorläuferfähigkeiten der „Theory of Mind“ gesehen. Die synchronische Imitation zeigen zweijährige Kinder während des Spiels mit einem Partner, indem sie dem Spielzeug des Partners ähnelndes Spielzeug wählen und in ähnlicher Weise nutzen. Diese drei Vorläuferfähigkeiten haben gemein, dass sie alle auf sekundären Repräsentationen basieren und so zur Entwicklung der Metarepräsentation bzw. ToM beitragen (Nielsen & Dissanayake, 2004). Da diese Bereiche nicht im Itempool enthalten sind, werden sie auch nicht genauer beschrieben.

6.2.2. Numerische Fähigkeiten

Bei der Entwicklung numerischer Fähigkeiten muss eine Unterscheidung getroffen werden: Laut Krajewski (2008) entwickelt das Kleinkind voneinander unabhängig numerische Basisfertigkeiten bezüglich Zahlen und Mengen.

Auch Bryant und Nuñez (2011) unterscheiden zwischen der Fähigkeit, zu zählen, und der Fähigkeit, die Ordinalität und Kardinalität der Zahlen zu begreifen. Unter Ordinalität versteht man das Wissen, dass die Zahlen gereiht sind und größere Zahlen eine größere Menge beschreiben. Die Kardinalität ist das Wissen, dass eine bestimmte Anzahl von Items derselben Anzahl anderer Objekte entspricht. Daher sollen die Entwicklung des Zählens und jene des Mengenverständnisses voneinander unabhängig beschrieben werden.

6.2.2.1. Zählen

Mit etwa zwei Jahren beginnen Kinder zu zählen (Gelman & Gallistel, 1978). Eine Studie von Wynn (1990) zeigte, dass zweijährige Kinder durchschnittlich vier bis fünf Objekte richtig zählen konnten. Den Kindern fiel es dabei schwerer, Geräusche und Aktivitäten (Sprünge einer Puppe) zu zählen als Objekte.

In der Literatur lassen sich zwei Ansätze zur Entwicklung des Zählens finden (Wynn, 1990) – die Kontinuitätshypothese und die Diskontinuitätshypothese (Le Corre, Van de Walle, Brannon & Carey, 2006).

Anhänger der *Kontinuitätshypothese* wie Gelman und Gallistel (1978) gehen davon aus, dass Kinder über ein angeborenes System an nonverbalen Symbolen verfügen, die sich nach bestimmten Prinzipien entwickeln (Le Corre et al. 2006). Gelman und Gallistel (1978) postulierten die Prinzipien der „Eins-zu-Eins Zuordnung“, der „stabilen Reihenfolge“ und des „Kardinalprinzips“. Beherrschen die Kinder das Prinzip der „Eins-zu-Eins Zuordnung“, weisen sie beim Zählen jedem Objekt je eine (und nur eine) Zahl bzw. ein Symbol zu. Dem Prinzip der „stabilen Reihenfolge“ wird entsprochen, wenn bei jedem Zählvorgang die gleiche, also eine stabile, Reihenfolge an Zahlen verwendet wird. Das „Kardinalprinzip“ bedeutet, dass die beim Zählen zuletzt genannte Zahl die Menge der gezählten Einheiten angibt.

Anhänger der Kontinuitätshypothese sind der Ansicht, dass die Kinder diese angeborenen Prinzipien beherrschen und darauf aufbauend das Zählen lernen. Zählen ist für die Kinder kein sinnloses Aufzählen von Worten, sondern sie erkennen aufgrund des Verständnisses der Prinzipien einen Sinn in dieser Tätigkeit. Sie imitieren nicht nur die Zahlenreihe, die sie von ihren Eltern gehört haben, sondern sie folgen den angeborenen Prinzipien des Zählens, indem sie die verbale Abfolge von Zahlen als den Prinzipien entsprechend sehen. Dies zeigt sich vor allem in individuellen, aber stabilen Zahlenreihenfolgen (Le Corre et al., 2006; Wynn, 1990).

Briars, Le Corre, Carey und Fuson zählen zu den Vertretern der *Diskontinuitätshypothese*. Sie sind der Überzeugung, dass Kinder aufgrund genereller kognitiver Fähigkeiten zuerst zählen lernen und erst in weiterer Folge ein Verständnis für Zahlen entwickeln. Die Kinder lernen Zahlen aufzusagen, indem sie ihre Mitmenschen imitieren, ohne darin vorerst einen Sinn zu erkennen. Erst wenn sie häufig

gezählt haben, erkennen sie Gemeinsamkeiten der einzelnen Zählvorgänge und verstehen letztlich die Prinzipien des Zählens (Le Corre et al., 2006). Auch Krajewski (2008) kann als Vertreter dieser Theorie gesehen werden. Er geht davon aus, dass Kinder zuerst die ersten beiden Prinzipien und erst später das Kardinalsprinzip entwickeln. Sie verwenden das Zählen, um Objekte in eine Reihenfolge zu bringen, können aber noch nicht das Zählen und die gezählte Menge miteinander in Verbindung bringen. Dies spricht somit gegen die Annahmen der Kontinuitätshypothese. Diverse Studien zum Mengenverständnis scheinen die Diskontinuitätshypothese zu unterstützen (Bryant & Nuñez, 2011).

6.2.2.2. Mengenverständnis

Schon Säuglinge scheinen Mengen unterscheiden zu können und über ein angeborenes Verständnis von Zahlen zu verfügen (Siegler, De Loache & Eisenberg, 2011). Kinder haben jedoch meist noch ein Jahr, nachdem sie zu zählen beginnen, Probleme das Kardinalsprinzip zu verstehen. Wynn zeigte dies schon 1990 in ihrer Studie. Sie bat zwei- und dreijährige Kinder, zu zählen, wie viele Objekte ihnen dargeboten wurden (*aktive Mengenerfassung*). Die meisten Kinder verwendeten eine stabile Zahlenfolge und beherrschten die Eins-zu-Eins Zuordnung. Jedoch nannten die jüngeren Kinder (Alter zwischen 2;4-3;3) signifikant weniger oft die letzte beim Zählen genannte Zahl, um die Menge zu beschreiben als die ältere Altersgruppe.

Bei den Aufgaben zur *passiven Mengenerfassung* bat Wynn (1990) die Kinder, einer Puppe eine bestimmte Anzahl an Objekten (eins bis sechs) zu geben („Gib eine Anzahl-Aufgabe“). Basierend auf der Weise, wie die Kinder die Objekte reichten, unterschied sie zwischen „Greifern“ und „Zählern“. Die „Greifer“ tendierten dazu, eine Handvoll von Objekten zu reichen. Sie lösten die Aufgabe richtig, wenn ein Objekt gefragt war und lösten die Aufgabe tendenziell richtig, wenn zwei Objekte gefragt waren. Bei höheren Zahlen reichten sie jedoch eine beliebige Anzahl von Objekten. Vier Kinder der ältesten Gruppe, die die Objekte zählten, („die Zähler“), reichten ein Objekt nach dem anderen und stoppten bei der verlangten Zahl. Somit zeigten diese vier Kinder ein Verständnis des Kardinalprinzips. Die Autorin fasste die Ergebnisse dahingehend zusammen, dass Kinder erst ab einem Alter von dreieinhalb Jahren das Kardinalsprinzip erwerben und somit zu „Zählern“ werden. Diese Ergebnisse entsprechen der

Diskontinuitätshypothese, da die Kinder zwar schon zählen können, jedoch nicht das Prinzip der Kardinalität verstehen.

Goswami (2008) betont jedoch, dass die Anfänge des Zählens dennoch von einem Verständnis, was eine Zahl bedeutet, geprägt sind. So konnte Wynn (1990, 1992) mit ihren Studien zeigen, dass jedes Kind aus der Gruppe der „Greifer“ die „Gib eine Anzahl- Aufgabe“ bis zu einer bestimmten Zahl lösen konnte und bei den Aufgaben zu jeder höheren Zahl scheiterte. Kinder scheinen zuerst die Bedeutung kleiner Zahlen schrittweise zu lernen (zuerst eins, dann zwei, dann drei) und dann die Bedeutung aller restlichen höheren Zahlen gleichzeitig zu verstehen. So verstehen die Kinder im Laufe der Entwicklung zuerst nur die Bedeutung von eins („one-knower“), danach von eins und zwei („two-knower) und später können sie die Mengen von eins, zwei und drei („three-knower“) unterscheiden. Sobald die Kinder jedoch die Bedeutung der Zahl vier verstanden haben, können sie auch alle höheren Anzahlen richtig rechen, die innerhalb ihrer Zählbandbreite liegen („cardinal-knower“ Barner, Chow & Yang, 2009).

Das passive Mengenverständnis wird erst ab einem Alter von dreieinhalb Jahren unabhängig von der Anzahl generalisiert (Wynn, 1990). Mit zwei Jahren sind die meisten Kinder „one-knowers“ und ab zweieinhalb „two-knowers“ und einige Monate später können sie auch drei Gegenstände rechen („three-knowers“; Huang, Spelke & Snedeker, 2010). Sie zeigen somit ein passives Verständnis für Kardinalität. Für dieses „knowers-Konzept“ spricht auch, dass die Kinder bei „Gib mir eine Anzahl-Aufgaben“ mit höheren Zahlen als Antworten jene Zahlen ausschließen, deren Bedeutung sie bereits kennen (z.B.: Ein „two-knower“ wird gebeten, fünf Äpfel zu rechen. Er würde nicht einen oder zwei Äpfel hergeben, sondern eine beliebige höhere Zahl; Sarnecke & Lee, 2009).

Wie kommt es, dass Kinder bei niedrigeren Zahlen (eins bis drei) das Kardinalprinzip zu verstehen scheinen und bei höheren nicht? Die Interpretation liegt nahe, dass die Kinder mittels schneller automatischer Prozesse kleine Mengen unterscheiden und hierfür nicht zählen. In der Literatur sind diese Prozesse der spontanen Mengenerfassung unter „Subitizing“ bekannt (Goswami, 2008).

Auch Studien mit älteren Kindern weisen darauf hin, dass „cardinal-knower“ eine qualitativ andere Repräsentation von Zahlen besitzen als Kinder, die das

Kardinalsprinzip nur hinsichtlich kleiner Zahlen verstanden haben. Dies weist auf das Zutreffen der Diskontinuitätstheorie hin (Le Corre, 2006). Es wird davon ausgegangen, dass es für unterschiedlich große Itemsets unterschiedliche Systeme der Repräsentation von Zahlen gibt, die schon bei Säuglingen existieren.

Das erste System, das bei größeren Sets an zu zählenden Items zum Einsatz kommt, ist dafür zuständig, Mengen im Vergleich zu anderen Anzahlen grob einzuschätzen. Säuglinge können daher bei Habituerungsstudien zwischen Mengen, die z.B. ein Verhältnis von eins zu zwei besitzen, unterscheiden. Weitere numerische Informationen bietet das System der genauen Repräsentation bei einer kleinen Anzahl individueller Objekte. Es werden bis zu drei individuelle Items gleichzeitig durch je ein mentales Symbol repräsentiert. Vor allem Repräsentationen zweiter Art helfen den Kindern beim Erlernen der Prinzipien des Zählens (Feigenson, Dahan & Spelke, 2004; Le Corre & Carey, 2007).

Die Kinder erwerben die Prinzipien des Zählens vorerst für kleinere Zahlen (eins bis drei), alle anderen Zahlen werden als „viel“ bezeichnet. Parallel dazu lernen Kinder die Zahlenreihe. Durch die Kenntnis der Zahlenreihe lernen sie, die Ordinalität und Kardinalität verstehen. (Goswami, 2008; Le Corre & Carey, 2007). Ob die grammatikalische Fähigkeit, zwischen Singular und Plural zu unterscheiden, einen Einfluss auf das Verständnis von eins hat, ist umstritten (Barner et al., 2009). Das frühe, häufige verwenden von Zahlwörtern durch die Eltern und gemeinsames Zählen sind für das Verstehen von Kardinalität förderlich (Levine et al., 2010). Dieser Ansatz des qualitativen Sprungs von drei zu vier wird in der Literatur stark diskutiert (Goswami, 2008).

6.2.3. *Schlussfolgerndes Denken*

Das Kind wird häufig als aktiver Problemlöser gesehen. Ein Problem besteht dann, wenn der Erreichung eines Ziels verschiedene Hindernisse im Weg stehen, die es zu überwinden gilt. Ein Weg, diese Probleme lösen zu können, ist das schlussfolgernde Denken. Formen des schlussfolgernden Denkens, die schon früh in der Entwicklung des Kindes auftauchen, sind das induktive, das analoge und das deduktive Denken (Goswami, 2001; Sodian, 2008). Hierzu liegen für das dritte Lebensjahr nur sehr wenige Studien vor. Das induktive und das deduktive Denken scheinen aber im Kindesalter

schon ähnlich wie im Erwachsenenalter zu funktionieren (Goswami, 2008). Da nur das deduktive Denken im spielbasierten Itempool enthalten ist, liegt der Schwerpunkt dieses Kapitels auf der Beschreibung der Entwicklung dieser Form.

6.2.3.1. Induktives Denken

Induktives Denken ist das Denken, das in der Feststellung der Gleichheit oder Verschiedenheit von Merkmalen oder Relationen besteht. Das induktive Denken führt im Gegensatz zum deduktiven Denken nicht zu einer möglichen logischen Folgerung, sondern zu einer Wahrscheinlichkeitsaussage. Während man beim deduktiven Denken von der Gesamtheit auf den Einzelfall schließt, schließt man beim induktiven Denken von einem bestimmten Einzelfall auf die Gesamtheit und generalisiert somit. Man entdeckt Regelmäßigkeiten und Gesetzmäßigkeiten (Meiser & Klauer, 2001).

Wichtig für induktive Schlüsse ist die Ähnlichkeit (z.B. Hunde haben eine Milz. Menschen haben eine Milz. Haben Hasen auch eine Milz? Aufgrund der Ähnlichkeit der Lebewesen würde man auf „Ja“ tippen). Das induktive Denken von Kleinkindern wird vor allem mittels Aufgaben zu Kategorisierungen getestet (Goswami, 2011). An einer Studie von Gelman und Coley (1990) nahmen zweijährige Kinder teil. Die Kategorie Vogel wurde zum Beispiel erfragt, indem dem Kind zuerst ein Bild von einem Vogel vorgelegt wurde und gesagt wurde, dass er in einem Nest lebt. Dann wurden Bilder von zwei Vögeln und zwei Dinosauriern vorgegeben, wobei jeweils ein Bild einen typischen Vertreter der Kategorie darstellte und eines einen untypischen Vertreter (z.B. Dodo Vogel). Das Kind wird zu jedem Tier gefragt, ob das jeweilige Tier in einem Nest lebt. Wenn die Kategorien benannt wurden, konnten die meisten Kinder die Fragen richtig beantworten. Ältere Zweijährige kamen ohne Benennung der Kategorie aus. Die Autoren betonen somit, dass die Induktion nicht nur auf äußerer Ähnlichkeit beruht, sondern auf Kategorisierungen (Gelman & Coley, 1990). Diese Form des induktiven Denkens wird unterschieden von Schlussfolgerungen durch Analogien (Goswami, 2008; Klauer, 2001).

Das analoge Denken bzw. die analoge Induktion hilft beim Lösen einer neuen Situation, indem man auf Wissen über ähnliche Probleme und deren Lösung zurückgreift. Aufgrund der Übereinstimmung von einer früheren Situation mit dem aktuellen Problem erkennt man eine Analogie zwischen den beiden Situationen (Goswami, 2008; Sodian, 2008). Während beim induktiven Denken an sich die äußere Ähnlichkeit für

Schlussfolgerungen notwendig ist, ist beim analogen Denken vor allem die relationale Ähnlichkeit bedeutsam. Hierbei ist nicht vordergründig, ob zwei Situationen äußerlich ähnlich sind (z.B. Eine Klette sieht einem Klettband ähnlich.), sondern ob sie einander verhältnismäßig ähneln (z.B. Klette und Klettband kleben an Gegenständen mithilfe ihrer Haken. Goswami, 2008).

Drei bis vierjährige Kinder können schon Aufgaben zu analogen Beziehungen zwischen Einzelementen lösen (z.B. Vogel:Nest = Hund:?: Sodian, 2008). Für jüngere Kinder sind diese Aufgaben jedoch nicht geeignet. Jüngeren Kindern wird ein konkretes Problem dargeboten, das sie lösen müssen. Zudem werden ihnen ein ähnliches Problem und dessen Lösung gezeigt (Goswami, 2008; Singer-Freeman, 2005). In einer Studie von Singer-Freeman (2005) wurden 72 24 und 30 Monate alte Kinder getestet. Die Aufgaben wurden so einfach wie möglich gehalten und die analogen Probleme wiesen eine hohe Ähnlichkeit in ihrer kausalen Relation auf; bezüglich ihres Aussehens waren sie jedoch nicht ähnlich. Die Kinder erhielten zum Beispiel einen Spielzeug-Baum, einen Stein und ein Band. Ihre Aufgabe war es, einen Vogel damit zu bewegen. Den Kindern in der Versuchsgruppe wurde die Lösung eines ähnlichen Problems gezeigt, der Kontrollgruppe nicht (z.B. Band zwischen zwei Objekte spannen, um eine Orange zu bewegen). Wenn notwendig, erhielten die Kinder eine Hilfestellung. Die meisten zweijährigen Kinder verstanden die Analogien nachdem die Hilfestellung gegeben wurde, einige sogar schon davor. Sie lösten die Probleme mithilfe der Analogie zu ähnlichen Teilen wie Erwachsene oder ältere Kinder (Singer-Freeman, 2005).

6.2.3.2. Deduktives Denken

Beim deduktiven Denken werden die Schlussfolgerungen eindeutig aus der logischen Kombination der Prämissen abgeleitet. Es gibt daher nur eine richtige Antwort. Je nach Aufgabe wird von propositionalem, syllogistischem oder rationalem Schließen gesprochen. Bei propositionalen Aufgaben werden die Prämissen im Sinne von „und“, „oder“ oder „wenn-dann“ verknüpft (z.B. Prämisse 1: Wenn es regnet, ist die Straße nass. Prämisse 2: Es regnet. Folgerung: Die Straße ist nass.). Syllogismen sind Aufgaben, die Quantoren wie „alle“, „einige“ und „keine“ verwenden (z.B.: Prämisse 1: Alle Hund sind gelb. Prämisse 2: Waldi ist ein Hund. Folgerung: Waldi ist gelb). Relationale Schlussfolgerungen beinhalten Prämissen, die Verhältnisse zwischen Objekten oder Ereignissen angeben (z.B. Prämisse 1: Max ist größer als Mimi. Prämisse

2: Mimi ist größer als Heinz. Folgerung: Max ist daher größer als Heinz. Meiser & Klauer, 2001). Auch wissenschaftliches Denken und Schlussfolgerungen über Ursache und Wirkungen basieren auf deduktivem Denken. Das syllogistische Schließen ist die einfachste Form des deduktiven Denkens und soll bezüglich der Entwicklung im Kleinkindalter daher näher beleuchtet werden (Goswami, 2011).

Da Aufgaben zu deduktivem Denken im Gegensatz zu induktivem Denken ohne Wissen über die Wirklichkeit gelöst werden können, wird das deduktive Denken häufig mit logischen Folgerung getestet, die nicht der Wahrheit entsprechen (z.B. Waldi ist gelb). Früher wurde jedoch angenommen, dass Kinder logische Schlussfolgerungen nicht unabhängig vom Wahrheitsgehalt der Aussagen bzw. ihrem Wissen über die Welt ziehen und daher diese Aufgaben nicht lösen können. Sie ziehen nicht die logische Folgerung, sondern wählen eine Folgerung, die ihren Überzeugungen entspricht (z.B. Ein Hund kann nicht gelb sein, daher ist Waldi nicht gelb. Sodian, 2008).

Dias und Harris (1990) konnten jedoch zeigen, dass vier- und fünfjährige Kinder Aufgaben zum deduktiven Denken unabhängig von deren Inhalt lösen können, wenn sie instruiert wurden, sich einen anderen Kontext (z.B. einen anderen Planeten) vorzustellen. Aufgaben in diesem „Stell-dir-vor“-Modus scheinen also die Vorstellungskraft der Kinder zu nutzen und ihnen zu erlauben, die Prämissen als gegeben anzunehmen. Richards und Sanderson (1990) konnten diese Ergebnisse auch für zwei-, drei- und vierjährige Kinder bestätigen. Selbst die zweijährigen Kinder erzielten signifikant bessere Leistungen als sie instruiert wurden, sich einen anderen Planeten vorzustellen als wenn sie nur gebeten wurden, sich vorzustellen, die Aussagen würden stimmen. Zwischen den Altersgruppen konnte kein signifikanter Unterschied gezeigt werden. Diese und viele andere Ergebnisse weisen darauf hin, dass sowohl deduktives als auch induktives Denken bei Kindern und Erwachsenen in ähnlicher Weise funktioniert (Goswami, 2011).

Lillard et al. (2013) fassen in ihrer Übersichtsarbeit die Ergebnisse über den Zusammenhang von „Als-ob-Spiel“ und syllogistischem Denken in dem Sinn zusammen, dass das „Als-ob-Spiel“ eines von vielen Merkmalen darstellt, die die Entwicklung von schlussfolgerndem Denken fördern. Kinder müssen beim Spielen so tun, als wäre etwas Fiktives wahr. Dies kommt der Entwicklung der Fähigkeit, Aufgaben zu lösen, die nicht der Wahrheit entsprechen, zugute.

Harris relativierte 2001 in einer Arbeit die Bedeutung des „Stell-dir-vor“-Modus für vier bis sechsjährige Kinder. Der Autor konnte feststellen, dass die Kinder eine Woche nach dem ihnen eine in den Phantasiekontext eingebettete, syllogistische Aufgabe vorgegeben wurde, die Aufgaben auch ohne diese Hilfestellung in ähnlichen Raten meisterten. Die besseren Ergebnisse basieren nicht allein auf der Vorstellungskraft der Kinder. Es scheint eine tiefgründige Änderung der Vorgehensweise zugrunde zu liegen. „Stell dir vor“-Aufgaben helfen Kindern nicht nur logische Schlussfolgerungen zu machen, wenn die Folgerungen der Wahrheit widersprechen sondern auch bei richtigen Folgerungen über die Kinder kein Wissen verfügen (Alle Weichtiere leben in einer Schale). „Stell dir vor“-Hilfestellungen kann daher nicht auf die Überwindung von Widersprüchen reduziert werden.

Harris (2001) geht daher davon aus, dass nicht die Vorstellungskraft für die besseren Ergebnisse in „Stell-dir-vor“ Aufgaben zuständig ist, sondern, dass die Kinder so motiviert werden, die erste Aussage als Startpunkt zu akzeptieren. So konnte er auch feststellen, dass es genügt die Kinder zu instruieren, darüber nachzudenken, wie es wäre, wenn die Prämisse stimmen würde.

6.3. Relevanz der Erhebung kognitiver Entwicklung

Die Gründe für die Erhebung der kognitiven Entwicklung sind den Zielen der allgemeinen Entwicklungsdiagnostik ähnlich. Kognitive Fähigkeiten werden erhoben, um die Stärken und Schwächen eines Kindes hinsichtlich der intellektuellen Entwicklung zu bestimmen und um Interventionen zur Optimierung dieser Entwicklung einzuleiten (Vig & Sanders, 2007).

Jede Entwicklungsdiagnostik schließt in der Regel eine Erhebung der allgemeinen intellektuellen Fähigkeiten ein. Der Erhebung dieses Bereichs kommt eine besondere Bedeutung zu, da basierend auf Informationen zu kognitiven Fähigkeiten die Ergebnisse von emotionalen, sozialen oder anderen leistungsbezogenen Aspekten interpretiert werden können (Ettrich, 2000).

Die kognitiven Fähigkeiten können erhoben werden, um Vorhersagen über die spätere Intelligenz bzw. spätere kognitive Fähigkeiten zu machen. Obwohl Entwicklungstests für Kleinkinder und spätere Intelligenztests für Schulkinder unterschiedliche Aufgaben beinhalten, stehen einige dieser Aufgaben miteinander im Zusammenhang. Die

Vorhersagekraft von Entwicklungstests ist jedoch oft eingeschränkt, da unter anderem viele Umweltfaktoren wie z.B. eine Therapie Einfluss auf diesen Zusammenhang haben (Reuner, et al., 2007).

So konnten zum Beispiel viele Studien numerische Vorläuferfähigkeiten im Kindergartenalter, vor allem aber im Vorschulalter, identifizieren, die mit einem erfolgreichen Erwerb von mathematischen Fähigkeiten im Grundschulalter zusammenhängen. Die Mengen-Zahlen-Kompetenz im Vorschulalter, die als Vorläuferfähigkeit zu nennen ist, hat Einfluss auf die späteren mathematischen Leistungen und eignet sich am besten für die Vorhersage derselben (Krajewski & Schneider, 2006). Da sich diese Kompetenz recht früh zu entwickeln beginnt (Krajewski, 2008; Wynn, 1990), scheinen eine frühe Diagnostik von numerischen Fähigkeiten und darauf aufbauende Präventionen sinnvoll.

6.4. Diagnostik kognitiver Entwicklung

Die kognitive Entwicklung wird meist mit sehr heterogenen Aufgaben erfasst (Sarimski, 2009). Im Folgenden soll nun genauer beschrieben werden, wie die kognitive Entwicklung im dritten Lebensjahr einerseits in Form von Subskalen bzw. Dimensionen in gängigen allgemeinen Entwicklungstests im deutschsprachigen Raum, andererseits als Gegenstand der Erhebung in deutschsprachigen Intelligenztests erfasst wird.

6.4.1. Diagnostik kognitiver Fähigkeiten in allgemeinen Entwicklungstests

Die „Münchener funktionelle Entwicklungsdiagnostik – zweites und drittes Lebensjahr“ (Hellbrügge, 1994) beinhaltet als einzigen kognitiven Funktionsbereich das „Perzeptionsalter“, mithilfe dessen die Entwicklung des Erfassens von Zusammenhängen erhoben werden soll. Für zweijährige Kinder beinhaltet dieser Bereich Aufgaben zum Öffnen eines Verschlusses, zu Form-, Farb- und Größenunterscheidungen und zum Nachbauen von Formen mit Würfeln.

Der „Entwicklungstest 6 Monate bis 6 Jahre“ (Petermann, Stein & Macha, 2008) beinhaltet neben fünf anderen Dimensionen eine Dimension zur „kognitiven Entwicklung“. Die Dimension besteht aus Aufgaben zur Erhebung von Gedächtnisleistungen, Handlungsstrategien, Kategorisierungsfähigkeit und Körperbewusstsein. Von besonderer Relevanz sind kognitive Fähigkeiten, die für den Erwerb späterer sozialer Fähigkeiten Voraussetzung sind.

Die Aufgaben zur Gedächtnisleistung werden erst Kindern ab drei Jahren vorgegeben. Bei den Aufgaben zu Handlungsstrategien müssen zweijährige Kinder planerisches Handeln und visuell-räumliche und räumlich-konstruktive Prozesse anwenden. Dies wird mithilfe von Puzzles und den Aufgaben, Gegenstände festzuhalten und Würfeln anzuordnen erhoben. Außerdem werden den Kindern Aufgaben zur Perspektivübernahme und Fragen zum Verhalten bei einer Ampel gestellt. Beim Kategorisieren müssen sie funktionsverwandte Gegenstände zu Paaren ordnen und zwischen den Geschlechtern unterscheiden. Außerdem müssen sie Objekte nach Größe, Form und Farbe kategorisieren. Die Kinder werden auch gebeten, Eins-zu-Eins-Zuordnungen vorzunehmen. Weitere Aufgaben zum Zählen werden als schwieriger eingestuft und daher älteren oder leistungsstarken jungen Kindern vorgegeben. Das Körperbewusstsein wird erfasst, indem das Kind gebeten wird verschiedene Körperteile zu zeigen oder sie zu benennen.

Die „*Bayley Scales of Infant Development – Second Edition*“ (Reuner, et al., 2007) besteht neben einer motorischen Skala aus einer sehr breit angelegten kognitiven Skala. In dieser kognitiven Skala werden Aufgaben zu frühen Gedächtnisleistungen, Habituation, Problemlösefähigkeit, frühen Zahlenkonzepten, Klassifikations- und Kategorisierungsfähigkeit, Vokalisation, sprachlichen Kompetenzen und sozial-kommunikativen Fähigkeiten subsumiert (Reuner et al. 2009).

Zweijährige Kinder sollen bei der kognitiven Skala unter anderem Aufgaben zur visuellen Diskrimination, Form-, Größen- und Gewichtsunterscheidung, zum Einstecken von Stäben in ein Brett, sowie zu Mustererkennung und Turmbauen bearbeiten. Bei den sprachlichen Aufgaben werden sie gebeten, Gegenstände zu benennen und es wird der Umfang des aktiven Wortschatzes beobachtet und erfragt. Außerdem wird die grammatikalische Fähigkeit beurteilt (Reuner et al. 2009). Aufgaben zum numerischen Bereich finden sich ebenfalls für Zweijährige in Form von Aufgaben zu quantitativem Denken und Zählen. Bei den jüngeren zweijährigen Kindern wird das „Konzept eins“ Verständnis erfragt. Außerdem wird das Kind aufgefordert, zu zählen. Bei älteren bzw. leistungsstärkeren zweijährigen Kindern wird erhoben, ob sie das Konzept „mehr“ verstehen und ob sie beim Zählen den im Kapitel 6.2.2. „Numerische Fähigkeiten“ beschriebenen Prinzipien von Gelman und Gallistel (1978)

folgen (Reuner et al. 2009). Es wird mittels einer Facette des Bayley II erfasst, ob das Kind die Perspektive einer anderen Person versteht (Reuner et al. 2009).

Die *“Bayley Scales of Infant and Toddler Development – Third Edition“* (Bayley, 2006) bestehen aus fünf Skalen, wobei die sprachliche von der kognitiven Skala getrennt wurde. Die Rolle des Spiels, der numerischen Konzepte bzw. des Zählens und der Informationsprozesse wurde bei der Neukonzeptionierung der kognitiven Skala hervorgehoben.

Für zweijährige Kinder beinhaltet das Verfahren unter anderem folgende unterschiedlich schwierige Aufgaben zur Testung der Problemlösefähigkeit: Finden von versteckten Objekten, Öffnen einer Flasche, Stäbchen in Loch stecken, Verwenden eines Stifts, um ein Objekt zu sich zu bewegen. Das Kind wird zum Spielen animiert und es wird beobachtet, ob Objekte in ihrer ursprünglichen oder einer anderen Funktion verwendet werden. Die Aufgaben zum numerischen Wissen, die auf den Prinzipien von Gelman und Gallistel (1978) beruhen, und eine Aufgabe zum Konzept „mehr“ sind im Startblock für ältere Kinder enthalten bzw. werden nur den fortgeschrittenen Zweijährigen vorgegeben.

6.4.2. Diagnostik kognitiver Fähigkeiten in Intelligenztests

Intelligenztests sind Verfahren zur Erfassung des kognitiven Entwicklungsstandes (Renner & Irblich, 2009). Die Erhebung der kognitiven Fähigkeiten in Intelligenztests soll anhand von drei Verfahren, die Aufgaben für Kinder ab zwei Jahren und sechs Monaten beinhalten, beschrieben werden (Renner & Irblich, 2009).

Die deutschsprachige *“Kaufmann-Assessment Battery for Children“* (K-ABC; Melchers & Preuß, 2006) erhebt gelernte Fähigkeiten mit der „Fertigkeitenskala“ und „intellektuelle Fähigkeiten“, die zum Umgang mit neuen Situationen notwendig sind, mit den Skalen „einzelheitliches“ und „ganzheitliches Denken“ bei 2;6 bis 12;5-jährigen Kindern. Das „ganzheitliche Denken“, das zum Lösen eines Problems die Integration aller Informationen zu einem Zeitpunkt verlangt, wird bei Zweijährigen (2;6-2;11) mit Aufgaben zum Erkennen von Bildausschnitten, zum Wiedererkennen von zuvor dargebotenen Gesichtern und zum Erkennen eines unvollständigen Bildes erfasst. Durch das Wiederholen einer Abfolge von Handbewegungen und das Nachsprechen von Zahlen wird das „einzelheitliche Denken“, welches einem seriellen und folgerichtigen

Denken entspricht, erhoben. Bei der „Fertigkeitenskala“ werden der Wortschatz und das Wissen über Orte und Gesichter erfragt.

Der „*Hannover-Wechsler-Intelligenztest für Kinder im Vorschulalter*“ (HAWIVA-III; Ricken, Fritz, Schuck & Preuß, 2007) erhebt die allgemeinen intellektuellen Fähigkeiten von Kindern im Alter von 2;6 bis 6;11 Jahren. Zweijährigen Kindern werden die zwei Untertests „passiver Wortschatz“ und „Allgemeinwissen“ aus dem Verbalteil, die beiden Untertests „Mosaik-Test“ und „Figuren legen“ aus dem Handlungsteil und optional Aufgaben zum „aktiven Wortschatz“ vorgegeben.

Der „*Snijders-Omen Nonverbaler Intelligenztest 2 ½ -7*“ (SON-R 2 ½ -7; Tellegen, Laros & Petermann, 2007) erhebt die Handlungsfähigkeit und die Denkfähigkeit von jungen Kindern. Mit den Subtests zu Kategorisierung und Analogien wird das abstrakte Denken, welches der Erkennung von Ordnungsprinzipien entspricht, erhoben. Aufgaben zu Puzzles und zum Ergänzen von fehlenden Bildteilen dienen der Erfassung des konkreten Denkens, d.h. der Fähigkeit, Menschen und Objekte miteinander in Verbindung zu bringen. Das Legen von Mustern (Mosaik) und das Nachzeichnen von Vorlagen dienen der Erhebung des räumlichen Denkens (nach Laschkowski et al., 2000).

6.5. *Geschlechtsspezifische Unterschiede der kognitiven Entwicklung*

Eine Vielzahl an Studien zur kognitiven Entwicklung zeigt, dass zwischen Mädchen und Buben ein Unterschied besteht. Mädchen scheinen im Alter zwischen 14 und 24 Monaten einen kognitiven Vorsprung zu erlangen (Reznick, Corley & Robinson, 1997).

In einer Studie unter Verwendung der Bayley-Scales mit Kindern im zweiten Lebensjahr erzielen Mädchen vor allem im Alter von 20 oder 24 Monaten in der mentalen Skala höhere Werte ($F(1,308)=24.06$, $p<.01$), wobei die Unterschiede bei sprachlichen Untertests ausgeprägter sind. Genetische und umweltbedingte Einflüsse bzw. Unterschiede im Temperament und in der Entwicklung scheinen die besseren Leistungen der Mädchen zu erklären (Reznick, Corley & Robinson, 1997).

Auch in den Normtabellen des „Entwicklungstests 6 Monate bis 6 Jahre“ zeigen zweijährige Mädchen in den kognitiven und sprachlichen Dimensionen tendenziell höhere Werte als Buben. Die Aufgaben zur Dimension „Gedächtnis“ werden Kindern

ab drei Jahren vorgegeben. Bei Kindern zwischen 36-42 Monaten sind die Mittelwerte der Mädchen (MW=4.19, SD=3.22) den Mittelwerten der Buben (MW=4.07, SD=2.68) überlegen. Auch in den kognitiven Dimensionen „Handlungsstrategien“, „Kategorisieren“ und „Körperbewusstsein“ sowie den sprachlichen Subtests erzielen zweijährige Mädchen durchschnittlich höhere Werte als gleichaltrige Buben (Petermann et al., 2008).

In einer Längsschnittstudie mit Elterneinschätzungen der sprachlichen Entwicklung von Kindern aus Taiwan konnten für 18 Monate alte Kinder signifikante Geschlechtsunterschiede zugunsten der Mädchen gefunden werden ($p < 0.001$). Die Studienautoren argumentieren, dass das Geschlecht auch die sprachlichen Fähigkeiten mit 36 bzw. 60 Monaten beeinflusst (Lung et al., 2011). Eine Bestätigung für die signifikant besseren sprachlichen Leistungen von 20 Monate alten amerikanischen Mädchen ($F(6, 177) = 3.59, p < .01$) fanden Bornstein und Haynes (1998) sowohl bei Elternbefragungen als auch Beobachtungen bzw. Testungen.

6.6. Zusammenfassung kognitive Entwicklung

Kognition ist ein sehr breites Feld. In dieser Arbeit wurden nur die Entwicklung der „Theory of Mind“, der numerischen Fähigkeiten und des deduktiven Denkens beschrieben, da diese Bereiche im spielbasierten Itempool enthalten sind.

Kinder ab vier Jahren besitzen eine „Theory of Mind“, jedoch entwickeln sich schon davor gewisse Vorläuferfähigkeiten wie das Verstehen der Absichten und Wünsche Anderer und das „Als-ob-Spiel“. Das Verständnis für die Überzeugungen anderer Menschen entsteht erst später, wenn sich die dafür notwendigen Repräsentationen entwickeln. Viele Autoren/Autorinnen gehen davon aus, dass auch das „Als-ob-Spiel“ auf Repräsentationen der Welt und einer fiktiven Situation basiert. Aus diesen Repräsentationen entwickelt sich die Metarepräsentation, die für die „Theory of Mind“ notwendig ist. Die empirischen Ergebnisse über den Zusammenhang zwischen ToM und „Als-ob-Spiel“ sind jedoch nicht eindeutig.

Bei den numerischen Fähigkeiten werden die Entwicklung des Zählens und jene des Mengenerfassens voneinander unabhängig gesehen. Die Theorien zur Entwicklung des Zählens unterscheiden sich darin, dass sie entweder davon ausgehen, dass sich basierend auf angeborenen Prinzipien das Zählen entwickelt oder dass Kinder zuerst das

Aufsagen der Zahlenreihe lernen und dann erst die Prinzipien des Zählens verstehen. Neben der „Eins-zu-Eins Zuordnung“ und dem „Prinzip der stabilen Reihenfolge“ ist das „Kardinalprinzip“ zu nennen, das vor allem für die Mengenerfassung relevant ist. Kinder scheinen erst ab dreieinhalb Jahren das Kardinalprinzip vollständig zu verstehen. Davor lernen sie jedoch die Bedeutung kleinerer Zahlen (eins bis drei).

Zur Entwicklung des schlussfolgernden Denkens liegen für zweijährige Kinder nur wenige Studien vor. Zu unterscheiden sind das induktive bzw. analoge und das deduktive Denken, wobei alle Formen des Denkens schon im dritten Lebensjahr vorhanden zu sein scheinen. Beim deduktiven Denken bewährte es sich vor allem, die syllogistischen Aufgaben in einen Phantasiekontext zu betten, um den Kindern so zu ermöglichen, die nicht der Wahrheit entsprechenden logischen Folgerungen zu ziehen.

Kognitive Fähigkeiten werden häufig erhoben, um Aussagen über Stärken und Schwächen des Kindes in diesem Bereich tätigen zu können und darauf basierend Interventionen anbieten zu können. Häufig werden sie aber auch erfasst, um Ergebnisse aus anderen Bereichen mit ihnen in Bezug zu setzen. Die Prognose späterer Intelligenz ist ein weiteres Ziel der Diagnostik kognitiver Fähigkeiten, wobei diesem Vorhaben oft methodische und entwicklungsspezifische Grenzen gesetzt sind.

Die unterschiedlichen Definitionen von Kognition und Intelligenz zeigen sich bei der Vielfalt an diagnostischen Aufgaben und Skalen zur kognitiven Entwicklung in allgemeinen Entwicklungs- und Intelligenztests. Die dargestellten Verfahren beinhalten in den Skalen zur kognitiven Entwicklung viele Aufgaben zu Sprache, Gedächtnis oder visueller Wahrnehmung. Diese Bereiche stellen im spielbasierten Itempool je eine eigene Skala dar. Die Mehrheit der Verfahren beinhaltet Aufgaben zu numerischen Fähigkeiten, die meist auf den Prinzipien von Gelman & Gallistel (1978) basieren. Zur „Theory of Mind“ beinhalten zwei Verfahren Aufgaben zur Perspektivenübernahme (Bayley II; ET 6-6). Die Bayley III beinhalten auch die Beobachtung des Spiels. Zusätzlich gibt es viele Aufgaben zu Problemlösefähigkeiten. Das deduktive Denken wird in keinem der aufgeführten Tests bei Kindern im dritten Lebensjahr erhoben.

Zweijährige Mädchen scheinen einen Vorsprung in der kognitiven Entwicklung im Vergleich zu gleichaltrigen Buben zu haben.

II. Empirischer Teil

7. Zielsetzung und Fragestellung

Angesichts der unzureichenden Erfüllung der Gütekriterien durch vorhandene Entwicklungstests für Kinder im dritten Lebensjahr ist es Ziel des Projektes von Ass.-Prof. Dr. Kastner-Koller und Ass.-Prof. Dr. Deimann, ein Verfahren zu entwickeln, das einerseits durch seinen spielerischen Ansatz die Motivation und die Aufmerksamkeit der Kinder hochhalten und andererseits den Gütekriterien gerecht werden soll. Der von Kuchler, Sapper, Deimann und Kastner-Koller (2011) entwickelte Itempool wurde bereits dreimal rund 20 Kindern vorgegeben, testtheoretisch überprüft und auf den Analysen basierend überarbeitet (Birngruber, 2012; Fuchs-Gaderer, 2012; Putzer, 2013; Sindelar, 2013).

Im Rahmen dieser Diplomarbeit wurden einige Änderungen am Itempool und an der Vorgabe der Aufgaben aufgrund der Analysen und Kritik der Vorgängerinnen (Putzer, 2013, Sindelar, 2013) vorgenommen und im Anschluss wurden mit dem weiterentwickelten Verfahren wiederum 20 Kinder getestet.

Mithilfe der erhobenen Daten soll beurteilt werden in wie fern das Verfahren den Gütekriterien entspricht.

Zuerst werden die Beobachtungen zur Zumutbarkeit des Itempools beschrieben und die Frage beantwortet, wie gut das Verfahren dem Hauptgütekriterium Objektivität entspricht.

Im Rahmen der nächsten Fragestellung wird die Reliabilität (Interne Konsistenz) der Skalen des Verfahrens betrachtet. Durch Analysen der Itemtrennschärfen und der Itemschwierigkeiten soll eine Basis für die Entscheidung geschaffen werden, welche Items den Entwicklungsstand der Kinder am adäquatesten abbilden und somit auch weiterhin im Itempool behalten werden sollen. Durch die Entfernung wenig trennscharfer Items soll die Reliabilität erhöht werden.

Als letzte Fragestellung ist die Validität des Verfahrens zu nennen. Die inhaltliche Validität wird basierend auf Beobachtungen und Erfahrungen mit dem Verfahren beschrieben.

Da der Anstieg der Häufigkeit der richtigen Antworten mit dem Alter eine Minimalanforderung an Entwicklungstests ist, soll die Korrelation des Alters und der

Testleistung als Maß für die Konstruktvalidität berechnet werden. Ein weiterer Zugang die Konstruktvalidität zu beurteilen, ist zu ermitteln, ob alters- und geschlechtsspezifische Unterschiede bezüglich der Lösungshäufigkeit der Items der Skalen bestehen. Des Weiteren sollen eventuelle Interaktionseffekte untersucht werden. Zu erwarten ist, dass ältere Kinder bzw. Mädchen bessere Leistungen erzielen als jüngere Kinder bzw. Buben.

Zu erfassen gilt es auch, inwiefern die Angaben der Eltern im Elternfragebogen mit den im Spiel beobachteten Fähigkeiten übereinstimmen. Die Übereinstimmung gilt als Maß für die Kriteriumsvalidität.

Im Folgenden werden die Bereiche „sprachliche Entwicklung“, „kognitive Entwicklung“ und „Entwicklung des Gedächtnisses“ behandelt. Für die Skalen „motorische Entwicklung“, „sozioemotionale Entwicklung“, „Entwicklung der visuellen Wahrnehmung“ und „Entwicklung der Aufmerksamkeit“ wird auf die Arbeit von Kronberger (in Vorbereitung) verwiesen.

8. Methode

In diesem Kapitel wird die für die Untersuchung verwendete Methode genauer beschrieben. Es werden die Stichprobe, der spielbasierte Itempool und das Setting sowie der Ablauf der Untersuchung ausführlich dargestellt.

8.1. Stichprobe

Die meisten Kinder wurden über das direkte Ansprechen der Eltern in einer Kinderarztpraxis in Wien rekrutiert. Im Freundes- und Bekanntenkreis sowie in einem Kindergarten in Wien wurde ein Elternbrief verteilt (Anhang A), der Informationen über die Studienintention, den Spielablauf, sowie Testort und Zeitraum enthielt. Dadurch konnten weitere Kinder, die an der Studie teilnahmen, ausfindig gemacht werden.

Tabelle 1: Verteilung des Alters in Monaten hinsichtlich des Geschlechts

	Alter in Monaten											gesamt
	24	25	26	27	28	29	30	31	33	34	35	
männlich	1	1	3	0	2	1	0	1	0	0	1	10
weiblich	0	1	1	2	1	0	1	2	1	1	0	10
gesamt	1	2	4	2	3	1	1	3	1	1	1	20

In Tabelle 1 wird dargestellt, wie sich das Alter hinsichtlich der beiden Geschlechter verteilt. Es wurden insgesamt 20 Kinder im Alter von 24 bis 35 Monaten getestet, wobei eine Hälfte der Kinder männlich und die andere weiblich ist. Die Stichprobe beinhaltet kein Kind, das 32 Monate alt ist. Das durchschnittliche Alter der Stichprobe beträgt 28.5 Monate (SD=3.15).

Tabelle 2: Häufigkeiten der Altersgruppen hinsichtlich des Geschlechts

		Altersgruppen		gesamt
		jüngere Kinder 24-29	ältere Kinder 30-35	
Geschlecht	männlich	8	2	10
	weiblich	5	5	10
gesamt		13	7	20

Es wurden zwei Altersgruppen gebildet, die hinsichtlich der Altersspannweite (6 Monate) gleich groß sind. 13 Kinder der Stichprobe sind der jüngeren Gruppe (24 und 29 Monate) zuzuordnen und 7 Kinder sind bereits der älteren Gruppe mit einem Alter zwischen 30 und 35 Monaten zugehörig (Tabelle 2). Neben dem ausgewogenen Geschlechterverhältnis (10 Buben/10 Mädchen) konnte mittels Chi-Quadrat Test festgestellt werden, dass sich die Kinder auch auf die Altersgruppen ($\chi^2=1.8$; $p=.26$) gleich verteilen. Ein Chi-Quadrat Test, der die beiden Variablen Geschlecht und Altersgruppen berücksichtigt, ergibt des Weiteren ein nicht signifikantes Ergebnis und weist somit auf die Gleichverteilung der Altersgruppen auf die Geschlechter bzw. die Unabhängigkeit der beiden Variablen hin ($\chi^2=1.98$; $p=.35$).

Mittels Fragebogen, der von den anwesenden Eltern ausgefüllt wurde, konnten noch weitere soziodemographische Informationen gewonnen werden:

Das durchschnittliche Alter der Väter beträgt 34.95 Jahre ($SD=7.08$). Der Mittelwert des Alters der Mütter liegt mit 32 ($SD=5.18$) Jahren etwas darunter.

Tabelle 3: Ausbildung der Väter und Mütter

	Pflichtschule	Lehre/Fachschule	Matura	Uni/FH	gesamt
Väter	1	8	4	7	20
Mütter	0	7	6	7	20
gesamt	1	15	10	14	40

Abgesehen von Müttern mit Pflichtschulabschluss konnten für alle Ausbildungsgruppen Vertreter/Vertreterinnen bei Müttern und Vätern gefunden werden. Fünf Prozent bzw. einer der Väter hat einen Pflichtschulabschluss. Acht Väter (40%) haben eine Lehre/Fachschule abgeschlossen und beinahe gleich viele Väter haben eine akademische Ausbildung (7 Väter bzw. 35%). Vier Väter haben die Matura als höchste abgeschlossene Ausbildung vorzuweisen. Die Anzahl der Väter in den einzelnen Ausbildungskategorien sind laut Chi-Quadrat Test gleichverteilt ($\chi^2=6$; $p=.12$). Je sieben Mütter (35%) weisen Abschlüsse einer Lehre/Fachschule bzw. Universitäts-/Fachhochschule auf. Sechs Mütter haben ihre Ausbildung mit Matura beendet (30%). Mittels Chi-Quadrat Test konnte auch hier die Gleichverteilung auf die Ausbildungskategorien bestätigt werden ($\chi^2=0.1$; $p=1$).

Alle 20 Kinder leben in einem gemeinsamen Haushalt mit der Mutter. Zwei Kinder wohnen getrennt vom Vater. 11 Kinder sind Einzelkinder und neun Kinder haben ein Geschwister, wobei drei jünger und sechs Geschwister älter als das zweijährige Testkind sind.

17 Kinder gehen bereits in einen Kindergarten oder zu einer Tagesmutter. Acht Kinder besuchen Sport- oder Freizeitkurse.

90 Prozent der Kinder lernen Deutsch als Muttersprache. Ein Kind spricht muttersprachlich serbisch und eines polnisch. Sieben Kinder wachsen mehrsprachig auf, wobei sechs Kinder zwei Sprachen lernen und ein Kind drei Sprachen lernt.

Bei einem Kind berichtete die Mutter von einer Risikoschwangerschaft und zwei Kinder waren Frühgeburten. Gesundheitliche Probleme ihres Kindes berichtete keine Bezugsperson.

8.2. *Untersuchungsinstrument*

Der spielbasierte Itempool wurde von Kuchler, Sapper, Deimann und Kastner-Koller (2011) entwickelt. In der aktuellen Fassung besteht er aus 149 Aufgaben zur Erhebung des allgemeinen Entwicklungsstandes zweijähriger Kinder. Er umfasst die acht Bereiche „motorische Entwicklung“, „sprachliche Entwicklung“, „sozioemotionale Entwicklung“ und „kognitive Entwicklung“ sowie die „Entwicklung der visuellen Wahrnehmung“, „Entwicklung des Gedächtnisses“ und „Entwicklung der Aufmerksamkeit“. Zusätzlich wird den Eltern ein Fragebogen zur Gewinnung von soziodemographischen Daten und ihrer Sichtweise der Entwicklung der Selbstständigkeit, des Sozial- und Spielverhaltens und des aktiven Wortschatzes ihrer Kinder vorgegeben (Kuchler et al., 2011).

Im Mittelpunkt des Verfahrens steht das gemeinsame Spiel von Kind und Testleiter/Testleiterin in der Puppenküche. Neben dieser befinden sich ein Kaufmannsladen, eine Stiege, einige Sessel und Tisch in angemessener Größe für Kinder und ein Sitzbereich für die Eltern im Testraum. Zusätzlich wurde ein Klebestreifen am Boden angebracht. Mit der im Raum befindlichen Videokamera wird die Spielsituation gefilmt, um die spätere Auswertung zu erleichtern (Kuchler et al., 2011).

Die meisten Aufgaben werden möglichst spielerisch während der Aktivitäten in der Puppenküche erfragt. Die Materialien aus der großen Spielkiste und die dazugehörigen Items sollen dabei in einer beliebigen Reihenfolge je nach Interesse und Motivation des Kindes in das Spiel eingebaut werden. Darin befinden sich unter anderem Utensilien zum Zeichnen, Bausteine, Perlen zum Auffädeln, Gedächtnisspiele, usw. (Kuchler et al., 2011).

Dem Manual von Kuchler et al. (2011) folgend wird bei allen Aufgaben das jeweils beste gezeigte Verhalten beurteilt. Löst das Kind ein Item mehrere Male nicht, bei einer späteren Vorgabe jedoch schon, so gilt dieses Item als gelöst. Die meisten Items sind dichotomer Natur. Für die Kodierung wird der Protokollbogen im Anhang D verwendet.

Im Weiteren wird die aktuelle Erhebung der einzelnen Bereiche erläutert. Die in dieser Arbeit betrachteten Skalen sind die „sprachliche Entwicklung“, „kognitive Entwicklung“ sowie die „Entwicklung des Gedächtnisses“. Außerdem wird der Elternfragebogen (siehe Anhang C) kurz dargestellt. Für die Beschreibung der anderen Bereiche wird auf Kronberger (in Vorbereitung) verwiesen. Anschließend werden die Änderungen, die vor der Untersuchung basierend auf den Analysen der vorangegangenen Diplomarbeiten (Birngruber, 2012; Fuchs-Garderer, 2012; Putzer, 2013; Sindelar, 2013) vorgenommen wurden, beschrieben

8.2.1. Erhebung der sprachlichen Entwicklung

Die sprachliche Entwicklung wird über Aufgaben zum aktiven und passiven Wortschatz sowie zur Grammatik erfasst (Kuchler et al., 2011).

Die aus 11 Items bestehende Skala „aktiver Wortschatz“ dient der Einschätzung des vom Kind produzierten Wortschatzes und wird erfasst, indem das Kind gebeten wird, Obst und Gemüse in der Puppenküche sowie Gegenstände im ToM-Buch („Theory of Mind“-Buch) zu benennen. Wenn eine Aufgabe in einer anderen Sprache als Deutsch nach Angaben der Mütter richtig gelöst wurde, wurde dies als richtig gewertet.

Beim „passiven Wortschatz“ soll das Sprachverständnis einzelner Wörter erhoben werden und das Kind wird gebeten, dem/der Testleiter/Testleiterin einen bestimmten Gegenstand im ToM-Buch zu zeigen. Hierzu werden 8 Items erfragt.

Die grammatikalischen Fähigkeiten werden vor allem mithilfe der Videos im Nachhinein kodiert. Die Skala besteht aus 24 Items, die unter anderem die Beurteilung der Satzlänge, der Bildung unterschiedlicher Fragen und Verneinungen beinhalten (Kuchler et al., 2011).

8.2.2. Erhebung der kognitiven Entwicklung

Bei der kognitiven Entwicklung werden sowohl die „Theory of Mind“, als auch die numerischen Fähigkeiten des Kindes erhoben (Kuchler et al., 2011).

8.2.2.1. Erhebung der „Theory of Mind“

Die „Theory of Mind“ oder mentalistische Alltagspsychologie ist als die Fähigkeit, sich selbst oder anderen Wünsche, Absichten oder Überzeugungen zuzuschreiben, zu verstehen (Sodian, 2008). Sie wird mithilfe der Vorläuferfähigkeiten „Als-ob-Spiel“ und „Hineinversetzen in Andere“ erhoben (Kuchler et al., 2011).

Das „Als-ob-Spiel“ ist die eigentliche kindliche Spielform. Das Kind schreibt den Spielsachen oder der Handlung eine neue Bedeutung zu und tut zum Beispiel so, als würde es kochen (Oerter, 2008). Diese Skala wird mithilfe der aufgezeichneten Videos im Nachhinein protokolliert. Es wird beurteilt, ob das Kind ein „Als-ob-Spiel“ zeigt und wenn ja, welches. Unterschieden werden „kein Als-ob-Spiel“, „Als-ob-Spiel anhand von realen Gegenständen“ (z.B. Kind tut so, als würde es aus dem leeren Becher trinken) und „Als-ob-Spiel gänzlich vorgestellt. Letzter Kategorie umfasst das „Als-ob-Spiel“, bei dem Objekte gänzlich vorgestellt (z.B. Kind stellt sich den Becher vor und tut so, als würde es trinken) oder als etwas Anderes verwendet werden (z.B.: Becher wird als Hut aufgesetzt; Kuchler et al., 2011).

Die fünf Items zum „Hineinversetzen in Andere“ werden mittels ToM-Buch vorgegeben. Den Kindern werden zu den Personen im Buch standardisierte Fragen zu Bedürfnissen und möglichen nächsten Handlungsschritten gestellt (wie z.B. Was macht der Mann, wenn er schmutzige Hände hat?), um zu überprüfen, ob sie sich in diese Personen hineinversetzen können (Kuchler et al., 2011).

8.2.2.2. Erhebung des numerischen Wissens

Die Erhebung der Entwicklung des numerischen Wissens handelt vom aktiven und passiven Mengenerfassen und dem Zählen (Kuchler et al., 2011).

Dem Manual folgend besteht die Skala „aktives Mengenerfassen“ aus zwei Items zur Benennung von Mengen. Die Kinder werden anhand von zwei bzw. drei Objekten gefragt, wie viele es seien.

Die „passive Mengenerfassung“ wird mithilfe von drei Items erhoben. Die Zweijährigen werden gebeten, dem Testleiter/der Testleiterin eins, zwei bzw. drei Objekte zu geben (Kuchler et al., 2011).

Beim „Zählen“ wird beurteilt, ob das Kind überhaupt zählt, ob es in der richtigen Reihenfolge zählt und ob es bereits die Prinzipien „Eins-zu-Eins-Zuordnung“, „stabile Reihenfolge“, „Kardinalprinzip“ nach Gelman & Gallistel (1978; Kapitel 6.2.2.1.) erworben hat. Handelt ein Kind den Prinzipien entsprechend, wird sein Zählen als richtig gewertet.

8.2.3. Erhebung der Entwicklung des Gedächtnisses

Laut Manual von Kuchler et al. (2011) wird in dieser Skala zwischen phonologischem und visuellem Gedächtnis unterschieden.

Die Skala „phonologisches Gedächtnis“ beinhaltet drei Items und soll die Merkfähigkeit des Kindes bei Vorgabe von Wörtern erheben. Der Testleiter/die Testleiterin nennt dem Kind ein bis drei verschiedene Gegenstände, die es sich merken und dem Testleiter/der Testleiterin reichen soll (z.B.: Gib mir bitte einen Apfel, eine Nuss und eine Banane).

Die Erhebung des „visuellen Gedächtnisses“ besteht ebenfalls aus drei Items und soll die Merkfähigkeit des Kindes bezüglich visuellen Materials erheben. Die Aufgaben werden mittels einer Schachtel mit vier Laden vorgegeben, die je eine Memorykarte beinhalten. Jede Lade wird geöffnet und gemeinsam mit dem Kind wird nachgesehen, welches Bild sich darin versteckt. Nach dem Schließen aller Laden werden dem Kind nacheinander vier identischen Memorykarten gezeigt und es wird gefragt, in welcher Lade sich z.B. die Sonne befindet (Kuchler et al., 2011).

8.2.4. Elternfragebogen

Der Fragebogen (Anhang C) wird den Eltern vorgegeben während ihre Kinder spielen. Zu Beginn werden Fragen zu den soziodemographischen Daten des Kindes gestellt. Die Eltern werden gebeten, Angaben zur familiären Situation, zur Muttersprache bzw. eventuellen Mehrsprachigkeit des Kindes und zur Geburt, Schwangerschaft sowie Gesundheit zu machen.

Der Fragebogen beinhaltet im Weiteren Fragen zur Selbstständigkeit und zum Sozial- und Spielverhalten. Für genauere Ausführungen hierfür muss auf die Arbeit von Kronberger (in Vorbereitung) verwiesen werden.

Zuletzt wird die anwesende Bezugsperson gebeten, anhand einer Liste von 42 Wörtern anzugeben, welche Wörter das Kind bereits häufig selbst verwendet hat. Dazu zählen auch Wörter, die das Kind etwas anders ausspricht (z.B. Afel statt Apfel). Es wird auch erfragt, ob das Kind ein anderes Wort verwendet (z.B. Heihei statt schlafen). Im Anschluss daran werden sechs Fragen zu den grammatikalischen Fähigkeiten des Kindes gestellt.

8.2.5. Änderungen

Im Sinne der Weiterentwicklung des Itempools wurden die testtheoretischen Überprüfungen und Kritikpunkte an der praktikablen Anwendung durch die Vorgängerinnen als Grundlage für die Überarbeitung des Itempools verwendet.

Da sich gezeigt hat, dass die meisten Kinder nicht die Motivation oder Aufmerksamkeit aufbringen können, drei Bücher anzusehen, wurde in der aktuellen Erhebung nur ein Buch vorgegeben (ToM-Buch – „Theory of Mind“-Buch). Durch das Weglassen des Magnetbuchs ergab sich eine Reduzierung um drei Items beim „aktiven Wortschatz“ und um 13 Items beim „passiven Wortschatz“. Ebenso wurde das Buch zum deduktiven Denken mit der gesamten Skala weggelassen, da viele Kinder in den Vorgängerarbeiten die Bearbeitung verweigerten (Putzer, 2013) und die Forschungslage für das deduktive Denken im dritten Lebensjahr unzureichend ist (siehe Kapitel 6.2.3.2.).

Manche Items wurden hinzugefügt, da die Literaturrecherche ergab, dass diese Aufgaben relevant für den Altersbereich sind:

Bei der Beurteilung der grammatikalischen Fähigkeiten wurde das Item „korrekte und vollständige Sätze“ hinzugefügt, da im dritten Lebensjahr auf Grund des steigenden Wortschatzes die Komplexität der Sätze steigt (Szagun et al., 2006). Bis zu einem Alter von zweieinhalb bis drei Jahren haben die Kinder die wichtigsten Regeln von Morphologie und Syntax gelernt (Weissenborn, 2000). Sie erwerben zunehmend die Fähigkeit, syntaktische Regeln wie Wortreihenfolge und grammatikalische Endungen bei Sätzen mit ihnen geläufigen Verben anzuwenden, sodass sie am Ende des dritten Lebensjahres verschiedenartige einfache Sätze mit bis zu 11 Wörtern bilden können (Behrens, 2004; Weinert, 2007).

Als neue und für den Altersbereich relevante Skala wurde auch das „Zählen“ eingeführt. Bei der Kodierung des Zählens dienten die Prinzipien von Gelman und Gallistel (1978) als Grundlage und es wurde versucht, die Kritik durch Wynn (1990) zu berücksichtigen. Bei Wynn (1990) stellte sich als wichtig heraus, ob ein Kind in der richtigen Reihenfolge zählen konnte. Daher wurde auch das Item „richtige Reihenfolge“ in den Itempool aufgenommen. Das „Kardinalprinzip“ wurde angesichts der Fähigkeit junger Kinder kleinen Zahlen (eins bis drei) zu verstehen, in den Itempool aufgenommen (siehe Kapitel 6.2.2.2.).

Es wurden auch einige Aufgaben entfernt. Die Literaturrecherche und die geringe Lösungshäufigkeit der Aufgaben „Gib mir 4“, „Wie viele sind das: Menge von 4“ und „Wie viele sind das: Menge von 5“ in den Analysen der vorangegangenen Arbeiten (Birngruber, 2012; Sapper, 2011) führte zur Entfernung dieser Items aus dem Itempool des numerischen Wissens.

Die Vorgabeweise der Items zum „visuellen Gedächtnis“ wurde modifiziert, um die Motivation und das Interesse des Kindes an der Aufgabe durch eine spielerischere Vorgabe zu erhöhen. Wurden zuerst nur die Memorykarten vorgelegt, so werden aktuell die Memorykarten in einer Schachtel mit vier Laden versteckt.

Die vorgenommenen Änderungen und Schwierigkeiten bei der Kodierung haben zur Modifizierung des Manuals und des Protokollbogens geführt.

8.3. Die Untersuchung

Im Folgenden wird genauer auf das Setting und den Ablauf der Untersuchung eingegangen.

8.3.1. Räumliche und zeitliche Rahmenbedingungen

Alle Kinder wurden innerhalb eines Monats (von 5. Februar 2013 bis 7. März 2013) getestet. Die Testung dauerte von der Begrüßung bis zur Verabschiedung durchschnittlich zwei Stunden ($SD=16.75$). Die längste Testung dauerte 160 Minuten und die kürzeste 96 Minuten.

Die Anlaufzeit, die als die Zeit definiert wird, bis das Kind sich mit den Materialien beschäftigt, dauerte durchschnittlich 4.08 Minuten ($SD=3.78$). Im Durchschnitt beschäftigten sich die Kinder $MW=99.42$ Minuten ($SD=11.43$) mit den Materialien und benötigten durchschnittlich nur 4.64 Minuten ($SD=5.13$) Pausenzeit. Die mittlere Pausenanzahl betrug 1.90 ($SD=2$). Die Verabschiedung dauerte im Durchschnitt 12.20 Minuten ($SD=6.32$).

Die Spielsituationen fanden im Kleinkindertestraum der Test- und Beratungsstelle an der psychologischen Fakultät der Universität Wien statt. Der Raum bot genügend Platz für die Kinder um sich frei bewegen und turnen zu können. Die Aufzeichnungen der im Raum installierten Videokamera konnten für spätere Auswertungen einzelner Skalen genutzt werden.

8.3.2. Anwesende Personen

Bei 17 Kindern war die Mutter als Bezugsperson anwesend, in einem Fall war nur der Vater und zwei Kinder wurden von beiden Elternteilen begleitet. Zusätzlich brachten vier Mütter ein Geschwisterkind mit. Drei Geschwisterkinder waren im Säuglingsalter und störten das Spiel nicht. Einem älteren Geschwisterkind mit sechs Jahren wurde die Situation erklärt. Daraufhin saß es die meiste Zeit ruhig am Tisch, las ein Buch und störte somit das Spiel ebenso wenig.

8.3.3. *Ablauf der Untersuchung*

Die Kinder und ihre jeweilige Bezugsperson wurden im Eingangsbereich der Fakultät für Psychologie empfangen. Wenn möglich, wurde zuerst das Kind auf Augenhöhe begrüßt, danach die Eltern. Da die meisten Kinder sehr schüchtern waren, wurde bis zum Eintreffen im Testraum vor allem mit den Eltern gesprochen. Den Kindern wurde erklärt, dass man in einen Raum mit sehr vielen Spielsachen gehe, um dort gemeinsam zu spielen.

Im Testraum wurden die Eltern zuerst um die Einwilligung für die Videoaufzeichnung gebeten, um sogleich mit der Aufnahme beginnen zu können. Sie wurden über die voraussichtliche Dauer und den Ablauf informiert und darüber aufgeklärt, dass sie, wenn es das Kind möchte, jederzeit mitspielen können. Die Eltern wurden außerdem gebeten, Aufgaben der Testleiterin zu wiederholen, wenn das Kind auf Fragen der Testleiterin nicht reagiere, jedoch keine eigenen Aufgaben zu stellen. Weiters wurden sie gebeten, weder Farben noch Obst zu benennen, da die Benennung zu den Aufgaben der Kinder zähle. Es wurde betont, dass herausgefunden werden solle, welche Aufgaben für zweijährige Kinder angemessen sind und dass daher einige Aufgaben noch sehr schwierig sein würden. So sollte den Eltern der Druck genommen werden, ihr Kind durch viele Fragen und Anweisungen zum bestmöglichen Verhalten zu motivieren und somit das Kind möglicherweise zu überfordern.

Beinahe alle Kinder interessierten sich sofort für die Puppenküche oder den sich daneben befindlichen Kaufmannsladen und der Einstieg ins gemeinsame Spiel gelang somit sehr schnell. Während des Spiels wurden die Items in individueller Reihenfolge dargeboten. Es wurde versucht, möglichst viele Items in das vom Kind präferierte Spiel zu integrieren. Zeichnete ein Kind gerne, wurde versucht, die Anzahl von gezeichneten Objekten zu erfragen, etc. Pausen wurden gemacht, wenn das Kind den Wunsch zu essen, trinken, etc. äußerte oder die Aufmerksamkeit merklich abnahm.

Wenn alle Items vorgegeben worden waren, das Kind nicht mehr zum Spielen motiviert werden konnte oder sichtlich müde war, wurde das Spiel beendet. Die Testleiterin hielt dann noch eine kleine Belohnung für das Kind bereit, um so das Ende der Spielsituation zu signalisieren. Da die meisten Kinder gerne weiterspielen wollten, wurde noch gemeinsam aufgeräumt.

9. Auswertung

Für die statistische Auswertung der Daten wurde das Statistikprogramm IBM SPSS Statistics 20 verwendet.

Zuerst sollen Beobachtungen zur Zumutbarkeit beschrieben werden. Insbesondere sollen Einschätzungen zur Attraktivität des Testmaterials für die Kinder vorgenommen werden. Danach wird das Verfahren hinsichtlich der Hauptgütekriterien beurteilt.

Für die Objektivität kann im Rahmen dieser Arbeit kein numerischer Wert angegeben werden. Sie wurde basierend auf inhaltlichen Überlegungen und Erfahrungen beurteilt.

Dann wird auf die Analysen der einzelnen Items und die Berechnungen der Reliabilität auf Skalenebene eingegangen. Nach der Itemselektion wurden Summenscores gebildet.

Die Summenscores wurden wiederum verwendet, um Kennwerte für die Validität zu berechnen. Hierzu wurden zuerst mögliche Geschlechts- und Altersunterschiede und Korrelationen der Skalen mit dem Alter ermittelt. Zum Schluss wurde die Übereinstimmung der Angaben am Elternfragebogen mit dem beobachteten Verhalten hinsichtlich „aktivem Wortschatz“ und „grammatikalischen Fähigkeiten“ eruiert.

In den folgenden Kapiteln soll auf die einzelnen Auswertungsschritte genauer eingegangen werden.

9.1. Auswertung: Objektivität

Wie schon im Kapitel Objektivität (4.1.1. Objektivität) erläutert, unterscheidet man bei der Objektivität zwischen Durchführungs-, Auswertungs- und Interpretationsobjektivität (Lienert und Raatz, 1998). Der Grad der Unabhängigkeit vom Untersucher/von der Untersucherin hinsichtlich der Durchführung und Auswertung soll in Bezug auf den spielbasierten Itempool beschrieben werden. Die Interpretationseindeutigkeit kann in dieser frühen Phase der Testentwicklung nicht eingeschätzt werden. Ein berechneter Übereinstimmungswert kann im Rahmen dieser Arbeit ebenfalls nicht gegeben werden.

9.2. Auswertung: Itemanalyse und Reliabilität

Als zweites Hauptgütekriterium wird für jede Skala die Reliabilität mittels Alphakoeffizienten nach Cronbach berechnet. Cronbachs Alpha ist ein Maß für die innere Konsistenz (siehe Kapitel 4.1.2. Reliabilität), die laut Kubinger (2009) angibt, wie sehr die Items dasselbe messen. Field (2009) betont, dass Cronbachs Alpha nur für Itemgruppen verwendet werden darf, für die schon vorab Eindimensionalität angenommen wird. Für die Skalen des spielbasierten Itempool ist dies der Fall.

Der Koeffizient wird bestimmt, indem über alle möglichen Testhalbierungen eine durchschnittliche Testhalbierungsreliabilität berechnet wird (Bortz und Döring, 2006). Für dichotome Variablen kommt die Kuder-Richardson-Formel zum Einsatz (Kubinger, 2009). Für das in der Diplomarbeit angewandte Verfahren können laut Kline (2000, S. 13) Werte ab .7 akzeptiert werden.

Neben Objektivität, Validität und der Reliabilität zählen auch die Schwierigkeit und Trennschärfe zu Kriterien einer guten Testaufgabe (Lienert und Raatz, 1998). Um die Frage zu beantworten, welche Items die einzelnen Entwicklungsbereiche adäquat abbilden, werden Itemtrennschärfen und Itemschwierigkeiten berechnet.

Die Itemschwierigkeit gibt die Rate der Lösungshäufigkeit an. Können viele Probanden ein Item nicht lösen, so gilt dieses Item als schwierig. Einfache Items werden von vielen Personen richtig gelöst (Bortz und Döring, 2006). Die Itemschwierigkeit eines mehrfach abgestuften Items wird mittels der Summe der erreichten Punkte dividiert durch die maximal zu erreichende Punktzahl bei dieser Aufgabe berechnet (Dahl, 1971). Die meisten Items des spielbasierten Itempools sind dichotom. Die Itemschwierigkeit eines dichotomen Items entspricht der Anzahl richtiger Antworten dividiert durch die Gesamtzahl der Antworten – also dem Mittelwert (Lienert & Raatz, 1998).

Der Index kann Werte zwischen 0 (schwierig) und 1 (einfach) annehmen (Bortz & Döring, 2006). Eine Schwierigkeit von .5 bedeutet, dass die Hälfte der Probanden das Item richtig gelöst hat und die andere Hälfte falsch. Items mit mittlerer Schwierigkeit besitzen besondere Relevanz, da Schwierigkeit und Trennschärfe zusammenhängen und bei Items mit einer Schwierigkeit von .5 die Trennschärfe ihr Maximum erreicht (Lienert & Raatz, 1998). Das bedeutet, dass Items mit Schwierigkeiten im mittleren Bereich zwischen .2 und .8 zu bevorzugen sind, da diese Personenunterschiede gut

sichtbar machen und informativer sind als Items, die von fast allen oder niemandem gelöst wurden (Bortz & Döring, 2006).

Die eben erwähnte Trennschärfe ist der wesentliche Kennwert, um die Frage zu beantworten, ob ein Item in der Skala verbleiben soll oder nicht. Unter dem Itemtrennschärfekoeffizienten versteht man die Berechnung der Korrelation jedes Items mit dem Gesamtwert. Er gibt somit an, wie gut das Item das Ergebnis der gesamten Skala bzw. des gesamten Tests repräsentiert. Wer bei einem trennscharfen Item zum Beispiel eine hohe Punktezahl erzielt, erreicht auch im Gesamtergebnis einen hohen Wert. Ein trennscharfes Item unterscheidet („trennt“) somit gut zwischen Personen mit hohen Werten und Personen mit niedrigen Werten die Skala betreffend (Bortz & Döring, 2006). Die in der Arbeit angegebenen korrigierten Item-Skala-Korrelationen berechnen den Trennschärfekoeffizient auf der Basis eines Gesamtwerts, der das jeweilige Item nicht inkludiert, um die Gefahr für Scheinkorrelationen zu verringern (Lienert & Raatz, 1998, Bortz & Döring, 2006).

Die Trennschärfe kann Werte zwischen -1 und +1 annehmen. Ein negativer Wert besagt, dass Testpersonen mit hohen Gesamtwerten bei diesem Item ein schlechtes Ergebnis erzielen und Menschen mit niedrigen Gesamtwerten ein gutes. Ein Koeffizient von 0 zeigt, dass das Item etwa gleich häufig von Personen mit hohen Gesamtwerten und niedrigen Gesamtwerten gelöst wird. Diese Aufgaben sind unbrauchbar. Wünschenswert sind daher möglichst hohe positive Koeffizienten (Lienert & Raatz, 1998). Trennschärfen unter .3 sind als niedrig, zwischen .3 und .5 als mittelmäßig und über .5 als hoch anzusehen (Weise, 1975, S. 219). In der vorliegenden Arbeit wurden daher Items mit einer Trennschärfe unter .3 aus den weiteren Berechnungen entfernt.

In den Analysen wurde folgendermaßen vorgegangen: Einige Buben verweigerten einen Großteil der Items beim Buchlesen. Daraus resultierten viele fehlende Werte beim „Wortschatz“ und „Hineinversetzen in Andere“. Um dem Datenverlust zu entgehen, wurden bei Skalen, bei denen die Daten von weniger als 18 Kindern in die Berechnungen eingehen würden, die fehlenden Werte auf Null gesetzt.

Die Itemkennwerte wurden für jedes Item und die Reliabilität wurde für jede Skala berechnet. Wenn die Analyse der Itemtrennschärfen zur Entfernung von Items aus der Skala führte, wurden für die bereinigte Skala wiederum die Reliabilität und die

Trennschärpen berechnet. Dieser Vorgang wurde so oft wiederholt, bis alle Trennschärpen über dem Wert von .3 lagen.

9.3. Validität

Unter Validität eines Tests ist zu verstehen, dass er tatsächlich jenes psychische Merkmal misst, welches es zu messen beansprucht (Kubinger, 2009, S.55). Zu ermitteln gilt es, in welchem Ausmaß ein Verfahren valide ist (Kubinger, 2009). Es wird eine Beurteilung der inhaltlichen Validität aufgrund von Erfahrungen und inhaltlichen Überlegungen vorgenommen.

Nachdem Items mit zu geringen Itemtrennschärpen aus den Skalen ausgeschlossen worden waren, wurde für diese neuen Skalen die Summe der einzelnen Items berechnet, die dann für die Berechnung der Validität herangezogen wurden. Zur Ermittlung der Konstruktvalidität wurden Alterstrends und Geschlechtsunterschiede berechnet. Um die Kriteriumsvalidität zu eruieren, wurden die Ergebnisse in den Untertests „aktiver Wortschatz“ und „Grammatik“ mit Einschätzungen dieser Bereiche durch die Eltern in einem Fragebogen korreliert. Die genaue Auswertung wird im Weiteren beschrieben.

9.3.1. Auswertung: Alterstrends und Geschlechtsunterschiede

Da der Anstieg der Häufigkeit der richtigen Antworten mit dem Alter eine Minimalanforderung an Entwicklungstests ist, wurde die Korrelation des Alters und der Testleistung als Maß für die Konstruktvalidität berechnet (Petermann & Macha, 2003).

Wenn die Voraussetzung der Normalverteilung für beide zu korrelierende Variablen erfüllt wurde, wurden Pearson-Korrelationen berechnet. War dem nicht der Fall, wurde das Kendall Tau (τ) berechnet. Die Prüfung der Normalverteilungs-Voraussetzung findet sich im Anhang F. Validitätswerte zwischen .40 und .90 sind als mittelmäßig anzusehen, Werte unter .40 als niedrig und über .90 als hoch (Weise, 1975). Es wurde zudem eine Gegenüberstellung der Mittelwerte der beiden Altersgruppen hinsichtlich der einzelnen Skalen vorgenommen.

Als alternative Methode wurde untersucht, ob Unterschiede zwischen der jüngeren und älteren Altersgruppe bestehen. Für die Berechnung der Altersunterschiede wurden zwei Altersgruppen gebildet. Die jüngeren Kinder sind zwischen 24 und 29 Monaten alt und die älteren Kinder zwischen 30 und 35 Monaten.

Von Interesse sind auch mögliche Geschlechtsunterschiede. Eine Vielzahl an Studien und die Normen des „Entwicklungstests 6 Monate bis 6 Jahre“ sowie die Ergebnisse vorangegangener Arbeiten zum spielbasierten Itempool (Birngruber, 2012; Kuchler, 2011) weisen auf tendenziell höhere Mittelwerte und teilweise signifikant bessere Ergebnisse der zweijährigen Mädchen bezüglich der kognitiven bzw. sprachlichen Entwicklung im Vergleich zu Buben hin. Daher gelten Geschlechtsunterschiede in der erwarteten Richtung als Hinweis für die Konstruktvalidität.

Sofern die Voraussetzungen dafür gegeben waren, wurden die alters- und geschlechtsspezifischen Unterschiede mittels zweifaktorieller Varianzanalyse untersucht. Die Diagramme und Werte zur Voraussetzungsprüfung befinden sich im Anhang G. Zur Prüfung der Normalverteilung wurden Histogramme, Schiefe und Kurtosis getrennt für die Gruppen betrachtet. Die Homogenität der Varianzen wurde mittels Levene-Test berechnet. Ein nicht signifikanter Wert ($p > .05$) spricht für die Homogenität der Varianzen. Bei Verletzung der Voraussetzung der Normalverteilung wurde ein Mann-Whitney-U-Test berechnet.

Beim „Als-ob-Spiel“ wurde aufgrund der kleinen Stichprobengröße der exakte Test nach Fisher (Chi²Test) angegeben, um die Vierfeldertafel auf Alters- bzw. Geschlechtsunterschiede zu überprüfen.

Da theoriegeleitet davon ausgegangen wird, dass ältere Kinder bzw. Mädchen bessere Leistungen als jüngere Kinder oder Buben zeigen, wurden die U-Tests und die Chi²-Tests einseitig getestet (Signifikanzniveau=.05). Aufgrund der geringen Stichprobengröße wurden die exakten Signifikanzwerte angeführt (Field, 2009).

9.3.2. Auswertung: Zusammenhang zwischen Elternfragebogen und beobachtetem Verhalten

Als Maß für die Kriteriumsvalidität wurden die im Elternfragebogen erhobenen Fähigkeitsbereiche „Wortschatz aktiv“ und „grammatikalische Fähigkeiten“ mit dem im Spiel beobachteten Verhalten in Beziehung gesetzt. Die Berechnung wurde mittels Korrelationen der Summenscores vorgenommen. Korrelationkoeffizienten im Bereich zwischen 0 und ± 0.1 gelten als kleine Effekte. Bei Werten zwischen $.1$ und $.3$ bzw. zwischen -0.1 und -0.3 spricht man von einem mittleren Effekt und ein Wert über $.5$ bzw. unter -0.5 spricht für einen großen Effekt (Field, 2009, S. 170).

10. Ergebnisse

10.1. Zumutbarkeit

Bevor auf die Hauptgütekriterien eingegangen wird, soll erläutert werden, wie sich die Durchführung hinsichtlich des Nebengütekriteriums Zumutbarkeit darstellte. Das Interesse und die Motivation an den Items bzw. die Verweigerung der Aufgaben durch die Kinder soll beurteilt werden.

Wie auch schon die Vorgängerarbeiten zeigen konnten, interessierten sich alle Kinder unabhängig von Geschlecht und Alter für das Spiel in der Puppenküche. Bis auf zwei Buben begaben sich alle Kinder sofort zur Puppenküche. Zwei Burschen begeisterten sich erst nach einer längeren Aufwärmphase für die Materialien in der Spielbox (siehe „Lösung von der Bezugsperson“ bei Kronberger, in Vorbereitung). Die meisten Kinder wollten nach Ende der Testung noch gerne weiterspielen, was dafür spricht, dass die Spieldauer zumutbar ist.

Viele Aufgaben, wie z.B. die Items zum numerischen Wissen, konnten gut in das gemeinsame Spiel in der Puppenküche eingebaut werden. Drei Buben konnten jedoch ein oder zwei Items der numerischen Aufgaben nicht vorgegeben werden. Bei den numerischen Fähigkeiten merkten einige Mütter an, dass ihr Kind diesbezüglich zuhause bessere Leistungen zeige.

Das Interesse der Kinder an bestimmten Aufgaben war individuell unterschiedlich. Eine Aufgabe gab es jedoch, für die sich alle Kinder sehr interessierten: die Schachtel zur Erhebung des visuellen Gedächtnisses. Bis auf ein Kind bearbeiteten alle diese Aufgaben mit großer Konzentration und einige spielten auch danach noch weiter mit der Box. Die Vorgabe der anderen beiden Items zum visuellen Gedächtnis mittels Bildkarten stieß hingegen auf weniger Interesse. Oft verweigerten die Kinder nach der Vorgabe des Vorzeigebilds, sich auf die Antwortbilder zu konzentrieren oder sich für eine Antwort zu entscheiden.

Schwieriger stellte sich auch die Vorgabe der Items aus dem Buch dar. Das gemeinsame Buchlesen und die darin enthaltenen Items zum „aktiven und passiven Wortschatz“ und zum „Hineinversetzen in Andere“ waren für einige Buben nicht interessant. Von den

insgesamt 24 Items verweigerten sechs Buben die Vorgabe von einem bis zu 13 Items ($MW=7$; $SD=4.15$).

10.2. Objektivität

Der Durchführungsobjektivität wird versucht, mithilfe möglichst hoher Standardisierung zu entsprechen. Die Vorgehensweise bei Begrüßung und Instruktion der Eltern ist im Manual ausführlich beschrieben. Auch das Spielmaterial, die Vorgabe der Items sowie die Instruktion der Kinder sind standardisiert und somit für alle Kinder gleich.

Im Manual wird nicht wie von Renner (2009) gefordert der Umgang mit schwierigen Situationen geregelt. Es fehlt eine standardisierte Anweisung im Manual, wie dem Kind die Spielsituation erklärt werden soll, wie mit Rückfragen des Kindes umgegangen werden soll und es bedarf einer Konkretisierung der Anweisungen zum Umgang mit verweigerndem Verhalten.

Probleme bezüglich der Durchführungsobjektivität ergeben sich aus den Besonderheiten der Testung von Kleinkindern (siehe Kapitel 3 und Kapitel 4.1.1). Den Kindern wurden die Items je nach individuellen Interessen und Motivation vorgegeben und nicht in einer standardisierten Reihenfolge. Darunter leidet die Durchführungsobjektivität, aber der fehlenden intrinsischen Motivation des Kindes, sein bestmögliches Verhalten zu zeigen, bzw. fehlendem Verständnis für die Situation kann nur so entgegengewirkt werden. Ziel ist es, alle Items vorzugeben. Wann jedoch Pausen gemacht werden oder wann das Spiel beendet wird, hängt stark von den Bedürfnissen des jeweiligen Kindes ab.

Außerdem kann die soziale Interaktion nicht, wie von Lienert und Raatz (1998) gefordert, auf ein Minimum reduziert werden, da man nur durch die individuelle Interaktion mit dem Kind sein Vertrauen für das gemeinsame Spiel gewinnen und die Motivation hochhalten kann.

Bei den Items, die für die Kinder eher uninteressant zu sein schienen und bei denen sie der Testleiterin nicht antworteten, wurde die Bezugsperson um die Vorgabe der Items gebeten (Kuchler, et al., 2011). Dieses Vorgehen stellte sich oft als hilfreich heraus.

Manche Eltern spielten jedoch auch mit, ohne dass das Kind dies forderte, und manche Eltern stellten häufig Aufgaben oder gar neue Aufgaben, die nicht im Manual enthalten

sind. Obwohl versucht worden war, dies durch Instruktion der Eltern unmittelbar vor der Spielsituation zu vermeiden (Kuchler et al., 2011), gelang dies nicht immer. Manche Eltern wollten ihre Kinder motivieren, bessere Leistungen zu zeigen. Als schwierig stellte sich auch heraus, dass einige anwesende Mütter die Items zum „Hineinversetzen in Andere“ umformulierten und das Kind fragten, was es denn machen würde, wenn es selbst z.B. schmutzige Hände habe.

Bei beiden Kindern, die nicht Deutsch als Muttersprache haben, wurden während der Spielsituation keinerlei Schwierigkeiten beim Verständnis der deutschen Anweisungen beobachtet.

Für die Auswertungsobjektivität gibt es klare Regeln, wie welches Verhalten zu kodieren ist und im Manual werden Lösungsbeispiele angeführt, die die Kodierung erleichtern. Bei der Kodierung stellten sich der Protokollbogen und die Videoaufnahme als unabdingbare Hilfsmittel heraus, die die Auswertung noch objektiver gestalten. Die „grammatikalischen Fähigkeiten“, das „Als-ob-Spiel“ und das „Zählen“ ließen sich sehr einfach mithilfe der aufgezeichneten Videos einschätzen. In der Testsituation selbst wäre dies nicht möglich gewesen, da man sich bei diesen Skalen auf zu viele Aspekte konzentrieren muss, um dies beiläufig während des Spiels zu machen. Da die meisten Items dichotom nach richtig und falsch verwertet werden, ist die Kodierung meist sehr eindeutig.

Wenn die Kodierregeln im Manual von Kuchler, Sapper, Deimann und Kastner-Koller (2011) Unklarheiten offen gelassen hatten, wurden die entsprechenden Skalen im Vorfeld überarbeitet und präzisiert, um die Standardisierung der Auswertung zu gewährleisten. So wurden unter anderem Definitionen der Zeitangaben und präzisere Definitionen der Ausprägungen von „Lösung von und Kontakt zur Bezugsperson“ vorgenommen.

10.3. Itemanalyse und Reliabilität

Trennschärfen, die unterhalb des kritischen Werts liegen (.3), und Schwierigkeiten, die außerhalb des mittleren Bereichs liegen (.2 bis .8), wurden in den Tabellen farblich markiert und kursiv gedruckt. Neben der jeweiligen Itembezeichnung wird in Klammer die jeweilige Anzahl Kinder angegeben, die die Bearbeitung verweigert haben oder für die die Werte aus anderen Gründen fehlen.

10.3.1. Sprachliche Entwicklung: Itemanalysen und Reliabilität

Die Ergebnisse für den Bereich der Sprache werden für die Skalen „aktiver Wortschatz“, „passiver Wortschatz“ und „grammatikalische Fähigkeiten“ getrennt voneinander dargestellt. Zum Schluss werden die Ergebnisse für die Gesamtskala Sprache angeführt.

Tabelle 4: Reliabilität und Itemtrennschärfen vor und nach der Selektion und Itemschwierigkeiten für die Skala "aktiver Wortschatz" (in Klammer – Anzahl der fehlenden Werte)

N=20 vor Selektion:17 Items nach erster S.:12 Items nach zweiter S.:11 Items	Schwierigkeit	Trennschärfe vor Selektion	Trennschärfe nach erster Selektion	Trennschärfe nach zweiter Selektion
Item 1: Apfel	.55	.27	-	-
Item 2: Banane	.75	.59	.54	.54
Item 3: Birne	.35	.49	.46	.48
Item 4: Karotte	.75	.59	.54	.54
Item 5: Orange	.40	.6	.56	.58
Item 6: Zitrone	.20	.41	.45	.47
Item 7: Badezimmer (5)	.25	.54	.57	.58
Item 8: Kasten (5)	0	-	-	-
Item 9: Kinderzimmer (3)	0	-	-	-
Item 10: Küche (2)	.10	.33	.14	-
Item 11: Lampe (1)	.60	.28	-	-
Item 12: Schlafzimmer (5)	.05	.45	.48	.49
Item 13: Schuh (1)	.70	.61	.6	.59
Item 14: Tasche (3)	.30	.49	.55	.53
Item 15: Torte/Käse	.45	.21	-	-
Item 16: Wasser (2)	.40	.66	.6	.58
Item 17: Wohnzimmer (2)	.05	.45	.48	.49

Reliabilität vor Selektion $\alpha=.838$

Reliabilität nach erster Selektion $\alpha=.835$

Reliabilität nach zweiter Selektion $\alpha=.841$

Fünf der 10 Buben verweigerten mindestens vier bis maximal acht Items ($MW=5.8$; $SD=1.79$) des „aktiven Wortschatzes“. Da in den Analysen mit dem ursprünglichen Datensatz diese fünf Kinder aufgrund der fehlenden Werte ausgeschlossen worden

wären, wurden die fehlenden Daten auf Null gesetzt. Die Reliabilität über alle 17 Items des aktiven Wortschatzes ist mit $\alpha=.838$ als zufriedenstellend einzuschätzen. Die Items „Kasten“ und „Kinderzimmer“ wurden von keinem Kind gelöst und auch die aktive Benennung von drei weiteren Räumen („Küche“, „Schlafzimmer“ und „Wohnzimmer“) erwies sich als zu schwierig. Die erste Berechnung deutet darauf hin, dass das Item „Küche“ im Itempool beibehalten werden kann. Berechnete man jedoch ein zweites Mal die Itemkennwerte, sank die Reliabilität leicht ($\alpha=.835$) und das Item „Küche“ fiel mit einer Trennschärfe von .14 aus der finalen Berechnung der Reliabilität. Bei der erneuten Itemanalyse nach der Selektion der sechs Items mit einer Trennschärfe unter .3 (inkl. „Küche“) erhöhte sich Cronbachs Alpha noch leicht auf $\alpha=.841$. Die Trennschärfen liegen alle über .3.

Tabelle 5: Reliabilität und Itemtrennschärfen vor und nach der Selektion und Itemschwierigkeiten für die Skala "passiver Wortschatz" (in Klammer – Anzahl der fehlenden Werte)

N=20 vor Selektion:8 Items nach Selektion:4 Items	Schwierigkeit	Trennschärfe vor Selektion	Trennschärfe nach Selektion
Item 1: Badewanne (1)	.85	.17	-
Item 2: Besen	.70	.43	.43
Item 3: Bild (2)	.55	.54	.58
Item 4: Fenster (3)	.15	.24	-
Item 5: Fernseher (1)	.55	.24	-
Item 6: Gitterbett (1)	.70	.27	-
Item 7: Stuhl (1)	.80	.36	.34
Item 8: Tür (1)	.90	.36	.34

Reliabilität vor der Selektion $\alpha=.606$

Reliabilität nach der Selektion $\alpha=.633$

Drei Buben verweigerten zwischen zwei und vier Items ($MW=3.33$; $SD=1.15$) des „passiven Wortschatzes“. Um dem Datenverlust entgegenzuwirken ($N=17$) wurden die fehlenden Werte auf Null gesetzt. Während das Item „Fenster“ sehr schwierig ist, wurden die meisten anderen Items eher häufig gelöst. Die Items „Tür“ und „Badewanne“ sind zu einfach. Das Alpha der Skala „passiver Wortschatz“ beträgt über alle acht Items berechnet $\alpha=.606$. Dies entspricht einem nicht akzeptablen Wert. Entfernte man aufgrund der geringen Itemtrennschärfe die vier Items, so ergab die

erneute Berechnung der Reliabilität ein ebenfalls niedriges Alpha von $\alpha=.633$. Zu beachten ist, dass diese geringe Reliabilität basierend auf nur vier Items berechnet wurde.

Tabelle 6: Reliabilität und Itemtrennschärfen vor und nach der Selektion und Itemschwierigkeiten für die Skala "grammatikalische Fähigkeiten"

N=20 vor Selektion:24 Items nach Selektion:16 Items	Schwierigkeit	Trennschärfe vor Selektion	Trennschärfe nach Selektion
Item 1: Satzlänge	.78	.88	.88
Item 2: korrekte Verbstellung in Aussagesätzen	.70	.88	.88
Item 3: Vorkommen von Verneinungen	.95	.29	-
Item 4: korrekte Verbstellung bei der Verneinung	.60	.85	.84
Item 5: Bildung der Verneinung	.80	.85	.88
Item 6: Vorkommen von Alternativfragen	0	-	-
Item 7: Wortstellung bei Alternativfragen	0	-	-
Item 8: Vorkommen von W-Fragen	.75	.67	.70
Item 9: Wortstellung bei W-Fragen	.60	.79	.80
Item 10: Fragewort bei W-Fragen	.68	.87	.88
Item 11: Vorkommen Ja-/ Nein-Fragen	.70	.42	.41
Item 12: Wortstellung Ja-/ Nein-Fragen	.30	.61	.62
Item 13: Frageintonation	.85	.39	.41
Item 14: Vorkommen der Vergangenheitsform	.65	.89	.88
Item 15: Perfekt mit ge- gebildet	.45	.60	.55
Item 16: Hilfsverb bei der Perfektbildung verwendet	.45	.67	.67
Item 17: korrekte Partizipendung	.55	.73	.73
Item 18: Kombination von Haupt- und Nebensatz	.10	.37	.37
Item 19: Anzahl der Verbindungswörter bei Haupt- und Nebensatz	0	-	-
Item 20: korrekte und vollständige Sätze	.55	.81	.80
Item 21: Relativsätze	0	-	-
Item 22: korrektes Relativpronomen	0	-	-
Item 23: Relativpronomen wo	0	-	-
Item 24: Relativsatz ohne Relativpronomen	0	-	-

Reliabilität vor der Selektion $\alpha=.942$

Reliabilität nach der Selektion $\alpha=.944$

Da die Items der Skala „grammatikalische Fähigkeiten“ im Nachhinein mittels Analysen der auf Video aufgezeichneten Äußerungen beurteilt wurden, ist der Datensatz vollständig. Die Reliabilität vor der Selektion der Items ist mit einem Wert von $\alpha=.942$ sehr hoch. Die Items „Vorkommen von Verneinungen“ und „Fragenintonation“ konnten von fast allen Kindern gelöst werden. Die Items betreffend die Alternativfragen, Haupt- und Nebensätze sowie Relativsätze wurden von keinem Kind gezeigt und somit aus den weiteren Analysen ausgeschlossen. Die erneute Berechnung der Reliabilität ergab ein sehr zufriedenstellendes Alpha von $\alpha=.944$.

Bei der Gesamtskala für Sprache (Tabelle siehe Anhang E) wurden die fehlenden Daten von sieben Buben mit Null ergänzt. Die sich für die 49 Items ergebende Reliabilität von $\alpha=.932$ ist sehr zufriedenstellend. Die Trennschärfekoeffizienten der meisten Items des „passiven Wortschatzes“ liegen unter dem kritischen Wert .3. Nach einer ersten Selektion der Items und erneuten Berechnung der Werte sind die Items „Gitterbett passiv“ und „Küche aktiv“ zu wenig trennscharf und schieden bei der finalen Berechnung aus. Die Reliabilität der 28 verbleibenden Items konnte letztlich auf .946 erhöht werden.

10.3.2. Kognitive Entwicklung: Itemanalysen und Reliabilität

Die Ergebnisse der kognitiven Fähigkeiten werden getrennt für die „Theory of Mind“ und das numerische Wissen dargestellt.

10.3.2.1. „Theory of Mind“: Itemanalysen und Reliabilität

Bei der „Theory of Mind“ wird das „Als-ob-Spiel“ und das „Hineinversetzen in Andere“ erhoben.

Die Skala „Als-ob-Spiel“ besteht aus nur einem Item. Daher konnten keine Itemkennwerte berechnet werden. Fünf Kinder zeigten während der Beobachtungssituation gänzlich vorgestelltes Spiel oder verwendeten ein Objekt in einer anderen als seiner vorgesehenen Funktion. Am häufigsten wurde ein „Als-ob-Spiel“ anhand von konkreten Objekten beobachtet. Jedes Kind zeigte irgendeine Form des „Als-Ob-Spiels“ (siehe Tabelle 13).

Tabelle 7: Reliabilität und Itemtrennschärfen vor und nach der Selektion und Itemschwierigkeiten für die Skala "Hineinversetzen in Andere" (in Klammer – Anzahl fehlender Werte)

N=20 vor Selektion:5 Items nach Selektion:4 Items	Schwierigkeit	Trennschärfe vor Selektion	Trennschärfe nach Selektion
Item 1: schmutzige Hände	.60	.46	.46
Item 2: hungrig	.45	.52	.60
Item 3: schlafen (2)	.15	.53	.50
Item 4: Pflaster (2)	.45	.07	-
Item 5: durstig (3)	.25	.56	.63

Reliabilität vor der Selektion $\alpha=.606$

Reliabilität nach der Selektion $\alpha=.633$

Ein bis drei fehlende Items ($MW=1.75$, $SD=0.96$) von vier Buben wurden mit Null ergänzt. Die Reliabilität beträgt $\alpha=.606$. Das Item „Pflaster“ fiel auf Grund der geringen Trennschärfe aus den weiteren Analysen hinaus. Die Reliabilität, die über die vier übrig gebliebenen Items berechnet wurde, ist mit einem Alpha von $\alpha=.633$ nicht akzeptabel.

10.3.2.2. Numerisches Wissen: Itemanalysen und Reliabilität

Beim „numerischen Wissen“ werden sowohl die „aktive“ und die „passive Mengenerfassung“ als auch das „Zählen“ erhoben. Außerdem werden die Ergebnisse für die Gesamtskala erläutert.

Tabelle 8: Reliabilität, Itemtrennschärfen und Itemschwierigkeiten für die Skala „aktive Mengenerfassung“ (in Klammer – Anzahl fehlender Werte)

N=18 2 Items	Schwierigkeit	Trennschärfe
Item 1: Wie viele sind das – Zwei	.44	.5
Item 2: Wie viele sind das – Drei (2)	.17	.5

Reliabilität vor der Selektion $\alpha=.649$

Die „aktive Mengenerfassung“ wird mithilfe von zwei Items erhoben. Die Reliabilität ist nicht zufriedenstellend ($\alpha=.649$). Es gilt zu beachten, dass nur zwei Items in die Berechnung eingegangen sind. Das Item „Wie viele sind das – 3“ ist zu schwierig.

Tabelle 9: Reliabilität und Itemtrennschärfen vor und nach der Selektion und Itemschwierigkeiten für die Skala "passive Mengenerfassung" (in Klammer – Anzahl fehlender Werte)

N=18 vor Selektion: 3 Items nach Selektion.: 2 Items	Schwierigkeit	Trennschärfe vor Selektion	Trennschärfe nach Selektion
Item 1: Gib mir Eins	.89	.47	.45
Item 2: Gib mir Zwei	.72	.54	.45
Item 3: Gib mir Drei (2)	.17	.26	-

Reliabilität vor der Selektion $\alpha=.597$

Reliabilität nach der Selektion $\alpha=.581$

Die Reliabilität der drei Items der „passiven Mengenerfassung“ beträgt $\alpha=.597$ (N=18).. Die Aufgabe „Gib mir Eins“ wurde von fast allen Kindern gelöst. Auch nach der Selektion des Items „Gib mir Drei“ ist die Reliabilität sehr niedrig ($\alpha =.581$).

Tabelle 10: Reliabilität, Itemtrennschärfen und Itemschwierigkeiten für die Skala "Zählen"

N=20 6 Items	Schwierigkeit	Trennschärfe
Item 1: zählen	.65	.72
Item 2: richtig zählen	.10	.46
Item 3: Eins zu Eins Zuordnung	.45	.61
Item 4: stabile Reihenfolge	.45	.68
Item 5: richtige Reihenfolge	.35	.39
Item 6: Kardinalprinzip	.55	.67

Reliabilität vor der Selektion $\alpha=.818$

Die Skala „Zählen“ weist mit einem Wert von .818 eine akzeptable innere Konsistenz auf. Das Item „richtig zählen“ ist für die zweijährigen Kinder zu schwierig. Die Trennschärfen liegen alle über .3, daher werden alle Items beibehalten.

Die Itemanalysen für die Gesamtskala „numerischer Fähigkeiten“ werden im Anhang E angeführt. Die Skala besteht aus elf Items. Fehlende Werte von vier Kindern mussten mit Null ergänzt werden. Die innere Konsistenz ist mit einem Alpha von .866 zufriedenstellend. Da alle Itemtrennschärfen über .3 liegen, können alle Items in der Skala behalten werden. Weitere Berechnungen sind nicht notwendig. Die Items „Gib mir Drei“, „Wie viele sind das: Drei“ und „richtig zählen“ wurden von sehr wenigen Kindern gelöst. Das Item „Gib mir Eins“ ist jedoch sehr einfach.

10.3.3. Gedächtnis: Itemanalysen und Reliabilität

Beim Gedächtnis werden das phonologische und das visuelle Gedächtnis erhoben. Außerdem wird versucht, einen Gesamtwert zu bilden.

Tabelle 11: Reliabilität, Itemtrennschärfen und Itemschwierigkeiten für die Skala "phonologisches Gedächtnis" (in Klammer – Anzahl fehlender Werte)

N=20 vor Selektion:3 Items nach Selektion:0 Items	Schwierigkeit	Trennschärfe
Item 1: Gib mir a	.90	.22
Item 2: Gib mir a-b (1)	.70	.17
Item 3: Gib mir a-b-c (4)	.25	.2

Reliabilität vor der Selektion $\alpha=.335$

Die Skala „phonologisches Gedächtnis“ beinhaltet drei Items. Vier Kinder verweigerten ein oder zwei Items. Die Reliabilität ist sehr gering und mit einem Wert von $\alpha=.335$ inakzeptabel. Alle Itemtrennschärfekoeffizienten liegen unter dem Wert von .3. Weitere Analysen sind daher nicht möglich und die Skala erweist sich als ungeeignet.

Tabelle 12: Reliabilität und Itemtrennschärfen vor und nach der Selektion und Itemschwierigkeiten für die Skala "visuelles Gedächtnis" (in Klammer – Anzahl fehlender Werte)

N=19 vor Selektion:5 Items nach erster Selektion:3 Items	Schwierigkeit	Trennschärfe vor Selektion	Trennschärfe nach Selektion
Item 1: Bilderset: Bär	.47	.41	.55
Item 2: Bilderset: Haus (1)	.63	.11	-
Item 3: Memory Sonne	.68	.33	.47
Item 4: Memory Haus	.89	.13	-
Item 5: Memory Baum	.68	.45	.47

Reliabilität vor der Selektion $\alpha=.511$

Reliabilität nach der Selektion $\alpha=.682$

Ein Bub verweigerte die Vorgabe des Items „Bilderset: Haus“. Die Daten von 19 Kindern gehen in die Analysen ein. Eine erste Reliabilitätsberechnung ergibt ein Alpha von $\alpha=.511$. Nach der Selektion der Items „Bilderset: Haus“ und „Memory Haus“ bleibt die Reliabilität für die drei übrig gebliebenen Items mit einem Wert von $\alpha=.682$ zu gering. Das Item „Memory Haus“ ist zu einfach.

Wenn man eine Gesamtskala über alle acht Items des Gedächtnisses bildet, ergibt sich eine sehr geringe Reliabilität von $\alpha=.497$. Die Items „Gib mir a“ und „Memory Haus“ wurden von fast allen Kindern gelöst und sind daher zu einfach. Bis auf die Itemtrennschärfen von „Bilderset: Bär“ und „Memory Haus“ liegen alle unter einem Wert von .3. Eine Gesamtskala über diese beiden Items des „visuellen Gedächtnisses“ ($\alpha=.623$) ist wenig sinnvoll.

10.4. Inhaltvalidität

Für die Erfüllung der inhaltlichen Validität spricht, dass die Items auf Basis von entwicklungspsychologischen Forschungsergebnissen konstruiert wurden und somit für das dritte Lebensjahr relevante Entwicklungsbereiche abbilden.

Die Items, die sich im Rahmen dieser Arbeit als zu schwierig oder zu einfach herausgestellt haben (Kapitel 10.3. Itemanalyse und Reliabilität), sollten in nachfolgenden Vorgaben aus dem Itempool genommen werden, um die inhaltliche Validität weiter zu erhöhen.

10.5. Konstruktvalidität

Zur Beurteilung der Konstruktvalidität wurden die Korrelation des Alters mit den Summenscores, sowie Alters- und Geschlechtsunterschiede berechnet.

10.5.1. Sprachliche Entwicklung: Alterstrends und Geschlechtsunterschiede

Tabelle 13: Mittelwerte und Standardabweichungen der beiden Altersgruppen und der Geschlechter hinsichtlich der Skalen der sprachlichen Entwicklung

Skala	jung		alt		männlich		weiblich	
	MW	SD	MW	SD	MW	SD	MW	SD
aktiver Wortschatz	3.54	2.47	5.43	3.55	2.9	2.33	5.5	3.03
passiver Wortschatz	2.69	1.18	3.43	1.12	2.3	1.25	3.6	0.7
grammatikalische Fähigkeiten	8.18	4.97	12.12	5.27	8.45	5.14	10.67	5.48
Sprache gesamt	11.87	7.16	19.55	8.36	11.35	7.99	17.77	7.62

In Tabelle 13 ist die deskriptive Statistik für die Skalen der sprachlichen Entwicklung angegeben. Die Mittelwerte der älteren Kinder und der Mädchen sind tendenziell größer als die der jüngeren Kinder bzw. Buben.

Tabelle 14: Überblick über die Ergebnisse zu den Geschlechts- und Altersunterschieden sowie der Korrelation zwischen dem Alter und den Summenscores der Skalen zur sprachlichen Entwicklung

Skala	Geschlechtsunterschied	Altersunterschied	Korrelation
aktiver Wortschatz	$F(1,16)=2.01$ $p=.175$	$F(1,16)=.85$ $p=.37$	$r=.42$ $p_{(einseitig)}=.031$
passiver Wortschatz	$U=21.5$ $z=-2.33$ $p_{(einseitig)}=.01$	$U=30$ $z=-1.33$ $p_{(einseitig)}=.116$	$\tau=.29$ $p_{(einseitig)}=.059$
grammatikalische Fähigkeiten	$U=40$ $z=-0.75$ $p_{(einseitig)}=.235$	$U=25$ $z=-1.63$ $p_{(einseitig)}=.054$	$\tau=.47$ $p_{(einseitig)}=.006$
Sprache gesamt	$F(1,16)=0.78$ $p=.389$	$F(1,16)=3.30$ $p=.088$	$r=.58$ $p_{(einseitig)}=.004$

Tabelle 14 gibt einen Überblick über die Ergebnisse zur Konstruktvalidität der „sprachlichen Entwicklung“. Die Ergebnisse der einzelnen Skalen werden nun genauer beschrieben.

Bei der Skala „aktiver Wortschatz“ erzielten die Mädchen mit 5.5 ($SD=3.03$) einen etwas höheren Mittelwert als die Buben ($MW=2.9$, $SD=2.33$) und die Älteren erreichten mit 5.43 ($SD=3.55$) einen höheren Mittelwert als die Jüngeren ($MW=3.54$, $SD=2.47$).

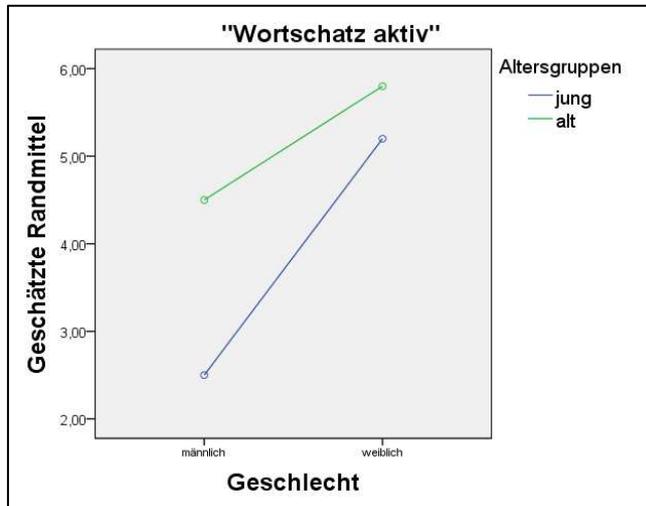


Abbildung 1: Interaktionsdiagramm zur Skala "Wortschatz aktiv" hinsichtlich Geschlecht und der Alter

Die Voraussetzungen für die zweifaktorielle ANOVA werden hier erfüllt (Anhang G). Es konnten mittels dieser Berechnungsmethode weder signifikante Unterschiede zwischen den Geschlechtern ($F(1,16)=2.01$, $p=.175$), noch zwischen den Altersgruppen ($F(1,16)=.85$, $p=.37$) festgestellt werden. Es konnten ebenfalls keine Interaktionseffekte

gefunden werden ($F(1,16)=.25$, $p=.626$). Die Abbildung 1 stellt das Interaktionsdiagramm dar.

Berechnet man hingegen die Korrelation zwischen dem Gesamtwert des „aktiven Wortschatzes“ mit dem Alter ergibt sich ein signifikanter, positiver Zusammenhang ($r=.42$, $p_{(einseitig)}=.031$), der für eine mittelmäßig hohe Validität spricht.

Da der Summenscore der Skala „passiver Wortschatz“ in Bezug auf die Alters- und Geschlechtsgruppen nicht normalverteilt ist, wurden zwei Mann-Whitney-U-Tests berechnet. Der Test zeigte bezogen auf das Geschlecht ein signifikantes Ergebnis ($U=21.5$, $z=-2.33$, $p_{(einseitig)}=.01$). Die Mädchen erzielten signifikant höhere Werte ($MW=3.6$, $SD=.7$) als die Buben ($MW=2.3$, $SD=1.25$). Zwischen den jüngeren ($MW=2.69$, $SD=1.18$) und den älteren Kindern ($MW=3.43$, $SD=1.13$) konnte kein signifikanter Unterschied festgestellt werden ($U=30$, $z=-1.33$, $p_{(einseitig)}=.116$). Auch die nicht signifikante Korrelation des „passiven Wortschatzes“ ($r=.29$, $p_{(einseitig)}=.059$) mit dem Alter spricht dafür, dass dieser Untertest nicht valide ist.

Auch bei der Skala „grammatikalische Fähigkeiten“ wurden aufgrund der Verletzung der Normalverteilungs-Voraussetzung Mann-Whitney-U-Tests gerechnet. Die Buben ($MW=8.45$, $SD=5.14$) und Mädchen ($MW=10.67$, $SD=5.48$) der Stichprobe unterschieden sich nicht signifikant voneinander ($U=40$; $z=-0.75$, $p_{(einseitig)}=.235$). Die jüngere ($MW=8.18$, $SD=4.97$) und die ältere Gruppe ($MW=12.12$, $SD=5.27$) unterschieden sich nur sehr knapp nicht signifikant voneinander ($U=25$, $z=-1.63$, $p_{(einseitig)}=.054$). Die Korrelation zwischen dem Alter und dieser Skala ergibt jedoch einen signifikanten, positiven Wert ($\tau=.47$, $p_{(einseitig)}=.006$).

Die Daten der Skala „Sprache gesamt“ erfüllen die Voraussetzungen für eine zweifaktorielle Varianzanalyse. Zwischen Buben ($MW=11.35$, $SD=7.99$) und Mädchen

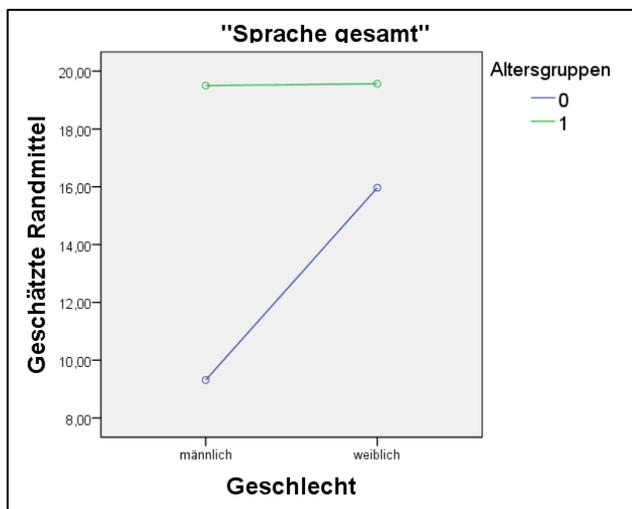


Abbildung 2: Interaktionsdiagramm zur Skala "Sprache gesamt" hinsichtlich Geschlecht und Alter

($MW=17.77$, $SD=7.62$) konnte kein signifikanter Unterschied festgestellt werden ($F(1,16)=0.78$, $p=.389$). Auch der Haupteffekt Alter war nicht signifikant ($F(1,16)=3.30$, $p=.088$). Die jüngeren Kinder erzielten einen durchschnittlichen Wert von 11.87 ($SD=7.16$) und die älteren erzielen einen höheren Mittelwert von 19.55 ($SD=8.36$).

Der Interaktionseffekt war ebenfalls nicht signifikant ($F(1,16)=.75$, $p=.389$). Abbildung 2 zeigt das Interaktionsdiagramm. Die Korrelation zwischen Alter und Summenscore spricht hingegen mit einem signifikanten Zusammenhang von $r=.58$ ($p_{(einseitig)}=.004$) für eine mittelmäßig hohe Validität.

10.5.2. Kognitive Entwicklung

10.5.2.1. „Theory of Mind“: Alterstrends und Geschlechtsunterschiede

Tabelle 15: Mittelwerte und Standardabweichung der beiden Altersgruppen und der Geschlechter hinsichtlich der Skalen der „Theory of Mind“

Skala	jung		alt		männlich		weiblich	
	MW	SD	MW	SD	MW	SD	MW	SD
Als-ob Spiel	.62	.22	.64	.24	.65	.24	.6	.21
Hineinversetzen in Andere	.77	.83	2.71	1.38	1	1.25	1.9	1.45

In Tabelle 15 ist die deskriptive Statistik für die Skalen der „Theory of Mind“ angegeben. Die Mittelwerte der älteren Kinder und der Mädchen sind beim „Hineinversetzen in andere“ tendenziell größer als die der jüngeren Kinder bzw. Buben. Beim „Als-ob Spiel“ ist dies nicht der Fall.

Tabelle 16: Überblick über die Ergebnisse zu den Geschlechts- und Altersunterschieden sowie der Korrelation zwischen dem Alter und den Summenscores der Skalen zur „Theory of Mind“

Skala	Geschlechtsunterschied	Altersunterschied	Korrelation
Als-ob Spiel	$\chi^2=.267$ $p_{(einseitig)}=.5$	$\chi^2=.073$ $p_{(einseitig)}=.594$	$\tau=.061$ $p_{(einseitig)}=.384$
Hineinversetzen in Andere	$U=32.5$ $z=-1.38$ $p_{(einseitig)}=.101$	$U=13$ $z=-2.68$ $p_{(einseitig)}=.004$	$\tau=.57$ $p_{(einseitig)}=.001$

Tabelle 16 gibt einen Überblick über die Ergebnisse zur Konstruktvalidität. Die Ergebnisse der einzelnen Skalen werden nun genauer beschrieben.

Das „Als-ob-Spiel“ wird nur mittels eines Items erhoben. Für die Kontingenztafel (Tabelle 17) wurde jeweils ein Chi² Test für Alter und Geschlecht berechnet.

Tabelle 17: Häufigkeiten der Antwortalternativen des Items „Als-ob-Spiel“ hinsichtlich Geschlecht und der Altersgruppen

N=20 1 Item	Mädchen	Buben	jüngere	ältere	gesamt
gänzlich vorgestellt/ anders verwendet (1)	2	3	3	2	5
anhand von konkreten Objekten (0,5)	8	7	10	5	15
kein Als-ob-Spiel (0)	0	0	0	0	0

Die jüngeren Kinder erzielten einen Mittelwert von .62 ($SD=.22$), der nur sehr gering unter dem Mittelwert der älteren Kinder liegt ($MW=.64$, $SD=.24$). Die Burschen erzielten ein geringfügig höheres durchschnittliches Ergebnis ($MW=.65$, $SD=.24$) als die Mädchen ($MW=.6$, $SD=.21$). Es konnte weder ein Geschlechts- ($\chi^2=.267$, $p_{(einseitig)}=.5$) noch ein Altersunterschied festgestellt werden ($\chi^2=.073$, $p_{(einseitig)}=.594$). Auch die Korrelation des Alters bzgl. des „Als-ob-Spiels“ ($\tau=.061$, $p_{(einseitig)}=.384$) war nicht signifikant.

Für die Skala „Hineinversetzen in Andere“ sind die Daten in den Gruppen nicht normalverteilt, weswegen U-Tests berechnet wurden. Die Buben ($MW=1$, $SD=1.25$) unterschieden sich nicht signifikant von den Mädchen ($MW=1.9$, $SD=1.45$; $U=32.5$, $z=-1.38$, $p_{(einseitig)}=.101$). Der altersbezogene U-Test ergab ein signifikant einseitiges Ergebnis ($U=13$, $z=-2.68$, $p_{(einseitig)}=.004$). Die älteren Kinder ($MW=2.71$, $SD=1.38$) erzielten signifikant höhere Werte als die jüngeren Kinder ($MW=.77$, $SD=.83$). Es konnte auch mittels Korrelation ein signifikanter, mittlerer Zusammenhang zwischen Alter und „Hineinversetzen in Andere“ bestätigt werden ($\tau=.57$, $p_{(einseitig)}=.001$).

10.5.2.2. Numerisches Wissen: Alterstrends und Geschlechtsunterschiede

Tabelle 18: Mittelwerte und Standardabweichung der beiden Altersgruppen und der Geschlechter hinsichtlich der Skalen des „numerischen Wissens“

Skala	jung		alt		männlich		weiblich	
	MW	SD	MW	SD	MW	SD	MW	SD
aktive Mengenerfassung	.31	.63	1	.82	.2	.63	.9	.74
passive Mengenerfassung	1.46	.66	1.71	.76	1.4	.7	1.7	.67
Zählen	1.92	1.93	3.71	1.89	1.9	1.91	3.2	2.1
numerisch gesamt	3.77	2.74	6.71	3.35	3.6	2.95	6	3.16

In Tabelle 18 ist die deskriptive Statistik für die Skalen des „numerischen Wissens“ angegeben. Die Mittelwerte der älteren Kinder und der Mädchen sind tendenziell größer als die der jüngeren Kinder bzw. Buben.

Tabelle 19: Überblick über die Ergebnisse zu den Geschlechts- und Altersunterschieden sowie der Korrelation zwischen dem Alter und den Summenscores der Skalen zum „numerischen Wissen“

Skala	Geschlechtsunterschied	Altersunterschied	Korrelation
aktive Mengenerfassung	$U=22.5$ $z=-2.37$ $p_{(einseitig)}=.01$	$U=23$ $z=-2.04$ $p_{(einseitig)}=.038$	$\tau=.39$ $p_{(einseitig)}=.042$
passive Mengenerfassung	$U=36.5$ $z=-1.21$ $p_{(einseitig)}=.168$	$U=33.5$ $z=-1.13$ $p_{(einseitig)}=.177$	$\tau=.25$ $p_{(einseitig)}=.197$
Zählen	$U=33$ $z=-1.31$ $p_{(einseitig)}=.1$	$U=22.5$ $z=-1.86$ $p_{(einseitig)}=.036$	$\tau=.35$ $p_{(einseitig)}=.045$
numerisch gesamt	$U=29.5$ $z=-1.56$ $p_{(einseitig)}=.062$	$U=20.5$ $z=-1.99$ $p_{(einseitig)}=.023$	$\tau=.37$ $p_{(einseitig)}=.016$

Tabelle 19 gibt einen Überblick über die Ergebnisse zur Konstruktvalidität des „numerischen Wissens“. Die Ergebnisse der einzelnen Skalen werden nun genauer beschrieben.

Da weder die Normalverteilung noch die Homogenität der Varianzen gegeben waren, wurden zur „aktiven Mengenerfassung“ U-Tests gerechnet. Zwischen den Geschlechtern besteht ein signifikanter Unterschied ($U=22.5$, $z=-2.37$, $p_{(einseitig)}=.01$), wobei die Mädchen ($MW=.9$, $SD=.74$) höhere Werte erzielten als die Buben ($MW=.2$, $SD=.63$). Es konnte auch ein signifikanter, einseitiger Unterschied zwischen den beiden Altersgruppen festgestellt werden ($U=23$, $z=-2.04$, $p_{(einseitig)}=.038$). Die älteren Kinder ($MW=1$, $SD=.82$) erbrachten durchschnittlich höhere Leistungen als die jüngeren Kinder ($MW=.31$, $SD=.63$). Die Korrelation zwischen Alter und Gesamtwert ergibt einen signifikanten, positiven Wert ($\tau=.39$, $p_{(einseitig)}=.042$).

Da die Daten der Skala „passive Mengenerfassung“ nicht normalverteilt sind, wurden U-Tests gerechnet. Zwischen Buben ($MW=1.4$, $SD=.7$) und Mädchen ($MW=1.7$, $SD=.67$) konnte kein signifikanter Unterschied bzgl. des Summenscores „passives Mengenverständnis“ festgestellt werden ($U=36.5$, $z=-1.21$, $p_{(einseitig)}=.168$). Auch die Kinder, die zwischen 24 und 29 Monaten alt sind ($MW=1.46$, $SD=.66$) und die Kinder, die zwischen 30 und 35 Monaten alt sind ($MW=1.71$, $SD=.76$), unterscheiden sich nicht signifikant voneinander ($U=33.5$, $z=-1.13$, $p_{(einseitig)}=.177$). Auch der nicht signifikante

Kendall-Tau Wert von $\tau=.25$ ($p_{(\text{einseitig})}=.197$) weist darauf hin, dass diese Skala nicht valide ist.

Beim „Zählen“ wurde ebenfalls aufgrund der Verletzung der Normalverteilung auf die parameterfreien U-Tests zurückgegriffen. Buben ($MW=1.9$, $SD=1.91$) und Mädchen ($MW=3.2$, $SD=2.1$) unterscheiden sich in ihren Werten nicht signifikant voneinander ($U=33$, $z=-1.31$, $p_{(\text{einseitig})}=.1$). Die älteren Kinder ($MW=3.71$, $SD=1.89$) erzielten signifikant höhere Leistungen als die jüngeren Kinder ($MW=1.92$, $SD=1.93$; $U=22.5$, $z=-1.86$, $p_{(\text{einseitig})}=.036$). Für eine geringe Validität spricht die signifikante Korrelation zwischen Alter in Monaten und der Skala „Zählen“ mit einem Werte von $\tau=.35$ ($p_{(\text{einseitig})}=.045$).

Dem Ergebnis der U-Tests zufolge unterscheiden sich die Buben ($MW=3.6$, $SD=2.95$) in Bezug auf die Skala „numerisch gesamt“ nicht signifikant von den Mädchen ($MW=6$, $SD=3.16$; $U=29.5$, $z=-1.56$, $p_{(\text{einseitig})}=.062$). Es konnte ein Altersunterschied festgestellt werden ($U=20.5$, $z=-1.99$, $p_{(\text{einseitig})}=.023$). Die älteren Kinder ($MW=6.71$, $SD=3.35$) sind den jüngeren Kindern ($MW=3.77$, $SD=2.74$) überlegen. Die Skala „numerisch gesamt“ steht in einem signifikanten, mittelmäßig hohen Zusammenhang mit dem Alter ($\tau=.37$, $p_{(\text{einseitig})}=.016$). Der Wert spricht für eine geringe Validität.

10.5.3. Gedächtnis: Alterstrends und Geschlechtsunterschiede

Tabelle 20: Mittelwerte und Standardabweichung der beiden Altersgruppen und der Geschlechter hinsichtlich der Skalen des „Gedächtnisses“

Skala	jung		alt		männlich		weiblich	
	MW	SD	MW	SD	MW	SD	MW	SD
Visuelles Gedächtnis	1.46	1.27	2.29	.76	1.5	1.27	2	1.05

In Tabelle 20 ist die deskriptive Statistik für die Skala „visuelles Gedächtnis“ angegeben. Die Mittelwerte der älteren Kinder und der Mädchen sind tendenziell größer als die der jüngeren Kinder bzw. Buben. Die Itemanalysen (Kapitel 10.3.3. Gedächtnis: Itemanalyse und Reliabilität) zeigten, dass sowohl die Skala „phonologisches Gedächtnis“ als auch die Skala „Gedächtnis gesamt“ ungeeignet ist. Daher wurden hier keine weiteren Berechnungen vorgenommen.

Tabelle 21: Überblick über die Ergebnisse zu den Geschlechts- und Altersunterschieden sowie der Korrelation zwischen dem Alter und den Summenscores der Skalen zum „Gedächtnis“

Skala	Geschlechtsunterschied	Altersunterschied	Korrelation
Visuelles Gedächtnis	$U=38.5$ $z=-.902$ $p_{(einseitig)}=.215$	$U=28.5$ $z=-1.4$ $p_{(einseitig)}=.095$	$\tau=.49$ $p_{(einseitig)}=.008$

Tabelle 21 gibt einen Überblick über die Ergebnisse zur Konstruktvalidität zum „visuellen Gedächtnis“.

Die Skala erfüllt die Voraussetzung der Normalverteilung nicht. Die U-Tests zeigten weder einen Geschlechts- ($U=38.5$, $z=-.902$, $p_{(einseitig)}=.215$), noch einen Altersunterschied ($U=28.5$, $z=-1.4$, $p_{(einseitig)}=.095$). Der Mittelwert der Mädchen liegt mit einem Wert von 2 ($SD=1.05$) knapp über dem der Burschen ($MW=1.5$, $SD=1.27$). Die älteren Kinder erzielten einen Mittelwert von 2.29 ($SD=.76$) und die jüngeren Kinder einen Wert von 1.46 ($SD=1.27$). Das „visuelle Gedächtnis“ steht jedoch in einem signifikanten, mittelmäßig hohen Zusammenhang mit dem Alter ($\tau=.49$, $p_{(einseitig)}=.008$). Der Wert spricht für eine mittelmäßig hohe Validität.

10.6. Kriteriumsvalidität

Im Folgenden werden die Ergebnisse zum „aktiven Wortschatz“ und zu „grammatikalischen Fähigkeiten“ dargestellt. Die Ergebnisse zur Selbstständigkeit finden sich bei Kronberger (in Vorbereitung).

Für die Signifikanzprüfung einer Pearson-Korrelation ist die Voraussetzung der Normalverteilung nicht gegeben (Anhang I). Da die Anzahl an gleichen Rängen hoch ist, wurde Kendall Tau berechnet.

Für den „aktiven Wortschatz“ ergab sich somit ein signifikanter mittlerer Zusammenhang mit mittlerer Effektstärke ($\tau= .431$, $p=.012$).

Die Informationen bezüglich der „grammatikalischen Fähigkeiten“ aus dem Fragebogen korrelieren signifikant und hoch mit den aus den Beobachtungen gewonnenen ($\tau=.701$, $p=.000$).

Beide Werte sprechen für eine mittelmäßig hohe Validität.

11. Diskussion

Es existieren nur sehr wenige deutschsprachige Verfahren zur Erfassung der Entwicklung zweijähriger Kinder (Quaiser-Pohl & Köhler, 2010). Die vorhandenen Verfahren erfüllen oft die Hauptgütekriterien nicht hinreichend. Viele Entwicklungstests haben veraltete oder amerikanische Normen, deren Repräsentativität für die aktuelle Anwendung bzw. für den deutschsprachigen Raum fraglich ist (Renner, 2009). In der testähnlichen Situation fällt es den Kleinkindern zudem häufig schwer, die Aufmerksamkeit aufrecht zu erhalten. Es mangelt ihnen an Verständnis für die Situation und intrinsischer Motivation, die Aufgaben zu bearbeiten. Aus diesem Grund wurden spielbasierte Ansätze entwickelt, die eine Beobachtung des Spiels des Kindes zur Erhebung der Entwicklung darstellen (Witzlack, 2001).

Das Ziel des Projekts von Kastner-Koller und Deimann ist daher die Entwicklung eines Itempools, der eine Kombination der beiden Ansätze darstellen soll. 2011 wurde die erste Version des Manuals von Kuchler, Sapper, Deimann und Kastner-Koller veröffentlicht.

Im Rahmen dieser Diplomarbeit wurde der spielbasierte Itempool basierend auf der Kritik und den testtheoretischen Analysen der Vorgängerinnen (Birngruber, 2012; Fuchs-Garderer, 2012; Kuchler 2011; Putzer, in Vorbereitung; Sapper, 2011; Sindelar, in Vorbereitung) in Zusammenarbeit mit Kronberger (in Vorbereitung) weiterentwickelt und 20 Kindern vorgegeben.

Da das Verfahren sowohl möglichst spielbasiert sein als auch die Gütekriterien erfüllen soll, wurden die gewonnenen Daten hinsichtlich der Zumutbarkeit, der Hauptgütekriterien und der Itemkennwerte beurteilt.

Die Ergebnisse werden im Folgenden für jeden Entwicklungsbereich getrennt diskutiert. Zuletzt wird eine zusammenfassende Bewertung des Itempools vorgenommen.

11.1. Sprachliche Entwicklung

Der Itempool beinhaltet zur Erhebung der sprachlichen Entwicklung die Skalen „aktiver Wortschatz“, „passiver Wortschatz“, „grammatikalische Fähigkeiten“ und „Sprache gesamt“.

Zu Beginn der Testung waren die meisten Kinder noch schüchtern und wortkarg. Im Laufe der Testung sprachen die Kinder jedoch viel, kommentierten das Spiel und antworteten meist auf Fragen. Die Mädchen waren häufig gesprächiger als die Buben.

Die meisten Items zum aktiven und passiven Wortschatz konnten gut in das gemeinsame Spiel eingebaut werden. Bei der Vorgabe der Items, die im ToM-Buch enthalten sind, verweigerten sechs Buben durchschnittlich sieben von insgesamt 24 Items. Die Aufgaben zum sprachlichen Bereich können als zumutbar angenommen werden, wobei jedoch die zu schwierigen Items (Benennung der Zimmer, Kasten, etc.) aus dem Buch entfernt werden sollten, um die Itemanzahl des Buches zu minimieren und die Zumutbarkeit und die Motivation der Kinder weiter zu steigern.

Die Vorgabe der Items und die Kodierung bzw. Auswertung der sprachlichen Skalen können als objektiv erachtet werden. Hierzu war die Videoaufzeichnung ein unverzichtbares Hilfsmittel. Die Instruktion der Eltern, Gegenstände nicht zu benennen, da die Benennung Aufgabe der Kinder sei, stellte sich als der Objektivität zuträglich heraus. Die Kinder, die nicht Deutsch als Muttersprache hatten, hatten keine Schwierigkeiten, die Anweisungen zu verstehen. Die Reihenfolge der Items ist sehr variabel und daher wenig objektiv. Diese Einbuße der Objektivität muss jedoch zugunsten einer unabdingbaren individuellen, auf die Interessen des Kindes angepassten Vorgehensweise hingenommen werden. Es wird davon ausgegangen, dass die Kinder mehr sprechen, wenn sie sich wohl fühlen und Vertrauen zur Testleiterin haben.

Die Reliabilität der Skala „aktiver Wortschatz“ ($\alpha=.841$; 11 Items) ist im akzeptablen ($>.7$), die der Skalen „grammatikalische Fähigkeiten“ ($\alpha=.944$; 16 Items) und „Sprache gesamt“ ($\alpha=.946$; 28 Items) im hohen Bereich ($>.9$). Die Reliabilität der Skala „passiver Wortschatz“ ergibt einen Wert von $.633$. Dieser Wert ist inakzeptabel, jedoch gilt es zu beachten, dass nach der Itemselektion nur vier Items in die Berechnung eingingen. Aus dem Itempool werden vor allem Items zur Benennung von Räumen und zum

Vorkommen von Alternativfragen und Relativsätzen entfernt, da sie sehr schwierig sind bzw. von keinem Kind gelöst wurde.

Auch bei den Vorgängerarbeiten wurden ähnlich hohe Reliabilitäten erzielt, wobei immer die Skala „aktiver Wortschatz“ eine mittelmäßige (.82-.893) Reliabilität aufweist und die Skalen „passiver Wortschatz“ (.93-.965) sowie „grammatikalische Fähigkeiten“ (.91-.93) eine hohe Reliabilität aufweisen (Fuchs-Garderer, 2012; Kuchler, 2011; Sindelar, in Vorbereitung).

Die Ergebnisse der Berechnung der geschlechts- und altersspezifischen Leistungsunterschiede sprechen dafür, dass die Skalen der sprachlichen Entwicklung nicht konstruktvalide sind. Nur bezüglich der Skala „passiver Wortschatz“ konnte ein signifikanter Geschlechtsunterschied zugunsten der Mädchen festgestellt werden. Über alle Skalen hinweg sind die Altersunterschiede nicht signifikant. Betrachtet man die Mittelwerte genauer, kann jedoch die Tendenz festgestellt werden, dass in allen Skalen die Mädchen höhere Werte als die Buben, und die älteren Kinder höhere Werte als die jüngeren erzielen. Die signifikant positiven Zusammenhänge zwischen dem Alter und den Skalen „aktiver Wortschatz“, „grammatikalische Fähigkeiten“ und „Sprache gesamt“ deutet auf mittelmäßig hohe Validitäten (.42-.58) dieser Skalen hin. Nur die Skala „passiver Wortschatz“ ist auch laut Korrelationsberechnungen nicht valide.

Die Ergebnisse dieser Studie und der Vorgängerarbeiten sind uneinheitlich. Im kompletten Gegensatz zu den soeben dargestellten Ergebnissen ergeben die Berechnungen von Fuchs-Garderer (2012) signifikante Alters- und Geschlechtsunterschiede hinsichtlich des „aktiven Wortschatzes“ und Alters-, aber keine Geschlechtsunterschiede hinsichtlich des „passiven Wortschatzes“. Bei Sindelar (in Vorbereitung) ergeben sich wiederum andere Ergebnisse. Hier bestehen signifikante Unterschiede zwischen den Geschlechtern hinsichtlich des „aktiven Wortschatzes“, aber es zeigt sich kein signifikanter Zusammenhang zwischen dem Alter und der Skala. Beim „passiven Wortschatz“ werden keine signifikanten Ergebnisse erzielt. In der Arbeit von Kuchler (2011) korrelieren alle Skalen – auch der „passive Wortschatz“ – signifikant mit dem Alter und weisen auf mittelmäßige Validitäten hin.

Alle Vorgängerarbeiten kommen bei den grammatikalischen Fähigkeiten zum selben Ergebnis: ältere Kinder erzielen signifikant höhere Leistungen und es besteht kein

signifikanter Unterschied zwischen den Geschlechtern. Gründe für die unterschiedlichen Ergebnisse können unter anderem in der begrenzten Generalisierbarkeit der der kleinen Stichprobe liegen und Unterschiede in den Stichproben. Zum Beispiel beinhalten die Stichproben der meisten Vorgängerarbeiten einen sehr hohen Anteil an Kindern von Eltern mit akademischer Ausbildung. In der vorliegenden Arbeit verteilen sich die Ausbildungen der Eltern gleich auf die unterschiedlichen Kategorien. Eine Schwierigkeit die Ergebnisse zu vergleichen, besteht weiters darin, dass der Itempool zahlreichen Überarbeitungen unterlegen ist und sich die Items vor allem beim „aktiven und passiven Wortschatz“ in den unterschiedlichen Erhebungen unterscheiden. Weitere einheitliche Erhebungen an einer großen Stichprobe sind daher von Nöten.

Bezüglich der Kriteriumsvalidität ergibt sich ein durchwegs positives Bild. Sowohl die vorliegende Arbeit als auch die Arbeit von Fuchs-Garderer (2012) ergeben signifikante Zusammenhänge zwischen den Skalen „aktiver Wortschatz“ bzw. „grammatikalische Fähigkeiten“ und den Angaben der Eltern im Fragebogen. Die Werte sprechen für eine mittelmäßig hohe Kriteriumsvalidität.

11.2. Kognitive Entwicklung

Da die „Kognition“ die gesamte Bandbreite psychologischer Prozesse, die mentalen Ereignissen unterliegen – wie z.B. die Wahrnehmung von Informationen, das Verstehen, das Gedächtnis und die Intelligenz – umfasst (Solso, 2008), beinhaltet der Bereich „kognitive Entwicklung“ des spielbasierten Itempools eine breite Spanne von Skalen.

Der ursprüngliche Itempool von Kuchler et al. (2011) beinhaltet neben den Skalen „Theory of Mind“ und „numerisches Wissen“ auch die Skala „deduktives Denken“. Letztgenannte Skala wurde bei der aktuellen Erhebung nicht vorgegeben, da in den Vorgängerarbeiten viele Kinder die Bearbeitung dieser Items verweigerten. In der Diskussion wird daher nur kurz auf diesen Bereich eingegangen

11.2.1. „Theory of Mind“

Die „Theory of Mind“ (ToM) oder „naive Alltagspsychologie“ ist die Fähigkeit, anderen oder sich selbst mentale Zustände wie Wünsche, Überzeugungen, Gefühle und Absichten zuzuschreiben (Goswami, 2008). Kinder scheinen erst ab einem Alter von vier Jahren über die eigentliche „Theory of Mind“ zu verfügen (Sodian, 2008). Daher werden im Rahmen des spielbasierten Itempools Aufgaben zu den Vorläuferfähigkeiten der ToM, nämlich „Als-ob Spiel“ und „Hineinversetzte in Andere“, vorgegeben.

Das „Als-ob Spiel“, das durch die Konstruktion einer neuen Realität und die Übernahme verschiedener Rollen charakterisiert wird (Lillard et al, 2013; Sheridan, 1999), wird im spielbasierten Itempool nur mithilfe eines Items erfasst (Kuchler et al, 2011). Angaben zu Itemkennwerten und Reliabilität können daher nicht gemacht werden. Die Variable konnte sehr leicht und objektiv mittels Videoanalysen kodiert werden.

Übereinstimmend mit den Ergebnissen von Nielsen und Dissanayake (2004), die davon ausgehen, dass zweijährige Kinder „Als-ob Spiel“ zeigen, konnte bei allen 20 Kindern der Stichprobe eine Form des „Als-ob Spiels“ beobachtet werden. Beobachtet wurde nicht nur, ob die Kinder „Als-ob Spiel“ zeigten, sondern auch ob sie das „Als-ob Spiel“ anhand eines konkreten Objekts zeigten und ihm Eigenschaften zuschrieben, die es eigentlich nicht besitzt (z.B. Kind trinkt aus einem leeren Becher.). Eine höhere Punktzahl erreichte das Kind, wenn es einen Gegenstand in einer anderen als seiner eigentlichen Funktion verwendete bzw. ein Objekt vollständig erfand (Kuchler et al, 2011). Es konnten für diese Variable jedoch keine signifikanten Alters- und Geschlechtsunterschiede ($\chi^2=.073$, $p_{(\text{einseitig})}=.594$; $\chi^2=.267$, $p_{(\text{einseitig})}=.5$) festgestellt werden. Der Mittelwert der Buben liegt wider erwarten etwas über jenem der Mädchen. Das Alter korreliert nicht mit der Variablen ($r=.061$, $p_{(\text{einseitig})}=.384$). Die Ergebnisse sprechen dafür, dass die Skala nicht valide ist.

Während Putzer (in Vorbereitung) zu den gleichen Ergebnissen kam (kein signifikanter Zusammenhang mit dem Alter und kein Geschlechtsunterschied), zeigte sich bei Birngruber (2012) zwar ebenfalls kein Geschlechtsunterschied, jedoch ebenso wie bei Sapper (2011) ein signifikanter Zusammenhang mit dem Alter. Die Kodierung der Variable war in allen drei Erhebungen gleich. Die unterschiedlichen Ergebnisse können sich daher unter anderem durch Unterschiede in den Stichproben ergeben. In einer

weiteren Erprobung des Itempools könnte versucht werden, eine präzisere Abstufung der Variable vorzunehmen (kein „Als-ob Spiel“/konkrete Objekte/anders verwendet/frei erfunden).

Darüber, dass das „Als-ob Spiel“ im Zusammenhang mit der ToM steht, scheinen sich die meisten Autoren einig zu sein. Studienergebnisse dazu, wie die Beziehung zwischen „Als-ob Spiel“ und der späteren „Theory of Mind“ aussieht, sind vor allem aufgrund von methodischen Mängeln der Studien und unterschiedlichen Definitionen des „Als-ob Spiels“ uneinheitlich (Lillard, et al., 2013). Hierzu ist noch weitere Forschung notwendig.

Die zweite Vorläuferfähigkeit der ToM stellt das Verstehen von Wünschen und Absichten dar, das mit der Skala „Hineinversetzen in Andere“ erhoben werden soll. Viele Autoren gehen davon aus, dass Kinder zwar ab dem ersten Lebensjahr die Absichten und Wünsche („desires“) anderer Menschen verstehen können, deren Überzeugungen („beliefs“) jedoch erst ab vier Jahren (Sodian, 2008; Wellman, 2011; Wellman & Wooley, 1990).

Den Kindern wurden in einem Buch Bilder von Menschen mit unterschiedlichen Absichten und Wünschen gezeigt und die Kinder sollten das Verhalten vorhersagen. Probleme in der Vorgabe ergaben sich vor allem durch das Desinteresse einiger Buben an der Bearbeitung des Buches. Es fiel den Kindern sichtlich schwer, genug Aufmerksamkeit für die Bearbeitung des Buches aufzubringen. Als schwierig stellte sich auch heraus, dass einige anwesende Mütter die Items umformulierten und das Kind fragten, was es denn machen würde, wenn es selbst z.B. schmutzige Hände habe. So vorgegebene Items wurden automatisch als nicht gelöst kodiert. Davon abgesehen, gestaltete sich die Vorgabe sehr objektiv und zumutbar.

Die Berechnung der Reliabilität über nur vier Items ergab eine sehr geringe Reliabilität von $\alpha=.633$. Bis auf ein zu schwieriges Item („schlafen“) liegen alle Aufgaben im angemessenen Schwierigkeitsbereich. Bei den Vorgängerinnen ist die Reliabilität mit Werten von .76 bis .79 höher (Birngruber, 2012; Putzer, in Vorbereitung; Sapper, 2011). Gründe hierfür liegen unter anderem in der Schwierigkeit, die Ergebnisse von kleinen Stichproben zu vergleichen.

Es konnten keine Geschlechtsunterschiede festgestellt werden ($U=32.5$, $z=-1.38$, $p_{(\text{einseitig})}=.101$), jedoch unterscheiden sich die Altersgruppen signifikant voneinander ($U=13$, $z=-2.68$, $p_{(\text{einseitig})}=.004$) und das Alter korreliert signifikant mit dem Gesamtwert der Skala ($r=.57$, $p_{(\text{einseitig})}=.001$), was für eine mittelmäßig hohe Validität spricht. Mit steigendem Alter steigt die Leistung in der Skala „Hineinversetzen in Andere“. Die Ergebnisse sind über alle Erhebungen hinweg übereinstimmend (Birngruber, 2012; Putzer, in Vorbereitung; Sapper, 2011).

Zusammenfassend über alle Erhebungen hinweg betrachtet, scheint die Skala „Hineinversetzen in Andere“ die Gütekriterien hinreichend zu erfüllen.

11.2.2. Numerisches Wissen

Da in der Theorie zwischen der Entwicklung des Zählens und der Mengenerfassung unterschieden wird (z.B. Krajewski, 2008), wird das „numerische Wissen“ im spielbasierten Itempool mittels der Skalen „Zählen“, „aktive Mengenerfassung“ und „passive Mengenerfassung“ erfasst. Die Skala „Zählen“ ist in der ursprünglichen Version des Manuals von Kuchler et al. (2011) noch nicht enthalten. Sie wurde im Rahmen dieser Arbeit entwickelt, da sie für das dritte Lebensjahr einen relevanten Entwicklungsbereich darstellt.

Die Items zum numerischen Wissen konnten gut in das gemeinsame Spiel in der Puppenküche eingebaut werden. Nur drei Buben konnten ein oder zwei Items der numerischen Aufgaben nicht vorgegeben werden. Einige Mütter merkten an, dass ihr Kind bezüglich des numerischen Wissens zuhause bessere Leistungen zeige.

Die Skala „aktive Mengenerfassung“ besteht aus nur zwei Items. Die Kinder werden hierbei gebeten, anhand von zwei bzw. drei Gegenständen zu sagen, um wie viele Objekte es sich handelt. In der aktuellen Erhebung stellte sich das Benennen der Anzahl von drei Objekten als etwas zu schwierig heraus. Die Vorgängerarbeiten beinhalteten auch die Benennung von einem bzw. vier oder fünf Gegenständen, diese Aufgaben waren jedoch zu einfach bzw. zu schwierig. Beim passiven Mengenerfassen werden drei Items vorgegeben und die Kinder werden gebeten, ein bis drei Gegenstände dem Testleiter/der Testleiterin zu reichen (Kuchler et al., 2011). Auch hier stellt sich das erste als zu einfach und das letzte Item (drei) als zu schwierig heraus, wobei

letztenanntes Item aufgrund der geringen Trennschärfe aus den weiteren Analysen selektiert werden musste.

Diese Ergebnisse lassen sich auch mit Studienergebnissen vereinbaren, die besagen, dass zweijährige Kinder zuerst das Kardinalsprinzip anhand von Zahlen kleiner als drei verstehen bzw. diese Anzahlen schätzen können ohne sie zu zählen (Wynn, 1990). Übereinstimmend mit den Ergebnissen von Wynn (1990) wurde ebenfalls beobachtet, dass der Großteil der Kinder eine Handvoll von Objekten reichte.

Die Reliabilitäten von $\alpha=.649$ (zwei Items) für die „aktive Mengenerfassung“ und $\alpha=.581$ (zwei Items) für die „passive Mengenerfassung“ liegen unter dem kritischen Wert von $.7$ und sind demnach inakzeptabel. Zu beachten ist aber die geringe Itemanzahl der Skalen. Auch in Birngrubers Arbeit (2012) liegen die Werte unter dem kritischen Wert. Vor allem für die „passive Mengenerfassung“ ist der Wert ($\alpha=.22$) jedoch viel geringer als der in dieser Arbeit erzielte. Die Kinder mussten hier im Gegensatz zur vorliegenden Arbeit eines bis vier Objekte reichen.

Die signifikanten Geschlechts- und Altersunterschiede ($U=22.5$, $z=-2.37$, $p_{(\text{einseitig})}=.01$; $U=23$, $z=-2.04$, $p_{(\text{einseitig})}=.038$) bezüglich der „aktiven Mengenerfassung“ und die signifikante Korrelation zwischen der Skala und dem Alter sprechen für die Validität. Der Kendall-Tau Wert von $\tau=.39$, ($p_{(\text{einseitig})}=.042$) steht für ein geringes Validitätsmaß. Für die „passive Mengenerfassung“ konnten keine signifikanten Unterschiede zwischen den Geschlechtern oder den Altersgruppen und kein signifikanter Zusammenhang mit dem Alter festgestellt werden. Dies spricht dafür, dass die Skala nicht valide ist.

Birngruber (2012) kam bei der „passiven Mengenerfassung“ zu dem gleichen Ergebnis, bei der Skala „aktive Mengenerfassung“ zeigt sich jedoch weder ein signifikanter Geschlechts- noch Altersunterschied. Unterschiede können sich unter anderem daraus ergeben, dass bei Birngruber (2012) in den Gesamtwert Items zur Benennungen der Anzahl von zwei bis vier Objekten eingingen.

Kritisch anzumerken ist, dass die richtige Antwort auf die Frage „Wie viele sind das?“ nicht unbedingt heißt, dass die Kinder auch die Bedeutung der Zahlen verstehen. Eine mögliche Erklärung ist, dass Kindern kleine Zahlen vertrauter sind und somit wahrscheinlicher ist, dass sie mit diesen Zahlen antworten. Werden sie nach dem Zählvorgang gefragt, um wie viele Objekte es sich handle, werden sie eher die

letztgenannte Zahl nennen, da sie im Gedächtnis am präsentesten ist (Wynn, 1990). Diese Kritik gilt jedoch nicht für die „passive Mengenerfassung“. Le Corre (2006) weist jedoch darauf hin, dass „Give a Number“-Aufgaben die numerischen Fähigkeiten der Kinder unterschätzen, da sie zu komplex sind. Weniger schwierig seien Aufgaben, in denen Kinder die Frage „Was ist auf dieser Karte“ beantworten müssen oder die Korrektheit des Zählvorgangs einer Puppe beurteilen müssen.

In jedem Fall lässt sich sagen, dass sich die Reduktion der Items der beiden Skalen als angemessen herausstellte, den Gütekriterien aber nicht hinreichend entsprochen wird.

Die meisten Kinder waren einfach für das Zählen zu begeistern. Beobachtet wurde, ob die Kinder den Prinzipien nach Gelman und Gallistel (1978) folgten. Die „Eins-zu-Eins Zuordnung“, die „stabile Reihenfolge“ und das „Kardinalsprinzip“ stellten sich dabei als angemessen schwierig heraus. Zusätzlich wurde beurteilt, ob die Kinder richtig zählten, d.h. die korrekte Reihenfolge an Zahlen verwendeten und jedem Objekt eine Zahl zuordneten. Dieses Item ist zu schwierig. Die Reliabilität der Skala „Zählen“ liegt mit $\alpha=.818$ im zufriedenstellenden Bereich.

Dass zweijährige Kinder schon neben den beiden anderen Prinzipien auch das Kardinalsprinzip – jedoch nur bei kleinen Anzahlen an Objekten – befolgen, ist mit der Literatur vereinbar (Wynn, 1990).

Es konnten zwar keine signifikanten Geschlechtsunterschiede ($U=33$, $z=-1.31$, $p_{(\text{einseitig})}=.1$) festgestellt werden, der Altersunterschied ($U=22.5$, $z=-1.86$, $p_{(\text{einseitig})}=.036$) und die Korrelation der Skala mit dem Alter ist jedoch signifikant. Mit einem Wert von $r=.35$ ($p_{(\text{einseitig})}=.045$) kann von einer geringen Validität ausgegangen werden.

Birngruber (2012), die nur erhob, ob die Kinder zählten oder nicht, erzielte ebenfalls keine Geschlechtsunterschiede aber einen signifikanten Zusammenhang mit dem Alter.

Die Gesamtskala „numerisch gesamt“ kann mit einem Wert von $\alpha=.866$ als reliabel erachtet werden. Alle Items konnten in der Berechnung der Reliabilität berücksichtigt werden.

Zwischen den Geschlechtern ist der Unterschied nicht signifikant ($U=29.5$, $z=-1.56$, $p_{(\text{einseitig})}=.062$). Der signifikante Altersunterschied ($U=20.5$, $z=-1.99$, $p_{(\text{einseitig})}=.023$) und

die Korrelation zwischen Alter und „numerisch gesamt“ spricht dafür, dass die Skala valide ist ($\tau=.37$, $p_{(\text{einseitig})}=.016$).

Zusammenfassend lässt sich sagen, dass die Gesamtskala als reliabel und valide erachtet werden kann hat und die Interpretation der Gesamtskala der Interpretation der Einzelskalen – vor allem der „aktiven und passiven Mengenerfassung“ – vorzuziehen ist.

11.2.3. Deduktives Denken

Zum deduktiven Denken im Kleinkindalter findet sich sehr wenig Literatur. Einige Studien haben gezeigt, dass Kleinkinder Aufgaben zum deduktiven Denken unabhängig von deren Inhalt lösen können, wenn sie instruiert wurden, sich einen anderen Kontext (z.B. einen anderen Planeten) vorzustellen (Dias & Harris, 1990; Richards & Sanderson, 1990). Da das syllogistische Schließen die einfachste Form des deduktiven Denkens ist (Goswami, 2011), wurden die Aufgaben zum spielbasierten Itempool in diesem Aufgabentyp von Kuchler et al. (2011) entworfen und in einen Phantasiekontext gebettet.

Während die Vorgängerarbeiten eine Reliabilität mit Werten zwischen .673 und .79 im nicht akzeptablen bis akzeptablen Bereich ergaben, lag das Problem nun vor allem darin, dass viele Kinder die Bearbeitung dieser Aufgaben, die mittels eines Buches vorgegeben wurden, verweigerten (Birngruber, 2012; Sapper, 2011; Putzer, in Vorbereitung). So lagen bei Putzer (in Vorbereitung) bei einem Item vier fehlende Werte und bei den restlichen vier Items je sieben fehlende Werte vor. Somit kann diese Vorgabe als wenig zumutbar eingeschätzt werden.

Eine überarbeitete Vorgabe von Aufgaben zum deduktiven Denken sollte nicht über ein Buch erfolgen, da viele Kinder desinteressiert daran sind oder die Aufmerksamkeit nicht so lange auf ein Buch richten können. Die Aufgaben sollten in das gemeinsame Spiel in der Puppenküche eingebaut werden. Absichten oder Wünsche könnten so zum Beispiel (Hand)-Puppen zugesprochen werden (z.B. Schau, die Puppe ist hungrig! – *Puppe streichelt ihren Bauch* – Was wird sie jetzt machen, wenn sie hungrig ist?).

11.3. Entwicklung des Gedächtnisses

Kuchler et al. (2011) ordnen der „Entwicklung des Gedächtnisses“ die Skalen „phonologisches Gedächtnis“ und „visuelles Gedächtnis“ unter.

Mit der Skala „phonologisches Gedächtnis“ soll die Merkfähigkeit des Kindes mittels der Vorgabe von Wörtern erhoben werden. Da alle drei Items eine Trennschärfe von unter .3 aufweisen, ist diese Skala als nicht sinnvoll anzusehen und es konnten keine weiteren Analysen berechnet werden.

Sapper (2011) und Sindelar (in Vorbereitung) konnten nur zwei Items (Gib-mir AB, Gib mir ABC) in die Berechnung der Reliabilität aufnehmen und erzielten geringe Reliabilitäten von .557 bzw. .68. Während die Ergebnisse von Sindelar (in Vorbereitung) und Birngruber (2012) darin übereinstimmen, dass sich kein signifikanter Alterstrend zeigt, weist die Arbeit von Birngruber (2012) auf Geschlechtsunterschiede hin, die Arbeit von Sindelar (in Vorbereitung) jedoch nicht. Die Vorgabe der Items war in allen Erhebungen gleich. Unterschiedliche Ergebnisse können unter anderem durch Unterschiede in den Stichproben bzw. Messungenauigkeit erklärt werden.

Die Skala „visuelles Gedächtnis“ soll die Merkfähigkeit des Kindes bezüglich visuellen Materials erheben. Fast alle Kinder zeigten sehr großes Interesse an der Schachtel zur Erhebung des visuellen Gedächtnisses. Bis auf ein Kind bearbeiteten alle diese Aufgaben mit großer Konzentration und einige spielten auch danach noch weiter mit der Box. Die Vorgabe der anderen beiden Items zum visuellen Gedächtnis mittels Bildkarten stieß hingegen auf weniger Interesse. Es war sehr schwierig, die Kinder nach der Vorgabe des Vorzeigebilds zur Betrachtung der Antwortbilder zu motivieren. Oft verstrichen zwischen der Vorgabe des Vorzeigebildes und der konzentrierten Betrachtung der Antwortbilder einige Minuten.

Werden zwei von den fünf Items aufgrund der geringen Itemtrennschärfen selektiert, ergibt sich eine Reliabilität von $\alpha=.682$, die deutlich geringer ist, als die Reliabilitäten der Vorgängerarbeiten (.76-.887; Birngruber, 2012; Sapper, 2011).

Die U-Tests zeigen weder einen Geschlechts- ($U=38.5$, $z=-.902$, $p_{(\text{einseitig})}=.215$), noch einen Altersunterschied ($U=28.5$, $z=-1.4$, $p_{(\text{einseitig})}=.095$). Jedoch konnte eine

signifikante Korrelation zwischen der Skala und dem Alter festgestellt werden ($r=.49$, $p_{(\text{einseitig})}=.008$), die für eine mittelmäßig hohe Validität spricht.

Auch die Vorgängerarbeiten kommen zu dem Ergebnis, dass kein Unterschied zwischen den Geschlechtern besteht. Hinsichtlich des Alters ergibt sich ein uneinheitliches Bild. Bei Birngruber (2012) besteht ein signifikanter Alterseffekt, bei Sindelar (in Vorbereitung) hingegen nicht.

Will man eine Skala „Gedächtnis gesamt“ bilden, so liegen alle bis auf zwei Itemtrennschärfen unter dem kritischen Wert. Die Bildung einer Gesamtskala ist daher nicht angemessen.

In Vorgängerarbeiten wurde eine Skala „Gedächtnis gesamt“ gebildet, wobei alle bis auf ein Item der Skala „phonologisches Gedächtnis“ selektiert werden mussten (Sapper, 2011). Bei Sindelar (in Vorbereitung) liegen alle Trennschärfen der Items zum „phonologischen Gedächtnis“ unter dem Wert von .3. Die Bildung dieser Gesamtskala ist daher auch wenig sinnvoll.

Die überarbeitete Vorgabe der Memory-Aufgaben zum „visuellen Gedächtnis“ mittels Schachtel stellte sich als sehr gelungen heraus, wenn auch den Gütekriterien nicht hinreichend entsprochen werden kann. Die Vorgabe mittels Schachtel sollte an einer größeren Stichprobe erprobt werden. Es sollte nun auch eine spielerischere, attraktivere Vorgabe der beiden Bilderset-Items zum „visuellen Gedächtnis“ und der Items zum „phonologischen Gedächtnis“ überlegt werden (z.B. Teddybär mit Aufnahmefunktion „bittet“ das Kind um Gegenstände).

11.4. Zusammenfassende Bewertung

Der spielbasierte Itempool in seiner aktuellen Version kann als sehr brauchbarer Zugang für die Erhebung der Entwicklung zweijähriger Kinder erachtet werden. Die Kinder zeigten unabhängig von Geschlecht und Alter Interesse an der Puppenküche und hatten offensichtlich Spaß am gemeinsamen Spiel. Die meisten Aufgaben konnten gut in das Spiel integriert werden und verloren somit ihren Testcharakter. Die meisten Kinder wollten nach Ende der Testung noch gerne weiterspielen, was dafür spricht, dass die Spieldauer zumutbar ist.

Das Interesse der Kinder an bestimmten Aufgaben war individuell unterschiedlich. In Einklang mit den Vorgängerarbeiten konnte beobachtet werden, dass die zweijährigen Kinder nur eine sehr geringe Aufmerksamkeitspanne oder wenig Interesse für die Betrachtung von Büchern aufbringen können. Dies gilt es in weiteren Überarbeitungen des Itempools zu beachten.

Als sehr tauglich stellten sich hingegen möglichst spielerische Zugänge heraus. Besonders viel Interesse hatten die Kinder an der Schachtel mit den versteckten Memorykarten. Um die Kinder zu Antworten zu motivieren, ist es auch hilfreich, wenn die anwesende Bezugsperson oder die beobachtende Person das Kind fragt oder man die Frage in einen spielerischen Kontext stellt. So kann zum Beispiel so tun als würde der Teddybär das Kind um einen Apfel und eine Banane bitten, da er großen Hunger hat.

Die Objektivität des Verfahrens kann für einen Entwicklungstest, der ein individuelles Vorgehen notwendig macht, als zufriedenstellend beschrieben werden. Hinsichtlich der Reliabilität und Validität ergibt sich für jede Skala ein unterschiedliches Bild. Während die meisten Skalen zur sprachlichen Entwicklung und zum numerischen Wissen zufriedenstellende Reliabilitätswerte aufweisen, sind es vor allem die Skalen mit geringer Itemanzahl („passiver Wortschatz“, „Hineinversetzen in Andere“, „aktive Mengenerfassung“, „passive Mengenerfassung“, „visuelles Gedächtnis“) die Reliabilitäten unter dem kritischen Wert von .7 erzielen. Die Bildung einer Skala „phonologisches Gedächtnis“ und „Gedächtnis gesamt“ stellten sich als nicht sinnvoll heraus. Die Berechnungen der Alterstrends wurden nach zwei verschiedenen Methoden (Unterschieds- und Zusammenhangsberechnung) vorgenommen, wobei die Ergebnisse hierbei oft divergieren. Es zeigt sich jedoch eine Tendenz über alle Skalen hinweg (ausgenommen dem „Als-ob Spiel“), dass stets der Mittelwert der älteren Kinder dem der jüngeren Kinder überlegen ist, und die Mädchen durchschnittlich höhere Werte als die Buben erzielen.

Die Ergebnisse zu den Gütekriterien sind oftmals nicht mit den Ergebnissen der Vorgängerinnen vereinbar, was deutlich macht, dass eine Erprobung des Itempools an einer größeren Stichprobe bzw. eine zusammenfassende Analyse der Daten aller Erhebungen notwendig ist.

12. Abkürzungsverzeichnis

α	Cronbach Alpha, Maß für innere Konsistenz
χ^2	Chi-Quadrat Teststatistik
τ	Kendall Tau, nichtparametrischer Korrelationskoeffizient
.....	
Bayley II.....	Bayley Scales of Infant Development – second Edition-deutsche Fassung
Df.....	Freiheitsgrade
ET 6-6.....	Entwicklungsdiagnostik 6 Monate bis 6 Jahre
F.....	F-Wert – Teststatistik in ANOVA
Me.....	Median
MFED 2-3.....	Münchener Funktionelle Entwicklungsdiagnostik für 2. Und 3. Lebensjahr
MW.....	Mittelwert
N.....	Stichprobengröße
p.....	p-Wert oder Signifikanzwert
SD.....	Standardabweichung
U.....	U-Wert – Teststatistik für den Mann-Whitney-Test
z.....	Wert der Standardnormalverteilung

13. Literaturangabe

- Albers, C. A. & Grieve, A. J. (2007). Test Review: Bayley, N. (2006). Bayley Scales of Infant and Toddler Development – third edition. San Antonio, TX-Harcourt Assessment. *Journal of Psychoeducational Assessment*, 25, 180-190.
- Bagnato, S. J. (2005). The Authentic Alternative for Assessment in Early Intervention: An Emerging Evidence-Based Practice. *Journal of Early Intervention*, 28, 17-22.
- Bagnato, S. J. & Neisworth, J. T. (1994). A national Study of the Social and Treatment "Invalidity" of Intelligence Testing for Early Intervention. *School Psychology Quarterly*, 9, 81-102.
- Barner, D., Chow, K. & Yang, S.-J. (2009). Finding one's meaning: A test of the relation between quantifiers and integers in language development. *Cognitive Psychology*, 58, 195-219.
- Bornstein, M. H. & Haynes, O. M. (1998). Vocabulary Competence in Early Childhood: Measurement, Latent Construct, and Predictive Validity. *Child Development*, 69 (3), 654-671.
- Bayley, N. (2006). *Bayley Scales of Infant and Toddler Development* (3rd ed.). San Antonio: Pearson.
- Behrens, H. (2004). Früher Grammatikerwerb. *Sprache Stimme Gehör*, 28 (1), 15-19.
- Bergen, D. (2013). Does Pretend Play Matter? Searching for Evidence: Comment on Lillard et al. (2013). *Psychological Bulletin*, 139 (1), 45-48.
- Birngruber, A. (2012). *Die Überarbeitung und testtheoretische Überprüfung eines spielbasierten Itempools zur Erfassung der Entwicklung im dritten Lebensjahr unter besonderer Berücksichtigung der kognitiven Entwicklung, der visuellen Wahrnehmung, des Gedächtnisses und der Aufmerksamkeit*. Unveröff. Dipl. Arbeit, Universität, Wien.
- Bortz, J. & Döring, N. (2006). *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler* (4. Aufl.). Heidelberg: Springer.
- Brähler, E., Holling, H., Leutner, D. & Petermann, F. (2002). *Brickenkamp Handbuch psychologischer und pädagogischer Test Band 1*(3., vollständig überarbeitete und erweiterte Auflage). Göttingen: Hogrefe
- Bryant, P. & Nuñez, T. (2011). Children's Understanding of Mathematics. In U. Goswami (Hrsg.), *The Wiley-Blackwell Handbook of Childhood Cognitive Development*. (S. 549-573). Oxford: Blackwell.
- Carpenter, M., Call, J. & Tomasello, M. (2002). Understanding "Prior Intentions" Enables Two-Year-Olds to Imitatively Learn a Complex Task. *Child Development*, 73, 1431-1441.
- Dahl G. (1971). Zur Berechnung des Schwierigkeitsindex bei quantitativ abgestufter Aufgabenbewertung. *Diagnostica*, 17, 139-142.
- Deimann, P. & Kastner-Koller, U. (2007). Entwicklungsdiagnostik. In M. Hasselhorn & W. Schneider (Hrsg.), *Handbuch der Entwicklungspsychologie* (S. 558-569). Göttingen: Hogrefe.
- Dias, M. G. & Harris, P. L. (1990). The Influence of the Imagination on Reasoning by Young Children. *British Journal of Developmental Psychology*, 8, 305-318.

- Dietrich, P. (2001). Der Scenotest. In D. Sturzbecher (Hrsg.), *Spielbasierte Befragungstechniken. Interaktionsdiagnostische Verfahren für Begutachtung, Beratung und Forschung* (S. 228-241). Göttingen: Hogrefe.
- Esser, G. & Petermann, F. (2010). *Entwicklungsdiagnostik*. (Kompendien Psychologische Diagnostik Bd. 13). Göttingen: Hogrefe.
- Ettrich, K. U. (2000). *Entwicklungsdiagnostik im Vorschulalter. Grundlagen, Verfahren, Neuentwicklungen, Screenings*. Göttingen: Vandenhoeck & Ruprecht.
- Farmer-Dougan, V. & Kaszuba, T. (1999). Reliability and Validity of Play-based Observations: relationship between the PLAY behavior observation system and standardized measures of cognitive and social skills. *Educational Psychology, 19* (4), 429-440.
- Feigenson, L., Dehaene, S. & Spelke, E. (2004). Core systems of number. *Trends in Cognitive Sciences, 8*, 307-314.
- Field, A. (2009). *Discovering statistics using SPSS* (3rd ed.). London: SAGE.
- Frey, B. (2012). Entwicklungsdiagnostik. In M. Cierpka (Hrsg.), *Frühe Kindheit 0-3. Beratung und Psychotherapie für Eltern mit Säuglingen und Kleinkindern* (S. 457-466). Springer Verlag: Berlin.
- Fuchs-Gaderer, M. (2012). *Die Überarbeitung und testtheoretische Überprüfung eines spielbasierten Itempools zur Erfassung der Entwicklung im dritten Lebensjahr unter besonderer Berücksichtigung der Bereiche Motorik, Sprache und sozioemotionale Entwicklung*. Unveröff. Dipl. Arbeit, Universität, Wien.
- Gelman, S. A. & Coley, J. D. (1990). The importance of knowing a dodo is a bird: Categories and inferences in 2-year-old children. *Developmental Psychology, 26*, 796-804.
- Gelman, R. & Gallistel, C. R. (1978). *The child's understanding of number*. Cambridge, MA: Harvard University Press.
- Goswami, U. (2001). *So denken Kinder. Einführung in die Psychologie der kognitiven Entwicklung*. Bern: Huber.
- Goswami, U. (2008). *Cognitive Development. The Learning Brain*. New York: Psychology Press.
- Goswami, U. (2011). Inductive and Deductive Reasoning. In U. Goswami (Hrsg.), *The Wiley-Blackwell Handbook of Childhood Cognitive Development*. (S. 399-419). Oxford: Blackwell.
- Hagmann von Arx, P., Meyer, C. S. & Grob, A. (2008). Intelligenz- und Entwicklungsdiagnostik im deutschen Sprachraum. *Kindheit und Entwicklung, 17*, 232-242.
- Harris, P. L. (2001). Thinking about the unknown. *Trends in Cognitive Science, 5* (11), 494-498).
- Hellbrügge, T. (Hrsg.) (1994). *Münchener Funktionelle Entwicklungsdiagnostik: zweites und drittes Lebensjahr. Durchführungs-, Beurteilungs- und Interpretationshinweise* (4., korr. u. erw. Aufl.). München: Deutsche Akademie für Entwicklungs-Rehabilitation.

- Huang, Y. T., Spelke, E. & Snedeker, J. (2010). When is four far more than three? Children's generalization of newly-acquired number words. *Psychological Science*, 21 (4), 600-606.
- Irblich, D. & Renner, G. (2009). Wie untersucht man Kinder? In D. Irblich & G. Renner (Hrsg.), *Diagnostik in der Klinischen Kinderpsychologie. Die ersten sieben Lebensjahre* (S. 21-32). Göttingen: Hogrefe.
- Kamphaus, R. W. (2005). *Clinical assessment of child and adolescent intelligence*. New York: Springer Science & Business Media.
- Kastner-Koller, U. & Deimann, P. (2009). Beobachtung und Befragung von Kindern. In D. Irblich & G. Renner (Hrsg.), *Diagnostik in der Klinischen Kinderpsychologie. Die ersten sieben Lebensjahre* (S. 97-107). Göttingen: Hogrefe.
- Kastner-Koller, U. & Deimann, P. (2011). Entwicklungstests. In L.F. Hornke, M. Amelang & M. Kersting (Hrsg.), *Enzyklopädie der Psychologie, Themenbereich B: Methodologie und Methoden. Serie II Psychologische Diagnostik, Bd. 3 Leistungs-, Intelligenz- und Verhaltensdiagnostik* (S. 275-304). Göttingen: Hogrefe.
- Kelly-Vance, L. & Ryalls, B. O. (2005). A Systematic, Reliable Approach to Play Assessment in Preschoolers. *School Psychology International*, 26, 398-412.
- Kelly-Vance, L., Ryalls, B. O. (2008). Best Practices in Play Assessment and Intervention. In A. Thomas & J. Grimes (Eds.), *Best Practices in School Psychology V* (pp. 549–560). Bethesda, MD: National Association of School Psychologists.
- Kelly-Vance, L., Ryalls, B. O. & Glover, K. G. (2002). The Use of Play Assessment to Evaluate the Cognitive Skills of Two- and Three-Year-Old Children. *School Psychology International*, 23, 169-185.
- Klauer, K. J. (2001). Training des induktiven Denkens. In K. J. Klauer (Hrsg.), *Handbuch Kognitives Training*. (2., überarbeitete und erweiterte Aufl., S. 165-209). Göttingen: Hogrefe.
- Kline, P. (2000). *The handbook of psychological testing (2nd ed.)*. London: Routledge.
- Krajewski, K. (2008). Prävention der Rechenschwäche. In W. Schneider & M. Hasselhorn (Hrsg.), *Handbuch der Pädagogischen Psychologie (S360-370)*. Göttingen: Hogrefe.
- Krajewski, K. & Schneider, W. (2006). Mathematische Vorläuferfähigkeiten im Vorschulalter und ihre Vorhersagekraft für die Mathematikleistungen bis zum Ende der Grundschulzeit. *Zeitschrift für Psychologie in Erziehung und Unterricht*, 53, 246-262.
- Kronberger, V. (in Vorbereitung). *Die Überarbeitung und testtheoretische Überprüfung eines spielbasierten Itempools zur Erfassung der Entwicklung im dritten Lebensjahr-unter besonderer Berücksichtigung der visuellen Wahrnehmung*. Unveröff. Dipl. Arbeit, Universität, Wien.
- Kubinger, K. D. (2009). Psychologische Diagnostik: *Theorie und Praxis psychologischen Diagnostizierens* (2., überarb. u. erw. Aufl.). Göttingen: Hogrefe.
- Kuchler, M. (2011). *Die Entwicklung und Erprobung eines Itempools zur spielbasierten Erfassung der Entwicklung Zweijähriger unter besonderer Berücksichtigung der Motorik, der Sprache und der sozialemotionalen Kompetenzen*. Unveröff. Dipl. Arbeit, Universität, Wien.

- Kuchler, M., Sapper, E., Deimann, P. & Kastner-Koller, U. (2011). *Manual zum Itempool zur spielbasierten Erfassung der Entwicklung Zweijähriger*. Unveröffentl. Manuskript.
- Laschkowski, W., Hermann, W., Mainka, D., Schütz, C., Schuster, D. & Titera, D. (2000). *SON. Snijders-Oomen Non-verbaler Intelligenztest 2 ½ -7*. Arbeitsgruppe SON: Erlangen.
- Le Corre, M. & Carey, S. (2007). One, two, three, four, nothing more: An investigation of the conceptual sources of the verbal counting principles. *Cognition*, 105 (2), 395-438.
- Le Corre, M., Van de Walle, G., Brannon, E. M. & Carey, S. (2006). Re-visiting the competence/performance debate in the acquisition of the counting principles. *Cognitive Psychology*, 52, 130-169.
- Leslie, A. M. (1987). Pretense and Representation: The Origins of "Theory of Mind". *Psychological Review*, 94 (4), 412-426.
- Levine, S. C., Whealton Suriyakham, L., Rowe, M. L., Huttenlocher, J. & Guderson, E. A. (2010). What counts in the development of young children's number. *Developmental Psychology*, 46 (5), 1309-1319.
- Lienert, G. A. & Raatz, U. (1998) *Testaufbau und Testanalyse* (6. Aufl.). Weinheim: PVU.
- Lifter, K., Mason, E. J. & Barton, E. E. (2011). Children's Play: Where We Have Been and Where We Could Go. *Journal of Early Intervention*, 33 (4), 281-297.
- Lillard, A. S. (1993). Pretend Play Skills and the Child's Theory of Mind. *Child Development*, 64, 348-371.
- Lillard, A. S., Pinkham, A. M. & Smith, E. (2011). Pretend Play and Cognitive Development. In U. Goswami (Hrsg.), *The Wiley-Blackwell Handbook of Childhood Cognitive Development*. (S. 285-311). Oxford: Blackwell.
- Lillard, A. S., Lerner, M. D., Hopkins, E. J., Dore, R. A., Smith, E. D. & Palmquist, C. M. (2013). The Impact of Pretend Play on Children's Development: A Review of the Evidence. *Psychological Bulletin*, 139 (1), 1-34.
- Lissmann, I., Domsch, H. & Lohaus, A. (2006). Zur Stabilität und Validität von Entwicklungstestergebnissen im Alter von sechs Monaten bis zwei Jahren. Eine Analyse am Beispiel des ET 6-6. *Kindheit und Entwicklung*, 15 (1), 35-44.
- Lung, F. W., Chiang, T. L., Lin, S. J., Feng, J. Y, Chen, P. F. & Shu, B. C. (2011). Gender differences of children's developmental trajectory from 6 to 60 months in the Taiwan Birth Cohort Pilot Study. *Research in Developmental Disabilities*, 32 (1), 100-106.
- Lücking, A. & Scheithauer, H. (2006). Entwicklungstest sechs Monate bis sechs Jahre (ET 6-6). Franz Petermann, Iris A. Stein, Thorsten Macha (2., veränderte Auflage 2004). [Frankfurt am Main: Harcourt Test Services, Preis des Gesamtsatzes: Euro 1.016,50]. *Diagnostica*, 52, 100-103.
- Macha, T., Proske, A. & Petermann, F. (2005). Allgemeine Entwicklungsdiagnostik. Validität von Entwicklungstests. *Kindheit und Entwicklung*, 14 (3), 150-162.

- Meiser, T. & Klauer, K. C. (2001). Training des deduktiven Denkens. In K. J. Klauer (Hrsg.), *Handbuch Kognitives Training*. (2., überarbeitete und erweiterte Aufl., S. 211-234). Göttingen: Hogrefe.
- Melchers, P. & Preuß, U. (2006). *K-ABC: Kaufman Assessment Battery for Children. Deutschsprachige Fassung*. Leiden: PITS.
- Montada, L. (2008). Fragen, Konzepte, Perspektiven. In R. Oerter & L. Montada (Hrsg.), *Entwicklungspsychologie* (6., vollständig überarbeitete Aufl., S. 3-18). Weinheim: Beltz.
- Moosbrugger, H. & Kelava, A. (2012). Qualitätsanforderungen an einen psychologischen Test (Testgütekriterien). In H. Moosbrugger & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (2., überarbeitete und aktualisierte Ausgabe) (S. 7-26). Heidelberg: Springer Medizin Verlag.
- Nielsen, M. & Dissanayake, C. (2004). Pretend play, mirror self-recognition and imitation: A longitudinal investigation through the second year. *Infant Behavior and Development*, 27, 342-365.
- Oerter, R. (1999). *Psychologie des Spiels. Ein handlungstheoretischer Ansatz*. Weinheim: Beltz
- Oerter, R. (2008). Kindheit. In R. Oerter & L. Montada (Hrsg.), *Entwicklungspsychologie*. (6., vollständig überarbeitete Aufl., S. 225-270). Weinheim: Beltz.
- Petermann, F. & Macha, T. (2003). Strategien in der testgestützten allgemeinen Entwicklungsdiagnostik. *Monatsschrift Kinderheilkunde*, 151, 6-13.
- Petermann, F. & Macha, T. (2008). Developmental Assessment. A General Framework. *Journal of Psychology*, 216 (3), 127-134.
- Perner, J. (1991). *Understanding the representational mind*. Cambridge, MA: MIT Press.
- Petermann, F., Stein, I. A. & Macha, T. (2008). *Entwicklungstest von sechs Monaten bis sechs Jahren (ET 6-6)* (3., veränderte Aufl.). Frankfurt: Pearson Assessment.
- Putzer, S. (2013). *Die Entwicklung und Erprobung eines Itempools zur spielbasierten Erfassung der Entwicklung Zweijähriger unter besonderer Berücksichtigung der Grobmotorik*. Unveröff. Dipl. Arbeit, Universität, Wien.
- Quaiser-Pohl, C. (2010). Was ist Entwicklungsdiagnostik? In C. Quaiser-Pohl & H. Rindermann (Hrsg.), *Entwicklungsdiagnostik*. (S. 85-100). München: Ernst Reinhardt.
- Quaiser-Pohl, C. & Köhler, A. (2010). Allgemeine Entwicklungstests. In C. Quaiser-Pohl & H. Rindermann (Hrsg.), *Entwicklungsdiagnostik*. (S. 85-100). München: Ernst Reinhardt.
- Rauh, H. (2008). Vorgeburtliche Entwicklung und frühe Kindheit. In R. Oerter & L. Montada (Hrsg.), *Entwicklungspsychologie* (6., vollständig überarbeitete Aufl., S. 436-479). Weinheim: Beltz.
- Renner, G. (2009). Testpsychologische Diagnostik bei Kindern. In D. Irblich & G. Renner (Hrsg.), *Diagnostik in der Klinischen Kinderpsychologie. Die ersten sieben Lebensjahre* (S. 73-85). Göttingen: Hogrefe.

- Renner, G. & Irblich, D. (2009). Intelligenzdiagnostik. In D. Irblich & G. Renner (Hrsg.), *Diagnostik in der Klinischen Kinderpsychologie. Die ersten sieben Lebensjahre* (S. 136-151). Göttingen: Hogrefe.
- Reuner, G. & Pietz, J. (2006). Entwicklungsdiagnostik im Säuglings- und Kleinkindalter. *Monatsschrift Kinderheilkunde*, 154, 305-313.
- Reuner, G., Rosenkranz, J., Pietz, J. & Horn, R. (2007). *Bayley Scales of Infant Development, Second Edition (Bayley II) – Deutsche Fassung*. Frankfurt/M.: Harcourt Test Services (Original erschienen 1993 von N. Bayley).
- Reznick J., Corley R. & Robinson J. A. (1997). Longitudinal twin study of intelligence in the second year of life. *Monographs of the Society for Research in Child Development*, 62, 155-160.
- Richards, C. A. & Sanderson, J. A. (1999). The role of imagination in facilitating deductive reasoning in 2-, 3- and 4-year-olds. *Cognition*, 72, B1-B9.
- Ricken, G., Fritz, A., Schuck, K. D. & Preuß, U. (2007). HAWIVA-III. Hannover-Wechsler-Intelligenztest für das Vorschulalter – III. Bern: Huber.
- Rindermann, H. & Kwiatkowski, V. (2010). Intelligenzdiagnostik im Kindesalter. In C. Quaiser-Pohl & H. Rindermann (Hrsg.), *Entwicklungsdiagnostik*. (S. 57-68). München: Ernst Reinhardt.
- Ritblatt, S. N. (2010). Children's Level of Participation in a False-Belief Task, Age, and Theory of Mind. *The Journal of Genetic Psychology*, 161 (1), 53-64.
- Rümmele, A. (2011). Inhalt und Ziele einer entwicklungsorientierten Diagnostik. In H. Keller (Hrsg.), *Handbuch der Kleinkindforschung* (4., vollständig überarbeitete Auflage, S. 982-961). Bern: Hogrefe.
- Sapper, E. (2011). *Die Entwicklung und Erprobung eines Itempools zur spielbasierten Erfassung der Entwicklung Zweijähriger unter besonderer Berücksichtigung der kognitiven Entwicklung, der visuellen Wahrnehmung, des Gedächtnisses und der Arbeitshaltungen*. Unveroff. Dipl. Arbeit, Universität, Wien.
- Sarimski, K. (2001). Entwicklungstest von 6 Monaten bis 6 Jahren (ET 6-6). Franz Petermann, Iris A. Stein (2000). [Frankfurt: Swets Test Services, 120 Seiten, Preis des Handbuchs: DM 3,-.Preis des kompletten Tests: DM 1800,-]. *Diagnostica*, 47 (2), 107-109.
- Sarimski, K. (2009). Entwicklungsdiagnostik. In D. Irblich & G. Renner (Hrsg.), *Diagnostik in der Klinischen Kinderpsychologie. Die ersten sieben Lebensjahre* (S. 136-151). Göttingen: Hogrefe.
- Sarnecka, B. W. & Lee, M. D. (2009). Levels of Number Knowledge in Early Childhood. *Journal of Experimental Child Psychology*, 103 (3), 325-337.
- Sheridan, M. D. (1999). *Play in Early Childhood. From birth to six years*. London: Routledge.
- Siegler, R., DeLoache, J. & Eisenberg, N. (2011). *Entwicklungspsychologie im Kindes- und Jugendalter*. (3.Aufl.). Heidelberg: Springer.
- Sindelar, I. (2013). *Die Entwicklung und Erprobung eines Itempools zur spielbasierten Erfassung der Entwicklung Zweijähriger unter besonderer Berücksichtigung der Feinmotorik und Visuomotorik*. Unveroff. Dipl. Arbeit, Universität, Wien.

- Singer-Freeman, K. E. (2005). Analogical reasoning in 2-year-olds: The development of access and relational inference. *Cognitive Development*, 20 (2), 214-234.
- Sodian, B. (2008). Entwicklung des Denkens. In R. Oerter & L. Montada (Hrsg.), *Entwicklungspsychologie* (6., vollständig überarbeitete Aufl., S. 436-479). Weinheim: Beltz.
- Solso, R. L., MacLin, O. H. & MacLin, M. K. (2008). *Cognitive Psychology* (8th Edition). New York: Pearson Education.
- Sturzbecher, D. (2001). Methodische Lösungsansätze zur Befragung jüngerer Kinder. In D. Sturzbecher (Hrsg.), *Spielbasierte Befragungstechniken. Interaktionsdiagnostische Verfahren für Begutachtung, Beratung und Forschung*. (S. 51-63). Göttingen: Hogrefe.
- Szagan, G., Steinbrink, C., Franik, M. & Stumper B. (2006). Development of vocabulary and grammar in young German-speaking children assessed with a German language development inventory. *First language*, 26 (3), 259-280.
- Vig, S. & Sanders, M. (2007). Cognitive Assessment. In M. R. Brassard & A. E. Boehm, *Preschool assessment: Principles and practices* (pp. 383-420). New York: Guilford Press.
- Walker, C. M. & Gopnik, A. (2013). Pretense and Possibility—A Theoretical Proposal About the Effects of Pretend Play on Development: Comment on Lillard et al. (2013). *Psychological Bulletin*, 139 (1), 40-44.
- Walker, R. F. & Muracher, T. (2012). Representation and Theory of Mind Development. *Developmental Psychology*, 48 (2), 509-520.
- Weinert, S. (2007). Spracherwerb. In M. Hasselhorn & W. Schneider (Hrsg.), *Handbuch der Entwicklungspsychologie* (S. 221-231). Göttingen: Hogrefe.
- Weise, G. (1975). *Psychologische Leistungstests*. Göttingen: Hogrefe.
- Weissenborn, J. (2000). *Der Erwerb von Morphologie und Syntax*. In H. Grimm (Hrsg.), *Sprachentwicklung. Enzyklopädie der Psychologie: Themenbereich C. Theorie und Forschung: Serie III. Sprache: Bd. 3*. (S. 141-169). Göttingen: Hogrefe.
- Wellman, H. M. (2011). Developing a Theory of Mind. In U. Goswami (Hrsg.), *The Wiley-Blackwell Handbook of Childhood Cognitive Development*. (S. 258-284). Oxford: Blackwell.
- Wellman, H. M. & Liu, D. (2004). Scaling of theory-of-mind tasks. *Child Development*, 75, 523-541.
- Wellman, H. M., & Woolley, J. D. (1990). From simple desires to ordinary beliefs: The early development of everyday psychology. *Cognition*, 35, 245-275.
- Witzlack, G. (2001). Spielanalytische Entwicklungsdiagnostik – historische Wurzeln und Ansätze in der DDR. In D. Sturzbecher (Hrsg.), *Spielbasierte Befragungstechniken. Interaktionsdiagnostische Verfahren für Begutachtung, Beratung und Forschung* (S. 218-225). Göttingen: Hogrefe.
- Wynn, K. (1990). Children's understanding of counting. *Cognition*, 36, 155-193.
- Wynn, K. (1992). Children's acquisition of the number words and the counting system. *Cognitive Psychology*, 24, 220-251.

14. *Abbildungsverzeichnis*

Abbildung 1: Interaktionsdiagramm zur Skala "Wortschatz aktiv" hinsichtlich Geschlecht und der Alter.....	94
Abbildung 2: Interaktionsdiagramm zur Skala "Sprache gesamt" hinsichtlich Geschlecht und Alter	95
Abbildung 3: Histogramm zur Normalverteilungsprüfung für die Variable „Alter in Monaten“	144
Abbildung 4: Histogramm zur Normalverteilungsprüfung für den Summenscore „aktiver Wortschatz“	145
Abbildung 5: Histogramm zur Normalverteilungsprüfung für den Summenscore „passiver Wortschatz“	145
Abbildung 7: Histogramm zur Normalverteilungsprüfung für den Summenscore „Sprache gesamt“	145
Abbildung 6: Histogramm zur Normalverteilungsprüfung für den Summenscore „grammatikalische Fähigkeiten“.....	145
Abbildung 8: Histogramm zur Normalverteilungsprüfung für den Summenscore „Hineinversetzen in Andere“.....	145
Abbildung 9: Histogramm zur Normalverteilungsprüfung für den Summenscore „passive Mengenerfassung“	145
Abbildung 11: Histogramm zur Normalverteilungsprüfung für den Summenscore „Zählen“ ..	146
Abbildung 10: Histogramm zur Normalverteilungsprüfung für den Summenscore „aktive Mengenerfassung“	146
Abbildung 12: Histogramm zur Normalverteilungsprüfung für den Summenscore „numerisch gesamt“	146
Abbildung 13: Histogramm zur Normalverteilungsprüfung für den Summenscore „numerisch gesamt“	146
Abbildung 14 a-d: Histogramme zur Normalverteilungsprüfung für die Skala „aktiver Wortschatz“ getrennt für Geschlecht (a männlich, b weiblich) und Altersgruppen (c jung, d alt).....	147
Abbildung 15 a-d: Histogramme zur Normalverteilungsprüfung für die Skala „passiver Wortschatz“ getrennt für Geschlecht (a männlich, b weiblich) und Altersgruppen (c jung, d alt).....	148
Abbildung 16 a-d: Histogramme zur Normalverteilungsprüfung für die Skala „grammatikalische Fähigkeiten“ getrennt für Geschlecht (a männlich, b weiblich) und Altersgruppen (c jung, d alt).....	149
Abbildung 17 a-d: Histogramme zur Normalverteilungsprüfung für die Skala „Sprache gesamt“ getrennt für Geschlecht (a männlich, b weiblich) und Altersgruppen (c jung, d alt).....	150
Abbildung 18: a-d: Histogramme zur Normalverteilungsprüfung für die Skala „Hineinversetzen in Andere“ getrennt für Geschlecht (a männlich, b weiblich) und Altersgruppen (c jung, d alt).....	151
Abbildung 19: a-d: Histogramme zur Normalverteilungsprüfung für die Skala „aktive Mengenerfassung“ getrennt für Geschlecht (a männlich, b weiblich) und Altersgruppen (c jung, d alt).....	152

Abbildung 20	a-d: Histogramme zur Normalverteilungsprüfung für die Skala „passive Mengenerfassung“ getrennt für Geschlecht (a männlich, b weiblich) und Altersgruppen (c jung, d alt).....	153
Abbildung 21	a-d: Histogramme zur Normalverteilungsprüfung für die Skala „Zählen“ getrennt für Geschlecht (a männlich, b weiblich) und Altersgruppen (c jung, d alt)	154
Abbildung 22	a-d: Histogramme zur Normalverteilungsprüfung für die Skala „numerisch gesamt“ getrennt für Geschlecht (a männlich, b weiblich) und Altersgruppen (c jung, d alt).....	155
Abbildung 23	a-d: Histogramme zur Normalverteilungsprüfung für die Skala „visuelles Gedächtnis“ getrennt für Geschlecht (a männlich, b weiblich) und Altersgruppen (c jung, d alt)	156
Abbildung 24	a-d: Histogramme zur Normalverteilungsprüfung für die Summenscores „aktiver Wortschatz“ (a), „aktiver Wortschatz – Fragebogen“ (b), „grammatikalische Fähigkeiten“ (c) und „grammatikalische Fähigkeiten – Fragebogen“ (d).....	158

15. Tabellenverzeichnis

Tabelle 1: Verteilung des Alters in Monaten hinsichtlich des Geschlechts	67
Tabelle 2: Häufigkeiten der Altersgruppen hinsichtlich des Geschlechts	67
Tabelle 3: Ausbildung der Väter und Mütter	68
Tabelle 4: Reliabilität und Itemtrennschärfen vor und nach der Selektion und Itemschwierigkeiten für die Skala "aktiver Wortschatz" (in Klammer – Anzahl der fehlenden Werte).....	85
Tabelle 5: Reliabilität und Itemtrennschärfen vor und nach der Selektion und Itemschwierigkeiten für die Skala "passiver Wortschatz" (in Klammer – Anzahl der fehlenden Werte).....	86
Tabelle 6: Reliabilität und Itemtrennschärfen vor und nach der Selektion und Itemschwierigkeiten für die Skala "grammatikalische Fähigkeiten"	87
Tabelle 7: Reliabilität und Itemtrennschärfen vor und nach der Selektion und Itemschwierigkeiten für die Skala "Hineinversetzen in Andere" (in Klammer – Anzahl fehlender Werte).....	89
Tabelle 8: Reliabilität, Itemtrennschärfen und Itemschwierigkeiten für die Skala „aktive Mengenerfassung" (in Klammer – Anzahl fehlender Werte)	89
Tabelle 9: Reliabilität und Itemtrennschärfen vor und nach der Selektion und Itemschwierigkeiten für die Skala "passive Mengenerfassung" (in Klammer – Anzahl fehlender Werte"	90
Tabelle 10: Reliabilität, Itemtrennschärfen und Itemschwierigkeiten für die Skala "Zählen"	90
Tabelle 11: Reliabilität, Itemtrennschärfen und Itemschwierigkeiten für die Skala "phonologisches Gedächtnis" (in Klammer – Anzahl fehlender Werte)	91
Tabelle 12: Reliabilität und Itemtrennschärfen vor und nach der Selektion und Itemschwierigkeiten für die Skala "visuelles Gedächtnis" (in Klammer – Anzahl fehlender Werte)	91
Tabelle 13: Mittelwerte und Standardabweichungen der beiden Altersgruppen und der Geschlechter hinsichtlich der Skalen der sprachlichen Entwicklung	93
Tabelle 14: Überblick über die Ergebnisse zu den Geschlechts- und Altersunterschieden sowie der Korrelation zwischen dem Alter und den Summenscores der Skalen zur sprachlichen Entwicklung.....	93
Tabelle 15: Mittelwerte und Standardabweichung der beiden Altersgruppen und der Geschlechter hinsichtlich der Skalen der „Theory of Mind“	96
Tabelle 16: Überblick über die Ergebnisse zu den Geschlechts- und Altersunterschieden sowie der Korrelation zwischen dem Alter und den Summenscores der Skalen zur „Theory of Mind“	96
Tabelle 17: Häufigkeiten der Antwortalternativen des Items „Als-ob-Spiel“ hinsichtlich Geschlecht- und der Altersgruppen	96
Tabelle 18: Mittelwerte und Standardabweichung der beiden Altersgruppen und der Geschlechter hinsichtlich der Skalen des „numerischen Wissens“.....	97
Tabelle 19: Überblick über die Ergebnisse zu den Geschlechts- und Altersunterschieden sowie der Korrelation zwischen dem Alter und den Summenscores der Skalen zum „numerischen Wissen“	98

Tabelle 20: Mittelwerte und Standardabweichung der beiden Altersgruppen und der Geschlechter hinsichtlich der Skalen des „Gedächtnisses“	99
Tabelle 21: Überblick über die Ergebnisse zu den Geschlechts- und Altersunterschieden sowie der Korrelation zwischen dem Alter und den Summenscores der Skalen zum „Gedächtnis“	100
Tabelle 22 Reliabilität und Itemtrennschärpen vor und nach der Selektion und Itemschwierigkeiten für die Skala „Sprache gesamt“ (in Klammer – Anzahl der fehlenden Werte)	140
Tabelle 23: Reliabilität, Itemtrennschärpen und Itemschwierigkeiten für die Skala „numerisch gesamt“ (in Klammer – Anzahl der fehlenden Werte)	142
Tabelle 24: Reliabilität, Itemtrennschärpen vor und nach der Selektion und Itemschwierigkeiten für die Skala „Gedächtnis gesamt“ (in Klammer- Anzahl fehlender Werte).....	143
Tabelle 25: Mittelwert, Standardabweichung, Schiefe und Kurtosis der Summenscores.....	144
Tabelle 26: Voraussetzungsprüfung „aktiver Wortschatz“	147
Tabelle 27: Voraussetzungsprüfung „passiver Wortschatz“	148
Tabelle 28: Voraussetzungsprüfung „grammatikalische Fähigkeiten“	149
Tabelle 29: Voraussetzungsprüfung „Sprache gesamt“	150
Tabelle 30: Voraussetzungsprüfung „Hineinversetzen in Andere“	151
Tabelle 31: Voraussetzungsprüfung „aktive Mengenerfassung“	152
Tabelle 32: Voraussetzungsprüfung „passive Mengenerfassung“	153
Tabelle 33: Voraussetzungsprüfung „Zählen“	154
Tabelle 34: Voraussetzungsprüfung „numerisch gesamt“	155
Tabelle 35: Voraussetzungsprüfung „visuelles Gedächtnis“	156
Tabelle 36: Mittelwert und Standardabweichung für den Summenscore der Skala „aktiver Wortschatz“ hinsichtlich Geschlecht- und Altersgruppen.....	157
Tabelle 37: Mittelwert und Standardabweichung für den Summenscore der Skala „Sprache gesamt“ hinsichtlich Geschlecht- und Altersgruppen.....	157
Tabelle 38: Überprüfung der Normalverteilung für die Summenscores „aktiver Wortschatz“, „aktiver Wortschatz– Fragebogen“, „grammatikalische Fähigkeiten“ und „grammatikalische Fähigkeiten – Fragebogen“	158

16. Anhang

ANHANG A: Elternbrief

ANHANG B: Einverständniserklärung Video

ANHANG C: Elternfragebogen

ANHANG D: Protokollbogen

ANHANG E: Tabellen Gesamtskalen

ANHANG F: Normalverteilungsprüfung für Validitätsberechnung

ANHANG G: Voraussetzungsprüfung ANOVA

ANHANG H: Deskriptive Statistiken ANOVA

ANHANG I: Voraussetzungsprüfung Korrelation Elternfragebogen

ANHANG J: Abstract (Deutsch)

ANHANG K: Lebenslauf

LIEBE ELTERN ALLER ZWEI- BIS DREIJÄHRIGEN

Im Rahmen unserer Diplomarbeiten sind wir, zwei Studentinnen der Psychologie, auf der Suche nach Eltern mit Kindern, die uns bei unserer Studie unterstützen möchten.

Worum es geht:

Am Institut für Entwicklungspsychologie und Psychologische Diagnostik der Universität Wien ist es Ziel, ein spielbasiertes Verfahren zur Erfassung der Entwicklung zwei- bis dreijähriger Kinder zu entwickeln. Wir wollen Informationen sammeln, welches Verhalten Zweijährige im Spiel üblicherweise zeigen.

Was passiert?

Wir spielen, lesen, turnen und lachen circa zwei Stunden lang mit Ihrem Kind in einem Testraum der Uni Wien. Sie können in Ruhe entspannen und zusehen.

Wann?

Wir würden gerne mit Ihnen und Ihrem Kind für Februar 2013 einen Termin vereinbaren.

Welchen Vorteil haben Sie davon:

Sie haben die Möglichkeit zu erfahren, welche Fähigkeiten eines zweijährigen Kindes als besonders wichtig gelten und wie diese Fähigkeiten erfasst werden.



Wenn Sie Interesse an einem spannenden Vormittag oder Nachmittag haben und zusätzlich noch einen Beitrag zu unserem Projekt leisten wollen, dann freuen wir uns sehr, Sie und Ihr Kind kennenzulernen!

Für alle Fragen stehen wir Ihnen jederzeit sehr gerne telefonisch oder per E-Mail zur Verfügung!

SIGRID PUNZ

a0705400@unet.univie.ac.at



VIKTORIA KRONBERGER

a0703387@unet.univie.ac.at



ANHANG B: Einverständniserklärung Video

Liebe Eltern,

da es für uns nicht möglich ist, während des Spiels mit Ihrem Kind alles zu erfassen, was für die Entwicklung unseres Verfahrens relevant ist, bitten wir Sie, die Spielsituation mit dem Kind auf Video aufzeichnen zu dürfen.

Wir versichern Ihnen, dass außer uns und unseren Diplomarbeitsbetreuerinnen, Ass.-Prof. Dr. Pia Deimann und Ass.-Prof. Dr. Ursula Kastner-Koller, niemand die Videos sehen wird und diese spätestens nach Beendigung unserer Diplomarbeiten gelöscht werden.

Vielen Dank

Viktoria Kronberger und Sigrid Punz

Ich, _____, erkläre mich damit einverstanden, dass meine Tochter/mein Sohn _____, geboren am _____ im Rahmen der Mitwirkung an der Diplomarbeitsstudie von Frau Kronberger und Frau Punz auf Video aufgezeichnet werden darf.

Datum

Unterschrift

Datum: _____

Elternfragebogen

Angaben zum Kind:

Vor- und Zuname Ihres Kindes: _____

Geschlecht: männlich weiblich

Geburtsdatum: _____

Besucht Ihr Kind zurzeit eine/n Kindergarten/Krippe? Ja Nein

Wenn ja: Seit wann: _____

Stunden pro Woche: _____

Besucht Ihr Kind Sport- oder Freizeitkurse? Ja Nein

Wenn ja, welche: _____

Angaben zur Familie des Kindes:

Eltern:

	Name	höchste abgeschlossene Ausbildung	Alter	gemeinsamer Haushalt
Vater:				<input type="checkbox"/> Ja <input type="checkbox"/> Nein
Mutter:				<input type="checkbox"/> Ja <input type="checkbox"/> Nein

Geschwister:

Name	Alter	Geschlecht	gemeinsamer Haushalt
		<input type="checkbox"/> M <input type="checkbox"/> W	<input type="checkbox"/> Ja <input type="checkbox"/> Nein
		<input type="checkbox"/> M <input type="checkbox"/> W	<input type="checkbox"/> Ja <input type="checkbox"/> Nein
		<input type="checkbox"/> M <input type="checkbox"/> W	<input type="checkbox"/> Ja <input type="checkbox"/> Nein
		<input type="checkbox"/> M <input type="checkbox"/> W	<input type="checkbox"/> Ja <input type="checkbox"/> Nein
		<input type="checkbox"/> M <input type="checkbox"/> W	<input type="checkbox"/> Ja <input type="checkbox"/> Nein

Wächst Ihr Kind mehrsprachig auf? Ja Nein

Wenn ja, welche Sprachen: _____

Muttersprache: _____

Angaben zu Geburt/Schwangerschaft/Gesundheit:

Risikoschwangerschaft: Ja Nein

Geburt in wievielter Woche: _____

Leidet Ihr Kind unter gesundheitlichen Beeinträchtigungen? Ja Nein

Wenn ja, welche: _____

Angaben zur Selbständigkeit:

Bitte beantworten Sie die folgenden Fragen nach Ihrer geschätzten Häufigkeit des Auftretens anhand der Zahlenkategorien 1 – 4.

	nie	manchmal	oft	immer
Mein Kind braucht tagsüber Windeln.	1	2	3	4
Mein Kind braucht in der Nacht Windeln.	1	2	3	4
Mein Kind sagt mir Bescheid, wenn es auf die Toilette gehen will.	1	2	3	4
Mein Kind kann sich 15 Minuten alleine beschäftigen	1	2	3	4
Mein Kind kann alleine ein Glas halten und daraus trinken.	1	2	3	4
Mein Kind teilt mir mit, wenn es Hunger oder Durst hat.	1	2	3	4
Mein Kind kann alleine mit einem Löffel essen	1	2	3	4
Mein Kind kann alleine mit einer Gabel essen.	1	2	3	4
Mein Kind geht alleine die Stiegen hinauf mit Festhalten	1	2	3	4
ohne Festhalten	1	2	3	4
Mein Kind geht alleine die Stiegen hinunter mit Festhalten	1	2	3	4
ohne Festhalten	1	2	3	4
Mein Kind sagt, wenn es etwas möchte.	1	2	3	4
Mein Kind sagt, wenn es etwas nicht möchte.	1	2	3	4
Mein Kind kann sich alleine anziehen.	1	2	3	4
Mein Kind kann den rechten vom linken Schuh unterscheiden.	1	2	3	4

Angaben zum Sozial- und Spielverhalten:

	nie	manchmal	oft	immer
Mein Kind spielt gerne mit anderen Kindern.	1	2	3	4
Mein Kind bevorzugt seine Eltern/Bezugsperson als Spielpartner.	1	2	3	4
Mein Kind kann einen Ball fangen.	1	2	3	4
Mein Kind kann einen Ball werfen.	1	2	3	4
Mein Kind turnt/klettert gerne.	1	2	3	4
Mein Kind kann im Spiel einfache Regeln befolgen.	1	2	3	4
Mein Kind spielt mit Konstruktionsspielzeug. (z.B.: Bauklötze, Duplo, Lego, ...)	1	2	3	4
Meinem Kind fallen viele Dinge ein, die es gerne spielen möchte.	1	2	3	4
Mein Kind imitiert im Spiel alltägliche Handlungen von Erwachsenen.	1	2	3	4
Mein Kind spielt Rollenspiele.	1	2	3	4

Angaben zur Sprache:

Hier finden Sie eine Wortliste. Bitte kreuzen Sie jene Wörter an, die Sie schon öfters von Ihrem Kind gehört haben. Bitte beachten Sie, dass Sie nur Wörter ankreuzen, die Ihr Kind selbst verwendet. Dazu zählen auch Wörter, die es etwas anders ausspricht (z.B.: „Nie“ statt „Knie“). Falls Ihr Kind für etwas ein anderes Wort benutzt, schreiben Sie es bitte daneben (z.B.: „heihei“ statt „schlafen“).

- | | | |
|------------------------------------|-------------------------------------|-------------------------------------|
| <input type="checkbox"/> schmutzig | <input type="checkbox"/> Auto | <input type="checkbox"/> Fenster |
| <input type="checkbox"/> Torte | <input type="checkbox"/> Apfel | <input type="checkbox"/> Bild |
| <input type="checkbox"/> Käse | <input type="checkbox"/> Birne | <input type="checkbox"/> Badezimmer |
| <input type="checkbox"/> Besen | <input type="checkbox"/> Banane | <input type="checkbox"/> Wasser |
| <input type="checkbox"/> hungrig | <input type="checkbox"/> Zitrone | <input type="checkbox"/> Badewanne |
| <input type="checkbox"/> Lampe | <input type="checkbox"/> Karotte | <input type="checkbox"/> Katze |
| <input type="checkbox"/> Fernseher | <input type="checkbox"/> Orange | <input type="checkbox"/> Besteck |
| <input type="checkbox"/> Tür | <input type="checkbox"/> Sessel | <input type="checkbox"/> Schlaf |
| <input type="checkbox"/> schlafen | <input type="checkbox"/> Gitterbett | <input type="checkbox"/> Fahrrad |

- | | | |
|-----------------------------------|---------------------------------------|----------------------------------|
| <input type="checkbox"/> Pflaster | <input type="checkbox"/> Schlafzimmer | <input type="checkbox"/> fliegen |
| <input type="checkbox"/> Bub | <input type="checkbox"/> durstig | <input type="checkbox"/> Elefant |
| <input type="checkbox"/> Knie | <input type="checkbox"/> Mädchen | <input type="checkbox"/> Hase |
| <input type="checkbox"/> Mann | <input type="checkbox"/> Kasten | <input type="checkbox"/> Eis |
| <input type="checkbox"/> Schuh | <input type="checkbox"/> Tasche | <input type="checkbox"/> Hund |

Mein Kind hat bereits begonnen, zwei Wörter miteinander zu verbinden.	<input type="checkbox"/> Ja <input type="checkbox"/> Nein
Mein Kind hat bereits begonnen, drei oder mehrere Wörter miteinander zu verbinden.	<input type="checkbox"/> Ja <input type="checkbox"/> Nein
Mein Kind verwendet andere Fragewörter außer „Wo“.	<input type="checkbox"/> Ja <input type="checkbox"/> Nein
Mein Kind verwendet bereits die Vergangenheitsform mit den Hilfsverben „haben“ und „sein“.	<input type="checkbox"/> Ja <input type="checkbox"/> Nein
Mein Kind verwendet die Vergangenheitsform schon richtig (z.B.: gegessen, weh getan, ...).	<input type="checkbox"/> Ja <input type="checkbox"/> Nein
Mein Kind verwendet bei der Verneinung das Wort „nicht“ (z.B.: nicht schlafen, nicht Zähneputzen, ...).	<input type="checkbox"/> Ja <input type="checkbox"/> Nein

Vielen Dank für Ihre Mitarbeit!

✓ Gelöst	X nicht gelöst
v. verweigert	- nicht vorgegeben

ANHANG D: Protokollbogen

Name des Testkinds: _____
 Name des Testleiters: _____
 Datum: _____
 Uhrzeit (Anfang/Ende): _____

Geburtsdatum: _____
 Alter: _____
 Anwesende Personen: _____
 Kontaktaufnahme zum Kind: _____

KIND-BEZUGSPERSON-BEZIEHUNG WORTSCHATZ PUPPENKÜCHE (OBST)

Ratingskala	Lösung von Bezugsperson (1- lange Zeit nicht möglich, 5- gleichgültig)				
	1	2	3	4	5
	Körperkontakt	mitspielen	Exploration	Verabschiedung	gleichgültig
	Kontakt zur Bezugsperson während Testung (1 -dauernd, 5 - gleichgültig)				
	1	2	3	4	5
	Körperkontakt	spielen	hingehen	Blickkontakt	gleichgültig

DESKRIPTIVE ANGABEN

Dauer	• Gesamt		
	• Anlaufzeit		
	• Beschäftigung mit Materialien		
	• Pause(n) – Anzahl		
	• Pause(n)–Minuten gesamt		
	• Verabschiedung		

Visuelle Wahrnehmung PUPPENKÜCHE (Obst sortieren nach Farbe)

Stabilität Farb-konzept Min 3	Rot		
	Grün		
	Gelb		
	orange		

Feinmotorik PUPPENKÜCHE (0/1 + Anzahl aufschreiben)

Feinmotorik	Teller einordnen (mind. 3)		
	Anzahl		
	Kochlöffel aufhängen (mind. 1)		
	Anzahl		
	Schraubverschlussglas öffnen		
	Schraubverschlussglas schließen		
	Aufkehren		
	Umfüllaufgabe		
	links & rechts umgefüllt (0/1)		
	nur mit einer Hand umgefüllt (0,5)		
ohne daneben schütten			

GEDÄCHTNIS - PUPPENKÜCHE

Gedächtnis	Gib mir A		
	Gib mir A + B		
	Gib mir A + B + C		

WORTSCHATZ (aktiv) PUPPENKÜCHE (OBST) + Farbe benennen (0/1)

& Farbdifferenzierung aktiv Wortschatz aktiv	Obst benennen		
	Apfel		
	Birne		
	Banane		
	Zitrone		
	Karotte		
	Orange		
	Farbe benennen		
	rot		
	grün		
	gelb		
	orange		

„Als-ob-Spiel“ - PUPPENKÜCHE

	Als-Ob-Spiel (1/0,5/0)		
	Kein Als-Ob-Spiel vorhanden (0)		
	Als-Ob-Spiel mit realen Dingen (0,5)		
	Als-Ob-Spiel gänzlich vorgestellt/ Objekt anders verwendet (1)		

NUMERISCHES WISSEN - PUPPENKÜCHE

Kognitive Entwicklung	Numerisches Wissen		
	Passive Mengenerfassung		
	Verständnis von 1: Gib mir 1...		
	Verständnis von 2: Gib mir 2 ...		
	Gibt eine Handvoll		
	gibt eines nach dem anderen		
	zählt offensichtlich		
	Verständnis von 3: gib mir 3		
	Gibt eine Handvoll		
	gibt eines nach dem anderen		
	zählt offensichtlich		
	Aktive Mengenerfassung		
	Wie viele sind das: Menge von 2		
	zählt offensichtlich		
	Wie viele sind das: Menge von 3		
	zählt offensichtlich		
	Zählen (0/1)		
	Anzahl Zählen		
	Zählen richtig (0/1)		
	Anzahl Zählen richtig		
	Eins zu eins Zuordnung (0/1)		
	Anzahl Eins zu eins Zuordnung		
	Stabile Zahlenfolge (0/1)		
Anzahl Stabile Zahlenfolge			
Richtige Reihenfolge (0/1)			
Anzahl Richtige Reihenfolge			
Kardinalsprinzip (0/1)			
Anzahl Kardinalsprinzip			

VERKLEIDEN – Selbstständigkeit und Feinmotorische Fähigkeiten (0/1)

Selbstständigkeit	Schuhe anziehen		
	Versucht selbstständig		
	Sucht Hilfe		
	Desinteresse		
	Schuhe anziehen geschafft		
	Hose anziehen		
	Versucht selbstständig		
	Sucht Hilfe		
	Desinteresse		
	In Hosenbein reinschlüpfen		
	Hose hochziehen		
	Hut aufsetzen		
	Versucht selbstständig		
	Sucht Hilfe		
	Desinteresse		
	Aufsetzten geschafft		
	T-Shirt anziehen		
	Versucht selbstständig		
	Sucht Hilfe		
	Desinteresse		
In Ärmel hineinschlüpfen			
Mit Kopf durchschlüpfen			
Shirt hinunterziehen			

ToM-BUCH – IN ANDERE HINEINVERSETZEN (0/1 verbal oder nonverbal)

ToM-Buch	Aufmerksamkeit (0/0,5/1)		
	Dauer bis zur 1. Unaufmerksamkeit (s)		
	Lässt sich zurückholen		
	Seiten aufmerksam angesehen (gesamt)	S.	
	Küche benennen		
	schmutzige Hände		
	Torte/Käse (aktiv)		
	Besen passiv		
	Wohnzimmer benennen		
	Hungrig		
	Lampe (aktiv)		
	Fernseher (passiv)		
	Tür (passiv)		
	Kinderzimmer benennen		
	Schlafen		
	Pflaster		
	Schuh (aktiv)		
	Stuhl/Sessel (passiv)		
	Gitterbett (passiv)		
	Schlafzimmer benennen		
	durstig		
	Kasten (aktiv)		
	Tasche (aktiv)		
	Fenster (passiv)		
	Bild (passiv)		
	Badezimmer (benennen)		
	Wasser (aktiv)		
	Badewanne (passiv)		

GROBMOTORIK STIEGE

Stiege	Hinaufgehen		
	mit anhalten (0,5)		
	ohne anhalten (1)		
	Nachstellschritt (0,5)		
	Wechselschritt (1)		
	Hinuntergehen		
	mit anhalten (0,5)		
	ohne anhalten (1)		
	Nachstellschritt (0,5)		
	Wechselschritt (1)		
	von letzter Stufe hüpfen (0/1)		

GROBMOTORIK BALL (MUSIK! 2 Meter 0/1)

Ballspielen (2 von 3)	Werfen		
	Anzahl geworfen		
	Fangen		
	Anzahl gefangen		
	Fußkick		

GROBMOTORIK STEHEN, HÜPFEN, LAUFEN (MUSIK! 1/0)

	Schmalere Weg		
	Vorwärts Balancieren-Linie		
	Seitwärts Gehen - Linie		
	Zehenspitzenangang-(Linie egal)		
	Beidbeiniges Hüpfen (3 Mal)		
	Stehen bleiben aus vollem Lauf		
	1. Versuch:		
	Sofort		
	2 Schritte		
	fällt um		
	2. Versuch		
	Sofort		
	2 Schritte		
	fällt um		
	3. Versuch		
Sofort			
2 Schritte			
fällt um			

FEINMOTORIK & FARBDIFFERENZIERUNG & VISUMOTORIK (1/0)

Zeichnen	Aufmerksamkeit (0/0,5/1)		
	Dauer bis 1. Unterbrechung		
	lässt sich wieder zurückführen		
	Stifthaltung:		
	Primitive Formen (0,5)		
	Erwachsene Stifthaltung (1)		
	Unterarm ruht auf Tisch		
	Hand zum Blatthalten verwendet		
	Linien zeichnen:		
	Horizontal		
	Vertikal		
	Kreis		
	Farben erkennen passiv (Stifte)		
	Grün		
	Gelb		
	Rot		
	Orange		
	Farben benennen aktiv (Stifte):		
	Blau		
	Weiß		
Lila/violett			
Blatt mit Schere schneiden (mind2)			

VISUMOTORIK – Formen

Turm	Aufmerksamkeit (0/0,5/1)		
	Dauer bis 1. Unterbrechung		
	lässt sich wieder zurückführen		
	Vertikal		
	Anzahl höchster Turm		
	Horizontal		
	Mauer		
	3D		
	Anzahl Bausteine gesamt		

Skizze von Turm

VISUMOTORIK – TISCHAUFGABEN

	Perlen fädeln (min 2)		
	Anzahl Perlen		
	Vorgezeigt		
	Aufmerksamkeit (0/0,5/1)		
	Dauer bis zur ersten Unterbrechung		
	Ließ sich zurückholen		
	Perlen Farben passiv – GIB MIR		
	blau		
	lila		
	weiß		
	Puzzle		
	Anzahl Puzzleteile (max 6)		
	Aufmerksamkeit (0/0,5/1)		
	Dauer bis zur ersten Unterbrechung		
	Ließ sich zurückholen		

VISUELLES GEDÄCHTNIS – TISCHAUFGABEN

Visuelles Ged.	Memory Sonne		
	Memory Haus		
	Memory Baum		
	Teddybär		
	Haus		

FARB-, FORM - UND GRÖßENDIFF. (BODENMATTEN)

	Form- und Größendifferenzierung		
	Formdiff. 1 (Kreis, Quadrat, Dreieck, Rechteck)		
	Formdiff. 2 (Vierecke)		
	Größendiff. 1 (Rechtecke)		
	Größendiff. 2 (Kreise)		

SPRACHENTWICKLUNG

Satzlänge	Einwortäußerungen		
	Zweiwortäußerungen		
	Dreiwortäußerungen		
	Mehrwortäußerungen		
	Satz korrekt (S, P, O)		
Haupt- und Nebensätze	Vorkommen von Verbindungen		
	Verbindung mit und/aber		
	Verbindung mit weil		
	Verbindungen mit dass		
	Verbindungen mit wenn		
	Verbindungen mit ob		
	Relativsätze		
	<ul style="list-style-type: none"> • mit korrektem Relativpronomen • mit wo • ohne Relativpronomen 		
Verben	Stellung im Aussagesatz		
	Endstellung		
	Verbstellung korrekt		
	Partizip Perfekt (kommt vor)		
	ohne ge- gebildet		
	mit ge- gebildet		
	Korrekte Partizipendung		
mit Hilfsverben (haben+sein)			

Verneinung	Verneinung		
	Korrekte Wortstellung		
	Bildung		
	mit nein		
	mit nicht		
Fragen	Fragenintonation		
	Ja/Nein-Fragen		
	Inversion von Subjekt und Prädikat		
	W-Fragen		
	Inversion von Subjekt und Prädikat		
	Fragewort vorhanden (außer wo?)		
	Wo? Vorhanden		
	Alternativfragen		
	Wortstellung korrekt		

ANHANG E: Tabellen Gesamtskalen

Tabelle 22 Reliabilität und Itemtrennschärfen vor und nach der Selektion und Itemschwierigkeiten für die Skala „Sprache gesamt“ (in Klammer – Anzahl der fehlenden Werte)

N=20 vor Selektion: 49 Items nach erster S.: 30 Items nach zweiter S.: 28 Items	Schwierigkeit	Trennschärfe	Trennschärfe nach erster Selektion	Trennschärfe nach zweiter Selektion
Item 1: Apfel aktiv	.55	.41	.44	.42
Item 2: Banane aktiv	.75	.54	.58	.58
Item 3: Birne aktiv	.35	.5	.5	.5
Item 4: Karotte aktiv	.75	.54	.58	.58
Item 5: Orange aktiv	.4	.6	.58	.56
Item 6: Zitrone aktiv	.2	.38	.34	.33
Item 7: Badezimmer aktiv(5)	.25	.66	.6	.58
Item 8: Kasten aktiv(5)	0	-	-	-
Item 9: Kinderzimmer aktiv (3)	0	-	-	-
Item 10: Küche aktiv(2)	.1	.33	.26	-
Item 11: Lampe aktiv(1)	.6	.13	-	-
Item 12: Schlafzimmer (5) aktiv	.05	.42	.38	.38
Item 13: Schuh aktiv(1)	.7	.59	.56	.56
Item 14: Tasche aktiv (3)	.3	.47	.44	.45
Item 15: Torte/Käse aktiv	.45	.25	-	-
Item 16: Wasser aktiv (2)	.4	.61	.57	.57
Item 17: Wohnzimmer aktiv(2)	.05	.42	.38	.38
Item 18: Badewanne passiv(1)	.85	.4	.41	.42
Item 19: Besen passiv	.7	.48	.46	.46
Item 20: Bild passiv (2)	.55	.15	-	-
Item 21: Fenster passiv (3)	.15	.2	-	-
Item 22: Fernseher passiv (1)	.55	.29	-	-
Item 23: Gitterbett passiv (1)	.7	.3	.28	-
Item 24: Stuhl passiv (1)	.8	.04	-	-
Item 25: Tür passiv (1)	.9	.06	-	-
Item 26: Satzlänge	.78	.8	.84	.86
Item 27: korrekte Verbstellung in Aussagesätzen	.7	.79	.84	.84

N=20 vor Selektion: 49 Items nach erster S.: 30 Items nach zweiter S.: 28 Items	Schwierigkeit	Trennschärfe	Trennschärfe nach erster Selektion	Trennschärfe nach zweiter Selektion
Item 28: Vorkommen von Verneinungen	1	-	-	-
Item 29: korrekte Verbstellung bei der Verneinung	.6	.76	.77	.78
Item 30: Bildung der Verneinung	.85	.8	.85	.85
Item 31: Vorkommen von Alternativ-fragen	0	-	-	-
Item 32: Wortstellung bei Alternativ-fragen	0	-	-	-
Item 33: Vorkommen von W- Fragen	.75	.63	.66	.67
Item 34: Wortstellung bei W- Fragen	.6	.77	.81	.82
Item 35: Fragewort bei W-Fragen	.68	.8	0,84	.85
Item 36: Vorkommen Ja-/ Nein-Fragen	.7	.27	-	-
Item 37: Wortstellung Ja-/ Nein-Fragen	.3	.56	.59	.6
Item 38: Fragen- intonation	.85	.25	-	-
Item 39: Vorkommen eines Perfekts	.65	.81	.84	.84
Item 40: Perfekt mit ge- gebildet	.45	.7	.66	.65
Item 41: Hilfsverb bei der Perfekt- bildung verwendet	.45	.62	.65	.67
Item 42: Korrekte Partizip-endung	.55	.7	.73	.73
Item 43: Kombination von Haupt- und Nebensatz	.1	.5	.47	.45
Item 44: Anzahl der Verbindungs-wörter bei Haupt- und Nebensatz	0	-	-	-
Item 45: Korrekte und vollständige Sätze	.55	.77	.78	.79
Item 46: Relativsätze	0	-	-	-
Item 47: korrektes Relativ-pronomen	0	-	-	-

N=20 vor Selektion: 49 Items nach erster S.: 30 Items nach zweiter S.: 28 Items	Schwierigkeit	Trennschärfe	Trennschärfe nach erster Selektion	Trennschärfe nach zweiter Selektion
Item 48: Relativ- pronomen wo	0	-	-	-
Item 49: Relativsatz ohne Relativ- pronomen	0	-	-	-

Reliabilität vor der Selektion $\alpha=.932$

Reliabilität nach erster Selektion $\alpha=.944$

Reliabilität nach zweiter Selektion $\alpha=.946$

**Tabelle 23: Reliabilität, Itemtrennschärfen und Itemschwierigkeiten für die Skala „numerisch
gesamt“ (in Klammer – Anzahl der fehlenden Werte)**

N=20 11 Items	Schwierigkeit	Trennschärfe
Item 1: Gib mir 1	.90	.43
Item 2: Gib mir 2	.65	.55
Item 3: Gib mir 3 (2)	.15	.38
Item 4: Wie viele sind das - 2	.40	.61
Item 5: Wie viele sind das - 3 (2)	.15	.53
Item 6: Zählt	.65	.76
Item 7: richtig gezählt	.10	.48
Item 8: Eins zu Eins Zuordnung	.45	.64
Item 9: stabile Reihenfolge	.45	.68
Item 10: richtige Reihenfolge	.35	.35
Item 11: Kardinalsprinzip	.55	.78

Reliabilität $\alpha=.866$

Tabelle 24: Reliabilität, Itemtrennschärfen vor und nach der Selektion und Itemschwierigkeiten für die Skala „Gedächtnis gesamt“ (in Klammer- Anzahl fehlender Werte)

N=20 vor Selektion: 8 Items nach Selektion: 2 Items	Schwierigkeit	Trennschärfe vor Selektion	Trennschärfe nach der Selektion
Item 1: Gib mir a	.9	.04	-
Item 2: Gib mir a-b (1)	.7	.25	-
Item 3: Gib mir a-b-c (4)	.25	.06	-
Item 4: Bilderset: Bär	.45	.37	.45
Item 5: Bilderset: Haus (1)	.60	.29	-
Item 6: Memory Sonne	.65	.23	-
Item 7: Memory Haus	.90	.15	-
Item 8: Memory Baum	.65	.39	.45

Reliabilität vor der Selektion $\alpha=.497$

Reliabilität nach der Selektion $\alpha=.623$

ANHANG F: Normalverteilungsprüfung für Validitätsberechnung

Tabelle 25: Mittelwert, Standardabweichung, Schiefe und Kurtosis der Summenscores

	MW	SD	Schiefe	Kurtosis
Alter	28.5	3.15	.63	-.54
aktiver Wortschatz	4.2	2.95	.43	.14
passiver Wortschatz	2.95	1.19	-.52	-1.4
grammatikalische Fähigkeiten	9.56	5.3	-.49	-1.31
Sprache gesamt	14.56	8.28	-.14	-1.13
Hineinversetzen in Andere	1.45	1.39	.63	-.97
aktive Mengenerfassung	.55	.76	1.02	-.37
passive Mengenerfassung	1.55	.69	-1.28	.54
Zählen	2.55	2.06	.12	-1.27
Numerisch gesamt	4.8	3.22	.00	-.97
visuelles Gedächtnis	1.75	1.16	-.35	-1.34

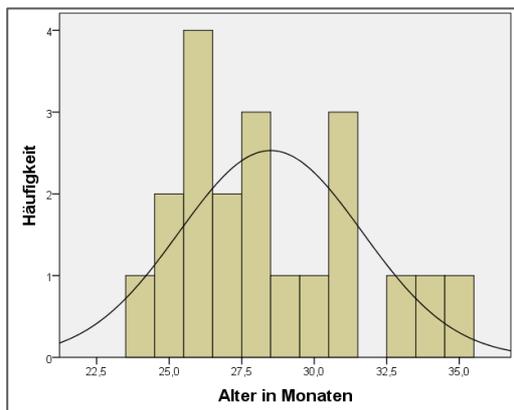


Abbildung 3: Histogramm zur Normalverteilungsprüfung für die Variable „Alter in Monaten“

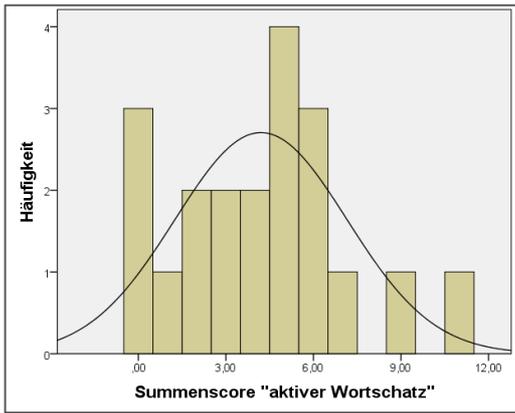


Abbildung 4: Histogramm zur Normalverteilungsprüfung für den Summenscore „aktiver Wortschatz“

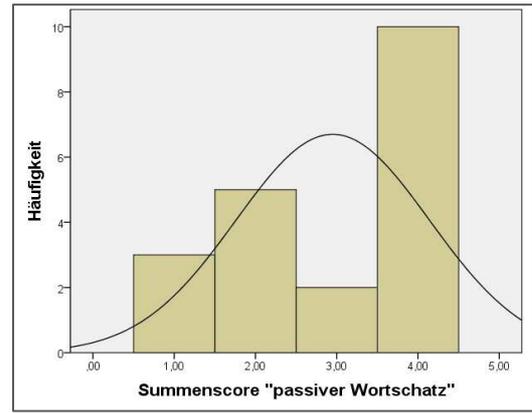


Abbildung 5: Histogramm zur Normalverteilungsprüfung für den Summenscore „passiver Wortschatz“

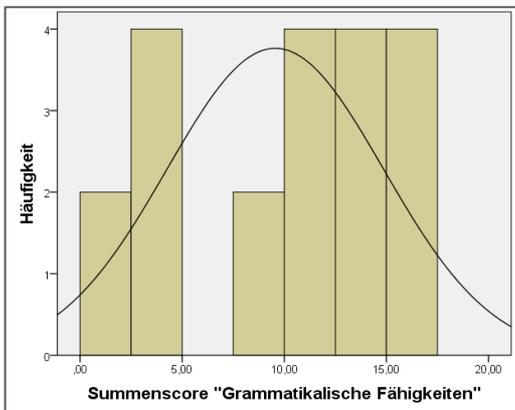


Abbildung 7: Histogramm zur Normalverteilungsprüfung für den Summenscore „grammatikalische Fähigkeiten“

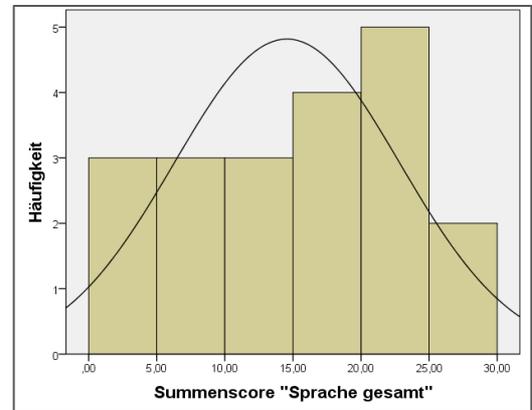


Abbildung 6: Histogramm zur Normalverteilungsprüfung für den Summenscore „Sprache gesamt“

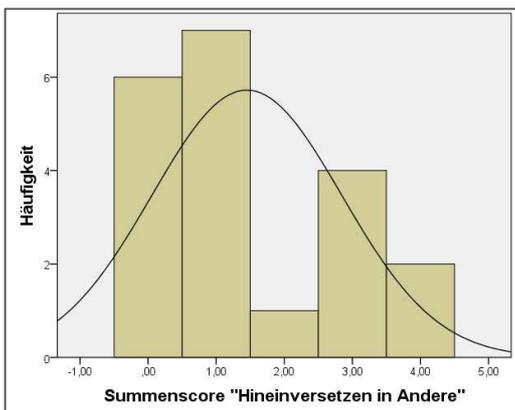


Abbildung 8: Histogramm zur Normalverteilungsprüfung für den Summenscore „Hineinversetzen in Andere“

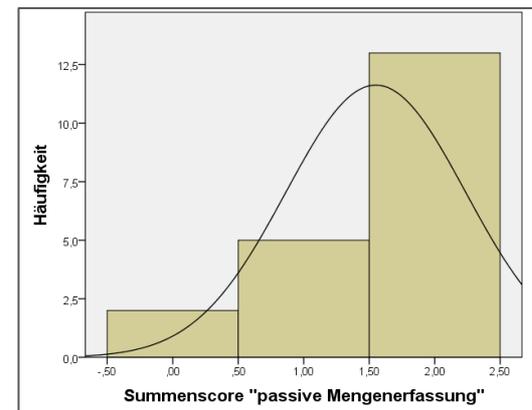


Abbildung 9: Histogramm zur Normalverteilungsprüfung für den Summenscore „passive Mengenerfassung“

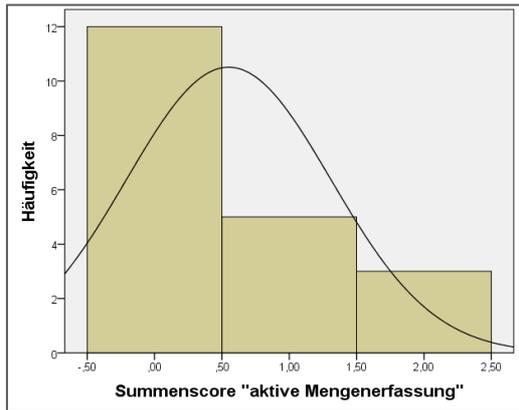


Abbildung 11: Histogramm zur Normalverteilungsprüfung für den Summenscore „aktive Mengenerfassung“

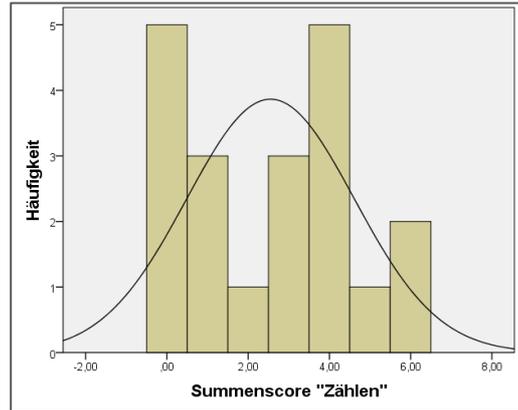


Abbildung 10: Histogramm zur Normalverteilungsprüfung für den Summenscore „Zählen“

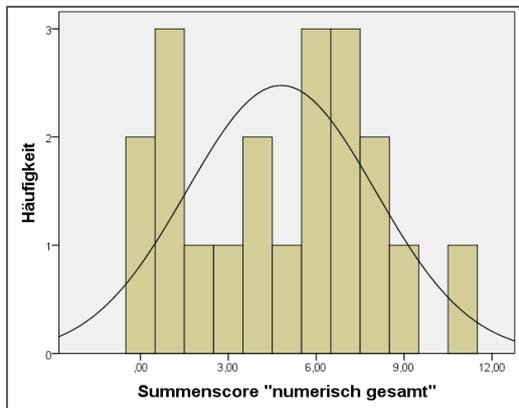


Abbildung 12: Histogramm zur Normalverteilungsprüfung für den Summenscore „numerisch gesamt“

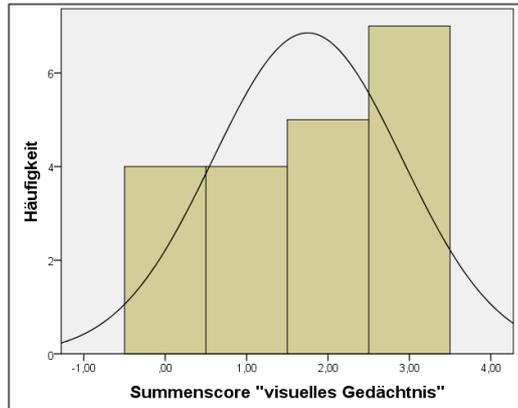
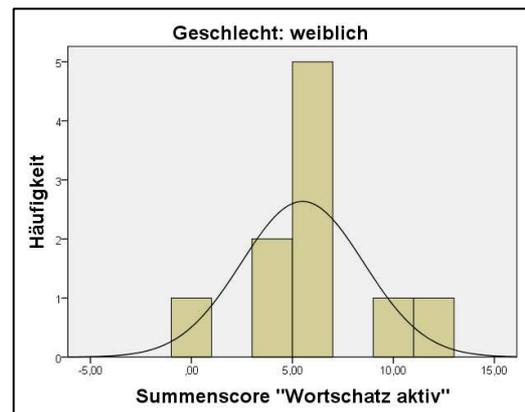
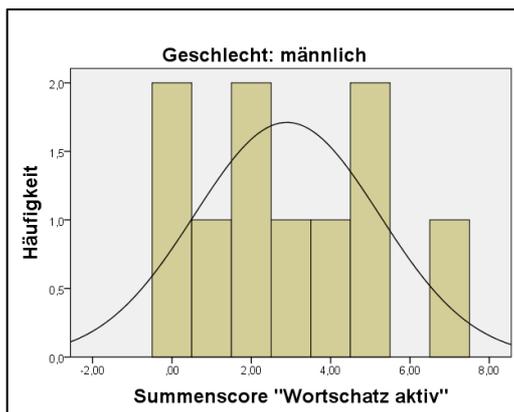


Abbildung 13: Histogramm zur Normalverteilungsprüfung für den Summenscore „numerisch gesamt“

ANHANG G: Voraussetzungsprüfung ANOVA

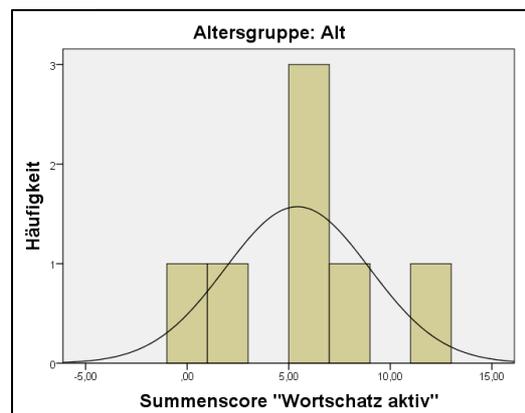
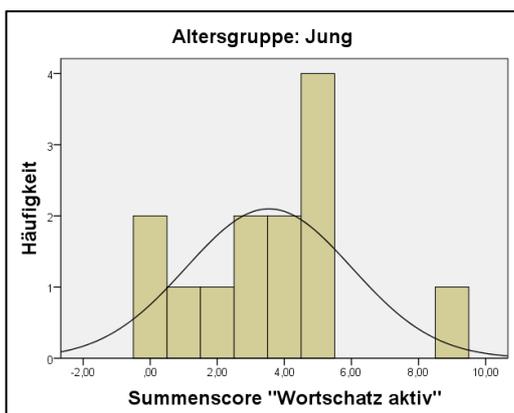
Tabelle 26: Voraussetzungsprüfung „aktiver Wortschatz“

	männlich	weiblich	Jung	Alt
N	10	10	13	7
MW	2.90	5.50	3.54	5.43
SD	2.33	3.03	2.47	3.55
Schiefe	.35	.12	.44	-.12
Kurtosis	-.8	.92	.77	.39
Levene	F(3,16)=.27; p=.843			



a)

b)



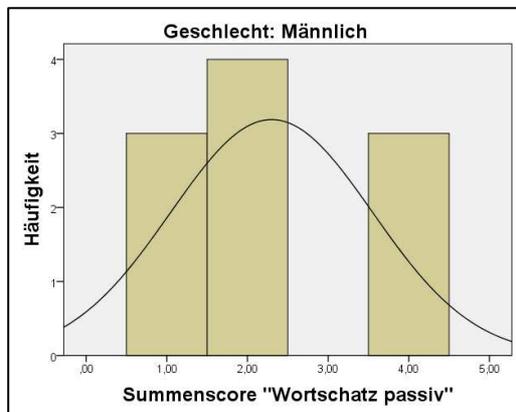
c)

d)

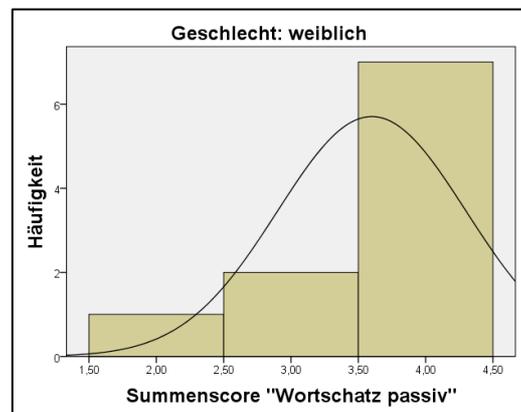
Abbildung 14 a-d: Histogramme zur Normalverteilungsprüfung für die Skala „aktiver Wortschatz“ getrennt für Geschlecht (a männlich, b weiblich) und Altersgruppen (c jung, d alt)

Tabelle 27: Voraussetzungsprüfung „passiver Wortschatz“

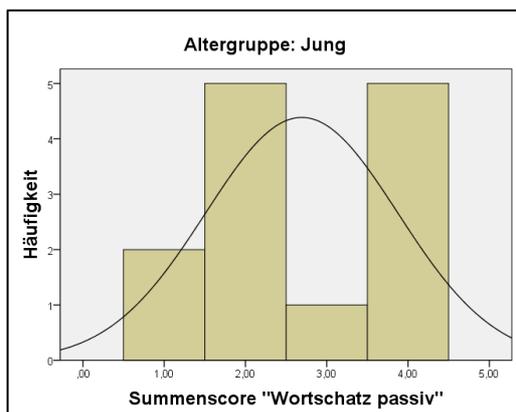
	männlich	weiblich	Jung	Alt
N	10	10	13	7
MW	2.3	3.6	2.69	3.43
SD	1.25	0.7	1.18	1.13
Schiefe	.57	-1.66	-.01	-2.16
Kurtosis	-1.35	2.05	-1.66	4.58
Levene	F(3,16)=2.72; p=.079			



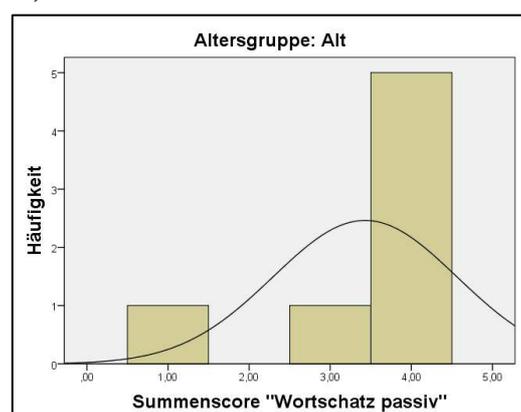
a)



b)



c)

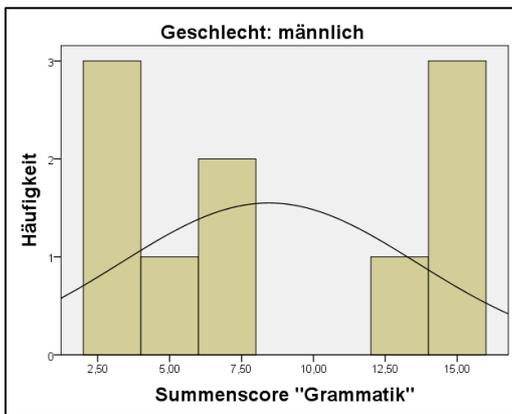


d)

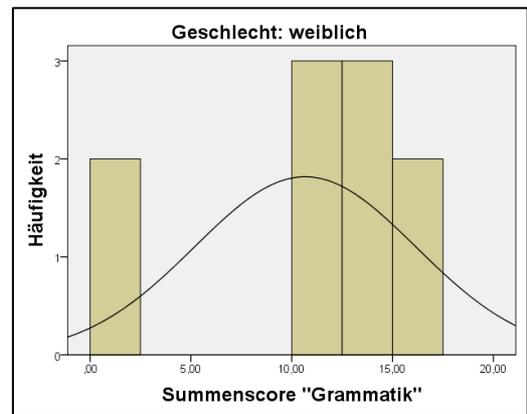
Abbildung 15 a-d: Histogramme zur Normalverteilungsprüfung für die Skala „passiver Wortschatz“ getrennt für Geschlecht (a männlich, b weiblich) und Altersgruppen (c jung, d alt)

Tabelle 28: Voraussetzungsprüfung „grammatikalische Fähigkeiten“

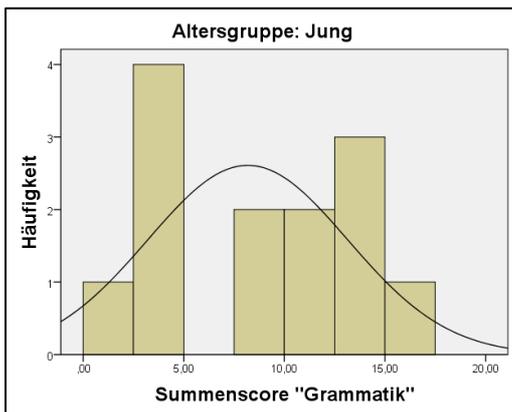
	männlich	weiblich	Jung	Alt
N	10	10	13	7
MW	8.45	10.67	8.18	12.12
SD	5.14	5.48	4.97	5.27
Schiefe	.20	-1.31	-.07	-2.07
Kurtosis	-1.87	.59	-1.64	4.74
Levene	F(3,16)=.4; p=.758			



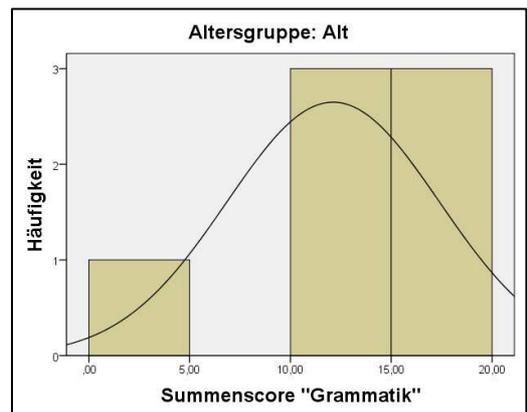
a)



b)



c)

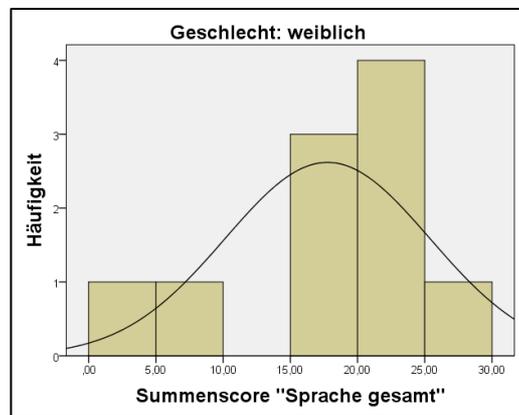
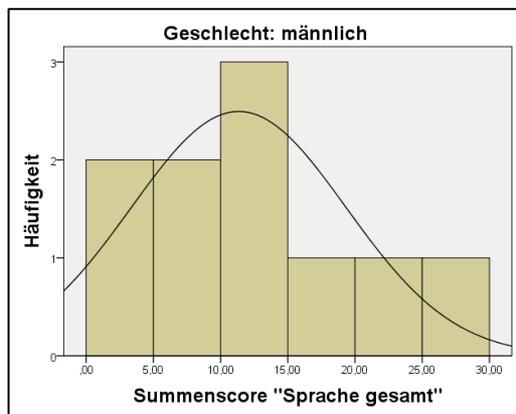


d)

Abbildung 16 a-d: Histogramme zur Normalverteilungsprüfung für die Skala „grammatikalische Fähigkeiten“ getrennt für Geschlecht (a männlich, b weiblich) und Altersgruppen (c jung, d alt)

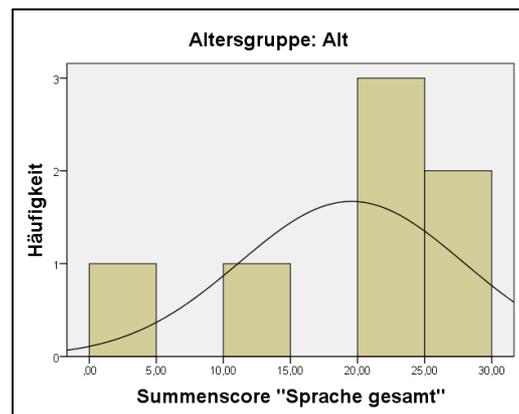
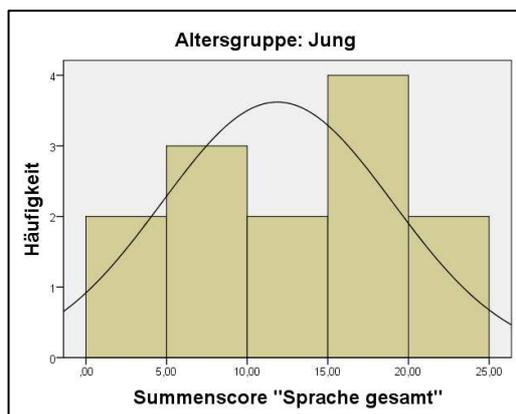
Tabelle 29: Voraussetzungsprüfung „Sprache gesamt“

	männlich	weiblich	Jung	Alt
N	10	10	13	7
MW	11.35	17.77	11.87	19.55
SD	7.99	7.62	7.16	8.36
Schiefe	.39	-.71	-.08	-1.18
Kurtosis	-.89	.20	-1.45	1.42
Levene	F(3,16)=.26; p=.857			



a)

b)



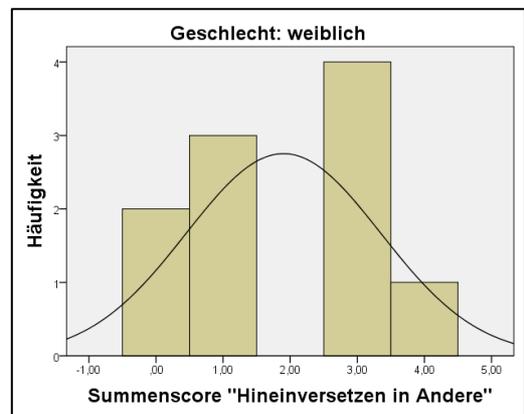
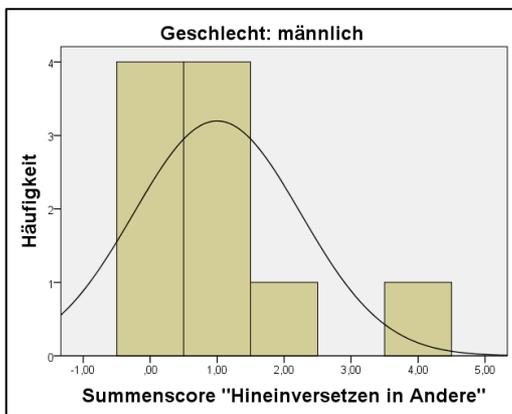
c)

d)

Abbildung 17 a-d: Histogramme zur Normalverteilungsprüfung für die Skala „Sprache gesamt“ getrennt für Geschlecht (a männlich, b weiblich) und Altersgruppen (c jung, d alt)

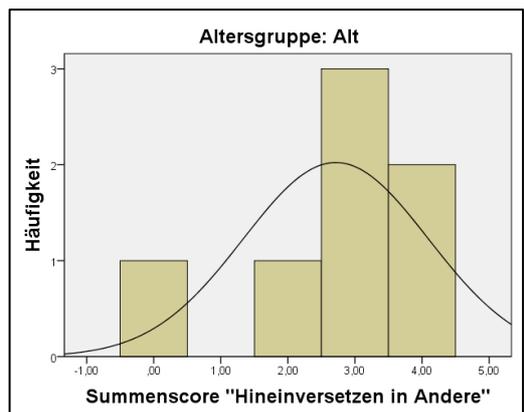
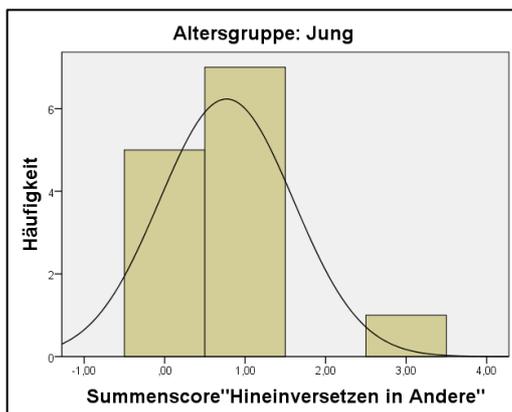
Tabelle 30: Voraussetzungsprüfung „Hineinversetzen in Andere“

	männlich	weiblich	Jung	Alt
N	10	10	13	7
MW	1	1.9	0.77	2.71
SD	1.25	1.45	0.83	1.38
Schiefe	1.72	-0.6	1.52	-1.42
Kurtosis	3.42	-1.7	3.71	2.32
Levene	F(3,16)=0.93; p=.448			



a)

b)



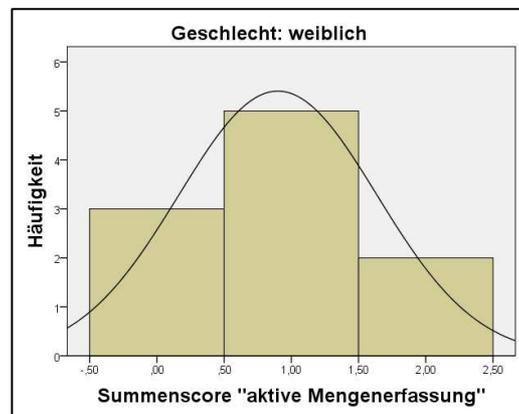
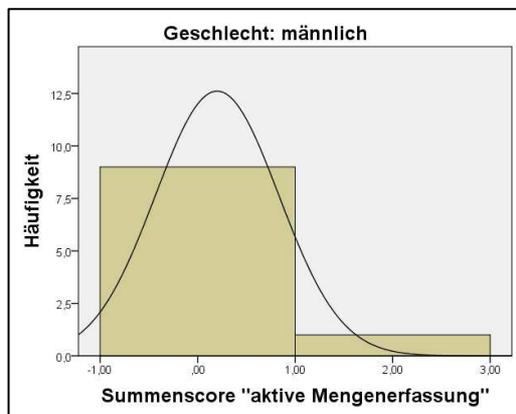
c)

d)

Abbildung 18: a-d: Histogramme zur Normalverteilungsprüfung für die Skala „Hineinversetzen in Andere“ getrennt für Geschlecht (a männlich, b weiblich) und Altersgruppen (c jung, d alt)

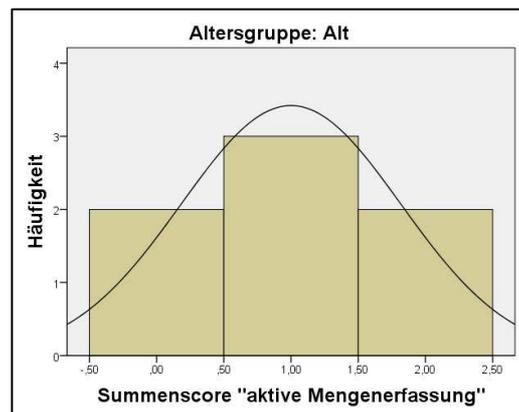
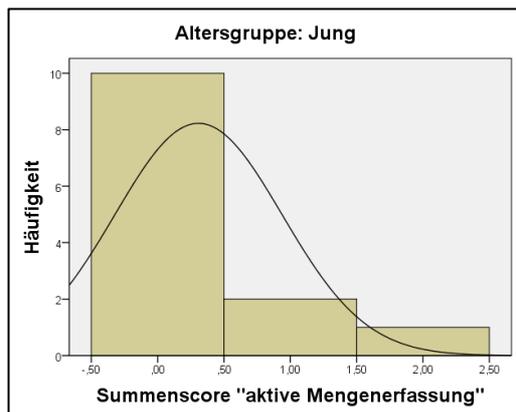
Tabelle 31: Voraussetzungsprüfung „aktive Mengenerfassung“

	männlich	weiblich	Jung	Alt
N	10	10	13	7
MW	.2	.9	.31	1
SD	.63	.74	.63	.82
Schiefe	3.16	.17	2.05	0
Kurtosis	10	-0.73	3.71	-1.2
Levene	$F(3,16)=6.16; p=.005$			



a)

b)



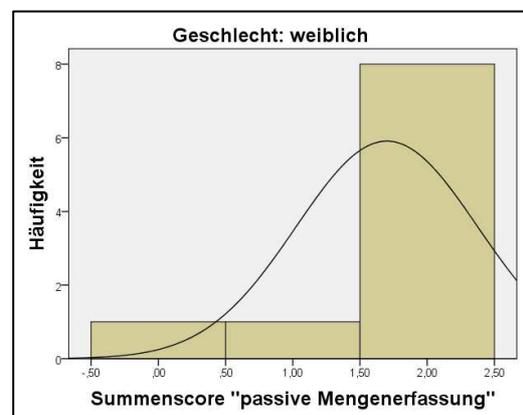
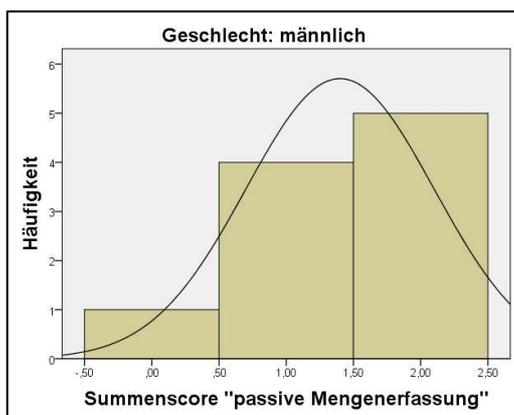
c)

d)

Abbildung 19: a-d: Histogramme zur Normalverteilungsprüfung für die Skala „aktive Mengenerfassung“ getrennt für Geschlecht (a männlich, b weiblich) und Altersgruppen (c jung, d alt)

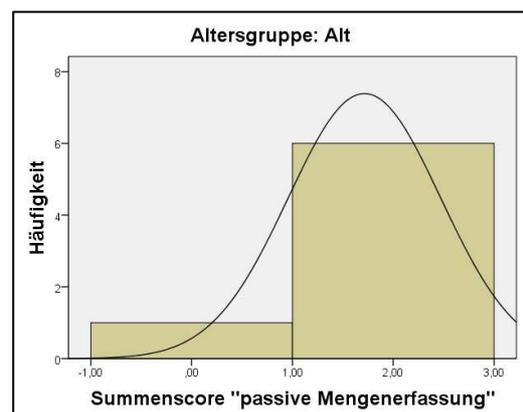
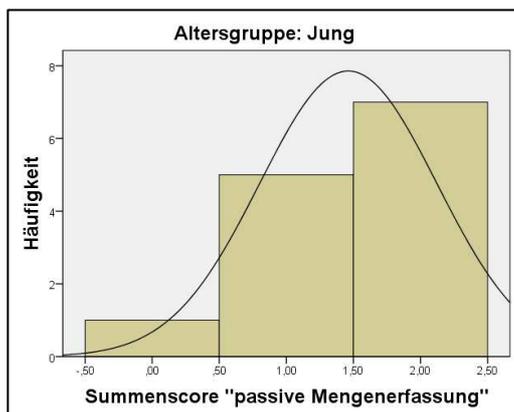
Tabelle 32: Voraussetzungsprüfung „passive Mengenerfassung“

	männlich	weiblich	Jung	Alt
N	10	10	13	7
MW	1.4	1.7	1.46	1.71
SD	0.7	0.67	0.66	0.76
Schiefe	-.78	-2.28	-.86	-2.64
Kurtosis	-.15	4.77	-.03	7
Levene	F(3,16)=1.7; p=.208			



a)

b)



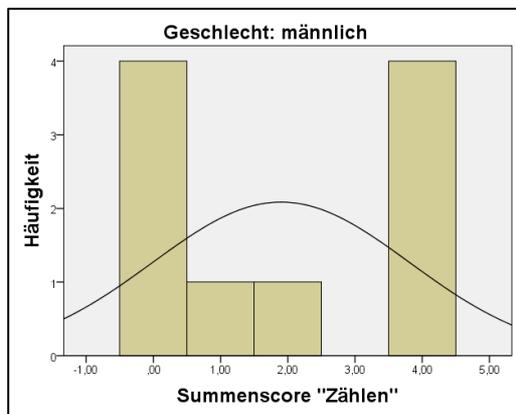
c)

d)

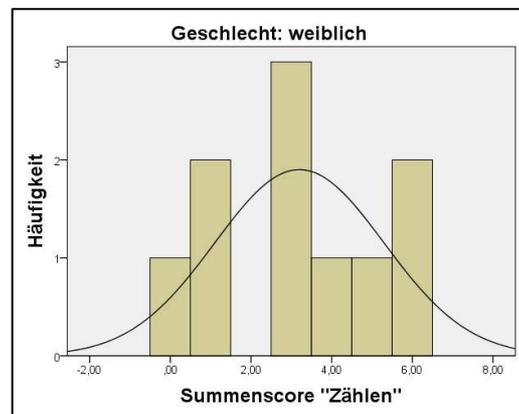
Abbildung 20 a-d: Histogramme zur Normalverteilungsprüfung für die Skala „passive Mengenerfassung“ getrennt für Geschlecht (a männlich, b weiblich) und Altersgruppen (c jung, d alt)

Tabelle 33: Voraussetzungsprüfung „Zählen“

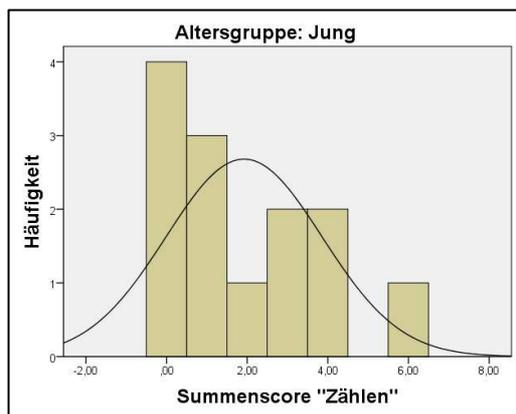
	männlich	weiblich	Jung	Alt
N	10	10	13	7
MW	1.9	3.2	1.92	3.71
SD	1.91	2.1	1.93	1.89
Schiefe	.18	-.06	.78	-1.3
Kurtosis	-2.21	-1.12	-.26	2.67
Levene	F(3,16)=1,275; p=0,317			



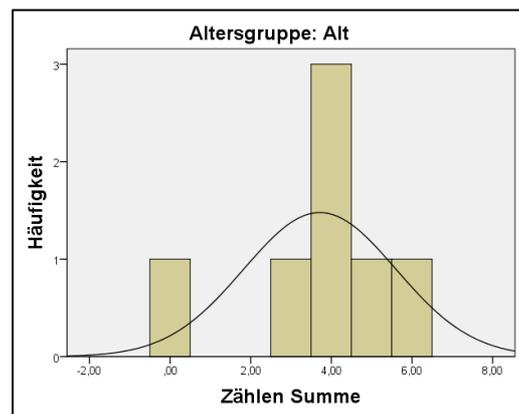
a)



b)



c)

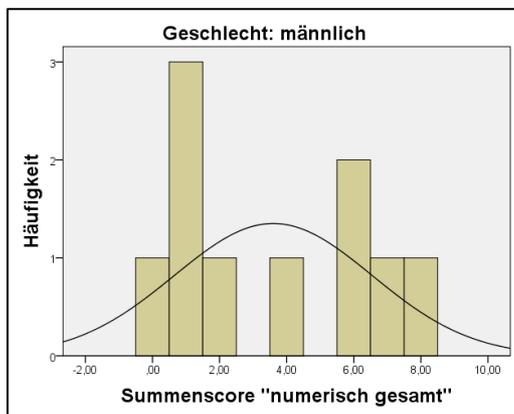


d)

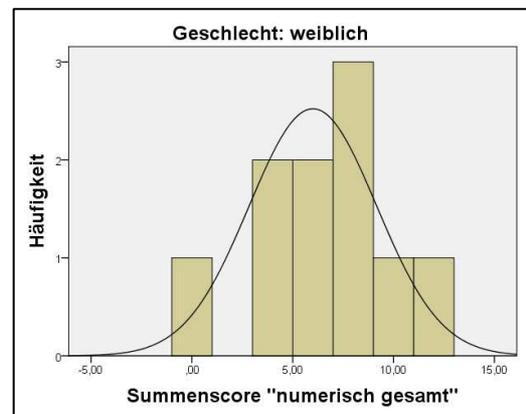
Abbildung 21 a-d: Histogramme zur Normalverteilungsprüfung für die Skala „Zählen“ getrennt für Geschlecht (a männlich, b weiblich) und Altersgruppen (c jung, d alt)

Tabelle 34: Voraussetzungsprüfung „numerisch gesamt“

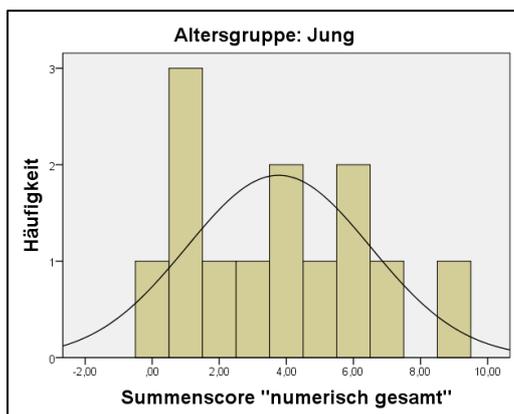
	männlich	weiblich	Jung	Alt
N	10	10	13	7
MW	3.6	6	3.77	6.71
SD	2.95	3.16	2.74	3.35
Schiefe	.26	-.4	.37	-1.36
Kurtosis	-1.79	.28	1.19	1.59
Levene	F(3,16)=.95; p=.438			



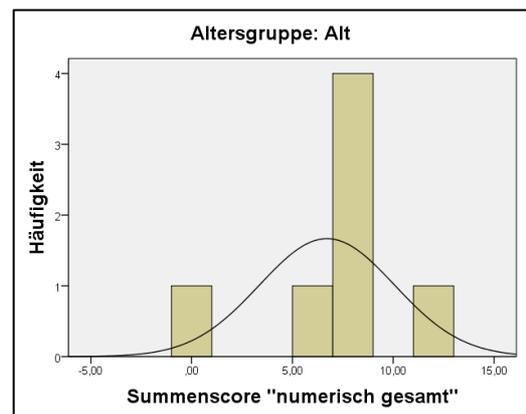
a)



b)



c)

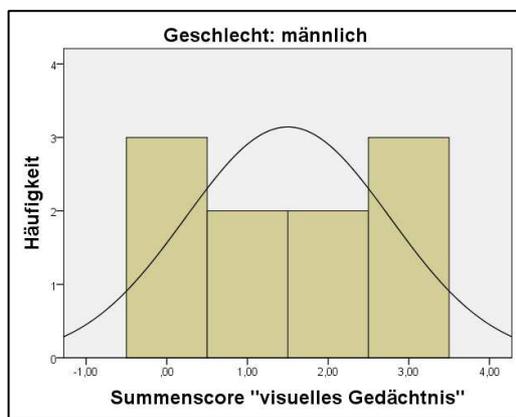


d)

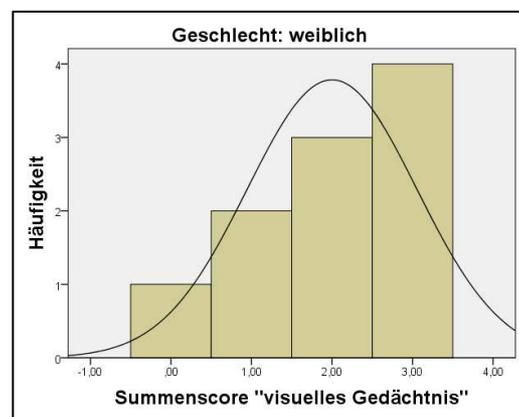
Abbildung 22 a-d: Histogramme zur Normalverteilungsprüfung für die Skala „numerisch gesamt“ getrennt für Geschlecht (a männlich, b weiblich) und Altersgruppen (c jung, d alt)

Tabelle 35: Voraussetzungsprüfung „visuelles Gedächtnis“

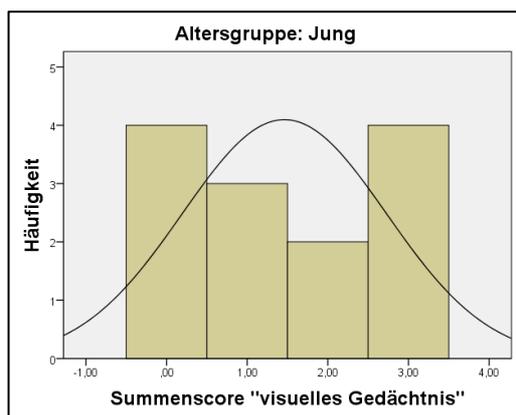
	männlich	weiblich	Jung	Alt
N	10	10	13	7
MW	1.5	2	1.46	2.29
SD	1.27	1.05	1.27	.76
Schiefe	0	-.71	.10	-.6
Kurtosis	-1.76	-.45	-1.73	-.35
Levene	F(3,16)=1.12; p=.371			



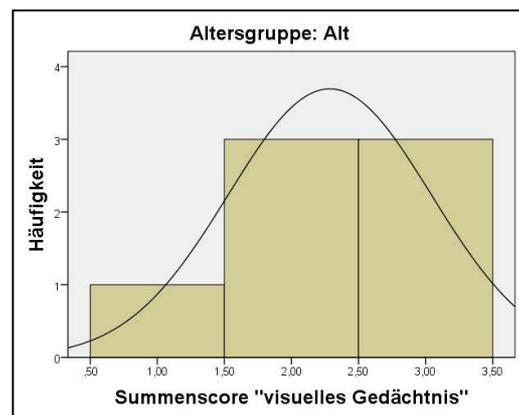
a)



b)



c)



d)

Abbildung 23 a-d: Histogramme zur Normalverteilungsprüfung für die Skala „visuelles Gedächtnis“ getrennt für Geschlecht (a männlich, b weiblich) und Altersgruppen (c jung, d alt)

ANHANG H: Deskriptive Statistiken ANOVA

Tabelle 36: Mittelwert und Standardabweichung für den Summscore der Skala „aktiver Wortschatz“ hinsichtlich Geschlecht- und Altersgruppen

	Geschlecht	Mittelwert	SD	N
jung	männlich	2.5	2.07	8
	weiblich	5.2	2.28	5
	gesamt	3.54	2.47	13
alt	männlich	4.5	4.54	2
	weiblich	5.8	3.9	5
	gesamt	5.43	3.55	7
gesamt	männlich	2.9	2.33	10
	weiblich	5.5	3.03	10
	gesamt	4.2	2.95	20

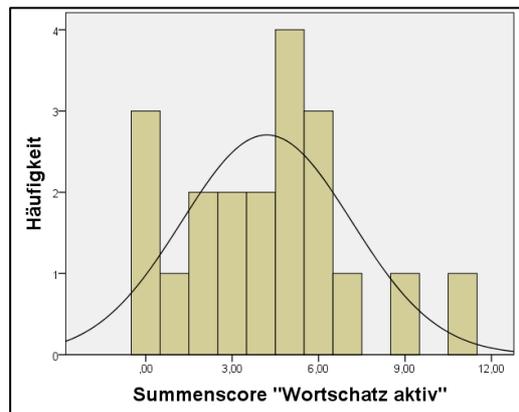
Tabelle 37: Mittelwert und Standardabweichung für den Summscore der Skala „Sprache gesamt“ hinsichtlich Geschlecht- und Altersgruppen

	Geschlecht	Mittelwert	SD	N
jung	männlich	9.31	7.06	8
	weiblich	15.97	5.72	5
	gesamt	11.87	7.16	13
alt	männlich	19.50	7.78	2
	weiblich	19.57	9.47	5
	gesamt	19.55	8.36	7
Gesamt	männlich	11.35	7.99	10
	weiblich	17.77	7.62	10
	gesamt	14.56	8.28	20

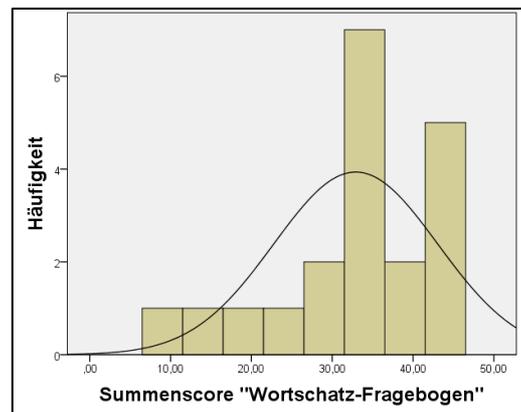
ANHANG I: Voraussetzungsprüfung Korrelation Elternfragebogen

Tabelle 38: Überprüfung der Normalverteilung für die Summenscores „aktiver Wortschatz“, „aktiver Wortschatz– Fragebogen“, „grammatikalische Fähigkeiten“ und „grammatikalische Fähigkeiten – Fragebogen“

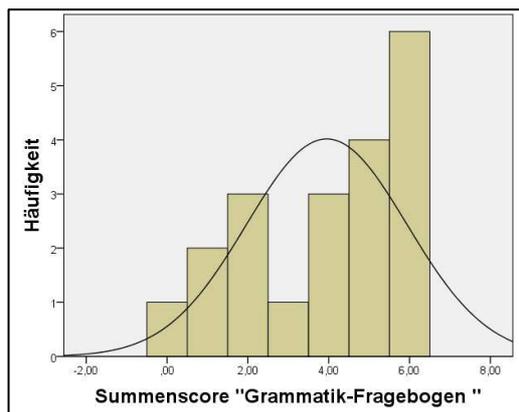
	aktiver Wortschatz	aktiver Wortschatz - Fragebogen	grammatikalische Fähigkeiten	grammatikalische Fähigkeiten- Fragebogen
N	20	20	20	20
MW	4.2	32.9	9.56	3.95
SD	2.95	10.14	5.3	1.99
Schiefe	.43	-1.01	-.49	-.6
Kurtosis	.14	.4	-1.31	-.99



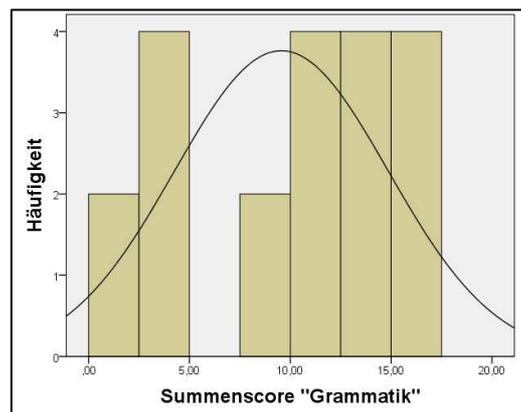
a)



b)



c)



d)

Abbildung 24 a-d: Histogramme zur Normalverteilungsprüfung für die Summenscores „aktiver Wortschatz“ (a), „aktiver Wortschatz – Fragebogen“ (b), „grammatikalische Fähigkeiten“ (c) und „grammatikalische Fähigkeiten – Fragebogen“ (d)

ANHANG J: Abstract (Deutsch)

Abstract. Da die wenigen vorhandenen allgemeinen Entwicklungstests oft die Gütekriterien nicht hinreichend erfüllen oder andere Mängel aufweisen, und auch fraglich ist, ob die alternativen spielbasierten Ansätze, die eine Beobachtung des Spiels des Kindes zur Erhebung der Entwicklung darstellen, den Gütekriterien entsprechen, ist das Ziel des Projekts von Kastner-Koller und Deimann die Entwicklung eines Itempools, welcher eine Kombination der beiden Ansätze darstellen soll. 2011 wurde die erste Version des Manuals von Kuchler, Sapper, Deimann und Kastner-Koller veröffentlicht. Im Rahmen dieser Diplomarbeit, deren Schwerpunkt bei der kognitiven Entwicklung liegt, wurde der spielbasierte Itempool in Zusammenarbeit mit Kronberger (in Vorbereitung) weiterentwickelt, 20 Kindern vorgegeben und hinsichtlich der Zumutbarkeit, der Hauptgütekriterien und der Itemkennwerte beurteilt. Insgesamt kann der spielbasierte Itempool in seiner aktuellen Version als sehr brauchbarer Zugang für die Erhebung der Entwicklung zweijähriger Kinder erachtet werden. Die Objektivität des Verfahrens kann für einen Entwicklungstest, der ein individuelles Vorgehen notwendig macht, als zufriedenstellend beschrieben werden. Hinsichtlich der Reliabilität und Validität ergibt sich für jede Skala ein unterschiedliches Bild. Die Erprobung des Itempools an einer größeren Stichprobe bzw. eine zusammenfassende Analyse der Daten aller Erhebungen ist notwendig.

Angaben zur Person

Name:	Sigrid Anna Punz
Geburtsort:	Scheibbs, NÖ

Schul- und Berufsbildung

Volksschule:	1995-1999 Volksschule St. Elisabeth
Gymnasium:	1999-2007 Sperlgymnasium – wirtschaftskundlicher Zweig Matura mit gutem Erfolg
Studium:	Seit März 2008 Studium der Psychologie an der Uni Wien

Berufserfahrung

Andritz AG:	2005 Ferialpraktikum; 03/2008-04/2010 geringfügige Beschäftigung Diverse Aufgaben im Bereich der Konstruktion/CAD
Krankenhaus St. Elisabeth:	2006, 2007 und 2011 je ein Monat Ferialpraktika als Abteilungshelferin
Caritas der Diözese St.Pölten:	2009 2 wöchiges Praktikum in der Werkstatt St.Leonhard, 2011 und 2012 je eine Woche Urlaubsaktion mit Menschen mit Behinderung
Österreichische Autistenhilfe	Seit April 2009 Begleitung eines autistischen Kindes in den Kindergarten und die Volksschule