



universität
wien

MASTERARBEIT / MASTER'S THESIS

Titel der Masterarbeit / Title of the Master's Thesis

„Retest-Reliabilität eines spielbasierten Inventars für
zweijährige Kinder unter besonderer Berücksichtigung
der kognitiven Entwicklung“

verfasst von / submitted by

Karin Prillinger, BSc

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of
Master of Science (MSc)

Wien, 2015 / Vienna 2015

Studienkennzahl lt. Studienblatt /
degree programme code as it appears on
the student record sheet:

A 066 840

Studienrichtung lt. Studienblatt /
degree programme as it appears on
the student record sheet:

Masterstudium Psychologie

Betreut von / Supervisor:

Ass.-Prof. Dr. Pia Deimann

Danksagung

Während des Verfassens dieser Masterarbeit haben mich viele Menschen unterstützt, denen ich an dieser Stelle meinen Dank aussprechen möchte.

Ein großes Dankeschön gilt Frau Ass.-Prof. Dr. Pia Deimann und Frau Ass.-Prof. Dr. Ursula Kastner-Koller für ihre hilfsbereite und konstruktive Betreuung dieser Masterarbeit sowie dem Team der Test- und Beratungsstelle für das Bereitstellen des Kleinkindertestraumes!

Zudem möchte ich mich ganz herzlich bei meiner gesamten Familie für die großartige Unterstützung während meines Studiums sowie beim Verfassen dieser Masterarbeit bedanken!

Ein besonderer Dank gilt auch meinen Kolleginnen Christina Jungwirth, Maria Pozniak und Magdalena Stark für die wertvolle Zusammenarbeit im Rahmen der Masterarbeit und für die schöne gemeinsame Zeit!

Außerdem möchte ich allen Kindern und ihren Eltern danken, die sich Zeit für die Teilnahme an unserer Studie genommen haben!

Inhaltsverzeichnis

1	Zusammenfassung	7
2	Abstract.....	8
3	Einleitung und theoretischer Hintergrund.....	9
3.1	Diagnostik bei Kleinkindern.....	9
3.2	Verfahren zur Entwicklungsdiagnostik	11
3.2.1	Verfahren zur Erfassung der allgemeinen Entwicklung	11
3.2.1.1	Bayley III.....	11
3.2.1.2	ET 6-6-R.....	12
3.2.1.3	MFED	13
3.2.2	Verfahren zur Erfassung der kognitiven Entwicklung.....	13
3.2.2.1	K-ABC.....	13
3.2.2.2	SON-R 2 1/2-7.....	14
3.3	Vorteile des spielbasierten Vorgehens	14
3.4	Spielbasierte Diagnostik.....	17
3.4.1	Transdisciplinary Play-Based Assessment 2.....	17
3.4.2	Play in Early Childhood Evaluation System	18
3.4.3	Assessment, Evaluation and Programming System for Eligibility	19
3.4.4	Early Problem Solving Indicator.....	20
3.4.5	Kritik am Spiel Assessment	20
3.5	Kognitive Entwicklung.....	21
3.5.1	Theory of Mind	22
3.5.1.1	Als-ob-Spiel.....	25
3.5.1.2	Als-ob und TOM	26
3.5.2	Numerisches Wissen	29
3.5.2.1	How-Many-Aufgaben.....	29
3.5.2.2	Give-a-number Aufgaben	30
3.5.2.3	Entwicklung des Zählens.....	31
3.5.2.4	Approximate Number System	32
3.6	Retest-Reliabilität	34
3.6.1	Intervall zwischen den Messzeitpunkten.....	36
4	Fragestellung	37
5	Methode	38
5.1	Untersuchungsinstrument	38
5.2	Untersuchungsablauf	39
5.3	Stichprobenbeschreibung.....	40
6	Ergebnisse.....	42

6.1	Testdauer und Pausenanzahl.....	43
6.2	Kognitive Entwicklung.....	44
6.3	Ergebnisse zur Theory of Mind	46
6.3.1	Als-ob-Spiel	46
6.3.2	Hineinversetzen in Andere	47
6.4	Ergebnisse zum numerischen Wissen.....	49
6.4.1	Zählen.....	49
6.4.2	Aktive Mengenerfassung.....	50
6.4.3	Passive Mengenerfassung	52
6.4.4	Mengenerfassung	53
7	Diskussion	54
7.1	Diskussion der Ergebnisse der Skala "numerisches Wissen".....	55
7.2	Diskussion der Ergebnisse der Skala "Theory of Mind"	57
7.3	Diskussion der Ergebnisse der Gesamtskala "kognitive Entwicklung"	59
7.4	Diskussion der Testdauer.....	59
7.5	Limitation und Ausblick.....	60
8	Literaturverzeichnis	63
9	Tabellenverzeichnis	73
10	Anhang.....	74
10.1	Anhang A: Berechnung der standardisierten Schiefe für die deskriptiven Angaben	74
10.2	Anhang B: Berechnung der standardisierten Schiefe für den kognitiven Bereich	75
10.3	Anhang C: Übersicht Retest-Reliabilität der Skalen und interne Konsistenz	76
10.4	Anhang D: Itemanalyse der Skala Theorie of Mind	77
10.5	Anhang E: Itemanalyse der Skala numerisches Wissen	78
10.6	Anhang F: Elternbrief.....	79
10.7	Anhang G: Elternfragebogen.....	80
10.8	Anhang H: Einverständniserklärung für die Videoauszeichnung	84
10.9	Anhang I: Protokollbogen	85
11	Lebenslauf.....	93

1 Zusammenfassung

Da herkömmliche Entwicklungstests meist nicht auf die Bedürfnisse von zweijährigen Kindern eingehen oder andere Mängel aufweisen und vorhandene spielbasierte Verfahren die Gütekriterien nicht hinreichend erfüllen, ist im Rahmen eines Projektes von Deimann und Kastner-Koller 2011 ein spielbasierter Entwicklungstest entstanden. Dieses neue Verfahren kombiniert den spielbasierten Ansatz mit den Ansprüchen an herkömmliche Entwicklungstests. Das Ziel dieser Masterarbeit ist die Ermittlung der Retest-Reliabilität des kognitiven Bereichs dieses spielbasierten Verfahrens für Zweijährige. Dazu wurden 21 Kinder (14 Mädchen und 7 Buben), die zwischen 24 und 36 Monate alt waren, im Abstand von vier bis sechs Wochen zweimal getestet. Die Gesamtskala der kognitiven Entwicklung wird aus den Skalen Theory of Mind und numerisches Wissen gebildet, umfasst insgesamt 18 Items und weist eine gute Retest-Reliabilität von $r_{tt} = .820$ auf. Für die einzelnen Skalen liegt sie zwischen $r_{tt} = .46$ und $r_{tt} = .90$. Zudem wurde für alle Skalen die interne Konsistenz berechnet und eine Itemanalyse durchgeführt. Die Gesamtskala der kognitiven Entwicklung erfüllt laut den Berechnungen die Anforderungen eines Entwicklungstests. Zur Absicherung der Ergebnisse wird die Ermittlung der Retest-Reliabilität an einer größeren und repräsentativeren Stichprobe empfohlen.

2 Abstract

As traditional development tests mostly do not respond to the needs of two-year-olds or have other deficiencies and existing play-based methods do not sufficiently meet the quality criteria, a play-based development test was developed within a project of Deimann and Kastner-Koller in 2011. This new procedure combines the play-based approach with the standards on traditional development tests. The aim of this master thesis is to determine the test-retest-reliability of the cognitive area of the play-based method for two-year-olds. For this purpose, 21 children (14 girls and 7 boys), between 24 and 36 months of age, were tested twice at intervals of four to six weeks. The total scale of cognitive development is formed from the scale theory of mind and numeric knowledge, comprises a total of 18 items and has good test-retest-reliability of $r_{tt} = .820$. For the individual scales test-retest-reliability lies between $r_{tt} = .46$ and $r_{tt} = .90$. In addition, the internal consistency was calculated for all scales and an item analysis was performed. According to the calculations, the total scale of cognitive development meets the requirements of a development test. In order to hedge the results a recalculation of the test-retest-reliability on a larger and more representative sample is recommended.

3 Einleitung und theoretischer Hintergrund

Im Rahmen eines Projektes von Ass.-Prof. Dr. Pia Deimann und Ass.-Prof. Dr. Ursula Kastner-Koller wurde 2011 an der psychologischen Fakultät ein spielbasierter Itempool entworfen. Das langfristige Ziel ist es ein Verfahren mit spielbasiertem Ansatz zur Erfassung der Entwicklung von zweijährigen Kindern, mit einer repräsentativen Normierung und entsprechenden Gütekriterien, zu entwickeln.

Vorausgehende Diplomarbeiten befassten sich mit der Entwicklung, Überarbeitung und Weiterentwicklung des Itempools sowie der testtheoretischen Überprüfung der Objektivität, Reliabilität und Validität. Diese Masterarbeit setzt die Überprüfung der Gütekriterien fort, wobei der Fokus auf die Retest-Reliabilität gelegt wird. Der vorhandene Itempool wird dazu einer Stichprobe von 21 Kindern in einem Intervall von vier bis sechs Wochen zweimal vorgelegt. Der Schwerpunkt dieser Arbeit liegt dabei auf der kognitiven Entwicklung.

In der vorliegenden Arbeit wird die Notwendigkeit eines Verfahrens mit spielbasiertem Ansatz und akzeptablen Gütekriterien dargestellt. Dabei wird auf vorhandene Verfahren und spielbasierte Methoden zur Entwicklungsdiagnostik eingegangen. Zudem wird die kognitive Entwicklung im Alter von zwei bis drei Jahren thematisiert. Anschließend werden die Ergebnisse der Untersuchung für die Retest-Reliabilität und die interne Konsistenz dargestellt und diskutiert.

Zur Darstellung der Ergebnisse der Bereiche Motorik und sozial-emotionale Entwicklung sei auf die Arbeit von Stark (in Vorbereitung) verwiesen, für die Bereiche visuelle Wahrnehmung und Gedächtnis auf die Arbeit von Pozniak (in Vorbereitung) und für die Bereiche Sprache und Aufmerksamkeit auf die Arbeit von Jungwirth (2015).

3.1 Diagnostik bei Kleinkindern

Die Konstruktion eines Tests wird durch die Altersgruppe und das Untersuchungssetting, das die jeweilige Altersgruppe erfordert, beeinflusst. So ist die Diagnostik bei Säuglingen durch die Beobachtung von gezeigten Verhaltensweisen geprägt. Sie erfordert vom Untersucher

bzw. von der Untersucherin hohe Flexibilität und Kenntnis der zu untersuchenden Fähigkeiten. Für Kinder im zweiten und dritten Lebensjahr finden sich in allen Entwicklungstests nur wenig mögliche Testaufgaben, da es für Kleinkinder nur sehr schwer möglich ist, ihre Aufmerksamkeit länger auf eine Aufgabe zu richten. Diese Aufgaben müssen deshalb einfach und kurz sein. Zudem führt die Impulsivität, die Ablenkbarkeit und der motorische Antrieb der Kinder zu Unruhe bei der Untersuchung (Macha, Proske, & Petermann, 2005). Die Aufgabenvorgabe an einem Tisch ist deshalb in diesem Alter kaum möglich. Die testende Person sollte hinreichend Erfahrung im Umgang mit Kleinkindern haben und die Sprache an die Kinder anpassen. Sie muss das Verfahren sowie die benötigten Materialien sehr gut kennen und vor der Testung vorbereiten, damit sie während der Testung beiläufig die Materialien holen kann, ohne die Aufmerksamkeit des Kindes zu verlieren und um die Abfolge der Aufgaben an die Reaktion des Kindes anpassen zu können (Esser & Petermann, 2010). Bei der Gestaltung des Raumes ist zu beachten, dass die Kinder ihre Umgebung erkunden können und zudem sollte die Ablenkung durch eine reizarme Gestaltung möglichst gering gehalten werden (Macha et al., 2005). Der Raum sollte mit Möbeln und Decken kindgerecht gestaltet sein (Esser & Petermann, 2010). Zu berücksichtigen sind auch Probleme bei der Standardisierung, da sich die Kinder in ihren sprachlichen Fähigkeiten unterscheiden. Zudem sind zur Motivation des Kindes für die Bearbeitung von Aufgaben oft verschiedene Vorgehensweisen und häufig einige Versuche erforderlich. Ab einem Alter von ungefähr drei Jahren ist es möglich, die Testung in einer standardisierten Situation durchzuführen (Macha et al., 2005). Bei sehr jungen Kindern wirkt sich die Anwesenheit einer Bezugsperson positiv auf die Untersuchung aus, da sie dem Kind Sicherheit gibt, um sich auf die unbekannte Situation und den fremden Erwachsenen einzulassen (Irblich & Renner, 2009). Der Zeitpunkt der Testung sollte an den Tagesrhythmus des Kindes angepasst sein und die Untersuchungsdauer sollte im Kleinkinderalter 30 bis 45 Minuten nicht überschreiten (Esser & Petermann, 2010; Quaiser-Pohl, 2010).

3.2 Verfahren zur Entwicklungsdiagnostik

Zur Entwicklungsdiagnostik stehen je nach Fragestellung verschiedene Verfahrenstypen zur Auswahl. Spezifische Entwicklungstests befassen sich mit einem bestimmten Bereich zum Beispiel der Sprache. Entwicklungs-Screenings sind Kurztestverfahren und stufen anhand von Cut-Off-Werten den Entwicklungsstand als auffällig oder unauffällig ein (Petermann & Macha, 2005). Sie haben zum Ziel Entwicklungsrisiken ökonomisch aufzudecken, bevor Symptome erkennbar sind (Deimann & Kastner-Koller, 2007). Allgemeine Entwicklungstests umfassen einen breiteren Bereich, ermöglichen präzisere Aussagen als Screenings und sind aufwändiger in der Durchführung (Petermann & Macha, 2005). Sie beanspruchen all jene Bereiche zu messen, in denen in einem bestimmten Altersabschnitt wichtige Entwicklungen stattfinden. Solche Verfahren gibt es hauptsächlich für Säuglinge, einige erfassen auch den Entwicklungsstand von Klein- und Vorschulkindern (Kastner-Koller & Deimann, 2011).

Die im Folgenden vorgestellten allgemeinen Entwicklungstests, entwicklungsdiagnostischen Inventare und spezifischen Entwicklungstests für den kognitiven Bereich können bei Kindern im Alter von zwei bis drei Jahren eingesetzt werden.

3.2.1 Verfahren zur Erfassung der allgemeinen Entwicklung

3.2.1.1 Bayley III

Eines der bekanntesten und international verbreitetsten Verfahren sind die Bayley Scales of Infant Development. Die deutschsprachige Version des Bayley III (Reuner & Rosenkranz, 2014) ist ein Verfahren zur Entwicklungsdiagnostik und beinhaltet eine kognitive Skala, eine Sprach-Skala und eine Motorik-Skala. Es ist für Kinder im Alter von einem bis zu 42 Monaten geeignet. Die Aufgaben werden in fester Reihenfolge vorgegeben, wodurch sich die Flexibilität der Testung verringert. Die Aufgabenanpassung an die situativen Voraussetzungen, an das momentane Interesse sowie an den Aufmerksamkeitsfokus des Kindes können dadurch nicht so einfach in die Testung eingebaut werden. Der Bayley III

wird als einer der anspruchsvollsten psychologischen Tests für UntersucherInnen bezeichnet und erfordert gut geschulte AnwenderInnen. Für Kinder im zweiten bis vierten Lebensjahr dauert die Untersuchung ungefähr 90 Minuten. Die Normstichprobe des Verfahrens setzt sich aus deutschen und niederländischen Kindern zusammen. Die interne Konsistenz liegt zwischen $\alpha = .77$ und $\alpha = .89$. Die Retest- Reliabilität wurde mit einer Risikogruppe im siebten und 24. Lebensmonat erhoben und die Werte lagen für die Untertests zwischen $r_{tt} = .20$ und $r_{tt} = .46$. Aufgrund des langen Intervalls dürfte die Messgenauigkeit des Verfahrens jedoch unterschätzt werden, da die gemessenen Merkmale eine geringe Stabilität aufweisen (Macha & Petermann, 2015). Mit dieser aktuellen Normierung bietet der Bayley III nun Normen, die für den deutschsprachigen Kulturkreis Geltung haben.

3.2.1.2 ET 6-6-R

Der Entwicklungstest für Kinder von sechs Monaten bis sechs Jahren - Revision (ET 6-6-R) von Petermann und Macha (2013) dient zur Erfassung des allgemeinen Entwicklungsstandes und dauert abhängig vom Alter der Kinder zwischen 20 bis 50 Minuten. Er umfasst die Bereiche Körper- und Handmotorik, kognitive Entwicklung, Sprachentwicklung, Sozialentwicklung und emotionale Entwicklung. Die verwendeten Materialien und Aufgaben sind kindgerecht und der Ablauf der Testung obliegt dem Untersucher/der Untersucherin. Die internen Konsistenzwerte liegen zwischen $\alpha = .53$ und $\alpha = .85$. Zur Retest-Reliabilität werden keine Angaben gemacht. Die Normierungsstichprobe pro Altersstufe ist mit 64 bis 125 Kindern eher klein und die Repräsentativität ist aufgrund des geringen Anteils an mehrsprachigen Kindern fraglich (Hasselhorn & Margraf-Stiksrud, 2015). Da Testanalysen wie eine Überprüfung der Messgenauigkeit der Entwicklungsskalen fehlen, kann nach Kastner-Koller und Deimann (2011) das Verfahren nicht als Entwicklungstest im engeren Sinne gesehen werden, sondern wird als entwicklungsdiagnostisches Inventar eingeordnet.

3.2.1.3 MFED

Ein weiteres entwicklungsdiagnostisches Verfahren ist die Münchner Funktionelle Entwicklungsdiagnostik. Diese gibt es für das 1. Lebensjahr (Hellbrügge, 2001) und für das 2. und 3. Lebensjahr (Hellbrügge, 1994). Die folgenden Beschreibungen beziehen sich auf das Inventar für das zweite und dritte Lebensjahr. Der Entwicklungsstand wird bei diesem Stufenleiterverfahren anhand von sieben Funktionsbereichen eingeschätzt, wobei die Aufgaben mit ansteigendem Schwierigkeitsgrad vorgegeben werden. Das Verfahren basiert auf dem Prinzip der ethologischen Entwicklungsdiagnostik. Es werden die Entwicklung der Körperbeherrschung, der Feinmotorik, des Erfassens von Zusammenhängen, der Lautäußerung und des Sprechens, des Sprachverständnisses, des sozialen Verhaltens und der Selbstständigkeit erfasst. Die Bearbeitungsdauer beträgt zwischen 40 und 60 Minuten. Die Angaben über die Normierungsstichprobe sind jedoch unzureichend. Ebenso werden von den Autoren keine Ergebnisse zur internen Konsistenz und zur Retest-Reliabilität angeführt. Aufgrund der Übersichtlichkeit und Praktikabilität ist dieses Verfahren trotz der oben genannten Kritik sehr populär und weit verbreitet (Esser & Petermann, 2010).

3.2.2 Verfahren zur Erfassung der kognitiven Entwicklung

3.2.2.1 K-ABC

Die deutsche Version der Kaufman Assessment Battery for Children (K-ABC) von Melchers und Preuß (2009) kann für Kinder zwischen 2;6 und 12;5 Jahren zur Bestimmung der Intelligenz beziehungsweise der lernabhängigen Fähigkeiten durchgeführt werden. Das Verfahren besteht aus 15 Untertests, die die Skalen einzelheitliches Denken, ganzheitliches Denken, eine Fertigkeitenskala und eine sprachfreie Skala bilden. Die Bearbeitungsdauer beträgt 40 bis 90 Minuten. Die Retest-Reliabilität wurde mit 24 Kindern zwischen 9;0 und 12;5 Jahren berechnet und beträgt für die Skalen zwischen $r_{tt} = .84$ bis $r_{tt} = .97$ und für die Untertests $r_{tt} = .57$ bis $r_{tt} = .95$. Die Normierungsstichprobe wurde 1986-89 erhoben und ist

nicht mehr aktuell (Rollett & Preckel, 2011). Rollett und Preckel (2011) sehen als Hauptproblem der deutschen Version des K-ABC, dass aktuelle Untersuchungen nicht berücksichtigt werden. Trotzdem wird das Verfahren häufig in der Praxis und Forschung verwendet (Kastner-Koller & Deimann, 2011). Die zweite deutschsprachige Auflage der Kaufman Assessment Battery for Children (K-ABC II) aus dem Jahr 2015 von P. Melchers und M. Melchers ist erst für Kinder ab 3 Jahren geeignet (Irblich, 2015).

3.2.2.2 SON-R 2 1/2-7

Ebenfalls etabliert ist der non-verbale Intelligenztest SON-R 2 1/2-7 aus dem Jahr 2007 von Tellegen, Laros und Petermann, der bei entwicklungsverzögerten und schwer zu testenden Kindern, sprach- und kommunikationsbeeinträchtigten Kindern oder Kindern, die nicht ausreichend Deutsch sprechen, angewendet wird. Es handelt sich um ein sprachfreies Verfahren für Kinder zwischen 2;6 und 7;11 Jahren. Die sechs Untertests werden den beiden Skalen Denk- und Handlungstest zugeordnet und erfordern abstraktes, konkretes oder räumliches Denken. Die Gedächtnisleistungen und die sprachlichen Fähigkeiten werden nicht erfasst. Zur Herstellung einer möglichst natürlichen Testsituation, die besonders für Kleinkinder wichtig ist, gibt die Testleiterin/der Testleiter dem Kind über die Antworten Rückmeldung und berichtigt falsche Antworten. Je nach Alter des Kindes dauert der Test zwischen 35 und 60 Minuten. Das Verfahren wurde 2005 in Deutschland normiert. Die gemittelte Reliabilität beträgt $\alpha = .90$ und die Retest-Reliabilität mit einem Intervall von einem Jahr beträgt für die Gesamtskala $rtt = .77$. Die Autoren empfehlen aufgrund von Boden- bzw. Deckeneffekten eine Anwendung vorwiegend bei 3;0 bis 6;11 jährigen Kindern (Vock, 2008).

3.3 Vorteile des spielbasierten Vorgehens

Die Testung mit herkömmlichen Verfahren entspricht nach Bagnato (2005) nicht den Bedürfnissen eines Kleinkindes. Diese Kinder in unnatürlichen Umgebungen an Tischen und

mit kleinen unmotivierenden Spielzeugen zu testen, erscheint nicht sinnvoll. Es werden andere Verfahren benötigt, die die Fähigkeiten in alltäglichen Settings erfassen.

Nach Sturzbecher (2001) soll sich das Kind aus Spaß und freiwillig an der Untersuchung beteiligen und von sich aus mit den Spielmaterialien befassen. Er versteht unter spielbasiert das Einbetten des diagnostischen Ziels in den Rahmen eines Spiels.

Diesen Forderungen entspricht das Spiel Assessment. Es wird beim Spielen, also in einer für das Kind bekannten Situation, in der es zahlreiche Stunden am Tag verbringt, ausgeführt. In den Spielsituationen beim Spiel Assessment dürfen die Kinder frei wählen, womit sie spielen wollen. Sie werden dabei nicht angeleitet, mit welchen Materialien sie spielen sollen (Kelly-Vance & Ryalls, 2005). Das Spiel Assessment ist so konzipiert, dass es die bestmögliche Leistung hervorruft, da es für die Kinder motivierend ist. Dahingegen wird den Kindern bei vielen hochstandardisierten Tests nicht die Möglichkeit gegeben, ihr Bestmögliches zu zeigen. Das Spiel Assessment hat im Vergleich zu dem steifen aufeinanderfolgenden Format beim standardisierten Testen ein flexibles Format und resultiert in einer Beschreibung von Stärken und interventionsbedürftigen Bereichen (Kelly-Vance, Ryalls, & Gill-Glover, 2002).

Standardisierte Tests hingegen erfordern eine standardisierte Vorgehensweise, die für die meisten Kinder fremd ist. Sie finden in einer fremden Umgebung mit meist uninteressanter Ausstattung statt. Eine unbekannte Untersucherin/ein unbekannter Untersucher gibt den Kindern Instruktionen, wie sie mit bestimmten Spielsachen spielen sollen und arbeitet mit einem hoch strukturierten Setting und einem Frage- und Antwort Format, mit dem die meisten Kinder keine Erfahrung haben (Kelly-Vance, Needelman, Troia, & Ryalls, 1999). Um die Standardisierung zu gewährleisten, dürfen die untersuchenden Personen nicht von dem im Manual vorgeschriebenen Vorgehen abweichen. Sie können sich daher nicht an die möglicherweise anderen Bedürfnisse der Kinder anpassen. Aufgrund der Einschränkungen von standardisierten Verfahren wurde das Spiel als praktikables Setting zum Sammeln von

Daten vorgeschlagen (Kelly-Vance et al., 1999). Es gibt einige Studien, die die ökologische Validität des Spiel Assessment stützen (z. B. Linder, 1993; Lowenthal, 1997). Besonders die Verwendung des freien Spiels in einer natürlichen Umgebung wurde von zahlreichen Forschern als angemessenes Mittel zur Beurteilung der kognitiven Funktionen genannt (z. B. Linder, 1993; Lowenthal, 1997).

Die Ergebnisse einer Studie, die den Zusammenhang des Bayley II und des Spiel Assessments zur Bestimmung der kognitiven Funktionen bei jungen Kindern untersucht hat, weisen darauf hin, dass Kinder ihre bestmöglichen Fähigkeiten während des Spiel Assessments zeigen können und vom standardisierten Format begrenzt werden. Eine mögliche Erklärung für die Ergebnisse ist, dass die Beschaffenheit der beiden Methoden beeinflusst wie die Kinder agieren. Das spielbasierte Verfahren ermöglicht mehr Flexibilität, um während der Untersuchung dem Spiel des Kindes zu folgen, wohingegen der Bayley II vom Testleiter gelenkt wird. Da Zweijährige oft Schwierigkeiten haben Anweisungen zu folgen, könnten sie bessere Leistungen zeigen, wenn sie sich die Aktivitäten selber aussuchen dürfen. Zudem beeinflussen Verhaltensschwierigkeiten und Verweigerungen den Wert des Bayley II wahrscheinlich negativer als das Spiel Assessment. Zu beachten ist auch, dass die Daten des Spiel Assessments weniger standardisiert und vom Beurteiler beeinflussbarer sind als die Daten des Bayley II. Zudem basieren die Bewertungen auf subjektiven Urteilen und könnten verzerrt sein. Auch kann zufällig auftretendes Spielverhalten die Werte des Spiel Assessments erhöhen (Kelly-Vance et al., 1999).

Zu erwähnen ist auch, dass die meisten projektiven Verfahren für Kinder im weitesten Sinne spielbasiert sind und eine Quelle für wertvolle Zusatzinformationen darstellen (Sturzbecher, 2001). Gegenwärtig werden projektive Verfahren, aufgrund der fehlenden psychometrischen Güte, nicht mehr als Testverfahren, sondern zur Generierung diagnostischer Hypothesen eingesetzt (Kastner-Koller & Deimann, 2009).

3.4 Spielbasierte Diagnostik

Die aktuelle Forschung beschäftigt sich sehr mit spielbasierten Interventionen (z. B. Brock, Kochanska, O'Hara, & Grekin, 2015; Wilkes-Gillan, Bundy, Cordier, & Lincoln, 2014). Jedoch finden sich kaum spielbasierte Methoden zur Entwicklungsdiagnostik. Vier spielbasierte Verfahren, die die kognitive Entwicklung erfassen, sind derzeit für den gewerblichen Gebrauch verfügbar (O'Grady & Dusing, 2015): das Play in Early Childhood Evaluation System (PIECES) von Kelly-Vance und Ryalls (2005), das Transdisciplinary Play-Based Assessment, 2nd edition (TPBA-2) von Linder et al. (2008), das Assessment, Evaluation and Programming System for Eligibility, 2nd edition (AEPS:E) von Bricker, Pretti-Frontczak, Johnson und Straka (2002) und der Early Problem Solving Indicator (EPSI-IGDI) von Greenwood, Walker, Carta und Higgins (2006). Zudem gibt es ähnliche Assessments und Vorläufer zu diesen Verfahren (O'Grady & Dusing, 2015). Auf die genannten spielbasierten Verfahren wird im Folgenden genauer eingegangen.

3.4.1 Transdisciplinary Play-Based Assessment 2

Ein spielbasiertes Verfahren, das die Altersgruppe der Zwei- bis Dreijährigen einschließt, ist das Transdisciplinary Play-Based Assessment, das 2008 in zweiter Version von Linder und Kollegen erschienen ist. Es handelt sich um einen auf Stärken basierenden, ganzheitlichen und funktionellen Ansatz zur Untersuchung des Entwicklungsstands von Kindern im Alter von einem Monat bis sechs Jahren (Linder & Bixby, 2010). Die bekannte Methode des Spiel Assessments kann zur Beurteilung der Fähigkeiten in den Hauptbereichen der Entwicklung einschließlich Sozialverhalten, kognitive Entwicklung, Sprache sowie Fein- und Grobmotorik verwendet werden (Kelly-Vance et al., 2002). Laut Crawford und Ogletree (2010) ermöglicht das Transdisciplinary Play-Based Assessment 2 die Erfassung von wertvollen Informationen, die viele standardisierte Tests nicht erkennen. Der Schwerpunkt ist nicht nur die Messung der Fähigkeiten des Kindes, sondern auch der Kontext in dem sie auftreten (Crawford & Ogletree, 2010). Das TPBA-2 ist ein Prozess bei dem Fachkräfte aus verschiedenen Disziplinen

zusammen mit Familienmitgliedern das Funktionieren des Kindes in unterschiedlichen Bereichen einschließlich der Schule und dem Zuhause untersuchen. Zuerst werden Informationen eingeholt und die Spielsituation wird geplant und anhand von Richtlinien bewertet. Die Ergebnisse der Beobachtung des Kindes werden diskutiert, die Familienmitglieder informiert und gegebenenfalls werden Interventionen geplant (Linder & Bixby, 2010). Das TPBA-2 besteht aus 118 Items. Es gibt keine Angaben zur Interrater-Reliabilität und zur Retest-Reliabilität, die Kriteriumsvalidität liegt bei .91 mit dem kognitiven Teil des Bayley III (Linan, 2009). Die methodische Güte dieser Messung der Kriteriumsvalidität bezeichnen O'Grady und Dusing (2015) jedoch aufgrund der geringen Stichprobengröße als schwach.

3.4.2 Play in Early Childhood Evaluation System

PIECES konzentriert sich nur auf den kognitiven Bereich und setzt sich aus einem Kernbereich, der 13 explorative und Als-ob-Spiel Verhaltensweisen beinhaltet, und fünf ergänzenden Bereichen, wie beispielsweise dem Problemlösen, zusammen. Da das explorative Spiel der Vorläufer des symbolischen Spiels ist, sind diese Fähigkeiten entlang eines Entwicklungskontinuums mit 13 Arten an Spielverhaltensweisen angeordnet. Insgesamt sind es 86 Items. Das Verfahren kann an jedem Ort durchgeführt werden, an dem das Kind sich wohlfühlt. Während des unstrukturierten spontanen Spiels, ohne die Anleitung eines Erwachsenen, wird das Verhalten für ungefähr 30 Minuten beobachtet und mittels Videoaufzeichnung kodiert. Die Spielmaterialien im Untersuchungsraum sind nach allgemeinen Themen wie zum Beispiel Küche angeordnet. Das Ziel ist die Ermittlung des höchsten gezeigten Spielverhaltens, um Interventionen für die Förderung von höheren Spielstufen setzen zu können. Die Stichprobe der Studie besteht aus 32 Kindern zwischen 19 und 52 Monaten. Die Retest-Reliabilität mit einem Intervall von ein bis drei Wochen liegt für normal entwickelte Kinder bei $r_{tt} = .48$ und für atypisch entwickelte bei $r_{tt} = .58$ (Kelly-Vance & Ryalls, 2005). Daten zur Interrater-Reliabilität und Kriteriumsvalidität liegen nicht

vor. Die methodische Güte der Retest-Reliabilität wird von O'Grady & Dusing (2015) aufgrund der kleinen Stichprobe als mittelmäßig beurteilt.

3.4.3 Assessment, Evaluation and Programming System for Eligibility

Das AEPS:E enthält 249 Items über die sechs Entwicklungsbereiche Grob- und Feinmotorik, kognitive, soziale sowie adaptive Entwicklung und soziale Kommunikation. In fünf Aktivitäten werden die Items nach einem im Manual vorgeschriebenen Ablauf und mit standardisierten Materialien vorgegeben. Das System ist aber flexibel und kann an die individuelle Routine des Kindes angepasst werden. Es findet in einem für das Kind bekannten Setting wie zum Beispiel Zuhause und bei einer vertrauten Aktivität statt. Zudem werden die Eltern und Bezugspersonen involviert. Die einzelnen Items werden auf der dreistufigen Skala mit "beherrscht", "aufkommende Fähigkeit" oder "nicht beherrscht" bewertet. Das Ziel ist eine zuverlässige Bewertung der beobachteten Fähigkeiten und die Nutzung der Ergebnisse zur Entwicklung von Zielen und als Leitfaden für Interventionen. Das AEPS:E ist für ein breites Spektrum an Bedürfnissen und Diagnosen für Kinder von einem Monat bis drei Jahren angemessen. Die Überprüfung der Gütekriterien fand anhand von 68 Kindern zwischen 18 und 36 Monaten statt. Es ergab sich eine Interrater-Reliabilität von .88 und eine Kriteriumsvalidität mit dem Battelle Development Inventory (BDI) von .65 (Macy, Bricker, & Squires, 2005). O'Grady und Dusing (2015) bewerten die Interrater-Reliabilität als mittelmäßig und die Validität aufgrund der Stichprobengröße als dürftig. Die Korrelation mit dem BDI war für Kinder, die älter als 24 Monate waren, geringer. Dies könnte daran liegen, dass ältere Kinder während einer Spielsituation mit einer begrenzten Anzahl an Spielmaterial und begrenztem Platz nicht die volle Bandbreite ihrer kognitiven Fähigkeiten zeigen, wohingegen ein traditionelles standardisiertes Verfahren die spezifischen Fähigkeiten des Kindes in allen Schwierigkeitsbereichen abfragt (O'Grady & Dusing, 2015).

3.4.4 Early Problem Solving Indicator

Das Verfahren stellt den kognitiven Teil des Individual Growth and Development Indicator (IGDI) dar. Das EPSI-IGDI basiert auf den vier Schlüsselementen Sehen, Explorieren, Funktion und Lösung, die sich aus dem konzeptuellen Framework des Problemlösens ableiten und ist bei Kinder von einem bis vier Jahren anwendbar. Die Testsituation findet in einem beliebigen Raum statt und das Kind spielt dabei gemeinsam mit der Untersucherin/dem Untersucher. Eine Bezugsperson kann zur Beruhigung des Kindes in Sichtweite sitzen und falls das Kind nicht mit der untersuchenden Person spielt, kann das Verfahren auch mit der Bezugsperson durchgeführt werden. Eine Session dauert zehn Minuten und dem Kind werden hintereinander für jeweils zwei Minuten fünf vorher festgelegte Spielzeuge dargeboten. Dabei wird das Vorkommen jedes Schlüsselements kodiert. Wenn das Kind beispielsweise ein Puzzleteil in die Hand nimmt, wird dies als Explorieren kodiert. Das Ziel ist es die individuelle Entwicklung zu überwachen und über das Erfordernis von Interventionen zu entscheiden. Anhand einer Stichprobe von 28 Kindern ergab sich eine Interrater-Reliabilität von .60 - .99 für die vier Fähigkeiten, eine Retest-Reliabilität von $r_{tt} = .88$ und eine Kriteriumsvalidität mit der kognitiven Skala des Bayley II von .42 (Greenwood et al., 2006). Alle drei Werte werden von O'Grady und Dusing (2015) als mittelmäßig eingestuft.

3.4.5 Kritik am Spiel Assessment

Das Spiel Assessment erhält als Messmethode von kognitiven Fähigkeiten bei jungen Kindern Aufmerksamkeit. Es ist verlockend, weil es als ökologisch valide, flexibel und motivierend für das Kind gesehen wird, besonders im Vergleich mit traditionellen standardisierten Instrumenten. Die Vorgehensweisen und Methoden beim Kodieren des kindlichen Verhaltens schwanken aber deutlich bei den AnwenderInnen. Deshalb sind weitere Untersuchungen für die Schaffung von eindeutigen Vorgaben zur Anwendung von Spiel Assessments notwendig. (Kelly-Vance & Ryalls, 2005).

Spielbasierte Verfahren haben das Potential reliable und valide Instrumente zu sein (O'Grady & Dusing, 2015). Allerdings fehlt in der vorhandenen Literatur zum Spiel Assessment oft die Ermittlung von psychometrischen Merkmalen (Kelly-Vance & Ryalls, 2005). Die Forscher müssen die Erhebung von Reliabilität und Validität bei spielbasierten Verfahren fortführen und bei Adaptionen die Psychometrie neu bewerten (O'Grady & Dusing, 2015).

Die vorgestellten spielbasierten Verfahren erfassen alle den Bereich der kognitiven Entwicklung. Zu den kognitiven Fähigkeiten von zwei- bis dreijährigen Kindern zählen auch die sich noch entwickelnde Theory of Mind und das numerische Wissen, die nachstehend dargestellt werden.

3.5 Kognitive Entwicklung

Als Kognition werden die Vorgänge definiert, die einem Organismus Kenntnis über seine Umwelt verschaffen. Beim Menschen sind das vorwiegend die Wahrnehmung, die Vorstellung, das Denken, das Urteilen und die Sprache (Edelmann & Wittmann, 2012). Kognitionen werden auch als mentale Prozesse bezeichnet und zu den kognitiven Fähigkeiten zählen unter anderem komplexere Denkprozesse, die Handlungsplanung und Handlungssteuerung, Lern- und Gedächtnisprozesse, die Informationsverarbeitungs- sowie Problemlösekompetenzen und der Wissenserwerb (Lohaus & Vierhaus, 2015).

Die kognitiven Prozesse bilden die Grundlage für viele Kompetenzen und Fähigkeiten und die Entwicklungen im kognitiven Bereich sind eng mit Entwicklungen in anderen Bereichen verbunden (Lohaus & Vierhaus, 2015). Im Folgenden wird auf die Entwicklungen in den kognitiven Bereichen Theory of Mind (TOM) und numerisches Wissen bei zweijährigen Kindern eingegangen.

3.5.1 Theory of Mind

Unter der Theory of Mind, die auch als Alltagspsychologie bezeichnet wird (Wellman, Cross, & Watson, 2001), versteht man die Fähigkeit bei der Menschen sich selbst und anderen Personen Wissen, Wünsche, Absichten, Gedanken, Emotionen und Überzeugungen zuschreiben und daraus Handlungen vorhersagen. Die Zuschreibung dieser mentalen Zustände entwickelt sich über mehrere Jahre. Personen mit dieser Fähigkeit verstehen, dass andere Personen möglicherweise nicht dieselben Erwartungen, Wünsche und Absichten haben wie sie selbst.

Ab einem Alter von ungefähr vier Jahren verfügen Kinder über eine Theory of Mind, die ihnen hilft menschliches Verhalten zu erklären. Zuvor entwickelt sich als erster Schritt das Verständnis von Wünschen und Absichten. Ab der Mitte des zweiten Lebensjahres können Kinder ihre Wünsche, Absichten und Emotionen von denen anderer Personen abgrenzen und unabhängig davon mental repräsentieren. Sie fangen an zwischen ihrem eigenen mentalen Zustand und dem anderer Personen zu differenzieren und sind in der Lage zwischen der realen und der fiktiven Welt zu unterscheiden. Danach werden als zweiter Schritt auch Überzeugungen erkannt. Die Kinder verstehen also, dass Handlungsentscheidungen von Zielen und Absichten einer Person abhängig sind und können umgekehrt mit dem Wissen über Ziele und Absichten der Person ihre Handlungen ankündigen (Sodian, 2008). Während sich ihre Theory of Mind noch entwickelt, können also bereits zwei- und dreijährige Kinder die Absichten von anderen Personen verstehen (Wellman, 2014).

Um zu erheben, wie ausgeprägt die Theory of Mind bereits entwickelt ist, werden False-Belief-Aufgaben verwendet, bei denen es um das Verstehen der Überzeugungen der handelnden Person geht. Bei False-Belief-Aufgaben erster Ordnung müssen die Kinder beispielsweise beantworten, wo eine Person glaubt, dass sich ein Gegenstand befindet, obwohl sie selbst wissen, dass der Gegenstand mittlerweile an einem anderen Ort ist. Ein bekanntes Beispiel ist die Aufgabe "Maxi und die Schokolade" von Wimmer und Perner aus

dem Jahr 1983. Bei den False-Belief-Aufgaben zweiter Ordnung ist es erforderlich zu erkennen, was Person A glaubt, dass Person B glaubt (Wyl, 2014).

Zweieinhalb- bis dreijährige Kinder können diese False-Belief-Aufgaben noch nicht korrekt lösen. Sie glauben beispielsweise, dass die Person den Gegenstand dort suchen wird, wo er sich mittlerweile befindet und nicht dort, wo sie ihn selbst versteckt hat. Zweieinhalbjährige Kinder antworten bei diesen Aufgaben so, als wüsste die Person über die Realität Bescheid. Vierjährige wissen bereits, dass die Handlung der Personen von ihrem Glauben (Belief) abhängt und nicht davon wie die Situation jetzt ist, da "Belief" und Realität sich voneinander unterscheiden (Wellman et al., 2001; Wellman, 2014).

Viele Studien verwenden ausschließlich False-Belief-Aufgaben zur Messung der Theory of Mind Fähigkeiten. Dieser Ansatz steht jedoch im Gegensatz zu dem Wissen, dass die TOM zahlreiche Fähigkeiten umfasst, die sich lange vor dem Verständnis von False-Belief-Aufgaben zu entwickeln beginnen (Wellman et al., 2001).

Hiller, Weber und Young (2014) untersuchten deshalb eine fünf Items umfassende TOM-Skala (Wellman & Liu, 2004), die Theory of Mind Fähigkeiten beurteilt, die dem False-Belief Verständnis vorausgehen. Insgesamt nahmen 68 Kinder zwischen zwei und fünf Jahren teil, darunter befanden sich 19 Kinder im Alter von 24 bis 35 Monaten. Die Aufgaben umfassen TOM Fähigkeiten erster Ordnung und die Schwierigkeit der Items nimmt mit ansteigender Reihenfolge zu, entsprechend der Fähigkeiten, die sich in der frühen Kindheit immer weiter entwickeln. Die Skala besteht aus 1) verschiedenen "Desires" (Verständnis, dass die Vorlieben der anderen von den eigenen abweichen können) 2) verschiedenen "Beliefs" (Verständnis, dass andere über dieselbe Situation etwas anderes denken/anderer Meinung sein können) 3) Wissen einschätzen (Fähigkeit, das Wissen anderer über ein Szenario beurteilen zu können) 4) False-Belief (False-Belief eines anderen über den Inhalt einer Box beurteilen können) 5) versteckte Emotionen (Verstehen, dass der Gesichtsausdruck einer Person nicht

immer ihrer Emotion entspricht). Die Aufgaben 1 bis 4 können die Kinder verbal oder durch das Zeigen auf ein Bild beantworten.

Die Ergebnisse zeigen, dass bereits zwei Jahre alte Kinder die TOM Fähigkeiten für ein Verständnis von "Desire" und "Belief" Aufgaben haben. Die Aufgaben 4 und 5 konnten dagegen von keinem Zweijährigen gelöst werden. Dementsprechend steigt die Wahrscheinlichkeit, dass ein Kind eine Aufgabe löst, mit dem Alter des Kindes an. Diese Skala ist geeignet um zwischen den frühen TOM Fähigkeiten von kleinen Kindern zu unterscheiden, bietet mit den leichteren Aufgaben eine umfassendere Untersuchung der TOM vor der Entwicklung der False-Belief Fähigkeit und erfasst als Messinstrument der frühen TOM mehr als mit False-Belief-Aufgaben möglich ist. Diese Skala könnte als longitudinale Bewertung eingesetzt werden, um frühzeitig jene Kinder zu erkennen, bei denen ein Risiko für die Entwicklung von sozialen Beeinträchtigungen besteht und die von einem Sozialkompetenztraining profitieren würden (Hiller et al., 2014).

Wie wichtig es ist, diese Kinder zu erkennen und zu fördern belegt auch eine aktuelle Studie, die zeigt, dass das Verständnis von Theory of Mind Aufgaben im Kindergartenalter ein starker Prädiktor für aktuelle und spätere Freundschaften ist. Kinder, die im Kindergarten bei False-Belief-Aufgaben schlecht abschnitten, hatten auch zwei Jahre später noch keine Freunde. Der Einfluss bleibt auch dann bestehen, wenn die Beliebtheit, Sprachkenntnisse und das Alter berücksichtigt werden (Fink, Begeer, Peterson, Slaughter, & de Rosnay, 2015). Viele Studien verbinden das Problem freudlos zu sein mit Problemen bei der psychischen Gesundheit. Ladd und Troop-Gordon (2003) fanden beispielsweise heraus, dass chronische Freundlosigkeit, die zwischen dem Kindergarten und der dritten Schulstufe entsteht, langfristig Einsamkeit und psychische Störungen vorhersagt, auch wenn andere Prädiktoren wie Ablehnung durch die Peergruppe statistisch kontrolliert wurden.

Die Theory of Mind Fähigkeiten der Kinder stehen auch im Zusammenhang mit dem Als-ob-Spiel, auf das nachstehend eingegangen wird (Wellman, 2014).

3.5.1.1 Als-ob-Spiel

Das Spiel ist eine nicht instrumentale Aktivität, die keinen nahen Zweck hat und keinem anderen unmittelbaren Ziel als dem Vergnügen dient (Weisberg, 2015). Das Als-ob-Spiel, als Subtyp des Spiels, unterscheidet sich von anderen Formen des Spiels, da es eine Form von Repräsentation oder so tun-als-ob umfasst. Die Verhaltensweisen oder Handlungen, die in einem Als-ob-Spiel stattfinden, sollen die Realität nicht genau widerspiegeln (Lillard, 1993).

Als-ob-Spiele beginnen sich mit 18 Monaten zu entwickeln (Lillard et al., 2013). Eine frühe Form ist die Zweckersetzung, bei der ein Objekt so verwendet wird, als wäre es ein anderes (Singer & Singer, 1992). In den frühen Kindergartenjahren beginnen die Kinder Als-ob mit unsichtbaren Objekten zu spielen, in denen das Als-ob vollständig in der kindlichen Phantasie geschieht. Diese Form ist komplexer, da es keine fassbaren Objekte gibt, die den Als-ob Objekten eine Grundlage in der echten Welt geben (Taylor & Carlson, 1997). Eine weitere Form des Als-ob-Spiels sind imaginäre Freunde (Burriss & Raif, 2015). Diese Form unterscheidet sich von anderen Arten insofern, da sie es erfordert, dass das Kind sich mit derselben Vorstellung über einen ausgedehnten Zeitraum also Wochen, Monate oder sogar Jahre beschäftigt. Der/die imaginäre Freund/Freundin kann völlig unsichtbar sein und nur in der Phantasie des Kindes existieren oder es kann ein Gegenstand sein, der mit einer Persönlichkeit erfüllt wird (Weisberg, 2015).

Alle Formen des Als-ob-Spiels sind im Alter von drei bis fünf Jahren, der sogenannten Hochsaison des Als-ob-Spiels, am markantesten (Singer & Singer, 1992). Während der Kindergartenzeit entwickeln sich die kognitiven und sozialen Fähigkeiten, die mit dem Als-ob-Spiel verbunden werden, am schnellsten (Weisberg, 2015).

Als-ob-Spiele können in einer Gruppe oder alleine stattfinden (Lillard et al., 2013). Das Als-ob-Spiel scheint eine Art Repräsentation des eigenen mentalen Zustands oder dem des Als-ob-Spielpartners zu erfordern. Um zu verstehen, was beim Spiel vorgeht, müssen die Handlungen als Repräsentationen abbildhaft und nicht tatsächlich interpretiert werden. Die

Banane ist beispielsweise nicht wirklich ein Telefon, aber die Mutter tut-so-als-ob. Das Als-ob ist daher eine Meta-Repräsentation, es erfordert das Repräsentieren der Repräsentation von jemand anderem über einen Zustand. Ohne diese Fähigkeit zu Meta-Repräsentationen würde man die Als-ob Handlung als unsinnig ansehen. Das Kind könnte sich zum Beispiel die Frage stellen, ob die Mutter die Banane zu ihrem Ohr hält und hinein spricht, weil es das ist, was man mit einer Banane tun soll (Weisberg, 2015).

Derselbe Meta-Repräsentationsprozess ist erforderlich beim Verständnis des Belief von anderen bzw. der Theory of Mind wie beispielsweise bei der Sally-Anne False-Belief-Aufgabe. Um sich angemessen verhalten zu können, ist in beiden Szenarien eine Abkopplung erforderlich. Man muss den wahren Zustand der Gegenstände hinter sich lassen, eine Repräsentation vom mentalen Zustand des anderen erzeugen (eine Meta-Repräsentation) und basierend auf dieser Meta-Repräsentation auf die Situation reagieren. Die Mutter spricht in die Banane, weil in Beachtung ihres mentalen Zustandes (tun-als-ob) die Banane ein Telefon ist und Sally soll in den Korb schauen, weil aufgrund ihres mentalen Zustandes (False-Belief) der Ball im Korb ist (Weisberg, 2015).

3.5.1.2 Als-ob und TOM

Einige Forscher glauben, dass dieselbe mentale Funktion dem Tun-als-ob und dem False-Belief Verständnis zugrunde liegt, da beide das Abkoppeln und Meta-Repräsentationen erfordern. Zudem fanden Studien Korrelationen zwischen einigen Aspekten der TOM und dem Als-ob-Spiel (Taylor & Carlson, 1997; Weisberg, 2015). Laut Lillard et al. (2013) sollte dieser Auffassung nach jegliches tun-als-ob im Zusammenhang mit der TOM stehen. Für Wellman (2014) sind die Zusammenhänge zwischen dem Als-ob-Spiel und besseren TOM Fähigkeiten eindeutig. Sie entstehen durch eine vermehrte Teilnahme der Kinder am Als-ob-Spiel, das einher geht mit dem Lernen über Personen und "Minds". Viele Wissenschaftler gehen auch davon aus, dass das Als-ob-Spiel der kognitiven Entwicklung nützt.

Smith (2010) diskutierte dazu drei theoretisch mögliche Zusammenhänge zwischen dem Als-ob-Spiel und positiven Entwicklungsergebnissen:

- 1) Das Als-ob-Spiel ist entscheidend für eine bestmögliche Entwicklung.
- 2) Das Als-ob-Spiel ist nur einer von vielen Wegen für positive Entwicklungsergebnisse. Es hilft einigen Entwicklungen, aber es ist nicht der einzig mögliche Weg. Andere Handlungen können genauso gut oder besser funktionieren.
- 3) Das Als-ob-Spiel ist eine Begleiterscheinung oder ein Nebenprodukt einer ausgewählten Fähigkeit, aber trägt an und für sich nicht zur Entwicklung bei. Stattdessen trägt die Handlung oder Bedingung, an der das Als-ob-Spiel manchmal anschließt zur Entwicklung bei.

Die Schlussfolgerung der Metaanalyse von Lillard et al. (2013) ist, dass es gegenwärtig keine Beweise gibt, die 1) unterstützen und die existierenden Studien, die in dieser Metaanalyse herangezogen wurden, unterstützen auch die Behauptung nicht, dass das Als-ob-Spiel die kognitive Entwicklung steigert bzw. dafür ausschlaggebend ist. Die Daten, die hier überprüft wurden, schlagen vor, dass das Als-ob-Spiel für sich alleine die Entwicklung nicht allgemein fördern könnte, aber dass das Spielen mit bestimmten Inhalten dies kann.

Im Zusammenhang mit dem Als-ob-Spiel fanden Lillard et al. (2013) heraus, dass Kinder, die fortgeschrittenere TOM Fähigkeiten hatten, sich auch mit fortgeschrittenerem Als-ob-Spiel beschäftigten. Dieser Zusammenhang ergibt sich aber möglicherweise aufgrund einer dritten Variable. Zum Beispiel könnten die Eltern mit ihren Kindern auf Arten interagieren, die das "Mental State Reasoning" und das Als-ob-Spiel fördern.

Es gibt auch die Theorie, dass das Als-ob-Spiel sich zugrundeliegende kognitive Strukturen neben der Theory of Mind noch mit anderen wichtigen kognitiven und sozialen Fähigkeiten einschließlich der Sprache, dem kontrafaktischen Denken und dem symbolischen Verständnis teilt. Diese Zusammenhänge weisen auf die Existenz einer einzelnen repräsentationalen Funktion hin, die all diesen Fähigkeiten zugrunde liegt. Diese Funktion

könnte bei der Erklärung, wie das Als-ob-Spiel die Entwicklung der anderen Fähigkeiten unterstützt und von ihnen unterstützt wird, den Schwerpunkt bilden (Weisberg, 2015).

Es wird weitere Forschung benötigt, um die mögliche Rolle des Als-ob-Spiels in der kindlichen Entwicklung und den Zusammenhang zwischen dem Als-ob-Spiel und den TOM Fähigkeiten aufzuklären (Lillard et al., 2013; Weisberg, 2015).

Um endgültige Belege dafür zu bekommen, ob das Als-ob-Spiel kausal notwendig für die Entwicklung von anderen Fähigkeiten ist, es diese nur fördert oder eine Begleiterscheinung ist, könnte allerdings, aufgrund von ethischen und praktischen Schwierigkeiten bei der Erstellung von "nicht spielenden" Kontrollgruppen und dies zudem über eine ausreichend lange Zeitspanne, um aussagekräftige Unterschiede feststellen zu können, unmöglich sein. Erschwerend kommt die Tatsache hinzu, dass viele Fähigkeiten im Zusammenhang stehen (Lillard et al., 2013; Weisberg, 2015). Zum Beispiel wirken sich Sprachkompetenzen auf die Entwicklung des Theory of Mind Verständnisses aus (Ebert, 2015). Das kindliche Als-ob-Spiel Niveau und die frühe sprachliche Entwicklung scheinen ebenfalls im Zusammenhang zu stehen, wobei das Als-ob-Spiel der Sprache vorausgeht (Lillard et al., 2013).

Dass Kinder Informationen über Objekte und deren Funktionen während des Als-ob-Spieles lernen können, behaupten die Ergebnisse einer aktuellen Studie von Hopkins, Dore und Lillard (2015). Dabei lernten die Kinder im Als-ob-Spiel Informationen über Objekte und deren Funktion. Anschließend konnten sie im realen Kontext daraus Schlussfolgerungen über für sie neue Objekte ziehen. Diese Ergebnisse stimmen mit anderen Studien überein, die behaupten, dass Kinder im Als-ob-Spiel neue Informationen lernen können, die sie dann in der wirklichen Welt anwenden können, obwohl sich das Lernen im Als-ob-Spiel vom Lernen in realen Kontexten unterscheidet (Hopkins et al., 2015).

Neben dem Als-ob-Spiel entwickeln sich auch die Fähigkeiten der Kinder im numerischen Bereich weiter. Wichtige Punkte bei der Entwicklung des numerischen Wissens werden im Folgenden beschrieben.

3.5.2 Numerisches Wissen

Von Gelman und Gallistel (1986) wurden drei "how-to-count" Grundsätze bestimmt: 1) Die Eins-zu-eins Zuordnung besagt, dass beim Zählen einer Reihe jedem Element in der Reihe nur eine Zahl zugeordnet wird. 2) Die stabile Reihenfolge bedeutet, dass Zahlwörter, die zum Zählen verwendet werden, bei jedem Zählen in der gleichen Reihenfolge gebraucht werden müssen. 3) Das Kardinalprinzip besagt, dass die Zahl, die für das letzte Element der Reihe verwendet wurde, die Anzahl an Elementen in der Reihe angibt. Diese Prinzipien werden laut Gelman und Gallistel (1986) intuitiv verstanden und bereits von Zweijährigen beim Zählen angewendet. Diese Ansicht ist als "Principles-first" oder "Principles-before-Skills" bekannt. Andere Studien konnten diese Behauptung jedoch nicht unterstützen, da beispielsweise Dreijährige oft die Eins-zu-eins Zuordnung missachten, indem sie ein Element auslassen oder doppelt zählen, oder indem sie dieselbe Zahl bei einer Zählung zweimal verwenden. Kinder missachten auch die stabile Reihenfolge, indem sie zu verschiedenen Zeiten unterschiedliche Zahlwörter verwenden (Miller, Smith, Zhu, & Zhang, 1995). Diese Ergebnisse haben dazu geführt, dass viele Personen davon ausgehen, dass die "how-to-count"-Prinzipien nicht von vornherein verstanden, sondern schrittweise gelernt werden. Dies ist bekannt als die "Principles-after" oder "Skills-before-Principles" Ansicht (Sarnecka & Carey, 2008).

3.5.2.1 How-Many-Aufgaben

Um die Kardinalprinzip-Fähigkeiten von Kindern zu erfassen, werden häufig How-Many-Aufgaben verwendet. Bei diesen Aufgaben wird den Kindern eine Anzahl an Elementen gezeigt und sie müssen antworten, wie viele Elemente es sind.

Einige Forschern behaupten, dass diese Aufgaben das Wissen der Kinder überschätzen, da einige Kinder die letzte Zahl beim Zählen wiederholen, ohne allem Anschein nach zu verstehen, dass diese Zahl angibt, aus wie vielen Elementen die Reihe besteht (Frye, Braisby, Lowe, Maroudas, & Nicholls, 1989).

Umgekehrt wurde auch behauptet, dass "How-Many"-Aufgaben das Wissen der Kinder unterschätzen, weil viele Kinder die Frage falsch beantworten, obwohl sie die Reihe zuvor korrekt gezählt haben. Sie beginnen auf diese Frage hin normalerweise die Reihe erneut zu zählen. Wenn man sie davon abhält, antworten sie entweder gar nicht oder sie antworten mit irgendeiner Zahl oder mit der letzten Zahl ihrer Zählung (Gelman, 1993). Unterstützer der "Principles-first" Ansicht argumentieren, dass dieses Verhalten zeigt, dass Kinder das Kardinalprinzip verstehen. Sie weisen darauf hin, dass es seltsam ist, direkt nach dem Zählen zu fragen "Wie viele sind das?". Um dies zu demonstrieren, führte Gelman (1993) eine "How-Many"-Aufgabe mit Studierenden durch. Als diese nach dem Zählen gefragt wurden "Wie viele?", begannen alle Studenten bis auf einen erneut zu zählen. Nur einer wiederholte die letzte Zahl seiner Zählung.

Es gibt also Uneinigkeit darüber, ob "How-Many"-Aufgaben das Wissen für das Kardinalprinzip der Kinder überschätzen, unterschätzen oder genau messen.

3.5.2.2 Give-a-number Aufgaben

Eine weitere Möglichkeit das Kardinalprinzip zu erfassen sind "Give-a-number" Aufgaben. Bei diesen Aufgaben soll das Kind beispielsweise einem Stofftier fünf Bananen geben (Negen & Sarnecka, 2015). Studien, die solche Aufgaben verwendet haben, fanden heraus, dass Kinder oft nicht in der Lage sind für Zahlen, die innerhalb ihres Zählbereichs liegen, eine Menge zu erstellen. Zum Beispiel können die Kinder bis fünf zählen, aber keine Menge von fünf Objekten zusammenstellen. Wenn das Kardinalprinzip Wissen anhand von "Give-a-number" Aufgaben untersucht wird und nicht mit "How-Many"-Aufgaben, scheinen Kinder

das Kardinalprinzip relativ spät zu erlangen, und erst nachdem sie die anderen beiden Zählprinzipien beherrschen (Sarnecka & Carey, 2008).

3.5.2.3 Entwicklung des Zählens

Die "Give-a-Number" Aufgaben haben auch eine neue Ansicht darüber hervorgebracht wie Zahlen gelernt werden. Es zeigte sich, dass die Leistung von Kindern bei diesen Aufgaben eine Reihe von vorhersagbaren Stufen durchmacht. Auf der ersten Stufe macht das Kind keinen Unterschied zwischen der Bedeutung von verschiedenen Zahlen. Bei einer "Give-a-Number" Aufgabe gibt das Kind entweder ein einzelnes Objekt oder eine ganze Handvoll, aber die Menge, die es gibt, ist ohne Bezug zu der verlangten Anzahl. Kinder auf dieser Stufe werden als "pre-numeral-knower" bezeichnet. Sie haben noch keiner der Zahlen, die sie gelernt haben, eine bestimmte Bedeutung zugeteilt. Auf der nächsten Stufe, die die meisten Kinder im Alter von zweieinhalb bis drei Jahren erreichen, weiß das Kind nur, dass Eins "Eins" bedeutet. In dieser Phase geben die Kinder ein Element, wenn sie nach einem gefragt werden und zwei oder mehr Elemente, wenn sie nach irgendeiner anderen Zahl gefragt werden. Diese Stufe wird als "one-knower" bezeichnet. Einige Monate später, wenn das Kind gelernt hat, dass Zwei "Zwei" bedeutet wird es ein "two-knower". Zu diesem Zeitpunkt gibt es ein Objekt, wenn es nach einem gefragt wird und zwei, wenn es nach zwei gefragt wird, aber es unterscheidet nicht zwischen den anderen Zahlen. Für alle anderen Zahlen nimmt das Kind einfach einige Objekte und übergibt sie. Die folgende Stufe wird als "three-knower" bezeichnet und es gibt auch Studien, die von einer "four-knower" Stufe berichten (Sarnecka & Carey, 2008). Kinder in diesen Stufen werden als "subset-knowers" bezeichnet (Le Corre & Carey, 2007; Le Corre, Van de Walle, Brannon, & Carey, 2006). Obwohl sie bereits die Zahlenreihe bis zehn oder höher auswendig gelernt haben, kennen sie die genaue Bedeutung nur für eine Teilmenge dieser Zahlen. Nachdem die Kinder einige Zeit, oft mehr als ein Jahr, als "subset-knower" verbracht haben, verändert sich ihre Leistung dramatisch und sie können die Kardinalität für die Zahlen fünf und darüber bilden. Während sich die Leistung durch die

"subset-knower" Stufen schrittweise verbessert hat, scheinen Kinder die Bedeutung der ihnen bekannten Zahlen ab der Zahl fünf, auf einmal zu erlangen. Diese Kinder werden "Kardinalprinzip-knower" genannt (Sarnecka & Carey, 2008). Studien zeigten, dass Kinder, die bei "Give-a-number" Aufgaben eine Menge von fünf bilden können, "Kardinalprinzip-knower" sind (Le Corre et al., 2006; Le Corre & Carey, 2007).

Kinder, die das Kardinalprinzip beherrschten, unterscheiden sich von "subset-knower" auch darin, dass letztere nicht zählen, wenn sie eine Give-a-number Aufgabe lösen, auch wenn ihnen explizit gesagt wurde, dass sie zählen sollen, wohingegen "Kardinalprinzip-knower" zählen. Die qualitativen Unterschiede im Zählverhalten von Kindern, die das Kardinalprinzip beherrschen und Kindern, die das nicht tun, weisen darauf hin, dass das was diese Gruppen letztlich trennt, nicht nur die Größe der Zahlenreihe ist, die sie beherrschen. Der Unterschied liegt darin, dass Kinder, die das Kardinalprinzip erworben haben, im Gegensatz zu den Kindern, die das noch nicht haben, verstehen, wie das Zählen funktioniert (Sarnecka & Carey, 2008).

Sarnecka und Carey (2008) fanden heraus, dass das "knower"-Level signifikant mit dem Alter zusammenhängt. Zweijährige fanden sich dabei sowohl in der "Kardinalprinzip-knower" Gruppe als auch in der "subset-knower" Gruppe. Die Leistung bei verschiedenen Wissensstufen ("knower-levels") ändert sich entlang von zwei Parametern: die Größe der Menge, die ein Kind auf Verlangen bilden kann und, ob es zählt oder nicht. Viele Kinder lernen die How-Many Regel, dass die Antwort auf diese Frage das letzte Wort beim Zählen ist, lange bevor sie das Kardinalprinzip verstehen. How-Many-Aufgaben sollten daher nicht verwendet werden, um das Verständnis des Kardinalprinzips zu untersuchen, weil sie das Wissen der Kinder darüber, wie das Zählen funktioniert, überschätzen.

3.5.2.4 Approximate Number System

Neben diesem System der exakten Zahlen, das Kinder bereits in der frühen Kindheit zu erlernen beginnen, gibt es auch das "approximate number system" (APN). Wenn

beispielsweise für 30 Personen nur 20 Sitzplätze zur Verfügung stehen, erlaubt uns dieses System, ohne zählen zu müssen, die Einschätzung, dass es zu wenig Plätze gibt. Die Genauigkeit bei der Abschätzung von einer ungefähren Anzahl wird als APN-Schärfe bezeichnet. Obwohl jeder ungefähre Anzahlen abschätzen kann, machen manche Personen exaktere Schätzungen als andere (Negen & Sarnecka, 2015). Verschiedene Studien berichten, dass individuelle Unterschiede in der APN-Schärfe bei jungen Kindern mit ihrem Wissen über exakte Zahlen wie beispielsweise dem Wort "sechs" zusammenhängen (Mussolin, Nys, Leybaert, & Content, 2012; Wagner & Johnson, 2011). Die Ergebnisse einer Studie von Negen und Sarnecka (2015) zeigen jedoch, dass diese Korrelationen aufgrund des Designs der Studien zustande gekommen sind. Bei den Aufgaben wurden den Kindern verschiedene Bilder mit Punkten gezeigt und sie mussten abschätzen, auf welcher Seite mehr Punkte sind, aber nicht antworten wie viele Punkte abgebildet sind. Das Vorgehen ermöglichte es den Kindern versehentlich, die richtige Antwort aufgrund der Größe der gezeigten Punkte und nicht deren Anzahl zu geben. Zudem wurde die Korrelation dadurch beeinflusst, dass junge Kinder, ohne ein Konzept darüber was eine exakte Kardinalzahl ist, die Frage "Welche Seite hat mehr Punkte" nicht als Frage nach der zahlenmäßig größeren Anzahl an Punkten aufgefasst haben, sondern zum Beispiel als Frage nach der Größe der Punkte oder der Größe der Fläche auf der die Punkte verteilt sind. In einer abgewandelten Untersuchung mit Kindern zwischen zwei und sechs Jahren, in der die beiden genannten Kritikpunkte durch eine Veränderung der ANS Aufgaben und Testversuche zur Überprüfung, ob die Kinder die Aufgabenstellung richtig verstanden haben, behoben wurden, konnte keine Korrelation zwischen dem Wissen über exakte Zahlen und der ANS Leistung gefunden werden (Negen & Sarnecka, 2015).

Personen können Zahlen also entweder genau durch ihr Wissen über exakte Zahlen, das durch die Sprache unterstützt wird, oder ungefähr durch die Verwendung des ANS darstellen. Erwachsene haben ein Mapping zwischen den beiden Systemen. Sie können aufgrund eines

ungefähren Gespürs für die Menge von Punkten, die sie kurz gesehen haben, eine exakte Anzahl abschätzen und umgekehrt auch eine Approximation erzeugen, wenn sie eine exakte Zahl hören, indem sie beispielsweise einen Vorgang in sehr schnellem Tempo ungefähr so häufig ausführen. Eine Studie beschäftigte sich damit, wie diese Systeme ursprünglich gebildet werden und, ob in der Entwicklung eine Abbildungsrichtung vor der anderen funktionell wird. Dazu wurden Kindergartenkinder von 2;7 bis 4;6 Jahren untersucht. Denn bevor Kinder Zahlwörter verstehen, ist eine Verbindung zwischen Zahlwörtern und dem ANS nicht möglich (Odic, Le Corre, & Halberda, 2015). Um das Alter von zwei Jahren lernen Kinder die Zahlen, aber sie fügen ihnen noch keine bestimmte numerische Bedeutung bei. Ungefähr mit zweieinhalb Jahren fangen sie an zu verstehen, was das Wort "eins" bedeutet (Le Corre & Carey, 2007).

Die Ergebnisse legen nahe, dass Kinder die "one-knower" sind, noch kein Mapping zwischen Zahlwörtern und dem ANS erlangt haben. Kinder, die bereits die Bedeutung der Wörter zwei, drei und des Kardinalprinzips kennen, haben erfolgreich eine Zuordnung von Zahlwörtern zum ANS aber nicht vom ANS zu Zahlwörtern. Dies geschieht signifikant früher. Diese Ergebnisse deuten auf einen Entwicklungsverlauf hin, demzufolge Kinder zuerst ein Mapping zwischen Zahlwörtern und dem ANS herstellen und dann langsam über die Zeit die Abbildungsgenauigkeit weiterentwickeln. Laut diesen Ergebnissen ist das Mapping zwischen dem ANS und Zahlwörtern also nicht sofort bidirektional. Vor dem Alter von vier Jahren konnten Kinder die Aufgabe, die die Funktion ANS zu Zahlwörtern erfordert, nicht erfolgreich meistern (Odic et al., 2015).

3.6 Retest-Reliabilität

Die Reliabilität gibt die Messgenauigkeit eines Verfahrens an, unabhängig davon, was gemessen wird (Kubinger, 2009). Wenn ein Test ein bestimmtes psychologisches Merkmal exakt, also ohne Messfehler misst, ist er perfekt reliabel (Schermelleh-Engel & Werner, 2012).

Um das Ausmaß der Reliabilität zu bestimmen, gibt es in der klassischen Testtheorie drei Vorgehensweisen: die Paralleltest-Reliabilität, die Retest-Reliabilität und die Berechnung der internen Konsistenz (Kubinger, 2009).

Zur Bestimmung der Retest-Reliabilität wird ein Testverfahren zu zwei unterschiedlichen Zeitpunkten vorgelegt und die Korrelation zwischen den beiden Testergebnissen berechnet. Diese Korrelation kann abhängig vom Zeitintervall zwischen den Messungen variieren, da sich verschiedene Einflüsse wie Übungs- oder Erinnerungseffekte, und eine Veränderung des gemessenen Merkmals auf die Reliabilität auswirken können (Schermelleh-Engel & Werner, 2012).

Bei der Erhebung der Retest-Reliabilität bei Entwicklungsverfahren im Kindesalter stellen Lerneffekte und der Fortschritt in der Entwicklung Schwierigkeiten dar. Beim zweiten Testzeitpunkt kann aufgrund dieser Faktoren eine höhere Testleistung erwartet werden (Petermann & Macha, 2005). Da Entwicklungsaspekte in Zeiträumen von Wochen oder Monaten wirksam werden, wird die Retest-Reliabilität in der Entwicklungsdiagnostik kritisch betrachtet (Petermann & Macha, 2003). Mit der Berechnung der Paralleltest-Reliabilität können reliabilitätsverändernde Einflüsse wie Übungs- oder Erinnerungseffekte und Merkmalsveränderungen kontrolliert werden (Schermelleh-Engel & Werner, 2012). Für Entwicklungstests liegen aber gewöhnlich keine vergleichbaren Testversionen zur Berechnung der Paralleltest-Reliabilität vor (Petermann & Macha, 2005). Da die Konstruktion von identischen Items zur Anwendung der Paralleltest-Reliabilität problematisch ist und die Entwicklungstheorien in ihren Grundannahmen von einer Veränderung über die Zeit ausgehen, was gegen die Verwendung der Retest-Reliabilität spricht, wird die Überprüfung der Reliabilität mit der Split-half Methode oder der Konsistenzanalyse empfohlen (Ettrich, 2000). Die Split-half-Reliabilität ist bei den meist heterogenen Skalen von Entwicklungstests allerdings wenig aussagekräftig. Die Ausprägung der internen Konsistenz ist ebenfalls von den erhobenen Merkmalen abhängig (Petermann & Macha, 2005). Zudem lässt sich die

Konsistenzanalyse, bei der für jedes Item ein Konsistenzwert berechnet wird, bei der Entwicklungsdiagnostik nur beschränkt einsetzen, da für die jeweiligen Altersstufen häufig nur sehr wenige Items zur Verfügung stehen und die Skalen oft nicht homogen sind. Die Homogenität der Skalen stellt eine Grundvoraussetzung für die Konsistenzanalyse dar (Ettrich, 2000).

In vielen Entwicklungsverfahren finden sich, aufgrund der oben beschriebenen Kritik, keine Angaben zur Retest-Reliabilität. Eine Reliabilitätsschätzung, die nur auf einer einmaligen Messung beruht, könnte die Zuverlässigkeit des Verfahrens überschätzen, die in der Praxis realisierbar ist, da bei dieser der situative Einfluss oder der Einfluss der Befindlichkeit des Kindes gering gehalten wird (Renner, 2009). Zudem stellen Angaben zur Retest-Reliabilität besonders bei heterogenen Merkmalen eine sinnvolle Überprüfung des Mindestniveaus der erreichten Messgenauigkeit dar (Hasselhorn & Margraf-Stiksrud, 2015). Mit dem Ziel zur psychometrischen Qualität des spielbasierten Verfahrens beizutragen, wird die Retest-Reliabilität erhoben. Die aktuelle Forschung hält an dem Ziel fest, gesicherte Angaben im Bezug auf die Retest-Reliabilität zu erlangen.

3.6.1 Intervall zwischen den Messzeitpunkten

Bei der Wahl des Intervalls zwischen den beiden Testungen spielen Veränderungen des wahren Merkmals eine wichtige Rolle (Schermelleh-Engel & Werner, 2012). Bei Kleinkindern finden Veränderungen innerhalb von Wochen oder Monaten statt, dabei variieren sowohl die Entwicklungsgeschwindigkeit als auch die Verlaufsform. Deshalb wird bei größeren Zeitabständen der Reliabilitätskoeffizient auch von Entwicklungsprozessen, die in der Zwischenzeit stattfanden, beeinflusst (Renner, 2009). Bei kürzeren Intervallen können sich Erinnerungs- und Übungseffekte auswirken (Esser & Petermann, 2010).

Die Veränderungen beim gemessenen Merkmal zwischen den beiden Testzeitpunkten können die Reliabilitätsschätzung verfälschen und werden in systematische und unsystematische Veränderungen unterteilt. Systematische Veränderungen des wahren Wertes,

die bei allen Personen im gleichen Ausmaß auftreten, wie ein identer Wissenszuwachs, können rechnerisch berücksichtigt werden und würden deshalb die Korrelation der beiden Messzeitpunkte nicht verändern. Ebenso würden situationsspezifische Einflüsse die Reliabilität nicht beeinflussen, insofern sie sich auf die Personen gleich auswirken. Es ist jedoch bei vielen psychologischen Merkmalen unwahrscheinlich, dass sie sich bei unterschiedlichen Personen auf die gleiche Weise verändern (Schermelleh-Engel & Werner, 2012).

Die Korrelation zwischen den beiden Messzeitpunkten ändert sich, wenn unsystematische Veränderungen der wahren Werte auftreten. Für verschiedene Personen können unterschiedlich große Übungeffekte oder Wissenszuwächse auftreten. Ebenso stellen verschiedene Reaktionen auf situationsspezifische Effekte und unterschiedliche Tagesformen unsystematische Veränderungen dar. Unsystematische Veränderungen bei instabilen Merkmalen mit interindividuell verschiedenen Entwicklungsverläufen führen bei der Berechnung der Retest-Reliabilität zur Unterschätzung der Reliabilität des Verfahrens. Umgekehrt können Erinnerungseffekte an die erste Testung die Reliabilität künstlich überhöhen. Diese Erinnerungseffekte werden bei sehr langen Intervallen kaum eine Rolle spielen, wohingegen unsystematische Merkmalsveränderungen die Reliabilität stark beeinflussen könnten. Bei einem kurzen Intervall ist von eher geringeren unsystematischen Merkmalsveränderungen auszugehen, dafür ist das Risiko von Erinnerungseffekten höher. Die Intervalle sollten umso kürzer sein, je instabiler das gemessene Merkmal ist (Schermelleh-Engel & Werner, 2012).

4 Fragestellung

Aufgrund der nicht zufriedenstellenden Situation an vorhandenen Verfahren zur Entwicklungsdiagnostik in diesem Altersbereich ist das Ziel des Projektes von Ass.-Prof. Dr. Deimann und Ass.-Prof. Dr. Kastner-Koller die Entwicklung eines Verfahrens für zwei- bis

dreijährige Kinder, das durch den spielbasierten Ansatz sowohl die Bedürfnisse der Zielgruppe berücksichtigt als auch den Gütekriterien entspricht.

Das Ziel dieser Arbeit ist die testanalytische Untersuchung des spielbasierten Inventars für Zwei- bis Dreijährige (Kuchler, Sapper, Deimann, & Kastner-Koller, 2011) hinsichtlich der Retest-Reliabilität. Der überarbeitete Itempool von Kronberger und Punz (2013) wird als vorläufiger Letztstand angenommen und zur Überprüfung der Retest-Reliabilität herangezogen. Den Schwerpunkt dieser Arbeit stellt der kognitive Entwicklungsbereich sowie die Darstellung und Diskussion der darauf bezogenen Ergebnisse der Retest-Reliabilität dar.

5 Methode

5.1 Untersuchungsinstrument

Zur spielbasierten Erfassung der Entwicklung von Zwei- bis Dreijährigen wurde 2011 von Kuchler, Sapper, Deimann und Kastner-Koller ein Itempool entwickelt. Dieser Itempool wurde mehrmals jeweils ungefähr 20 Kindern vorgegeben und testtheoretisch überprüft. Die darauf basierenden Analysen und Änderungsvorschläge wurden in den folgenden Arbeiten berücksichtigt und die abgeänderten Items wurden erneut vorgelegt und hinsichtlich der Gütekriterien diskutiert (Birngruber, 2012; Fuchs-Gaderer, 2012; Kronberger, 2013; Liszt, 2014; Punz, 2013; Putzer, 2013; Sindelar, 2013; Tscherne, 2014). Aufgrund der Ergebnisse von Tscherne und Liszt (2014) sind keine weiteren Veränderungen mehr notwendig und der adaptierte Itempool von Kronberger und Punz (2013) wurde unverändert für diese Untersuchung herangezogen. Er umfasst die kognitive, sprachliche, motorische und sozial-emotionale Entwicklung, sowie die Entwicklung der visuellen Wahrnehmung, der Aufmerksamkeit und des Gedächtnisses.

Der kognitive Bereich, mit dem sich diese Masterarbeit befasst, besteht aus der Skala "Theory of Mind", die sich aus den Skalen "Hineinversetzen in andere" und "Als-ob-Spiel" zusammensetzt, und der Skala "numerisches Wissen", die die Skalen "aktive

Mengenerfassung", "passive Mengenerfassung" und "Zählen" umfasst. Insgesamt besteht die Gesamtskala der kognitiven Entwicklung aus 18 Items. Zur Erhebung der Skala "Hineinversetzen in andere" wurde ein Buch verwendet, das gemeinsam mit den Kindern angesehen wurde. Das Als-ob-Spiel wurde mittels Beobachtung kodiert und die Items zur Erfassung des numerischen Wissens wurden variabel, mit jeweils zur Spielsituation passenden Objekten, in den Untersuchungsablauf eingebaut.

Der Protokollbogen (Anhang I) gibt Aufschluss über die einzelnen Items. Außerdem wird ein Elternfragebogen (Anhang G) zur Erhebung der soziodemographischen Daten und der Einschätzung der Eltern in den Bereichen Selbständigkeit, Sozial- und Spielverhalten und Sprachentwicklung vorgegeben.

5.2 Untersuchungsablauf

Alle Testungen fanden von Dezember 2014 bis April 2015 im Kleinkindtestraum der Fakultät für Psychologie der Universität Wien statt. Die Testungen wurden zwischen 9:00 Uhr und 18:00 Uhr durchgeführt. Bei den beiden Terminen waren immer dieselbe Testleiterin und Beobachterin anwesend.

Die Testleiterin holte die Bezugsperson und das Kind beim Eingang der Fakultät für Psychologie ab und brachte sie zum Kleinkindtestraum. Im Testraum befanden sich immer gleich angeordnet ein Kaufmannsladen, eine Puppenküche, drei Stufen, ein kindgerechter Tisch mit Sesseln sowie eine Sitzgelegenheit für die Bezugsperson und die Beobachterin. Eine Videokamera für die Auswertung wurde am Ende des Raumes positioniert und während der Testung gegebenenfalls von der Beobachterin neu ausgerichtet. Die zusätzlich benötigten Spielmaterialien waren im Raum verteilt, damit das Kind eine Auswahlmöglichkeit hatte und selbst entscheiden konnte, mit welchen Materialien es zu welchem Zeitpunkt spielen möchte. Die Bezugsperson wurde vor der Testung gebeten, sich während der Spielsituation zurückzuhalten, dem Kind keine Antworten vorzusagen und, falls nötig, zu Beginn gemeinsam mit dem Kind die Materialien zu explorieren. Die meisten Kinder waren an der

Puppenküche sehr interessiert und die Exploration erfolgte ohne die Bezugsperson. Dabei erfolgte sehr häufig auch die Vorgabe der ersten Items. Die Reihenfolge der Itemvorgabe wurde an die Interessen des Kindes angepasst. Während der Testung füllte die Bezugsperson den Elternfragebogen und die Erlaubnis zur Videoaufzeichnung (Anhang H) aus.

Am Ende jeder Spielsituation durfte sich das Kind als Belohnung bzw. als Motivation für einen zweiten Termin ein Geschenk aus der Schatztruhe aussuchen. Diese war mit verschiedenen altersgerechten Spielzeugen gefüllt und auch die Bezugsperson bekam ein kleines Dankeschön überreicht.

5.3 Stichprobenbeschreibung

Zur Rekrutierung der Stichprobe wurde ein Elternbrief (Anhang F) verfasst. Dieser wurde in Kindergärten, Kinderarztpraxen und Kinderspielgruppen sowie im Freundes- und Bekanntenkreis verteilt. Zusätzlich wurden Kinder durch das direkte Ansprechen der Eltern bei Indoor Spielplätzen und Kinderbuchvorlesungen rekrutiert. Außerdem wurde die Stichprobe durch das Schneeballprinzip erweitert, da teilnehmende Eltern die Informationen an ihren Freundes- und Bekanntenkreis weitergaben.

Interessierte Eltern wurden entweder persönlich, telefonisch oder per E-Mail über den Ablauf und die Aufzeichnung mit einer Videokamera informiert. Anschließend wurden Termine für den ersten und zweiten Testzeitpunkt vereinbart. Die Eltern wurden bei der Terminvereinbarung auf die Beachtung des üblichen Tagesrhythmus ihres Kindes hingewiesen. Zudem wurde darauf geachtet, dass der erste und der zweite Testtermin möglichst zur selben Uhrzeit stattfinden. Einige Tage vor den vereinbarten Terminen bekamen die Eltern ein Erinnerungsmail mit den wichtigsten Informationen.

Aufgrund von Erkrankungen der Kinder oder Bezugspersonen gab es einige Ausfälle vor dem ersten und vor dem zweiten Testtermin.

Die Stichprobe zum ersten Testzeitpunkt bestand aus 24 Kindern. Für die Berechnungen wurden nur jene Kinder herangezogen, die zu beiden Testungen kamen. Die Stichprobe

reduzierte sich dadurch auf 21 Kinder (sieben Buben und 14 Mädchen). Tabelle 1 zeigt die Altersverteilung der Kinder zum ersten und zweiten Testzeitpunkt sowie die Geschlechtsverteilung. Das durchschnittliche Alter der Stichprobe betrug beim ersten Testzeitpunkt 28.52 Monate ($SD = 3.108$) und beim zweiten Testzeitpunkt 29.76 Monate ($SD = 3.048$). Die genaue Altersverteilung ist in Tabelle 1 dargestellt. Kein Kind war zu einem der Testzeitpunkte 31 Monate alt. Zwischen der ersten und zweiten Testung lagen mindestens 28 und maximal 42 Tage.

Tabelle 1

Alter in Monaten

		Alter in Monaten												Summe
		24	25	26	27	28	29	30	32	33	34	35	36	
1. Testung	weiblich	1	1	2	0	3	4	0	1	0	2	0	0	14
	männlich	1	0	1	2	1	0	0	1	0	1	0	0	7
	Summe	2	1	3	2	4	4	0	2	0	3	0	0	21
2. Testung	weiblich		1	0	2	1	2	5	0	1	0	1	1	14
	männlich		0	1	1	2	1	0	0	1	0	1	0	7
	Summe		1	1	3	3	3	5	0	2	0	2	1	21

Aufgrund der Angaben aus dem Elternfragebogen ergeben sich die folgenden Informationen: Das durchschnittliche Alter der Mütter lag bei 34.2 Jahren, wobei die jüngste Mütter 25 und die älteste 46 Jahre alt war. Bei den Vätern lag das durchschnittliche Alter mit 37.45 Jahren etwas darüber. Der jüngste Vater war 26 und der älteste 52 Jahre alt. Eine Mutter und ein Vater machten keine Angaben zu ihrem Alter. Alle Kinder leben mit dem Vater und der Mutter im gemeinsamen Haushalt. Bei der Ausbildung der Eltern zeigt sich ein hoher Anteil an AkademikerInnen. Bei den Vätern haben zwölf (57.1 %) einen Universitäts- oder Fachhochschulabschluss und bei den Müttern fünfzehn (71.4 %). Fünf Väter (23.8 %) haben eine Lehre oder Fachschule besucht und vier (19 %) haben die Matura. Bei den

Müttern haben zwei (9.5 %) eine Lehre oder Fachschule absolviert und vier (19 %) schlossen mit Matura ab.

Von den 21 Kindern besuchten zum ersten Untersuchungszeitpunkt vierzehn (66.7 %) den Kindergarten und elf (52.4 %) nahmen an Sport- oder Freizeitkursen teil. Dreizehn Kinder (61.9 %) sind Einzelkinder, sechs (28.6 %) haben einen Bruder oder eine Schwester und jeweils ein Kind (4.8 %) hat zwei bzw. drei Geschwister. Mehrsprachig wachsen sechs (28.6 %) der Kinder auf. Bei zwei Kindern (9.5 %) wurde von einer Risikoschwangerschaft berichtet. Gesundheitliche Probleme der Kinder wurden keine angegeben.

6 Ergebnisse

Im Folgenden werden die Ergebnisse des kognitiven Bereichs dargestellt. Neben der Retest-Reliabilität und der internen Konsistenz werden auch die Itemschwierigkeiten und Trennschärfen berichtet.

Vor der Berechnung wurden die Daten mittels standardisierter Schiefe (Field, 2009) auf Normalverteilung überprüft. Dafür wurden bei allen Skalen die Differenzen aus den beiden Messzeitpunkten berechnet und für diese wurde die standardisierte Schiefe errechnet. Eine Normalverteilung der Differenzen kann angenommen werden, wenn der z-Wert unter 1.96 liegt. Alle Differenzen der Messzeitpunkte des kognitiven Bereichs waren normalverteilt (Anhang B) und zur Berechnung wurden deshalb die parametrischen Verfahren Korrelation nach Pearson und der T-Test für verbundene Stichproben herangezogen. Bei den deskriptiven Daten weisen die Pausendauer und die gesamte Testdauer (Anhang A) keine Normalverteilung auf. Für diese Daten wurde daher mit dem nichtparametrischen Wilcoxon-Vorzeichen-Rang-Test gerechnet. Die Signifikanz wurde 2-seitig mit einem Signifikanzniveau von $p \leq 0.05$ geprüft.

6.1 Testdauer und Pausenanzahl

Die Gesamtdauer betrug bei der ersten Testung durchschnittlich 107 Minuten ($SD = 15.9$) und beinhaltet die komplette Zeitspanne vom Betreten des Testraumes bis zur Verabschiedung. Die kürzeste Ersttestung dauerte 79 Minuten und die längste 133 Minuten. Die zweite Testung dauerte mit durchschnittlich 102.57 Minuten ($SD = 12.432$) signifikant kürzer als die erste Testung ($U = -2.204, p = .028$). Dabei benötigte die kürzeste Testung 80 Minuten und die längste 131 Minuten.

Die Anlaufzeit bei der ersten Testung lag durchschnittlich bei 3.05 Minuten ($SD = 2.061$) und umfasst die Zeit, die das Kind bis zur Exploration des Spielmaterials mit der Testleiterin benötigte. Beim zweiten Zeitpunkt betrug die Anlaufzeit ebenfalls durchschnittlich 3.05 Minuten ($SD = 2.5$). Bei der durchschnittlichen Anlaufzeit gab es bei den beiden Terminen keine signifikanten Unterschiede.

Die "Beschäftigung mit den Materialien" berechnet sich aus der gesamten Zeit ausschließlich der Pausen, der Anlaufzeit und der Verabschiedung. Sie beschreibt also die Zeit, die sich das Kind mit den Testmaterialien beschäftigte und lag bei der ersten Testung durchschnittlich bei 86.29 Minuten ($SD = 13.290$). Die kürzeste Beschäftigung mit den Materialien dauerte 62 Minuten und die längste 107. Die "Beschäftigung mit den Materialien" beträgt bei der zweiten Testung im Durchschnitt 80.05 Minuten ($SD = 6.830$) und ist damit signifikant kürzer als beim ersten Termin ($t(20) = 2.150, p = .044$), wobei die Zeitspanne von 69 bis 94 Minuten reicht.

Bei der ersten Testung benötigten die Kinder zwischen keiner und drei Pausen, wobei der Durchschnittswert bei 0.9 lag ($SD = 0.889$). Im Durchschnitt dauerten die Pausen 4.4 Minuten ($SD = 5.295$). Als Pausen wurden die Zeitspannen kodiert, in denen sich das Kind nicht mit dem Testmaterial beschäftigte wie zum Beispiel Essen, Wickeln, WC. Die Pausenanzahl zum zweiten Termin liegt durchschnittlich bei 0.81 ($SD = 0.873$) und die

Pausen dauerten im Mittel 5 Minuten ($SD = 6.943$). Signifikante Unterschiede zwischen den Terminen gibt es dabei nicht.

Die Verabschiedung bezeichnet die Zeitspanne zwischen der Beendigung der Spielsituation durch die Testleiterin und dem Verlassen des Testraumes durch das Kind und die Bezugsperson. Dies dauerte beim ersten Termin durchschnittlich 13.24 Minuten ($SD = 8.654$) und beim zweiten Termin belief sich die Verabschiedung auf durchschnittlich 14.67 Minuten (7.024), dabei zeigt sich kein signifikanter Unterschied.

6.2 Kognitive Entwicklung

Die Gesamtskala "kognitive Entwicklung" setzt sich aus den beiden Skalen "Theory of Mind" und "numerisches Wissen" zusammen. Bei der Berechnung der Retest-Reliabilität ergibt sich mit $r_{tt} = .820$, $p < .001$ ein signifikantes Ergebnis mit einem guten Retest-Reliabilitätskoeffizienten.

Bezüglich der Mittelwerte zeigt sich von Testzeitpunkt 1 ($M = 6.48$, $SD = 4.05$) zu Testzeitpunkt 2 ($M = 8.21$, $SD = 4.29$) eine signifikante Verbesserung $t(20) = 3.169$, $p = .005$.

Zwischen den Mädchen und Buben zeigen sich weder beim ersten Testzeitpunkt $t(19) = 3.16$, $p = .755$ noch beim zweiten Testzeitpunkt $t(19) = 0.053$, $p = .959$ signifikante Unterschiede.

Die Gesamtskala "kognitive Entwicklung" weist beim ersten Testzeitpunkt ein Cronbachs Alpha von $\alpha = .849$ auf und hat auch beim zweiten Testzeitpunkt mit einem etwas höheren Wert von $\alpha = .864$ einen guten Reliabilitätskoeffizienten.

Die in Tabelle 2 dargestellte Itemanalyse für die Gesamtskala weist zum ersten Zeitpunkt für die Items "schlafen", "Wie viele sind das - zwei", "Wie viele sind das - drei" und "Kardinalprinzip" eine hohe Schwierigkeit auf. Die Lösungswahrscheinlichkeit für das Item "Gib mir 1" ist hingegen hoch. Als wenig trennscharf zum ersten Zeitpunkt zeigen sich die Items "Als-Ob-Spiel", "schlafen", "schmutzige Hände", "Gib mir 1" und "Gib mir 3". Beim zweiten Zeitpunkt zeigen die drei Items "Gib mir 3", "Wie viele sind das - zwei" und "Wie

viele sind das - drei" eine sehr hohe Itemschwierigkeit. Dagegen wird das Item "Gib mir 1" von allen Kindern gelöst. Alle anderen Items liegen beim ersten und zweiten Zeitpunkt im mittleren Schwierigkeitsbereich, der sich laut Bortz und Döring (2006) von .2 bis .8 erstreckt. Die Items "hungrig" und "Gib mir 2" weisen eine geringe Trennschärfe auf, die übrigen Items haben mittlere bis hohe Werte.

Tabelle 2

Ergebnisse der Itemanalyse der Gesamtskala "kognitive Entwicklung" beim ersten und zweiten Testzeitpunkt

Item	1. Zeitpunkt		2. Zeitpunkt	
	Item-schwierigkeit	korrigierte Trennschärfe	Item-schwierigkeit	korrigierte Trennschärfe
Als-Ob-Spiel	.571	.085	.548	.300
schmutzige Hände	.524	.200	.619	.389
hungrig	.476	.437	.476	.251
schlafen	.095	.156	.333	.456
Pflaster	.524	.705	.571	.511
durstig	.286	.463	.476	.450
Wie viele sind das - zwei	.190	.396	.190	.729
Wie viele sind das - drei	.143	.621	.190	.468
Gib mir 1	.905	.194		
Gib mir 2	.476	.303	.667	.297
Gib mir 3	.333	.078	.190	.532
zählen	.571	.657	.667	.634
richtig zählen	.238	.733	.381	.602
Eins-zu-eins				
Zuordnung	.333	.609	.571	.498
stabile Reihenfolge	.381	.628	.476	.647
richtige Reihenfolge	.238	.733	.524	.575
Kardinalprinzip	.190	.742	.333	.662

6.3 Ergebnisse zur Theory of Mind

Der Fähigkeitsbereich "Theory of Mind" umfasst die beiden Skalen "Als-ob-Spiel" und "Hineinversetzen in Andere" und besteht aus sechs Items. Die Retest-Reliabilität der Skala weist mit $r_{tt} = .887$, $p < .001$ einen guten Wert auf.

Bei der Berechnung der Mittelwerte von Testzeitpunkt 1 ($M = 2.48$, $SD = 1.62$) und Testzeitpunkt 2 ($M = 3.02$, $SD = 1.77$) zeigt sich eine signifikante Verbesserung $t(20) = 3.060$, $p = .006$.

Die Prüfgröße des T-Tests zeigt keine signifikanten Geschlechtsunterschiede beim ersten $t(19) = -0.05$, $p = .963$ und beim zweiten Testzeitpunkt $t(19) = 0.82$, $p = .422$.

Der Reliabilitätskoeffizient Cronbachs Alpha liegt beim ersten Zeitpunkt mit $\alpha = .673$ nicht im zufriedenstellenden Bereich und erhöht sich beim zweiten Zeitpunkt auf einen akzeptablen Wert von $\alpha = .709$. Die Ergebnisse der Itemanalyse dieser Skala sind zur Übersicht im Anhang D beigefügt.

6.3.1 Als-ob-Spiel

Die Retest-Reliabilität dieses Items zeigt ein signifikantes Ergebnis ($r_{tt} = .596$, $p = .004$), liegt aber nicht im akzeptablen Bereich. Zwischen den beiden Mittelwerten von Testzeitpunkt 1 ($M = 0.57$, $SD = 0.24$) und Testzeitpunkt 2 ($M = 0.55$, $SD = 0.15$) lassen sich keine signifikanten Veränderungen feststellen ($t(20) = -0.568$, $p = .576$).

Zwischen den Mädchen und Buben ergeben sich weder beim ersten $t(19) = -0.50$, $p = .621$ noch beim zweiten Zeitpunkt $t(19) = -0.97$, $p = .346$ signifikante Geschlechtsunterschiede.

Zum ersten Testzeitpunkt weist das Item eine mittlere Schwierigkeit von $.571$ auf. Dabei zeigten vier Kinder ein gänzlich vorgestelltes Als-ob-Spiel oder verwendeten Objekte als etwas anderes, als sie wirklich sind. Ein Als-ob-Spiel anhand von konkreten Objekten konnte bei 16 Kindern beobachtet werden. Ein Kind zeigte kein Als-ob-Spiel. Zum zweiten Testzeitpunkt liegt die Schwierigkeit mit $.548$ ebenfalls im mittleren Bereich und alle Kinder

zeigten eine Form des Als-ob-Spiels. Bei 19 Kindern wurde ein Als-ob-Spiel an konkreten Objekten beobachtet und bei zwei Kindern ein gänzlich vorgestelltes Als-ob-Spiel.

6.3.2 Hineinversetzen in Andere

Die Skala "Hineinversetzen in Andere" besteht aus fünf Items und weist eine gute Retest-Reliabilität auf ($r_{tt} = .901, p < .001$). Die Mittelwerte zwischen den beiden Testzeitpunkten, Testzeitpunkt 1 ($M = 1.90, SD = 1.55$) und Testzeitpunkt 2 ($M = 2.48, SD = 1.72$), haben sich signifikant verbessert $t(20) = 3.508, p = .002$.

Bezüglich der Geschlechtsunterschiede fallen die Berechnungen zum ersten Zeitpunkt $t(19) = 0.10, p = .924$ und zum zweiten Zeitpunkt $t(19) = 0.89, p = .384$ nicht signifikant aus.

Das Cronbachs Alpha dieser Skala liegt beim ersten Testzeitpunkt bei nicht zufriedenstellenden $\alpha = .680$ und erhöht sich beim zweiten Testzeitpunkt auf akzeptable $\alpha = .717$.

In Tabelle 3 sind die Ergebnisse der Itemanalyse dargestellt. Beim ersten Zeitpunkt weist das Item "schlafen" die höchste Schwierigkeit auf (.10). Die anderen Items liegen im mittleren Schwierigkeitsbereich. Eine hohe Trennschärfe weist das Item "durstig" auf (.654). Die Trennschärfe des Items "schlafen" ist beim ersten Testzeitpunkt nicht zufriedenstellend (.156). Beim zweiten Testzeitpunkt ist "schlafen" ebenfalls das schwierigste Item (.33), aber es befindet sich nun im mittleren Bereich. Die korrigierte Trennschärfe des Items "schlafen" ist beim zweiten Termin im Gegensatz zum ersten hoch und es repräsentiert zusammen mit dem Item "hungrig" das Gesamtergebnis der Skala gut. In Tabelle 3 sind die Itemschwierigkeiten und Trennschärfen aller Items dargestellt.

Tabelle 3

Ergebnisse der Itemanalyse zur Skala "Hineinversetzen in Andere" beim ersten und zweiten Testzeitpunkt

Item	1. Zeitpunkt		2. Zeitpunkt	
	Item-schwierigkeit	korrigierte Trennschärfe	Item-schwierigkeit	korrigierte Trennschärfe
schmutzige Hände	.52	.456	.62	.484
hungrig	.48	.447	.48	.567
schlafen	.10	.156	.33	.596
Pflaster	.52	.456	.57	.343
durstig	.29	.654	.48	.405

Bei der Überprüfung der Skala auf Missings, dies sind Items, die von den Kindern nicht bearbeitet wurden, zeigt sich, dass beim ersten Termin ein Kind alle fünf Items verweigerte. Einige Kinder verweigerten ein Item auch zu beiden Zeitpunkten. Zwei Kinder verweigerten das Item "schmutzige Hände" zu beiden Terminen, eines das Item "schlafen" und zwei das Item "durstig". Die fehlenden Werte aller Items dieser Skala zeigt Tabelle 4.

Tabelle 4

Fehlende Werte der Skala "Hineinversetzen in Andere"

Item	1. Zeitpunkt		2. Zeitpunkt	
	Häufigkeit	Prozent	Häufigkeit	Prozent
schmutzige Hände	3	14,3 %	2	9,5 %
hungrig	1	4,8 %	1	4,8 %
schlafen	2	9,5 %	2	9,5 %
Pflaster	1	4,8 %	1	4,8 %
durstig	5	23,8 %	5	23,8 %

6.4 Ergebnisse zum numerischen Wissen

Das numerische Wissen setzt sich aus den Skalen "aktive Mengenerfassung" sowie "passive Mengenerfassung" und der Skala "Zählen" zusammen und besteht aus elf Items. Der Reliabilitätskoeffizient liegt für die Retest-Reliabilität mit $r_{tt} = .677$, $p = .001$ nicht im akzeptablen Bereich.

Bei der Analyse der Mittelwerte zeigen sich signifikante Verbesserungen von Testzeitpunkt 1 ($M = 4.00$, $SD = 3.07$) zum Testzeitpunkt 2 ($M = 5.19$, $SD = 3.20$), $t(20) = 2.163$, $p = .043$.

Mädchen und Buben unterscheiden sich beim ersten Zeitpunkt $t(19) = 0.44$, $p = .662$ und beim zweiten Zeitpunkt $t(19) = -0.38$, $p = .710$ nicht signifikant voneinander.

Bei diesem Fähigkeitsbereich liegt das Cronbachs Alpha zum ersten Zeitpunkt bei $\alpha = .848$ und beim zweiten Zeitpunkt bei $\alpha = .871$ und weist damit gute Werte auf. Die Ergebnisse der Itemanalyse dieser Skala sind zur Übersicht im Anhang E beigefügt.

6.4.1 Zählen

Die Skala "Zählen" umfasst sechs Items und zeigt mit $r_{tt} = .595$, $p = .004$ eine Korrelation die außerhalb des akzeptablen Bereiches liegt. Beim T-Test zeigen sich signifikante Verbesserungen des Mittelwertes $t(20) = 2.160$, $p = .043$ zwischen dem ersten Testzeitpunkt ($M = 1.95$, $SD = 2.38$) und dem zweiten Testzeitpunkt ($M = 2.95$, $SD = 2.33$).

Keine signifikanten Unterschiede zeigen sich zwischen den Geschlechtern beim ersten Zeitpunkt $t(19) = 0.32$, $p = .755$ und beim zweiten Zeitpunkt $t(19) = -0.85$, $p = .404$.

Für diese Skala ergibt sich ein Cronbachs Alpha für den ersten Zeitpunkt von $\alpha = .928$. Beim zweiten Testzeitpunkt liegt das Cronbachs Alpha mit $\alpha = .871$ etwas darunter. Dennoch weisen beide Zeitpunkte einen guten Reliabilitätskoeffizienten auf.

Bei der Itemschwierigkeit zeigen sich, wie in Tabelle 5 dargestellt, beim ersten Zeitpunkt durchwegs Werte im mittleren Bereich. Nur das Kardinalprinzip liegt mit einem Wert von .19 im schwierigeren Bereich. Beim zweiten Termin weisen alle Itemschwierigkeiten einen Wert

im mittleren Bereich auf. Bezüglich der Trennschärfe ergeben sich bei allen Items beim ersten und zweiten Zeitpunkt hohe Werte. Die Items unterscheiden damit gut zwischen Personen mit hoher und niedriger Merkmalsausprägung.

Tabelle 5

Ergebnisse der Itemanalyse zur Skala "Zählen" beim ersten und zweiten Testzeitpunkt

Item	1. Zeitpunkt		2. Zeitpunkt	
	Item-schwierigkeit	korrigierte Trennschärfe	Item-schwierigkeit	korrigierte Trennschärfe
zählen	.57	.602	.67	.620
richtig zählen	.24	.893	.38	.748
Eins-zu-eins Zuordnung	.33	.827	.57	.560
stabile Reihenfolge	.38	.789	.48	.778
richtige Reihenfolge	.24	.893	.52	.603
Kardinalprinzip	.19	.786	.33	.721

6.4.2 Aktive Mengenerfassung

Die Skala "aktive Mengenerfassung" besteht aus zwei Items und die Retest-Reliabilität liegt im akzeptablen Bereich $r_{tt} = .719$, $p < .001$.

Zwischen dem ersten Testzeitpunkt ($M = 0.33$, $SD = 0.66$) und dem zweiten Testzeitpunkt ($M = 0.38$, $SD = 0.67$), $t(20) = 0.439$, $p = .666$ ergeben sich keine signifikanten Mittelwertsunterschiede.

Zwischen den getesteten Mädchen und Buben zeigt sich beim ersten Zeitpunkt $t(19) = 0.23$, $p = .821$ und beim zweiten Zeitpunkt $t(19) = 0.45$, $p = .656$ kein signifikanter Unterschied.

Das Cronbachs Alpha dieser Skala ist nicht zufriedenstellend. Es liegt bei der ersten Spielsituation bei $\alpha = .659$ und weist bei der zweiten Spielsituation einen Wert von $\alpha = .553$ auf.

Die in Tabelle 6 dargestellte Itemanalyse zeigt zu beiden Zeitpunkten eine hohe Schwierigkeit bei beiden Items. Die Trennschäfte liegt bei der ersten und zweiten Spielsituation im mittleren Bereich.

Tabelle 6

Ergebnisse der Itemanalyse zur Skala "aktive Mengenerfassung" beim ersten und zweiten Testzeitpunkt

Item	1. Zeitpunkt		2. Zeitpunkt	
	Item-schwierigkeit	korrigierte Trennschärfe	Item-schwierigkeit	korrigierte Trennschärfe
Wie viele sind das - zwei	.19	.495	.19	.382
Wie viele sind das - drei	.14	.495	.19	.382

In Tabelle 7 sind die Missings dieser Skala aufgeführt. Bei genauerer Betrachtung fällt auf, dass vier Kinder beide Items zu beiden Testzeitpunkten verweigerten. Die fehlenden Werte von jeweils neun zum ersten Zeitpunkt stellen die höchsten in der Skala der kognitiven Entwicklung dar.

Tabelle 7

Fehlende Werte der Skala "aktive Mengenerfassung"

Item	1. Zeitpunkt		2. Zeitpunkt	
	Häufigkeit	Prozent	Häufigkeit	Prozent
Wie viele sind das - zwei	9	42,9 %	4	19,0 %
Wie viele sind das - drei	9	42,9 %	8	38,1 %

6.4.3 Passive Mengenerfassung

Die aus drei Items bestehende Skala zeigt für die Retest-Reliabilität eine Korrelation, die nicht im akzeptablen Bereich liegt $r_{tt} = .468$, $p = .032$. Signifikante Unterschiede zwischen Testzeitpunkt 1 ($M = 1.71$, $SD = 0.90$) und Testzeitpunkt 2 ($M = 1.86$, $SD = 0.73$) lassen sich mittels T-Test keine $t(20) = 0.767$, $p = .452$ feststellen.

Die Prüfgröße des T-Tests ergibt zwischen dem ersten Zeitpunkt $t(19) = 0.50$, $p = .621$ und dem zweiten Zeitpunkt $t(8,6) = 0.54$, $p = .602$ keine signifikanten Unterschiede zwischen den getesteten Buben und Mädchen.

Nicht zufriedenstellend ist auch der Alpha-Koeffizient nach Cronbach der beim ersten Testzeitpunkt bei $\alpha = .421$ und beim zweiten Testzeitpunkt bei $\alpha = .505$ liegt.

Wie in Tabelle 8 ersichtlich, war das Item "Gib mir 1" das leichteste Item der Skala und weist zudem die geringste Trennschärfe auf. Beim zweiten Testzeitpunkt konnten alle Kinder das Item "Gib mir 1" lösen. Das Item "Gib mir 3" weist beim ersten Zeitpunkt eine unzureichende Trennschärfe auf und stellt zum zweiten Zeitpunkt ein schwieriges Item dar.

Tabelle 8

Ergebnisse der Itemanalyse zur Skala "passive Mengenerfassung" beim ersten und zweiten Testzeitpunkt

Item	1. Zeitpunkt		2. Zeitpunkt	
	Item-schwierigkeit	korrigierte Trennschärfe	Item-schwierigkeit	korrigierte Trennschärfe
Gib mir 1	.90	.126		
Gib mir 2	.48	.475	.67	.343
Gib mir 3	.33	.206	.19	.343

Die Missings sind in Tabelle 9 dargestellt. Es gab zu beiden Zeitpunkten jeweils zwei fehlende Werte.

Tabelle 9

Fehlende Werte der Skala "passive Mengenerfassung"

Item	1. Zeitpunkt		2. Zeitpunkt	
	Häufigkeit	Prozent	Häufigkeit	Prozent
Gib mir 1	0	0,0 %	0	0,0 %
Gib mir 2	1	4,8 %	0	0,0 %
Gib mir 3	1	4,8 %	2	9,5 %

6.4.4 Mengenerfassung

Die Skala "Mengenerfassung" setzt sich aus der "aktiven Mengenerfassung" und der "passiven Mengenerfassung" zusammen und beinhaltet fünf Items. Diese Skala weist einen akzeptablen Reliabilitätskoeffizienten auf $r_{tt} = .749, p < .001$.

Im T-Test können zwischen Testzeitpunkt 1 ($M = 2.05, SD = 1.24$) und Testzeitpunkt 2 ($M = 2.24, SD = 1.22$) keine Mittelwertsunterschiede festgestellt werden $t(20) = 1.000, p = .329$.

Zwischen den Geschlechtern zeigen sich beim ersten Testtermin $t(19) = 0.61, p = .550$ und beim zweiten Testtermin $t(19) = 0.62, p = .541$ keine signifikanten Unterschiede.

Unzureichend ist die Reliabilität sowohl bei der ersten Spielsituation mit $\alpha = .542$ als auch mit dem höheren Wert von $\alpha = .690$ beim zweiten Termin.

Zum ersten Zeitpunkt weisen die Items "Wie viele sind das - zwei" und "Wie viele sind das - drei" eine sehr hohe Itemschwierigkeit auf, wohingegen die Lösungswahrscheinlichkeit für das Item "Gib mir 1" sehr hoch ist (siehe Tabelle 10). Als nicht trennscharf erweisen sich die Items "Wie viele sind das - drei", "Gib mir 1" und "Gib mir 3". Zum zweiten Zeitpunkt zeigt nur das Item "Gib mir 2" eine Itemschwierigkeit im mittleren Bereich. Das Item "Gib mir 1" wurde von allen Testpersonen gelöst und die übrigen drei Items weisen eine geringe Lösungswahrscheinlichkeit auf. Die Trennschärfe zeigt sich bei allen Items im mittleren bis hohen Bereich, mit Ausnahme des von allen gelösten Items "Gib mir 1".

Tabelle 10

Ergebnisse der Itemanalyse zur Skala "Mengenerfassung" beim ersten und zweiten Testzeitpunkt

Item	1. Zeitpunkt		2. Zeitpunkt	
	Item-schwierigkeit	korrigierte Trennschärfe	Item-schwierigkeit	korrigierte Trennschärfe
Wie viele sind das - zwei	.19	.438	.19	.649
Wie viele sind das - drei	.14	.292	.19	.341
Gib mir 1	.90	.185		
Gib mir 2	.48	.452	.67	.317
Gib mir 3	.33	.196	.19	.649

7 Diskussion

Zur Ermittlung der Retest-Reliabilität des spielbasierten Entwicklungstests wurde der Itempool einer Stichprobe von 21 Kindern im Intervall von vier bis sechs Wochen zweimal vorgelegt. Für die Gesamtskala der kognitiven Entwicklung ergibt sich eine gute Retest-Reliabilität und für die einzelnen Skalen liegt sie zwischen $r_{tt} = .46$ und $r_{tt} = .90$, dabei fielen alle Korrelationen signifikant aus. Die Gesamtskala kann aufgrund der guten Ergebnisse als reliabel erachtet werden.

Auf den Reliabilitätskoeffizienten wirken sich Veränderungen der Fähigkeiten aus und bei Kleinkindern können diese innerhalb von Wochen entstehen. Deshalb spielt die Wahl des Intervalls zwischen den beiden Testungen eine wichtige Rolle (Renner, 2009; Schermelleh-Engel & Werner, 2012). Übungseffekte oder Wissenszuwächse, die nicht für alle Personen gleich sind, stellen unsystematische Veränderungen dar. Solche unsystematischen Veränderungen bei instabilen Merkmalen mit interindividuell verschiedenen Entwicklungsverläufen führen bei der Berechnung der Retest-Reliabilität zur Unterschätzung

der Reliabilität des Verfahrens. Die Intervalle bei der Messwiederholung sollten daher umso kürzer sein, je instabiler das gemessene Merkmal ist (Schermele-Engel & Werner, 2012). Da mit dem untersuchten Verfahren instabile Merkmale gemessen werden, wurde für die Testungen ein Intervall von vier bis sechs Wochen gewählt.

Die Reliabilität von manchen Skalen könnte sich auch durch eine Itemselektion erhöhen. In dieser Arbeit wurde jedoch auf eine Itemselektion verzichtet, da diese zu einem großen Datenverlust geführt hätte, denn aufgrund der Symmetrie hätten die entsprechenden Items zu beiden Zeitpunkten ausgeschlossen werden müssen. Zudem waren die Ergebnisse der Itemanalyse der beiden Testzeitpunkte oft nicht einheitlich.

Die Ergebnisse der einzelnen Skalen des kognitiven Bereichs werden im Folgenden diskutiert und mit den Ergebnissen von Tscherne (2014) verglichen, die denselben Itempool vorgab.

Dabei fanden sich in dieser Stichprobe weder zum ersten noch zum zweiten Zeitpunkt signifikante **Geschlechtsunterschiede** in einer der untersuchten Skalen. Diese Ergebnisse decken sich mit denen von Tscherne (2014), bei der sich ebenfalls keine signifikanten Unterschiede bezüglich der Leistungen zwischen Mädchen und Buben zeigten.

7.1 Diskussion der Ergebnisse der Skala "numerisches Wissen"

Am problematischsten zeigt sich die Skala "**passive Mengenerfassung**". Sie weist die niedrigste Retest-Korrelation ($r_{tt} = .468$) und auch die geringste Reliabilität ($\alpha = .421$ bzw. $\alpha = .505$) der gesamten kognitiven Skala auf. Diese Werte liegen noch unter dem Ergebnis von Tscherne (2014) mit $\alpha = .678$. Vor allem das Item "Gib mir 1" sticht mit einer geringen Trennschärfe (.126) und sehr hohen Lösungswahrscheinlichkeit hervor. Da dieses Item bei Tscherne (2014) einen hohen Wert bei der Trennschärfe (.590) aufweist, liegt ein Stichprobeneffekt nahe.

Beim zweiten Zeitpunkt konnten alle Kinder dieses Item lösen. Da diese Skala nur aus drei Items besteht und es zur Erfassung, ob die Kinder bereits "one-knower" sind, dient, sollte

es in der Skala verbleiben. Auf dieser Stufe können Kinder auf Verlangen ein Objekt geben, bei jeder anderen Zahl geben sie jedoch eine beliebige Menge. Diese Stufe erreichen die Kinder im Alter von zweieinhalb bis drei Jahren (Sarnecka & Carey, 2008). Dies entspricht auch dem durchschnittlichen Alter der Stichprobe, das beim zweiten Zeitpunkt bei 29,76 Monaten ($SD = 3,048$) lag. Die Kinder bleiben einige Monate auf einer Stufe, bevor sie die folgenden Aufgaben "Gib mir 2" bzw. "Gib mir 3" lösen können (Sarnecka & Carey, 2008). Dafür spricht auch, dass es zwischen den beiden Testterminen keine signifikanten Mittelwertsunterschiede gibt.

Im annehmbaren Bereich lag die Retest-Reliabilität der Skala "**aktive Mengenerfassung**" ($r_{tt} = .719$). Die zwei Items dieser Skala konnten im Spiel mit den Kindern einfach eingebaut werden, da zwei bzw. drei beliebige Objekte verwendet werden können. Jedoch gaben viele Kinder auf die Frage "Wie viele sind das?" keine Antwort. Diese Skala weist mit neun Missings pro Item zum ersten Zeitpunkt und vier bzw. acht zum zweiten Zeitpunkt auch die meisten Verweigerungen des kognitiven Bereichs auf, zudem ist die Itemschwierigkeit bei beiden Zeitpunkten hoch. Die Reliabilität liegt mit $\alpha = .659$ und $\alpha = .553$ über der von Tscherne (2014) ($\alpha = .392$), aber nicht im akzeptablen Bereich.

Aufgrund der nicht zufriedenstellenden Werte der Skalen "aktive Mengenerfassung" und "passive Mengenerfassung" wurde die Berechnung der Skala "**Mengenerfassung**", die sich aus den beiden Skalen zusammensetzt, vorgenommen. Diese neue Skala weist eine akzeptable Retest-Reliabilität auf ($r_{tt} = .749$), die Reliabilität bleibt jedoch zu beiden Zeiten im nicht zufriedenstellenden Bereich und liegt mit $\alpha = .542$ zum ersten Zeitpunkt unter und mit $\alpha = .690$ zum zweiten Zeitpunkt über dem berechneten Wert von Tscherne (2014) ($\alpha = .645$).

Die Skala "**Zählen**" erfasst die "how-to-count" Prinzipien von Gelman und Gallistel (1986). In die Berechnung fließen die Eins-zu-eins Zuordnung, die stabile sowie richtige Reihenfolge, das Kardinalprinzip, ob das Kind mindestens zweimal zählt (dabei ist es unerheblich, ob es richtig zählt und womit es zählt) sowie das richtige Zählen, also die

Einhaltung der richtigen Reihenfolge und die Eins-zu-eins Zuordnung ein. Diese aus den angeführten sechs Items bestehende Skala weist mit $\alpha = .928$ eine hohe interne Konsistenz auf und übersteigt noch den von Tscherne (2014) berechneten Wert von $\alpha = .906$. Auch beim zweiten Zeitpunkt ist der Reliabilitätskoeffizient mit $\alpha = .871$ gut. Die Retest-Reliabilität ist jedoch unzureichend ($r_{tt} = .595$), dies könnte durch unsystematische Veränderungen der Fähigkeiten der Kinder zustande gekommen sein. Die Entwicklungsschritte können bei den Kindern interindividuell verschieden sein und führen bei der Berechnung der Retest-Reliabilität zur Unterschätzung der Reliabilität des Verfahrens (Schermelleh-Engel & Werner, 2012). Die Veränderung der Fähigkeiten der Kinder zeigt sich auch durch eine signifikante Verbesserung des Mittelwertes vom ersten ($M = 1.95, SD = 2.38$) zum zweiten Testzeitpunkt ($M = 2.95, SD = 2.33$).

Die Skala "**numerisches Wissen**", die sich aus der "aktiven Mengenerfassung" und "passiven Mengenerfassung" sowie der Skala "Zählen" zusammensetzt, weist zu beiden Zeitpunkten mit $\alpha = .848$ bzw. $\alpha = .871$ eine gute interne Konsistenz auf, die nur leicht unter dem Wert von Tscherne ($\alpha = .883$) liegt. Die Retest-Reliabilität liegt mit $r_{tt} = .677$ nicht mehr im akzeptablen Bereich. Dies lässt im Zusammenhang mit der signifikanten Verbesserung zwischen den Testungen auf unsystematische Veränderungen des gemessenen Merkmals schließen. Weitere mögliche Einflüsse auf die Korrelation zwischen dem ersten und zweiten Testzeitpunkt, die neben einer individuell unterschiedlichen Entwicklungsgeschwindigkeit beachtet werden müssen, sind besonders bei Kleinkindern Schwankungen in der Motivierbarkeit für die Untersuchungen, unterschiedliche physische und psychische Befindlichkeiten wie Ängste, Erkrankungen oder Müdigkeit zum Zeitpunkt der Untersuchungen (Ettrich, 2000).

7.2 Diskussion der Ergebnisse der Skala "Theory of Mind"

Als-ob-Spiele beginnen mit 18 Monaten zu entstehen (Lillard et al., 2013). Das Item "**Als-ob-Spiel**" wurde nicht erhoben, sondern dessen Auftreten wurde durch Beobachtung kodiert.

Unterschieden wurde dabei zwischen einem gänzlich vorgestellten Als-ob-Spiel und einem Als-ob-Spiel anhand von konkreten Objekten. Die Form bei der ein Objekt so verwendet wird, als wäre es ein anderes, tritt früher auf, als jene bei der Kinder beginnen Als-ob mit unsichtbaren Objekten zu spielen (Singer & Singer, 1992). Bei dieser komplexeren Form gibt es keine realen Objekte, die den Als-ob Objekten eine Grundlage in der echten Welt geben (Taylor & Carlson, 1997). Aufgrund der Literatur wäre zu erwarten gewesen, dass alle Kinder eine Form des Als-ob-Spiels zeigen. Allerdings zeigte ein Kind zum ersten Zeitpunkt kein Als-ob-Spiel, und erst beim zweiten Zeitpunkt konnte es bei allen Kindern beobachtet werden. Dabei zeigten weit mehr Kinder ein Als-ob-Spiel anhand von konkreten Objekten und nur wenige ein gänzlich vorgestelltes. Die Skala lag bei beiden Zeitpunkten im mittleren Schwierigkeitsbereich (.571 und .548). Bei Tscherne (2014) wird eine Schwierigkeit von .05 berichtet, was einem sehr schwierigen Item entsprechen würde, obwohl auch bei dieser Stichprobe alle Kinder in der Spielsituation eine Form des Als-ob-Spiels zeigten. Signifikante Mittelwertsunterschiede finden sich im T-Test keine. Die Retest-Reliabilität mit $r_{tt} = .596$ muss als unzureichend bezeichnet werden.

Das Als-ob-Spiel der Kinder beeinflusst auch ihre Theory of Mind. Ab der Mitte des zweiten Lebensjahres können Kinder ihre Wünsche, Absichten und Emotionen von denen anderer Personen abgrenzen und unabhängig davon mental repräsentieren (Sodian, 2008). Sogar zwei und drei Jahre alte Kinder können mit einer sich noch entwickelnden Theory of Mind die Absichten von anderen Personen verstehen (Wellman, 2014).

Mithilfe des TOM-Buchs, bei dem die Kinder aufgrund von Zuständen anderer Personen, die ihnen erzählt wurden, die Absichten der Personen voraussagen mussten, wurden die fünf Items der Skala "**Hineinversetzen in andere**" erfasst. Mit $r_{tt} = .901$ weist diese Skala die höchste Retest-Reliabilität der kognitiven Skala auf. Die Mittelwerte zwischen den beiden Testzeitpunkten haben sich signifikant verbessert. Die Reliabilität liegt beim ersten Zeitpunkt unter dem akzeptablen Bereich $\alpha = .680$ und befindet sich beim zweiten im

zufriedenstellenden Bereich $\alpha = .717$, aber unter dem von Tscherne (2014) errechneten ($\alpha = .755$). Das schwierigste Item stellte zu beiden Zeitpunkten das Item "schlafen" mit .10 bzw. .33 dar. Auch die Trennschärfe ist für dieses Item zum ersten Zeitpunkt niedrig.

Aus den Skalen "Hineinversetzen in andere" und "Als-ob-Spiel" setzt sich die Skala "**Theory of Mind**" zusammen und die Retest-Reliabilität stellt mit $rtt = .887$ einen guten Wert dar. Das Cronbachs Alpha liegt mit $\alpha = .673$ knapp unter dem und zum zweiten Zeitpunkt mit $\alpha = .709$ gerade noch im zufriedenstellenden Bereiche. Beide Werte liegen jedoch unter dem von Tscherne (2014) berechneten Wert ($\alpha = .726$).

7.3 Diskussion der Ergebnisse der Gesamtskala "kognitive Entwicklung"

Für die **Gesamtskala der kognitiven Entwicklung** ergibt sich eine gute Retest-Reliabilität von $rtt = .820$ und zwischen den beiden Messzeitpunkten zeigen sich signifikante Verbesserungen der Leistungen der Kinder. Ebenso kann die Reliabilität zu beiden Zeitpunkten mit $\alpha = .849$ und $\alpha = .864$ als gut bezeichnet werden. Ein vergleichbarer Wert der Gesamtskala der kognitiven Entwicklung wird bei Tscherne (2014) nicht berichtet.

Zur Interpretation sollte aufgrund der Ergebnisse dieser Arbeit unbedingt die Gesamtskala der kognitiven Entwicklung herangezogen werden, die sowohl bei der Retest-Reliabilität als auch bei der internen Konsistenz durchwegs gute Werte von $>.8$ aufweist. Auf eine Interpretation der einzelnen Skalen sollte gegenwärtig verzichtet werden.

7.4 Diskussion der Testdauer

Im Vergleich zu Tscherne (2014), die denselben Itempool vorgab, und eine durchschnittliche Testdauer von 89.6 Minuten ($SD = 10.6$) beschreibt, zeigt sich in dieser Untersuchung eine etwas längere Testdauer von durchschnittlich 107 Minuten ($SD = 15.9$). Dabei dauerte die kürzeste Ersttestung 79 Minuten und die längste 133 Minuten, verglichen mit 74 Minuten bzw. 120 Minuten bei Tscherne (2014). Ein Teil der Unterschiede kann auf die Jahreszeit der Testung zurückgeführt werden, da die Zeit ab dem Betreten des Testraumes gezählt wird und die Zeit für das Aus- und Anziehen der Kinder berücksichtigt werden muss, was bei Tscherne

(2014) nicht der Fall war. Dies zeigt sich auch in einer längeren Dauer der Anlaufzeit und Verabschiedung von durchschnittlich 3.05 bzw. 13.24 Minuten bei der aktuellen Untersuchung im Vergleich zu 1.78 bzw. 11.95 Minuten bei Tscherne (2014). Da die Materialien und Rahmenbedingungen sonst ident sind, lassen sich die restlichen Unterschiede auf Stichprobeneffekte zurückführen.

Die signifikant kürzere Spielzeit beim zweiten Termin mit durchschnittlich 102.57 Minuten ($SD = 12.432$) lässt sich durch die Vertrautheit der Kinder mit dem Ablauf und der Testleiterin erklären.

Nach Quaiser-Pohl (2010) sollte die Untersuchung bei Kleinkindern nicht länger als 30 bis 45 Minuten dauern. Diese Zeit wird mit dem aktuellen Verfahren weit überschritten. Eine Aufteilung des Verfahrens auf zwei Termine erscheint daher überlegenswert. Dafür spricht auch, dass die Reliabilität bei sechs der acht berechneten Skalen zum zweiten Zeitpunkt höher ausfällt. Nur bei den Skalen "aktive Mengenerfassung" und "Zählen" ist das Gegenteil zu beobachten, diese sind jedoch zu beiden Zeitpunkten entweder nicht zufriedenstellend oder gut. Im Gegensatz zu herkömmlichen Verfahren verfolgt diese Untersuchung einen spielbasierten Ansatz und das Spiel stellt für das Kind eine vertraute Situation dar, in der es zahlreiche Stunden am Tag verbringt (Kelly-Vance & Ryalls, 2005). Zudem sollen sich die Kinder aus Spaß und freiwillig an der Untersuchung beteiligen und von sich aus mit Spielmaterialien befassen (Sturzbecher, 2001). Dies konnte bei dieser Untersuchung beobachtet werden und trug zur Motivation der Kinder bei, die oft nachdem alle Items kodiert waren, die Spielsituation nicht verlassen wollten. Vor allem da den Kindern Pausen gewährt wurden, in denen sie essen und trinken konnten, ist die Untersuchungszeit nicht als zu lang anzusehen.

7.5 Limitation und Ausblick

Einschränkend ist zu erwähnen, dass der Anteil an AkademikerInnen bei den Eltern der getesteten Kinder mit 57.1 % bei den Vätern und 71.4 % bei den Mütter, weit über den

Angaben für die Gesamtbevölkerung liegt, der für Männer 14.7 % und für Frauen 16.7 % beträgt (Benedik et al., 2015). Auch die Geschlechterverteilung ist mit sieben Buben und 14 Mädchen nicht ausgeglichen und die Untersuchungsstichprobe mit nur 21 Kindern nicht ausreichend groß. Da die Reliabilitätsschätzungen stichprobenabhängig sind (Renner, 2009) sollten weitere Untersuchungen mit einer repräsentativen Stichprobe sowie eine Analyse über alle vorhandenen Daten angeschlossen werden.

Zudem ist die Durchführung des Verfahrens mit einer Testleiterin und einer Beobachterin und die anschließende Nachkodierung mithilfe der Videoaufzeichnung sehr zeitaufwändig und nicht ökonomisch. Die Videoaufzeichnung wurde zur Nachkodierung des kognitiven Bereichs nicht benötigt, jedoch ist sie für andere Funktionsbereiche wie die Aufmerksamkeit unbedingt erforderlich. Bei sehr guter Kenntnis der Spielmaterialien sowie des Protokollbogens und großer Vertrautheit mit der Diagnostik bei kleinen Kindern ist die Vorgabe und gleichzeitige Kodierung durch dieselbe Person möglich.

Neben der Testleiterin und der Beobachterin war während der Testungen immer eine Bezugsperson des Kindes anwesend und dies wirkt sich bei sehr jungen Kindern positiv auf die Untersuchung aus, da die Bezugsperson dem Kind Sicherheit gibt (Irblich & Renner, 2009). Es kann jedoch, obwohl die Items in das Spiel mit dem Kind eingebaut wurden, nicht ausgeschlossen werden, dass einige Bezugspersonen zwischen den beiden Testterminen mit dem Kind gewisse Items geübt haben. Dieser Übungseffekt bzw. Wissenszuwachs würde zu einer unsystematischen Veränderung des wahren Wertes bei einigen Kindern führen, was eine Verringerung der Retest-Reliabilität zur Folge hätte.

Vorhandene entwicklungsdiagnostische Verfahren für zweijährige Kinder verfolgen oft eine feste Reihenfolge bei der Vorgabe (Bayley III) oder geben für den Altersbereich der Zweijährigen (K-ABC) bzw. für das gesamte Verfahren (ET 6-6; MFED 2-3) keine Retest-Reliabilität an. Auch sind das K-ABC und der SON-R 2 1/2-7 nicht für die Anwendung des ganzen dritten Lebensjahres, sondern erst für Kinder ab 2;6 Jahren genormt. Als einziges

entwicklungsdiagnostisches Verfahren berichtet der Bayley III eine Retest-Reliabilität. Diese wurde mit einer Risikogruppe im siebten und 24. Lebensmonat erhoben und die Werte lagen für die Untertests zwischen $r_{tt} = .20$ und $r_{tt} = .46$ (Macha & Petermann, 2015). Aufgrund des langen Altersintervalls und des Alters sowie der Zusammensetzung der Stichprobe können diese Werte nicht mit den hier berechneten in Bezug gesetzt werden.

Die spielbasierten Ansätze gehen hingegen auf die Erfordernisse der Altersgruppe ein und haben ein flexibles Format (Kelly-Vance et al., 2002). Die Angaben bezüglich der Gütekriterien sind jedoch oft unzureichend, so wird beim TPBA-2 nur die Kriteriumsvalidität berichtet (O'Grady & Dusing, 2015). Für PIECES liegt nur eine Untersuchung für die Retest-Reliabilität vor (Kelly-Vance & Ryalls, 2005) und der AEPS:E berichtet nur die Interrater-Reliabilität und eine Kriteriumsvalidität (Macy et al., 2005). Der EPSI-IGDI stellt Werte für die Interrater-Reliabilität, die Retest-Reliabilität und die Kriteriumsvalidität dar (Greenwood et al., 2006). Die Ermittlung von weiteren psychometrischen Eigenschaften fehlt in der vorhandenen Literatur zum Spiel Assessment (Kelly-Vance & Ryalls, 2005).

Dieses Verfahren geht mit dem spielbasierten Zugang auf die Bedürfnisse von Zweijährigen ein und erlaubt mit der variablen Itemvorgabe die Berücksichtigung der Interessen des Kindes. Zudem weist das Verfahren mit den Ergebnissen der aktuellen Untersuchung gute Werte für die Retest-Reliabilität des kognitiven Bereichs auf. Dieser Entwicklungstest könnte unter Berücksichtigung der fehlenden Normierung und der Ergebnisse aus den anderen Funktionsbereichen in Zukunft als spielbasierter Zugang für die Erfassung der Entwicklung von zweijährigen Kindern eingesetzt werden.

8 Literaturverzeichnis

- Bagnato, S. J. (2005). The Authentic Alternative for Assessment in Early Intervention: An Emerging Evidence-Based Practice. *Journal of Early Intervention*, 28(1), 17–22. doi:10.1177/105381510502800102
- Benedik, O., Bönisch, M., Edelhofer, E., Gumpoldsberger, H., Lehner, D., Martinschitz, S., ... Wisbauer, A. (2015). *Bildung in Zahlen 2013/14 - Schlüsselindikatoren und Analysen*. Wien: Statistik Austria.
- 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*. Unveröffentlichte Diplomarbeit, Universität Wien.
- Bortz, J., & Döring, N. (2006). *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler* (4., überar.). Heidelberg: Springer.
- Bricker, D., Pretti-Frontczak, K., Johnson, J., & Straka, E. (2002). *Assessment, Evaluation, and Programming System for Infants and Children (AEPS), Second Edition*. Baltimore: Paul H. Brookes.
- Brock, R. L., Kochanska, G., O'Hara, M. W., & Grekin, R. S. (2015). Life Satisfaction Moderates the Effectiveness of a Play-Based Parenting Intervention in Low-Income Mothers and Toddlers. *Journal of Abnormal Child Psychology*, 43, 1283–1294. doi:10.1007/s10802-015-0014-y
- Burris, C. T., & Raif, K. (2015). Make Believe Unmakes Belief?: Childhood Play Style and Adult Personality as Predictors of Religious Identity Change. *The International Journal for the Psychology of Religion*, 25(2), 91–106. doi:10.1080/10508619.2014.916590

- Crawford, K. C., & Ogletree, B. T. (2010). Book Review. *Focus on Autism and Other Developmental Disabilities*, 25(1), 55–57. doi:10.1177/1088357609350368
- Deimann, P., & Kastner-Koller, U. (2007). Entwicklungsdiagnostik. In M. Hasselhorn & W. Schneider (Eds.), *Handbuch der Entwicklungspsychologie* (pp. 558–569). Göttingen: Hogrefe.
- Ebert, S. (2015). Longitudinal Relations Between Theory of Mind and Metacognition and the Impact of Language. *Journal of Cognition and Development*, 16(4), 559–586. doi:10.1080/15248372.2014.926272
- Edelmann, W., & Wittmann, S. (2012). *Lernpsychologie*. Weinheim: Beltz.
- Esser, G., & Petermann, F. (2010). *Entwicklungsdiagnostik*. Göttingen: Hogrefe.
- Ettrich, K. U. (2000). *Entwicklungsdiagnostik im Vorschulalter*. Göttingen: Vandenhoeck & Ruprecht.
- Field, A. (2009). *Discovering Statistics using SPSS* (Third Edit.). London: SAGE.
- Fink, E., Begeer, S., Peterson, C. C., Slaughter, V., & de Rosnay, M. (2015). Friendlessness and theory of mind: a prospective longitudinal study. *The British Journal of Developmental Psychology*, 33, 1–17. doi:10.1111/bjdp.12060
- Frye, D., Braisby, N., Lowe, J., Maroudas, C., & Nicholls, J. (1989). Young Children's Understanding of Counting and Cardinality. *Child Development*, 60(5), 1158–1171.
- Fuchs-Gaderer, M. (2012). *Die Überarbeitung und testtheoretische Überprüfung eines spielbasierten Itempools zur Erfassung der Entwicklung im dritten Lebensjahr*. Unveröffentlichte Diplomarbeit, Universität Wien.
- Gelman, R. (1993). A rational-constructivistic account of early learning about numbers and

- objects. In D. L. Medin (Ed.), *The Psychology of Learning and Motivation* (Volume 30., pp. 61–96). London: Academic Press.
- Gelman, R., & Gallistel, C. R. (1986). *The Child's Understanding of Number*. Cambridge: Harvard University Press.
- Greenwood, C. R., Walker, D., Carta, J. J., & Higgins, S. K. (2006). Developing a General Outcome Measure of Growth in the Cognitive Abilities of Children 1 to 4 Years Old: The Early Problem-Solving Indicator. *School Psychology Review*, 35(4), 535–551.
- Hasselhorn, M., & Margraf-Stiksrud, J. (2015). TBS-TK Rezension: Entwicklungstest für Kinder von sechs Monaten bis sechs Jahren - Revision (ET 6-6 R). *Psychologische Rundschau*, 66, 208–210.
- Hellbrügge, T. (1994). *Münchener Funktionelle Entwicklungsdiagnostik, zweites und drittes Lebensjahr (MFED 2-3)*. München: Universität München, Institut für soziale Pädiatrie und Jugendmedizin.
- Hellbrügge, T. (2001). *Münchener Funktionelle Entwicklungsdiagnostik, erstes Lebensjahr (MFED 1)*. Lübeck: Hansisches Verlagskontor.
- Hiller, R. M., Weber, N., & Young, R. L. (2014). The Validity and Scalability of the Theory of Mind Scale With Toddlers and Preschoolers. *Psychological Assessment*, 26(4), 1388–1393.
- Hopkins, E. J., Dore, R. A., & Lillard, A. S. (2015). Do children learn from pretense? *Journal of Experimental Child Psychology*, 130, 1–18. doi:10.1016/j.jecp.2014.09.004
- Irblich, D. (2015). Neuere Testverfahren. KABC-II. *Praxis der Kinderpsychologie und Kinderpsychiatrie*, 64, 635–648.
- Irblich, D., & Renner, G. (2009). Wie untersucht man Kinder? In D. Irblich & G. Renner

- (Eds.), *Diagnostik in der Klinischen Kinderpsychologie* (pp. 21–32). Göttingen: Hogrefe.
- Jungwirth, C. (2015). *Testtheoretische Analyse eines spielbasierten Entwicklungsverfahrens für Zweijährige unter besonderer Berücksichtigung der Sprachentwicklung und der Aufmerksamkeit*. Unveröffentlichte Masterarbeit, Universität Wien.
- Kastner-Koller, U., & Deimann, P. (2009). Beobachtung und Befragung von Kindern. In D. Irblich & G. Renner (Eds.), *Diagnostik in der Klinischen Kinderpsychologie* (pp. 97–107). Göttingen: Hogrefe.
- Kastner-Koller, U., & Deimann, P. (2011). Entwicklungstests. In L. Horncke, M. Amelang, & M. Kersting (Eds.), *Enzyklopädie der Psychologie. Leistungs-, Intelligenz- und Verhaltensdiagnostik* (pp. 275–304). Göttingen: Hogrefe.
- Kelly-Vance, L., Needelman, H., Troia, K., & Ryalls, B. O. (1999). Early Childhood Assessment: A Comparison of the Bayley Scales of Infant Development and Play-Based Assessment in Two-Year Old At-Risk Children. *Developmental Disabilities Bulletin*, 27(1), 15.
- Kelly-Vance, L., & Ryalls, B. O. (2005). A Systematic, Reliable Approach to Play Assessment in Preschoolers. *School Psychology International*, 26(4), 398–412. doi:10.1177/0143034305059017
- Kelly-Vance, L., Ryalls, B. O., & Gill Glover, K. (2002). The Use of Play Assessment to Evaluate the Cognitive Skills of Two- and Three-Year-Old Children. *Journal of Composite Materials*, 23(2), 169–185. doi:0803973233
- Kronberger, M. V. (2013). *Die Überarbeitung und testtheoretische Überprüfung eines spielbasierten Itempools zur Erfassung der Entwicklung zweijähriger Kinder unter besonderer Berücksichtigung der visuellen Wahrnehmung*. Unveröffentlichte

Diplomarbeit, Universität Wien.

- Kubinger, K. D. (2009). *Psychologische Diagnostik: Theorie und Praxis psychologischen Diagnostizierens* (2., überar.). Göttingen: Hogrefe.
- Kuchler, M., Sapper, E., Deimann, P., & Kastner-Koller, U. (2011). *Manual zum Itempool zur spielbasierten Erfassung der Entwicklung Zweijähriger*. Universität Wien: unveröffentlichtes Manuskript.
- Ladd, G. W., & Troop-Gordon, W. (2003). The Role of Chronic Peer Difficulties in the Development of Children's Psychological Adjustment Problems. *Child Development*, 74(5), 1344–1367. doi:10.1111/1467-8624.00611
- 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. doi:10.1016/j.cognition.2006.10.005
- Le Corre, M., Van de Walle, G., Brannon, E., & Carey, S. (2006). Re-visiting the competence/performance debate in the acquisition of the counting principles. *Cognitive Psychology*, 52(2), 130–169. doi:10.1016/j.cogpsych.2005.07.002
- Lillard, A. S. (1993). Pretend Play Skills and the Child's Theory of Mind. *Child Development*, 64(2), 348–371.
- 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. doi:10.1037/a0029321
- Linás, K. E. (2009). *Concurrent Validity of the Transdisciplinary Play Based Assessment-2*. Dissertation, University of Denver.
- Linder, T. (1993). *Transdisciplinary play-based assessment: A functional approach to*

- working with young children*. Baltimore: Paul H. Brookes.
- Linder, T., & Bixby, B. (2010). Transdisciplinary Play-Based Assessment and Intervention in Primary Years. In A. A. Drewes & C. E. Schaefer (Eds.), *School-based Play Therapy* (Second Edi., pp. 123–144). New Jersey: John Wiley & Sons.
- Linder, T. W., Anthony, T. L., Bundy, A. C., Charlifue-Smith, R., Hafer, J. C., Hancock, F., & Rooke, C. C. (2008). *Transdisciplinary play-based assessment* (second edi.). Baltimore: Paul H. Brookes.
- Liszt, S. (2014). *Überprüfung der Testgütekriterien eines spielbasierten Inventars für Zweibis Dreijährige unter besonderer Berücksichtigung der Bereiche: Motorik, Sprache und sozial-emotionale Entwicklung*. Unveröffentlichte Diplomarbeit, Universität Wien.
- Lohaus, A., & Vierhaus, M. (2015). Kognition. In A. Lohaus & M. Vierhaus (Eds.), *Entwicklungspsychologie des Kindes- und Jugendalters für Bachelor* (pp. 116–130).
- Lowenthal, B. (1997). Useful early childhood assessment: Play-based, interviews and multiple intelligences. *Early Child Development and Care*, 129, 43–49.
- Macha, T., & Petermann, F. (2015). Bayley Scales of Infant and Toddler Development, Third Edition – Deutsche Fassung. *Zeitschrift für Psychiatrie, Psychologie und Psychotherapie*, 63(2), 139–143. doi:10.1024/1661-4747/a000232
- Macha, T., Proske, A., & Petermann, F. (2005). Validität von Entwicklungstests. *Kindheit und Entwicklung*, 14(3), 150–162. doi:10.1026/0942-5403.14.3.150; 10.1026/0942-5403.14.3.150
- Macy, M. G., Bricker, D. D., & Squires, J. K. (2005). Validity and Reliability of a Curriculum-Based Assessment Approach to Determine Eligibility for Part C Services. *Journal of Early Intervention*, 28(1), 1–16. doi:10.1177/105381510502800101

Melchers, P., & Melchers, M. (2015). *KABC-II. Kaufman Assessment Battery for Children – II*. Frankfurt: Pearson.

Melchers, P., & Preuß, U. (2009). *K-ABC: Kaufman – Assessment Battery for Children* (8., unverä.). Frankfurt: Pearson.

Miller, K. F., Smith, C. M., Zhu, J., & Zhang, H. (1995). Preschool Origins of Cross-National Differences in Mathematical Competence: The Role of Number-Naming Systems. *Psychological Science*, 6(1), 56–60. doi:10.1111/j.1467-9280.1995.tb00305.x

Mussolin, C., Nys, J., Leybaert, J., & Content, A. (2012). Relationships between approximate number system acuity and early symbolic number abilities. *Trends in Neuroscience and Education*, 1(1), 21–31. doi:10.1016/j.tine.2012.09.003

Negen, J., & Sarnecka, B. W. (2015). Is there really a link between exact-number knowledge and approximate number system acuity in young children? *The British Journal of Developmental Psychology*, 33(1), 92–105. doi:10.1111/bjdp.12071

O’Grady, M. G., & Dusing, S. C. (2015). Reliability and Validity of Play-Based Assessments of Motor and Cognitive Skills for Infants and Young Children: A Systematic Review. *Physical Therapy*, 95(1), 25–38. doi:10.2522/ptj.20140111

Odic, D., Le Corre, M., & Halberda, J. (2015). Children’s mappings between number words and the approximate number system. *Cognition*, 138, 102–121. doi:10.1016/j.cognition.2015.01.008

Petermann, F., & Macha, T. (2003). Strategien in der testgestützten allgemeinen Entwicklungsdiagnostik. *Monatsschrift für Kinderheilkunde*, 151(1), 6–13. doi:10.1007/s00112-002-0621-x

Petermann, F., & Macha, T. (2005). Entwicklungsdiagnostik. *Kindheit und Entwicklung*,

14(3), 131–139. doi:10.1026/0942-5403.14.3.131

Petermann, F., & Macha, T. (2013). *Entwicklungstest für Kinder von sechs Monaten bis sechs Jahren (ET 6 – 6-R)*. Frankfurt: Pearson.

Pozniak, M. (in Vorbereitung). *Die Retest-Reliabilität eines spielbasierten Entwicklungstests für zweijährige Kinder unter besonderer Berücksichtigung des Gedächtnisses und der visuellen Wahrnehmung*. Unveröffentlichte Masterarbeit, Universität Wien.

Punz, S. A. (2013). *Die Überarbeitung und testtheoretische Überprüfung eines spielebasierten Itempools zur Erfassung der Entwicklung im dritten Lebensjahr*. Unveröffentlichte Diplomarbeit, Universität Wien.

Putzer, S. (2013). *Die Entwicklung und Erprobung eines Itempools zur spielbasierten Erfassung der Entwicklung Zweijähriger mit besonderer Berücksichtigung der Grobmotorik*. Unveröffentlichte Diplomarbeit, Universität Wien.

Quaiser-Pohl, C. (2010). Wie teste ich Kinder: Zur Gestaltung der diagnostischen Situation. In C. Quaiser-Pohl & H. Rindermann (Eds.), *Entwicklungsdiagnostik* (pp. 57–68). München: Ernst Reinhardt.

Renner, G. (2009). Testpsychologische Diagnostik bei Kindern. In D. Irblich & G. Renner (Eds.), *Diagnostik in der Klinischen Kinderpsychologie* (pp. 73–85). Göttingen: Hogrefe.

Reuner, G., & Rosenkranz, J. (2014). *Bayley-Scales of Infant and Toddler Development, Third Edition – Deutsche Version*. Frankfurt: Pearson Assessment.

Rollett, B., & Preckel, F. (2011). TBS-TK Rezension: K-ABC: Kaufman - Assessment Battery for Children. *Psychologische Rundschau*, 63, 139–141. doi:10.1026/0932-4089/a000188

Sarnecka, B. W., & Carey, S. (2008). How counting represents number: What children must

learn and when they learn it. *Cognition*, 108(3), 662–674.
doi:10.1016/j.cognition.2008.05.007

Schermelleh-Engel, K., & Werner, C. S. (2012). Methoden der Reliabilitätsbestimmung. In H. Moosbrugger & A. Kelava (Eds.), *Testtheorie und Fragebogenkonstruktion* (2., aktual., pp. 119–141). Heidelberg: Springer.

Sindelar, I. (2013). *Die Entwicklung und Erprobung eines Itempools zur spielbasierten Erfassung der Entwicklung Zweijähriger unter besonderer Berücksichtigung der Feinmotorik und der Visuomotorik*. Unveröffentlichte Diplomarbeit Universität Wien.

Singer, D. G., & Singer, J. L. (1992). *The House of Make-Believe: Children's Play and the Developing Imagination*. Cambridge: Harvard University Press.

Smith, P. K. (2010). *Children and Play*. West-Sussex: Wiley-Blackwell.

Sodian, B. (2008). Entwicklung des Denkens. In R. Oerter & L. Montada (Eds.), *Entwicklungspsychologie* (6., vollst., pp. 436–479). Weinheim: Beltz.

Stark, M. (in Vorbereitung). *Retest-Reliabilität eines spielbasierten Entwicklungstests für zweijährige Kinder unter besonderer Berücksichtigung der Motorik und sozial-emotionaler Entwicklung*. Unveröffentlichte Masterarbeit, Universität Wien.

Sturzbecher, D. (2001). Methodische Lösungsansätze zur Befragung jüngerer Kinder. In D. Sturzbecher (Ed.), *Spielbasierte Befragungstechniken* (pp. 51–63). Göttingen: Hogrefe.

Taylor, M., & Carlson, S. M. (1997). The Relation between Individual Differences in Fantasy and Theory of Mind. *Child Development*, 68(3), 436–455. doi:10.1111/j.1467-8624.1997.tb01950.x

Tellegen, P. J., Laros, J. A., & Petermann, F. (2007). *Non-verbaler Intelligenztest (SON-R 2½-7)*. Göttingen: Hogrefe.

- Tscherne, N. (2014). *Überprüfung der Testgütekriterien eines spielbasierten Inventars für zwei- bis dreijährige Kinder*. Unveröffentlichte Diplomarbeit, Universität Wien.
- Vock, M. (2008). Testinformation. Non-verbaler Intelligenztest. *Diagnostica*, 54(2), 112–115.
- Wagner, J. B., & Johnson, S. C. (2011). An association between understanding cardinality and analog magnitude representations in preschoolers. *Cognition*, 119(1), 10–22. doi:10.1016/j.cognition.2010.11.014
- Weisberg, D. S. (2015). Pretend play. *Wiley Interdisciplinary Reviews: Cognitive Science*, 6(3), 249–261. doi:10.1002/wcs.1341
- Wellman, H. M. (2014). *Making Minds: How Theory of Mind develops*. New York: Oxford University Press.
- Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-analysis of theory-of-mind development: the truth about false belief. *Child Development*, 72(3), 655–684. doi:10.1111/1467-8624.00304
- Wellman, H. M., & Liu, D. (2004). Scaling of theory-of-mind tasks. *Child Development*, 75(2), 523–541. doi:10.1111/j.1467-8624.2004.00691.x
- Wilkes-Gillan, S., Bundy, A., Cordier, R., & Lincoln, M. (2014). Eighteen-month follow-up of a play-based intervention to improve the social play skills of children with attention deficit hyperactivity disorder. *Australian Occupational Therapy Journal*, 61(5), 299–307. doi:10.1111/1440-1630.12124
- Wyl, A. Von. (2014). Mentalisierung und Theory of Mind. *Praxis der Kinderpsychologie und Kinderpsychiatrie*, 63, 730–737.

9 Tabellenverzeichnis

Tabelle 1:	Alter in Monaten	41
Tabelle 2:	Ergebnisse der Itemanalyse der Gesamtskala kognitive Entwicklung beim ersten und zweiten Testzeitpunkt	45
Tabelle 3:	Ergebnisse der Itemanalyse zur Skala Hineinversetzen in Andere beim ersten und zweiten Testzeitpunkt	48
Tabelle 4:	Fehlende Werte der Skala Hineinversetzen in Andere	48
Tabelle 5:	Ergebnisse der Itemanalyse zur Skala Zählen beim ersten und zweiten Testzeitpunkt.....	50
Tabelle 6:	Ergebnisse der Itemanalyse zur Skala aktive Mengenerfassung beim ersten und zweiten Testzeitpunkt	51
Tabelle 7:	Fehlende Werte der Skala aktive Mengenerfassung.....	51
Tabelle 8:	Ergebnisse der Itemanalyse zur Skala passive Mengenerfassung beim ersten und zweiten Testzeitpunkt	52
Tabelle 9:	Fehlende Werte der Skala passive Mengenerfassung	53
Tabelle 10:	Ergebnisse der Itemanalyse zur Skala Mengenerfassung beim ersten und zweiten Testzeitpunkt.....	54

10 Anhang

10.1 Anhang A: Berechnung der standardisierten Schiefe für die deskriptiven Angaben

Skala	Schiefe	Standardfehler	Standardisierte Schiefe
gesamte Testdauer	-1.619	0.501	-3.232
Anlaufzeit	-0.559	0.501	-1.116
Beschäftigung mit Materialien	-0.332	0.501	-0.663
Pausendauer	-1.280	0.501	-2.555
Verabschiedung	-0.367	0.501	-0.733

10.2 Anhang B: Berechnung der standardisierten Schiefe für den kognitiven Bereich

Skala	Schiefe	Standardfehler	Standardisierte Schiefe
Hineinversetzen	0.130	0.501	0.259
AOS	-0.561	0.501	-1.120
Theorie of Mind	0.509	0.501	1.016
aktive Mengenerfassung	0.130	0.501	0.259
passive Mengenerfassung	0.773	0.501	1.543
Mengenerfassung	0.095	0.501	0.190
Zählen	0.834	0.501	1.665
numerisches Wissen	0.610	0.501	1.218
Kognitive Gesamtskala	0.543	0.501	1.084

10.3 Anhang C: Übersicht Retest-Reliabilität der Skalen und interne Konsistenz

Skala	Retest-Reliabilität	1. Testzeitpunkt Cronbachs Alpha	2. Testzeitpunkt Cronbachs Alpha
Gesamtskala kognitive Entwicklung	.820, $p < .001$.849	.864
Theorie of Mind	.887, $p < .001$.673	.709
Als-ob Spiel	.596, $p = .004$		
Hineinversetzen in andere	.901, $p < .001$.680	.717
numerisches Wissen	.677, $p = .001$.848	.871
aktive Mengenerfassung	.719, $p < .001$.659	.553
passive Mengenerfassung	.468, $p = .032$.421	.505
Mengenerfassung	.749, $p < .001$.542	.690
Zählen	.595, $p = .004$.928	.871

10.4 Anhang D: Itemanalyse der Skala Theorie of Mind

Item	1. Zeitpunkt		2. Zeitpunkt	
	Item- schwierigkeit	korrigierte Trennschärfe	Item- schwierigkeit	korrigierte Trennschärfe
Als-Ob-Spiel	.571	.222	.548	.294
schmutzige Hände	.524	.445	.619	.494
hungrig	.476	.443	.476	.586
schlafen	.095	.189	.333	.585
Pflaster	.524	.403	.571	.325
durstig	.286	.722	.476	.428

10.5 Anhang E: Itemanalyse der Skala numerisches Wissen

Item	1. Zeitpunkt		2. Zeitpunkt	
	Item- schwierigkeit	korrigierte Trennschärfe	Item- schwierigkeit	korrigierte Trennschärfe
Wie viele sind das - zwei	.19	.377	.19	.729
Wie viele sind das - drei	.14	.719	.19	.456
Gib mir 1	.90	.177		
Gib mir 2	.48	.227	.67	.226
Gib mir 3	.33	.081	.19	.590
zählen	.57	.529	.67	.597
richtig zählen	.24	.863	.38	.731
Eins-zu-eins Zuordnung	.33	.701	.57	.455
stabile Reihenfolge	.38	.669	.48	.804
richtige Reihenfolge	.24	.863	.52	.678
Kardinalprinzip	.19	.808	.33	.690

10.6 Anhang F: Elternbrief



Liebe Eltern aller zwei- bis dreijährigen Kinder!

Im Rahmen unserer Masterarbeiten sind wir, vier Psychologiestudentinnen der Universität Wien, auf der Suche nach Eltern mit Kindern im Alter von 24 bis 36 Monaten, die uns bei unserer Forschung unterstützen können.

Worum es geht:

Am Institut für Angewandte Psychologie: Gesundheit, Entwicklung und Förderung der Universität Wien läuft unter der Leitung von Dr. Deimann und Dr. Kastner-Koller ein Projekt mit dem Ziel, ein spielbasiertes Verfahren zur Erfassung der Entwicklung zwei- bis dreijähriger Kinder zu entwickeln. Zurzeit sammeln wir Informationen über die Genauigkeit dieses neuen Verfahrens.

Was passiert?

In einer angenehmen Atmosphäre spielen, lesen, turnen und lachen wir an zwei Terminen für circa ein bis zwei Stunden mit Ihrem Kind. Sie sind herzlich eingeladen, bei der Spielsituation dabei zu sein.

Wo?

Die Spielsituation findet im Spielzimmer des Instituts für Angewandte Psychologie der Universität Wien statt (Liebiggasse 5, 1010 Wien).

Wann?

Wir würden gerne für Dezember oder Jänner den ersten Termin vereinbaren und einen zweiten Termin vier bis sechs Wochen später.

Welchen Vorteil haben Sie davon?

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

So erreichen Sie uns:

Wenn Sie Interesse an zwei spannenden Vormittagen oder Nachmittagen haben und zusätzlich noch einen Beitrag für unser Projekt leisten wollen, dann freuen wir uns sehr Sie und Ihr Kind kennenzulernen. Kontaktieren Sie uns bitte so bald wie möglich, entweder telefonisch oder per E-Mail, damit uns genügend Zeit zur Planung und Koordination der Termine bleibt.

Für alle Fragen stehen wir Ihnen jederzeit gerne zur Verfügung.

Christina Jungwirth, BSc
0664 5409252

Maria Pozniak, BSc
0650 9906412

Karin Prillinger, BSc
0664 1582930

Magdalena Stark, BSc
0664 1662251

anmeldung-masterarbeit@gmx.at

10.7 Anhang G: Elternfragebogen

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!

10.8 Anhang H: Einverständniserklärung für die Videoauszeichnung

EINVERSTÄNDNISERKLÄRUNG

Liebe Eltern,

da es uns während des Spiels mit Ihrem Kind nicht möglich ist, alles zu erfassen, was für die Entwicklung unseres Verfahrens relevant ist, bitten wir Sie um Ihr Einverständnis, die Spielsituationen mit Ihrem Kind auf Video aufzuzeichnen.

Die Aufzeichnungen werden im Rahmen der Masterarbeiten ausschließlich von den unten angeführten Personen, sowie den Betreuerinnen der Masterarbeit Ass. Prof. Dr. Pia Deimann und Ass. Prof. Dr. Ursula Kastner-Koller, zu Analysezwecken verwendet. Das Videomaterial wird keinesfalls an Dritte weitergegeben und nach Beendigung der Masterarbeiten gelöscht.

Vielen Dank,

Christina Jungwirth, Maria Pozniak, Karin Prillinger & Magdalena Stark

Ich, _____, erkläre mich damit einverstanden, dass meine Tochter / mein Sohn _____, geboren am _____ im Rahmen der Mitwirkung an der Masterarbeitsstudie von Frau Jungwirth, Frau Pozniak, Frau Prillinger und Frau Stark auf Video aufgezeichnet werden darf.

Datum

Unterschrift

✓ Gelöst	X nicht gelöst
v. verweigert	- nicht vorgegeben

10.9 Anhang I: Protokollbogen

Name des Testkinds: _____

Name des Testleiters: _____

Datum: _____

Uhrzeit (Anfang/Ende): _____

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 Farbkonzept Min 3	Rot		
	Grün		
	Gelb		
	orange		

Geburtsdatum: _____

Alter: _____

Anwesende Personen: _____

Kontaktaufnahme zum Kind: _____

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

GEDÄCHTNIS - PUPPENKÜCHE

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

Wortschatz aktiv & Farbdifferenzierung 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		
	Aufsetzen 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 (2 Meter 0/1)

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

GROBMOTORIK STEHEN, HÜPFEN, LAUFEN (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)

VISUMOTORIK – Formen

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)			

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 Blume		
	Memory Haus		
	Memory Bär		
	Memory Ball		
	Teddybär		
	Haus		

FARB-, FORM - UND GRÖßENDIFF. (BODENMATTEN)

	Form- und Größendifferenzierung		
	Formdiff. 1 (Kreis, Quadrat, Dreieck, Rechteck) → rot		
	Formdiff. 2 (Vierecke) → grün		
	Größendiff. 1 (Rechtecke) → gelb		
	Größendiff. 2 (Kreise) → orange		

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		

Protokoll der vom Kind getätigten Aussagen:

