



universität
wien

MASTERARBEIT

Titel der Masterarbeit

Testtheoretische Analyse eines spielbasierten Entwicklungsverfahrens für Zweijährige

unter besonderer Berücksichtigung der Sprachentwicklung und der
Aufmerksamkeit

Verfasserin

Christina Jungwirth, BSc

Angestrebter akademischer Grad

Master of Science (MSc)

Wien, 2015

Studienkennzahl lt. Studienblatt:

A 066 840

Studienrichtung lt. Studienblatt:

Masterstudium Psychologie

Betreuerin / Betreuer:

Ass.-Prof. Dr. Ursula Kastner-Koller

Für meinen Vater Alois Jungwirth

Danksagung

„Das nächste Ziel mit Lust und Freude und aller Kraft zu verfolgen ist der einzige Weg, das Fernste zu erreichen.“ (Friedrich Hebbel)

Nach fünf Jahren Studium bin ich meinem beruflichen Ziel ein ganzes Stück näher gekommen. Ich blicke auf eine sehr spannende und lehrreiche Zeit zurück, die aber aufgrund meiner Berufstätigkeit neben dem Studium ungemein fordernd war. An dieser Stelle möchte ich allen Menschen danken, die mich in dieser Zeit unterstützt, ermutigt und begleitet haben.

Zuerst möchte ich meinen Professorinnen Dr. Pia Deimann und Dr. Ursula Kastner-Koller für die Möglichkeit der Teilhabe an Ihrem Forschungsprojekt danken. Im Zuge der Testungen konnte ich viele wertvolle Erfahrungen sammeln. Neben den fachlichen Anregungen danke ich ihnen für ihr Verständnis, wenn ich Termine aufgrund meiner Berufstätigkeit nicht wahrnehmen konnte.

Bei meinen Mitstreiterinnen Maria Pozniak, Karin Prillinger und Magdalena Stark möchte ich mich für die gute Zusammenarbeit, den konstruktiven Austausch, aber auch für die Freundschaft, die im Zuge unserer gemeinsamen Testungen entstanden ist, danken.

Ferner danke ich meiner Studienkollegin und guten Freundin Kerstin Huber für die wertvollen Tipps, den fachlichen Austausch und die vielen motivierenden Worte. Auch meinen Studienkollegen und Freunden Simon Lentner und Christoph Naturski sei gedankt für die lustigen Stunden und unvergesslichen Momente. Eure Freundschaft hat meine Studienzeit sehr bereichert.

Besonderer Dank gebührt meiner Kollegin und Mentorin Ruth Mutz. Sie hatte immer ein offenes Ohr für mich und stets einen guten Ratschlag und aufmunternde Worte parat. Egal ob bei beruflichen, universitären oder privaten Sorgen, ich konnte immer auf sie zählen.

Meiner Familie danke ich für den Halt den sie mir gab, den Zuspruch und die Unterstützung.

Unverzichtbarer Begleiter auf meinem Weg ist mein Lebensgefährte Johannes, der mich vorbehaltlos unterstützt. Ihm verdanke ich meine Energie und Lebensfreude. An seiner Seite bin ich für alles gewappnet und kann mich den kommenden Herausforderungen stellen. Seine Geduld, seine Ratschläge und wertvollen Tipps haben wesentlich zum Gelingen dieser Arbeit beigetragen.

Nicht zuletzt danke ich all den Kindern und Familien, die an den Testungen teilgenommen haben. Erst mit ihrer Hilfe konnte diese Studie verwirklicht werden. Namentlich sei an dieser Stelle meine liebe Nichte Fabienne erwähnt, die ebenso das Verfahren erprobte.

Meine Masterarbeit widme ich meinem Vater, der mir ein großes Vorbild ist und auf dessen Unterstützung ich immer zählen kann. Seinen Ehrgeiz und Arbeitseifer hat er mir vorgelebt und an mich weitergegeben. Er ermöglichte mir in jungen Jahren die Ausbildung zur Kindergartenpädagogin, den Grundstein meiner beruflichen Laufbahn.

Wien, im Juli 2015

Christina Jungwirth

Inhaltsverzeichnis

1. THEORETISCHER HINTERGRUND.....	1
1.1. ENTWICKLUNGSDIAGNOSTIK	1
1.2. ENTWICKLUNGSDIAGNOSTIK IM KLEINKINDALTER	4
1.3. SPIELBASIERTE DIAGNOSTIK.....	8
1.4. MEILENSTEINE IN DER ENTWICKLUNG ZWEIJÄHRIGER	11
1.4.1. <i>Entwicklung der Sprache</i>	12
1.4.2. <i>Entwicklung der Aufmerksamkeit</i>	15
1.5. QUALITÄTSANFORDERUNGEN AN ENTWICKLUNGSTESTS.....	17
1.5.1. <i>Objektivität</i>	18
1.5.2. <i>Validität</i>	18
1.5.3. <i>Reliabilität</i>	19
2. KONTEXT UND ZIEL DER VORLIEGENDEN MASTERARBEIT	23
3. FRAGESTELLUNGEN	25
4. METHODE.....	27
4.1. VORGEHEN	27
4.2. SETTING	28
4.3. STICHPROBE	28
4.4. ERHEBUNGSTRUMENT	30
4.4.1. <i>Erfassung der sprachlichen Entwicklung</i>	31
4.4.2. <i>Erfassung der Aufmerksamkeit</i>	32
5. ERGEBNISDARSTELLUNG	33
5.1. ERGEBNISSE DER DESKRIPTIVEN ANGABEN.....	33
5.2. ANALYSE FEHLENDER WERTE	35
5.3. RETEST-RELIABILITÄTEN DER BEIDEN MESSZEITPUNKTE	36
5.4. VERGLEICH DER BEIDEN MESSZEITPUNKTE.....	37
5.5. SKALEN- UND ITEMANALYSE	38
5.5.1. <i>Fähigkeitsbereich Sprache</i>	38
5.5.2. <i>Fähigkeitsbereich Aufmerksamkeit</i>	44
6. DISKUSSION.....	47
LITERATURVERZEICHNIS.....	59
TABELLENVERZEICHNIS	65
ANHANG	I
ANHANG A. ELTERNBRIEF	I

ANHANG B. EINVERSTÄNDNISERKLÄRUNG	III
ANHANG C. PROTOKOLLBogen	V
ANHANG D. ELTERNFRAGEBOGEN	XIII
ANHANG E. ÜBERPRÜFUNG DER NORMALVERTEILUNG	XVII
<i>Anhang E.1. Prüfung der Messwertdifferenzen deskriptiver Angaben.....</i>	<i>XVII</i>
<i>Anhang E.2. Prüfung der Messwertdifferenzen der Skala „Sprache“</i>	<i>XVII</i>
<i>Anhang E.3. Prüfung der Messwertdifferenzen der Skala „Aufmerksamkeit“.....</i>	<i>XVII</i>
ANHANG F. ZUSAMMENHANG UND UNTERSCHIEDE DER ANLAUF-, VERABSCHIEDUNGSZEIT UND DER PAUSENDAUER	XIX
<i>Anhang F.1. t-Test für abhängige Stichproben</i>	<i>XIX</i>
<i>Anhang F.2. Spearman – Rangkorrelation</i>	<i>XIX</i>
<i>Anhang F.3. Pearson – Korrelation</i>	<i>XIX</i>
<i>Anhang F.4. Wilcoxon – Vorzeichen – Rangtest.....</i>	<i>XIX</i>
ANHANG G. CURRICULUM VITAE	XXI

Zusammenfassung

Psychologische Diagnostik im Kleinkindalter erfordert eine besondere Herangehensweise und stellt erhöhte Anforderungen an die Durchführung. Kleinkinder verfügen über eine kurze Aufmerksamkeitsspanne, ihr Verhalten ist wenig vorhersehbar und ihre Kooperationsbereitschaft sprunghaft. Es gibt kaum Verfahren für Kinder im zweiten Lebensjahr, die den Bedürfnissen eines Kleinkindes gerecht werden. Bestehende Entwicklungstests erfüllen die Testgütekriterien oft unzureichend und weisen einige Limitationen auf. Als Alternative zu herkömmlichen Verfahren ist die spielbasierte Diagnostik, die einen spielerischen Zugang zur Testung verfolgt, entstanden. Das Spielinteresse des Kindes wird dabei ausreichend berücksichtigt, um die höchste Leistung des Kindes zu entfesseln.

Ziel dieser Arbeit ist die Erprobung und testtheoretische Untersuchung eines neu entwickelten Itempools zur spielbasierten Diagnostik zweijähriger Kinder. Der Itempool wird derselben Stichprobe zweimal vorgegeben, um Aussagen zur Retest-Reliabilität zu gewinnen. Dabei werden Unterschiede und Zusammenhänge zwischen den Ergebnissen der beiden Messzeitpunkte ermittelt. Im Zuge der Überprüfung werden die Aufgaben einer Skalen- und Itemanalyse unterzogen.

Die Ergebnisse zeigen, dass die Skala zur Erfassung der sprachlichen Kompetenzen die methodischen Anforderungen weitgehend erfüllt. Die Kennwerte der Skala zur Erfassung der Aufmerksamkeitsfähigkeit fallen wenig zufriedenstellend aus. Es bedarf weiterer testtheoretischer Analysen und gegebenenfalls einer Modifikation dieser Skala.

Das entwickelte Verfahren besticht durch seine vorzügliche praktische Anwendbarkeit und den hohen Aufforderungscharakter der Materialien.

Schlüsselwörter: Entwicklungsdiagnostik, Entwicklungstest, frühe Kindheit, spielbasierte Diagnostik, Testgüte

Abstract

Psychological assessment of toddlers requires a particular approach and makes great demands regarding performing the assessment. Toddlers' attention span is very short, their behaviour is only foreseeable to a small extent and their willingness to cooperate is rather volatile. There is only a limited number of assessments tailored to the needs of 2-year-old toddlers. These existing assessment cannot fulfil the criteria for test quality and have some limitations. As an alternative to common assessments, play-based assessment – which takes a playful approach – has been developed and adequately takes the child's interest in playing into consideration, thus unleashing their highest potential.

The aim of this work is the test-theoretical evaluation of a newly-created item pool for the play-based-assessment of children at the age of two. In order to gain reliable propositions about the retest reliability of the item pool, the same sample is tested twice. This allows for capturing the differences and connections between the results of each measurement, performed at different points in time. During the assessment, each exercise was scrutinized by both, a scaling and an item analysis.

Results show that the scale regarding language competence fulfils the methodological requirements. Contrary to this, the characteristic values of the scale of attention are not satisfactory. There is an urgent need for further test-theoretical analyses which may lead to a modification of the scale.

The newly developed assessment shows an outstanding practical applicability and the used materials are very appealing for the tested children.

Key words: developmental assessment, developmental test, early childhood, play-based assessment, test quality

1. Theoretischer Hintergrund

1.1. Entwicklungsdiagnostik

Die psychologische Diagnostik als Teildisziplin der Psychologie dient der Beantwortung von Fragestellungen im Zusammenhang mit psychischen Merkmalen und Verhaltensweisen (Kubinger, 2009). Ein Teilgebiet der Diagnostik ist die Entwicklungsdiagnostik, die sich auf Entwicklungsvariablen bezieht und sich überwiegend mit der Kindheit beschäftigt (Quaiser-Pohl, 2010a).

Fälschlicherweise wird der Begriff „Entwicklungsdiagnostik“ oftmals als Synonym für Entwicklungstests verwendet (Esser & Petermann, 2010). In der Psychologie und der Medizin spricht man üblicherweise erst dann von Entwicklungsdiagnostik, wenn zumindest ein standardisiertes psychometrisches Testverfahren, also ein sogenannter Entwicklungstest, zum Einsatz kommt (Quaiser-Pohl, 2010a). Ziel einer Entwicklungsdiagnostik ist es, den aktuellen Entwicklungsstand eines Kindes abzubilden, Entwicklungsabweichungen und -defizite, aber auch Ressourcen zu identifizieren. Darüber hinaus ist es Aufgabe der Entwicklungsdiagnostik, zukünftige Prognosen zu treffen und weitere Therapie- und Fördermaßnahmen abzuleiten (Esser & Petermann, 2010). Quaiser-Pohl (2010a) spricht sogar von Förderdiagnostik als ethisches Prinzip, das dem Zweck dient, Kinder optimal in allen Entwicklungsstufen zu unterstützen. Um Kinder ihren Fähigkeiten entsprechend einschätzen zu können, sind Referenzwerte notwendig, die durch wissenschaftlich entwickelte psychologische Testverfahren gewonnen werden (Quaiser-Pohl, 2010a).

Entwicklungstests zielen darauf ab, den aktuellen Entwicklungsstand eines Menschen zu erfassen, Rückschlüsse auf die bisherige Entwicklung zu ziehen und Prognosen über die zukünftige Entwicklung zu treffen. Vor allem förderungsorientierte diagnostische Verfahren eignen sich zur Feststellung von Entwicklungsdefiziten, um frühzeitig entsprechende Interventionsmaßnahmen abzuleiten (Deimann & Kastner-Koller, 2011a).

In der Literatur werden mit dem Begriff „Entwicklung“ unterschiedliche Bedeutungen verknüpft (Esser & Petermann, 2010). Das Verständnis der Entwicklung unterlag in der Vergangenheit einem ständigen Wandel und wurde aus der Perspektive der vorherrschenden Paradigmen betrachtet. Heutzutage wird der Umwelt eine entscheidende Rolle bei der menschlichen Entwicklung zugeschrieben. Sowohl genetische Dispositionen als auch Umweltbedingungen spielen eine zentrale Rolle in der Entwicklung und beeinflussen sich

gegenseitig. Modernen entwicklungsdiagnostischen Verfahren liegt ein kontextualistisches und interaktionistisches Verständnis zugrunde (Deimann & Kastner-Koller, 2011a). Gegenwärtig betrachtet man Entwicklung als einen lebenslangen Prozess, der den Erwerb und die Differenzierung von Handlungskompetenzen, unter Berücksichtigung der Interaktion zwischen dem Individuum und der Umwelt, erfasst. Eine optimale Passung zwischen Individuum und Lernumwelt ist Voraussetzung dafür, dass sich ein Kind Handlungskompetenzen aneignet und somit optimal entwickelt (Kastner-Koller & Deimann, 2012). Unter Entwicklung versteht man demnach intraindividuelle Veränderungen des Verhaltens und Erlebens (Esser & Petermann, 2010), dessen Vielfältigkeit die Vielgestaltigkeit moderner Lebenswelten widerspiegelt (Petermann & Macha, 2005). Die Veränderungen lassen sich nur multikausal erklären (Esser & Petermann, 2010) und folgen nicht festgelegten Schritten, wie der reifungstheoretische Ansatz behauptet. Der Verlauf ist nicht linear, sondern oft sprunghaft (Petermann & Macha, 2005). Die Annahme eines End- oder Reifezustandes wird heutzutage ebenfalls in Frage gestellt. Endzustände lassen sich für die Entwicklung eines Menschen nicht definieren. Entwicklung ist vielmehr multidirektional und während der gesamten Lebensdauer möglich (Esser & Petermann, 2010). Sie läuft in zumindest zwei Richtungen und beschränkt sich nicht auf verbesserte Leistungen. Demnach kann Entwicklung sowohl Ausdruck von Wachstum als auch von Rückgang sein. Gemäß der Lebensspannenperspektive wird die Plastizität der Entwicklung in jedem Alter betont (Berk, 2005).

Trotz der in der Psychologie vorherrschenden Ansicht, dass die Entwicklung eines Menschen über die gesamte Lebensdauer stattfindet, existieren Entwicklungsverfahren in erster Linie für die frühe Kindheit und nur wenige, die entwicklungsbedingte Veränderungen im hohen Alter erfassen (Deimann & Kastner-Koller, 2007). Je nachdem welche Entwicklungsaspekte operationalisiert werden, unterscheidet man zwischen allgemeinen und speziellen Entwicklungsverfahren.

Spezielle Entwicklungstests fokussieren auf einzelne Entwicklungsbereiche (z.B. Motorik, Wahrnehmung, Sprache, kognitive oder sozial-emotionale Entwicklung), um diese differenziert zu erfassen. Allgemeine Entwicklungstests erheben hingegen den Anspruch, das gesamte Verhalten zu erfassen (Deimann & Kastner-Koller, 2011a). Überdies gibt es Entwicklungsscreenings, die dazu vorgesehen sind, ohne größeren testpsychologischen Aufwand Risikogruppen zu identifizieren und Entwicklungsabweichungen festzustellen. Sie

sind zeitökonomisch (Grob, 2008), bieten aber lediglich Hinweise auf Auffälligkeiten, die im zweiten Schritt durch eine Diagnostik abgeklärt werden müssen (Michaelis, Berger, Nennstiel-Ratzel, & Krägeloh-Mann, 2013).

Die Anzahl an diagnostischen Verfahren, die für Klein- und Vorschulkinder geeignet sind, hat sich in den letzten Jahren erhöht. Die verschiedenen Entwicklungstests unterscheiden sich jedoch bei den Zielsetzungen und Gültigkeitsbereichen (Petermann & Macha, 2005).

Entwicklungstests, mit deren Hilfe vor allem leistungsbezogene Daten des Kindes erfasst werden, stellen dabei jedoch nur einen Teilbereich der Entwicklungsdiagnostik dar. Auch kindbezogene Merkmale, wie das Temperament oder Verhaltensdispositionen, müssen bei einer differenzierten Diagnostik berücksichtigt werden (Macha & Petermann, 2008). Den kontextualistischen Entwicklungstheorien folgend ist es für das diagnostische Urteil, insbesondere um Prognosen über die weitere Entwicklung treffen zu können, bedeutsam, familiäre, soziale oder wirtschaftliche Faktoren der Umwelt des Kindes zu kennen (Deimann & Kastner-Koller, 2011a; Macha & Petermann, 2008).

Eltern, Verwandte und wichtige Bezugspersonen aus dem sozialen Umfeld des Kindes, wie zum Beispiel Pädagoginnen/Pädagogen aus dem Kindergarten, dienen als Informationsquelle. Je jünger Kinder sind, desto mehr Informationen werden von Dritten gewonnen. Neben einer Anamnese mit den Eltern werden auch Fremdbeurteilungsbögen zur Einschätzung des Entwicklungsstandes und zur Erhebung von Verhaltensweisen des Kindes im diagnostischen Prozess eingesetzt (Deegener, 2009). Elternbefragungen über die Entwicklung und Verhaltensweisen des Kindes werden neuerdings sogar anstatt der Anwendung psychologischer Tests bei der Diagnostik von Kleinkindern verwendet. Diese Interviews nehmen wesentlich weniger Zeit in Anspruch und sind daher ökonomischer als Entwicklungstests. Zudem können sie Aufschluss über soziale und kommunikative Fähigkeiten des Kindes geben, die in einer standardisierten Testung nicht beobachtet werden können. Es wird jedoch kontrovers diskutiert, ob die dadurch gewonnenen Erkenntnisse zuverlässig sind. Studien belegen, dass Eltern dazu neigen, die Entwicklung ihres Kindes zu überschätzen (Deimann, Kastner-Koller, Benka, Kainz, & Schmidt, 2005), selbst dann, wenn Entwicklungsprobleme beim Kind offenkundig sind (Deimann & Kastner-Koller, 2011b). Untersuchungen zeigen, dass eine zuverlässige Einschätzung des Kindes durch die Mütter

hinsichtlich verbaler und motorischer Fähigkeiten besonders schwierig ist (Deimann et al., 2005; Willinger & Eisenwort, 2005).

Zudem können Elternbefragungen durch die Fragetechniken der Testleiterin/des Testleiters verzerrt werden (Deimann & Kastner-Koller, 2011b). Laut Deegener (2009) sind Diagnostikerinnen/Diagnostiker nicht ausreichend zur Durchführung von Anamnesen und Explorationen ausgebildet.

Es existiert keine allgemein anerkannte Theorie, die als Grundlage für Entwicklungstests dienen könnte. Aus diesem Grund basieren die vorhandenen standardisierten Entwicklungstests auf unterschiedlichen Konzepten. Deren Ergebnisse sind daher kaum miteinander vergleichbar. Ebenso kann die Zusammenstellung von Aufgaben in Entwicklungstests nicht auf eine theoretische Anleitung zurückgreifen, sondern muss auf praktische Erfahrungswerte gestützt werden (Esser & Petermann, 2010). Eine fachgerechte kinderpsychologische Diagnostik im Kleinkindalter stellt eine besondere Herausforderung für die Untersucherin/den Untersucher dar. Der folgende Abschnitt geht auf die Hürden ein, die dabei überwunden werden müssen.

1.2. Entwicklungsdiagnostik im Kleinkindalter

Während Vorschulkinder zunehmend in der Lage sind, den Testanforderungen nachzukommen, und somit in einer standardisierten Situation „getestet“ werden können, benötigen Kleinkinder eine andere Herangehensweise. Der motorische Antrieb, die Impulsivität und die Ablenkungsbereitschaft von Kleinkindern erfordern ein Untersuchungsinstrument mit kurzen und einfachen Aufgaben. Kleinkinder können ihre Aufmerksamkeit noch nicht über längere Zeit einer Aufgabe widmen, schon gar nicht an einem Tisch sitzend (Petermann, Proske, & Macha, 2005). Sie sind in ihrer Kooperationsbereitschaft sprunghaft, und ihr Verhalten ist wenig vorhersehbar (Irblich & Renner, 2009). Aufmerksamkeitsschwankungen und Vorlieben für bestimmte Materialien können die Ergebnisse verfälschen (G. Reuner & Pietz, 2006).

Hinsichtlich der Untersuchungszeit einer Diagnostik müssen der Tagesablauf des Kindes und vorangegangene Aktivitäten berücksichtigt werden (Quaiser-Pohl, 2010b). Die Leistungen von Kleinkindern sind stark von der Tagesverfassung und dem körperlichen Befinden abhängig (Sarimski, 2009). Im Kleinkindalter sollte eine Untersuchung maximal 30 bis 40 Minuten in Anspruch nehmen (Quaiser-Pohl, 2010b).

Es gibt kaum Verfahren für Kinder im zweiten Lebensjahr, die den Bedürfnissen eines Kleinkindes gerecht werden. In dieser Altersspanne spielt auch das Untersuchungssetting eine wichtige Rolle. Eine reizarme Situation soll der Ablenkbarkeit des Kindes entgegenwirken, dennoch sollte der Raum kindgerecht und ansprechend eingerichtet sein. Zudem ist in dieser Altersspanne eine große Variabilität des Sprachentwicklungsstandes zu beobachten. Damit Kleinkinder bei den Testaufgaben mitmachen und die Instruktionen verstehen, sind häufig unterschiedliche Vorgehensweisen, mehrere Versuche oder Anpassungen notwendig. Dies wiederum hat Einfluss auf den Grad der Standardisierung und kann die Objektivität beeinträchtigen (Petermann et al., 2005; Sarimski, 2009). Die Diagnostik junger Kinder stellt daher eine besondere Herausforderung für die Untersucherin/den Untersucher dar und erfordert viel psychologisches Geschick und Flexibilität (Irblich & Renner, 2009). Um flexibel auf das Kind reagieren zu können, ist es notwendig, dass die Aufgaben sicher beherrscht und nicht mithilfe des Handbuches durchgeführt werden. Die Diagnostikerin/der Diagnostiker muss über ein umfassendes Wissen situativer kindlicher Verhaltensweisen verfügen (G. Reuner & Pietz, 2006). Grundlegende Voraussetzung für eine valide Diagnostik ist eine gelungene Kontaktaufnahme zum Kind. Die Zuwendung soll von einer freundlichen und kindgemäßen Ausdrucksweise und einem wertschätzenden Umgang geprägt sein. Die Untersucherin/der Untersucher muss über ein großes Verhaltensrepertoire verfügen, um mit Kindern in Kontakt zu treten und mit ihnen zu interagieren (Irblich & Renner, 2009).

Im Kindes- und Jugendalter vollzieht sich die Entwicklung einzelner Fertigkeiten sehr schnell und unabhängig voneinander (G. Reuner & Pietz, 2006). Besonders bei Kleinkindern werden rasante Entwicklungsschritte innerhalb kurzer Zeit sichtbar (Quaiser-Pohl, 2010a), die interindividuell jedoch sehr unterschiedlich verlaufen können. Daher ist es bei allgemeinen Entwicklungstests schwierig, eine normabweichende Entwicklung oder spezifische Störungen festzustellen (Petermann & Macha, 2003; G. Reuner & Pietz, 2006). Die prognostische Aussagekraft von entwicklungsdiagnostischen Verfahren ist aufgrund der raschen entwicklungsbedingten Veränderungen begrenzt (Sarimski, 2009). Sowohl Beobachtungen des Kindes als auch Befragungen der Bezugspersonen sind notwendig, um die frühkindliche Entwicklung adäquat zu bestimmen. Der alleinige Einsatz eines standardisierten Verfahrens ist für eine entwicklungspsychologische Abklärung demnach nicht ausreichend (G. Reuner & Pietz, 2006).

Im Folgenden werden die beiden gängigsten allgemeinen Entwicklungsverfahren für die Altersgruppe der zwei- bis dreijährigen Kinder im Überblick dargestellt.

Der „Entwicklungstest für Kinder von 6 Monaten bis 6 Jahren – Revision“ (ET 6 - 6R; Petermann & Macha, 2013) erfasst die Entwicklungsbereiche *Körper- und Handmotorik*, *kognitive Entwicklung*, *Sprachentwicklung* sowie mittels eines Elternfragebogens die *sozial-emotionale Entwicklung*. Der Entwicklungstest beurteilt spezifische Teilleistungen und überprüft die Grenzsteine der Entwicklung (Michaelis, Berger & Nennstiel-Ratzel, 2013). Das Verfahren kann sowohl als Entwicklungsscreening wie auch als differenzierter Entwicklungstest eingesetzt werden (Petermann et al., 2013). Die Testmaterialien und Aufgaben sind kindgemäß, und die Testdauer ist an die spezifischen Voraussetzungen der Kinder der jeweiligen Altersspanne angepasst. Im Kleinkindalter ist die Dauer mit 30 bis 40 Minuten bemessen (Petermann & Macha, 2013). Die Kritikpunkte, dass die Normierungsstichprobe eher klein ist und die Gütekriterien Objektivität, Zuverlässigkeit und Validität nur teilweise gegeben sind (Hasselhorn & Margraf-Stiksrud, 2012), wurden bei der Revision eingeräumt. Repräsentative Altersnormen liegen vor und die inhaltliche Validität ist gegeben. Zur Reliabilitätsbestimmung werden interne Konsistenzen angeführt, die für die einzelnen Skalen zwischen .66 und .77 liegen. Durch den hohen Standardisierungsgrad ist die Objektivität des Verfahrens gegeben. Dennoch kann der Testablauf an die Bedürfnisse des Kindes angepasst werden, was in dieser Altersgruppe sehr wichtig ist, um die Motivation des Kindes zu erhalten (Petermann & Macha, 2013).

Die „Bayley Scales of Infant and Toddler Development - Third Edition“ (BSID III; Bayley, 2006) erfassen die frühkindliche Entwicklung und die Funktionen von Kleinkindern im Alter von 1 bis 42 Monaten. Das Verfahren besteht aus den drei Skalen *Cognitive Scale*, *Language Scale* und *Motor Scale*. Für die drei Skalen stehen standardisierte Werte, Prozentränge und Konfidenzintervalle zur Verfügung. Zusätzlich enthält der Entwicklungstest eine *Social-Emotional-Scale* und eine *Adaptive Behavior Scale*, die mittels Fragebogen der Erziehungsberechtigten erhoben werden. Die Dauer der Testung beträgt für Kinder ab 12 Monaten 90 Minuten (Bayley, 2006). Für diese Subskalen können standardisierte Werte und das Entwicklungsalter bestimmt werden. Die Testung erlaubt die Erstellung eines Entwicklungsprofils des Kindes. Die empirische Basis der Bayley III ist zufriedenstellend (Macha, 2009), und die Gütekriterien, insbesondere Normierung, Validität, Reliabilität und Objektivität, sind in hohem Maße erfüllt (Albers & Grieve, 2007). Zudem bietet das Verfahren ein sehr ansprechendes Material für junge Kinder, und es stehen detaillierte Instruktionen im Manual zur Verfügung (Albers & Grieve, 2007). Längsschnittstudien sind jedoch

notwendig, um die Gültigkeit der gewonnen Entwicklungseinschätzungen der dritten Auflage der Bayley Scales zu beurteilen.

In einer Studie erfolgte für Kindern mit gesundheitlichen Beeinträchtigungen ein Vergleich der Ergebnisse der Entwicklungseinschätzung von Bayley II mit Bayley III. Die Untersuchung zeigte, dass die durch die dritte Auflage gewonnenen Werte signifikant höher sind als die Ergebnisse der Bayley II. Es bleibt offen, ob die Bayley II die Fähigkeiten der Kinder unterschätzt oder die dritte Auflage diese überschätzt (Acton et al., 2011).

Seit 2015 gibt es eine deutsche Bearbeitung der dritten Auflage der Bayley III. Das Verfahren wurde an einer deutschen ($N = 878$) und holländischen Stichprobe ($N = 131$) normiert. Die Reliabilitätskoeffizienten der Skalen liegen zwischen $r = .86$ und $r = .88$, und die der Subskalen zwischen $r = .77$ und $r = .89$. Die inhaltliche Validität und die Konstruktvalidität sind gegeben, und für die kriterienbezogene Validität liegen erste Nachweise vor (Bayley, 2015).

Ob die Nützlichkeit der „Bayley Scales of Infant and Toddler Development - Third Edition“ (Bayley, 2006) gesteigert werden konnte, muss erst festgestellt werden. Studien zur prognostischen Validität und Genauigkeit fehlen bislang. Kritisiert wird außerdem, dass aus den gewonnenen Ergebnissen keine Interventionsmaßnahmen abgeleitet werden können. Zukünftige Forschungen sind nicht nur hinsichtlich der Bayley III notwendig, sondern allgemein für die testtheoretische Beurteilung diagnostischer Verfahren (Albers & Grieve, 2007; Nelson-Gray, 2003).

Wie die Ausführungen in Abschnitt 1.1 zeigen, existiert kein einheitlicher Entwicklungsbegriff, und das Verständnis von „Entwicklung“ unterlag einem zeitlichen Wandel. Den „Bayley Scales of Infant and Toddler Development - Third Edition“ (Bayley, 2006) liegt ein hierarchisch determiniertes Entwicklungskonzept zugrunde. Die Grundannahme, dass Entwicklung nach festgelegten Stufen verläuft, wurde durch die heutige Sichtweise einer individuell und adaptiv verlaufenden Entwicklung im Zusammenhang mit der Umwelt ersetzt. Der einzige allgemeine Entwicklungstest für diese Altersspanne, der auf dieser theoretischen Basis beruht, ist der „ET 6-6R“ (Michaelis et al., 2013; Petermann & Macha, 2013).

Demnach stehen zur Beurteilung der Entwicklung in den ersten Lebensjahren kaum geeignete Verfahren zur Verfügung. Grimm und Aktas (2002) sprechen von einer „diagnostischen Lücke“. Praktikerinnen/Praktiker sind daher gezwungen, trotz der erwähnten Einschränkungen nicht standardisierte Instrumente zu verwenden. Aufgrund der bestehenden Mängel und anhaltender Kritik an vorhandenen testpsychologischen Verfahren wurde ein neuer spielbasierter Ansatz entwickelt, der im folgenden Kapitel vorgestellt wird.

1.3. Spielbasierte Diagnostik

Der spielbasierte Ansatz ist in der psychologischen Diagnostik als neues Instrumentarium zur Erfassung der Entwicklung junger Kinder entstanden. Ausgehend von der Problemlage und den Schwächen herkömmlicher psychologischer Testverfahren für Kleinkinder sollen das Konzept und die Vorteile einer spielbasierten Diagnostik dargestellt werden.

Bereits in den 90er Jahren war man sich der Problematik bewusst. Aus diesem Grund befragten Bagnato und Neisworth (1994) Psychologinnen und Psychologen, die mit Kleinkindern arbeiten. Fast 60 Prozent der Befragten verwenden standardisierte Tests, wobei angegeben wurde, dass 43 Prozent der Kinder bei Verwendung dieser Verfahren als „nicht testbar“ gelten. Aus diesem Grund gaben 80 Prozent der Psychologinnen und Psychologen an, alternative Strategien einzusetzen. An erster Stelle wurde die Verhaltensbeobachtung beim Spielen genannt, und an zweiter Stelle die Befragung der Eltern.

Je jünger die Kinder sind, desto wichtiger ist es, ergänzende Informationen durch Interviews mit den Eltern oder weiteren Bezugspersonen zu erhalten. Ein diagnostisches Urteil sollte sich aber nicht ausschließlich auf Fremdbeurteilungen stützen. Die Befragung, aber vor allem die Beobachtung des kindlichen Verhaltens, insbesondere bei sehr jungen oder entwicklungsverzögerten Kindern, wird für unverzichtbar im diagnostischen Prozess erachtet (Kastner-Koller & Deimann, 2009). Nur wenige psychologische Verfahren sind vorhanden, die die neuropsychologischen Funktionen des Kleinkindes erfassen können. Ergänzende Methoden, wie beispielsweise die Verhaltensbeobachtung, sind notwendig (Dykeman, 2008).

Bei Kindern mit gravierenden Entwicklungsstörungen liefern Entwicklungstests keine brauchbaren Hinweise für die Therapie- und Förderplanung. Eine Untersuchung sollte jedoch feststellen, über welche Schlüsselkompetenzen in den Bereichen der Kognition (Spielinitiative, Kognition), der Kommunikation (dialogischer Austausch, der geteilten Aufmerksamkeit) sowie der sozial-emotionalen Entwicklung (Freude am Spiel, Ausdauer, Kooperation) ein Kind verfügt. Diese Beurteilungen können mithilfe eines standardisierten Verfahrens nicht gewonnen werden. Es bedarf einer Spielsituation, in der die Diagnostikerin/der Diagnostiker die Rolle der Spielpartnerin/des Spielpartners einnimmt. Durch das Spielen können qualitative Einschätzungen gewonnen und Fördermaßnahmen abgeleitet werden (Sarimski, 2009).

Der spielbasierte Ansatz stellt eine Alternative dar, um den Entwicklungsstand eines Kleinkindes abbilden und spezielle Bedürfnisse feststellen zu können (Dykeman, 2008).

Die unnatürliche Situation einer Testung mit herkömmlichen Verfahren entspricht nicht den Bedürfnissen eines Kleinkindes (Bagnato, 2005). Ein spielerischer Zugang zur Testung für Kinder unter vier Jahren wird daher empfohlen. Bei Kindern dieser Altersspanne muss oftmals während der Testsituation das momentane Spielinteresse berücksichtigt werden, wodurch der standardisierte Testablauf vernachlässigt wird. Eine gewisse Arbeitshaltung und Leistungsmotivation kann erst bei vier- oder fünfjährigen Kindern erwartet werden (Irblich & Renner, 2009). Der neue Ansatz der spielbasierten Diagnostik nimmt sich dieser Problematik an und versucht die Nachteile bestehender Testverfahren zu vermeiden.

Ein spielbasierter Test beruht auf der Beobachtung verschiedener Fähigkeiten des Kindes während des Spiels. Aufgrund der ökologischen Validität, der hohen Flexibilität und der Motivation des Kindes wird der spielbasierte Ansatz bevorzugt. Junge Kinder spielen gerne, daher erscheint es sinnvoll, die einzelnen Entwicklungsbereiche des Kindes im Spiel zu erfassen (Kelly-Vance & Ryalls, 2005).

Das „freie Spiel“ in einer natürlichen Umgebung wird von vielen als angemessener Kontext erachtet, um die geistigen Fähigkeiten von jungen Kindern zu evaluieren. Dementsprechend wurden spielbasierte Verfahren in der natürlichen Spielumgebung des Kindes, in der sich das Kind mehrere Stunden pro Tag aufhält, durchgeführt (Kelly-Vance, Ryalls, & Glover, 2002).

Der spielbasierte Ansatz erlaubt dem Kind die maximale Freiheit, um in einer natürlichen und flexiblen Umgebung das höchste Level an Fähigkeiten zu zeigen (Dykeman, 2008). Eltern empfinden die Situation einer spielbasierten Diagnostik weniger stressig und sind der Ansicht, dass diese Methode das typische Verhalten des Kindes besser abbilden kann als standardisierte Verfahren (Gagnon & Nagle, 2004).

Im Vergleich zu standardisierten Methoden weisen spielbasierte Verfahren laut Kelly-Vance et al. (2002) folgende Vorteile auf:

- (a) Der Test wird in der natürlichen Umgebung des Kindes durchgeführt, im Gegensatz zu einer sterilen und hoch strukturierten Testumgebung bei Verwendung eines Frage-Antwort-Formats.
- (b) Spielbasierte Verfahren wurden entworfen, um die optimale Leistung des Kindes zu erfassen – im Gegensatz zu vielen standardisierten Tests, die den Kindern nicht die Gelegenheit geben ihr Optimum zu erreichen.
- (c) Im Unterschied zu dem strengen Format standardisierter Tests haben spielbasierte Verfahren einen flexiblen Ablauf.

- (d) Das Ergebnis spielbasierter Verfahren beschreibt die Stärken des Kindes und die Bereiche, in denen eine Verbesserung notwendig ist. Standardisierte Skalen hingegen liefern numerische Werte, die nur minimale Deskriptoren der Leistung des Kindes enthalten.

Die Qualität spielbasierter Verfahren birgt das Potential, eine genauere Bewertung der geistigen Fähigkeiten junger Kinder zu erreichen. Da Kinder Vergnügen beim Spielen empfinden und bei der Auswahl der Aktivitäten spontan entscheiden können, wirkt das Spielen motivierend und animiert daher die Kinder ihre Höchstleistung zu erbringen (Kelly-Vance et al., 2002).

Die Resultate von spielbasierten Verfahren werden, im Gegensatz zu standardisierten Tests, von Entwicklungstabellen abgeleitet und im Bezug zu Altersspannen präsentiert. Trotzdem können mithilfe spielbasierter Verfahren Werte gewonnen werden, die es erlauben, ein Urteil über den Entwicklungsstand von Kleinkindern zu bilden. Weiters besitzen die Ergebnisse unmittelbare Relevanz für neuropsychologische Probleme des Denkens, der Aufmerksamkeit, der Speicherkapazität, der senso-motorischen Funktionen, der visuellen-räumlichen Wahrnehmung, der Sprache und der exekutiven Funktionen.

Darüber hinaus stellen spielbasierte Verfahren für klinische Psychologinnen/Psychologen ein breites Beobachtungsspektrum des Verhaltens zur Verfügung, welches ein vollständigeres Bild des neurologischen Zustandes bietet als Skalen, die bei standardisierten Tests zum Einsatz kommen (Dykeman, 2008).

Erste Ansätze für die Entwicklung spielbasierter Verfahren liegen vor, wobei noch kaum Forschungen hinsichtlich Nutzen und Gütekriterien existieren (Kelly-Vance et al., 2002). Nur wenige Studien beschäftigen sich mit der psychometrischen Qualität von spielbasierten Entwicklungstests. Kelly-Vance und Ryalls (2005) erweiterten die Forschung von Kelly-Vance et al (2002) und untersuchten die Reliabilität, und als Ergänzung zu bisherigen Forschungen die Retest-Reliabilität, des entwickelten spielbasierten Verfahrens „Play in Early Childhood Evaluation System“. Die Studie verfolgte das Ziel, Informationen bezüglich Urteilerübereinstimmung (Interrater-Reliabilität) und Retest-Reliabilität zu erlangen. Die Stichprobe bestand aus 25 normal entwickelten und sieben entwicklungsverzögerten Kindern. Die Ergebnisse zeigten, dass das Spielverhalten bei normal entwickelten Kindern ($r = .482$; $p = .015$) über die beiden Messzeitpunkte hinweg moderat stabil ist. Bei den Kindern mit

Entwicklungsverzögerungen zeigte sich der gleiche Zusammenhang, jedoch war das Ergebnis nicht signifikant ($r = .575$; $p = .177$).

Kinderpsychologinnen/Kinderpsychologen sehen das Spiel eher als eine Behandlungstechnik, nicht als eine Bewertungsstrategie (Gagnon & Nagle, 2004). Die natürliche Beobachtung des Spielverhaltens findet sich jedoch schon in der frühen Literatur, beispielsweise bei Piaget (1951), als Methode, um sich ein Urteil über die kognitive Entwicklung des Kindes zu bilden.

Das spielbasierte Verfahren erfordert von der Untersucherin/vom Untersucher eine gute Beobachtungsfähigkeit und ein umfassendes Wissen über die kindliche Entwicklung. Die Psychologin/der Psychologe muss vom beobachteten Verhalten des Kindes auf seine Fähigkeiten schließen können (Gagnon & Nagle, 2004).

Besonders für Kinder in den ersten Lebensjahren werden neue Testverfahren benötigt, um ihre Entwicklung verlässlich feststellen zu können. Ziel entwicklungspsychologischer Forschungen ist, die Entwicklung neuer Verfahren, die für junge Kinder geeignet sind und dennoch den testtheoretischen Anforderungen genügen. Die Verfolgung eines spielbasierten Ansatzes erscheint zielführend.

Erstmals im deutschsprachigen Raum wurde ein Itempool zur spielbasierten Entwicklungsdiagnostik zweijähriger Kinder entwickelt (Kuchler, Sapper, Kastner-Koller, & Deimann, 2011), dessen testtheoretische Bewertung Ziel dieser Arbeit ist.

1.4. Meilensteine in der Entwicklung Zweijähriger

Der spielbasierte Entwicklungstest, der in der vorliegenden Arbeit testtheoretisch evaluiert werden soll, ist nur für eine sehr kurze Zeitspanne, nämlich das dritte Lebensjahr, bestimmt. In diesem Lebensabschnitt machen Kinder gewaltige Fortschritte und entwickeln ihre Fähigkeiten rasant weiter. In diesem Abschnitt werden die wichtigsten Meilensteine der Entwicklung zweijähriger Kinder in den Bereichen *Sprache* und *Aufmerksamkeit*, überblicksmäßig dargestellt. Auf die Erläuterung der Vorläuferfähigkeiten wird bewusst verzichtet und auf einschlägige Literatur verwiesen (Berk, 2005; Szagun, 2007).

Meilensteine sind bedeutsame Entwicklungsschritte, die sich in Verhaltensweisen und Fähigkeiten ausdrücken (Pauen, Frey, & Ganser, 2012). Die meisten Aufgaben von Entwicklungstests orientieren sich am Konzept der Meilensteine. Demzufolge müssen

Entwicklungsschritte in einem bestimmten Zeitraum stattfinden, um von altersgerechter Entwicklung sprechen zu können (Sarimski, 2009).

1.4.1. Entwicklung der Sprache

Besonders bei der Sprachentwicklung zeigt sich ein bemerkenswerter Leistungszuwachs im Alter von zwei bis sechs Jahren. Während ein zweijähriges Kind einen Wortschatz von 200 Wörtern aufweist, umfasst der Wortschatz eines Sechsjährigen bereits 10 000 Wörter (Berk, 2005). Bei der Sprachentwicklung besteht eine hohe Variabilität (Michaelis et al., 2013; von Suchodoletz, 2012). Das erste Wort wird im Durchschnitt im Alter von 12 Monaten gesprochen, mit einer Varianz von 8 bis 18 Monaten. Sobald ein Kind erste Wörter gesprochen hat, entwickeln sich die sprachlichen Fähigkeiten schnell weiter (Berk, 2005). Der Verlauf des Spracherwerbes ist sehr unterschiedlich. Manche Kinder zeigen kontinuierliche Fortschritte bei der Entwicklung der Sprache, während sie bei anderen sprunghaft und ungleichmäßig verläuft. Je jünger das Kind ist, desto größer sind die individuellen Unterschiede und desto schwieriger ist es, diagnostische Prognosen bezüglich der folgenden Sprachleistungen zu treffen (von Suchodoletz, 2012). Spätere Erwerbsprozesse bauen auf den bereits erworbenen Wissensstrukturen auf. Defizite wirken sich daher auch auf nächste Entwicklungsschritte aus (Grimm, 2003).

1.4.1.1. Wortschatz

Die Sprachproduktion stellt eine wichtige Errungenschaft in der frühkindlichen Entwicklung dar und ist das Endergebnis der Lautentwicklung (Grimm, 2003). Voraussetzung für das Erlernen der produktiven Sprache ist die rezeptive Sprache (Papoušek, 2012).

Wie bereits erwähnt, umfasst der Wortschatz eines zweijährigen Kindes ungefähr 200 Wörter. Ein Prozess namens „Schnellzuordnung“ (fast mapping) ist dafür verantwortlich, dass Kinder ihren Wortschatz rasant ausbauen. Durch diesen Prozess sind Kinder in der Lage, ein neues Wort mit einem zugrunde liegenden Konzept zu verbinden (Berk, 2005). Die Kinder erkennen, dass Wörter eine Bedeutung und alle Dinge einen Namen haben (Grimm, 2003). Vorschulkinder setzen neue Wörter in Kontrast zu jenen, die sie bereits kennen. Auch soziale Hinweise werden von den Kindern verwendet, um Wortbedeutungen zu verstehen. Wie genau Kinder begreifen, welches Konzept jedes Wort umfasst, ist noch nicht vollständig geklärt (Berk, 2005).

Bezeichnungen von Gegenständen werden von westlichen Kindern besonders schnell erlernt, weil sie über ein großes Wissen über konkrete Gegenstände verfügen. Nach und nach werden diese Bezeichnungen um Handlungswörter, wie zum Beispiel „gehen“, erweitert. Die Bedeutung von Eigenschaftswörtern wird erst mit zunehmendem Alter erlernt. Zweijährige können bereits den Unterschied zwischen „groß“ und „klein“ begreifen, noch nicht aber feine Unterscheidungen wie „lang/kurz“ oder „hoch/niedrig“ (Berk, 2005). Laut Anglin (1993) erlernt ein Kind täglich fünf neue Wörter.

Erst durch den zunehmenden Wortschatz können Inhalts- und Funktionswörter unterschieden und in späterer Folge Sätze gebildet werden. Mit 18 Monaten sollte ein Kind über einen Wortschatz von 50 Wörtern verfügen. Hat es diesen Schwellenwert erreicht, kommt es zur sogenannten Wortexplosion. Das Kind ist dann bereit, seinen Wortschatz schnell zu vergrößern (Grimm, 2003). Das Erreichen dieses Meilensteines hängt stark von der kognitiven Entwicklung des Kindes ab (Papoušek, 2012). Ab dem zweiten Lebensjahr nimmt nicht nur die Größe des Wortschatzes zu, sondern er wird auch inhaltlich neu organisiert. Den Nomen und sozial-affektiven Regulatoren werden prädikative hinzugefügt. Ein Kind verwendet in dieser Altersstufe bereits einfache Wortkombinationen wie „Mädchen schlafen“. Kleinkinder verbinden jedoch Wörter mit anderen Bedeutungen und weisen ihnen einen erweiterten bzw. engeren Geltungsbereich zu. Man spricht in diesem Zusammenhang von Übergeneralisierung und Überdiskriminierung (Grimm, 2003). Die Wörter „Ja“ und „Nein“ werden schon sehr früh im Sprachgebrauch verwendet (Szagun, 2007).

Aussprachefehler, wie die Vereinfachung von Konsonantenverbindungen, sind typisch für Kleinkinder. Erst am Ende der Vorschulzeit werden Wörter vollständig und korrekt ausgesprochen (Grimm, 2003).

Das Sprachverständnis ist der Produktion von Sprache in Quantität und Komplexität überlegen. Kleinkinder verstehen mehr als sie bereits sprechen können. Im Alter zwischen 24 und 36 Monaten sind Kinder zunehmend in der Lage komplexe Sätze zu verstehen. Mit ca. 30 Monaten können Kinder selbst Sätze bilden. Ein umfangreicher Wortschatz ist die Voraussetzung für den Grammatikerwerb (Grimm, 2003).

1.4.1.2. Grammatik

Mit zunehmenden Wortschatz erwerben Kinder erste grammatikalische Regeln. Sie gewöhnen sich an die Gesetzmäßigkeiten der Wortfolge und bilden erste Sätze mit Subjekt, Verb und

Objekt. Sie lernen den Plural zu bilden, benutzen erste Präpositionen und bilden verschiedene Zeitformen. Auch hierbei kommt es zur Übergeneralisierung, indem ein Kind beispielsweise den Plural immer durch das Anhängen eines „-s“ bildet (Berk, 2005). Kinder entwickeln im Zuge des Grammatikerwerbs eigene Regeln, die sich von denen der Erwachsenen unterscheiden (Aktas & Doil, 2009). Die ersten Sätze, die Kinder bilden, beziehen sich auf das Vorhandensein von Gegenständen, Tieren oder Personen. Sie verwenden das Wort „da“ mit dem jeweiligen Substantiv. In Zweiwortsätzen platzieren sie das Verb am Ende des Satzes und verwenden anfangs noch keine Modalverben. Sobald Kinder Mehrwortsätze gebrauchen, lernen sie, die richtigen Wortstellungen relativ rasch. Dieser Entwicklungsschritt vollzieht sich zwischen dem zweiten und dritten Lebensjahr (Szagun, 2007).

Ebenso wie der passive Wortschatz der Kinder größer ist als der aktive, verstehen Kinder komplexere grammatische Strukturen als sie selbst produzieren können (Aktas & Doil, 2009).

Sowohl Wortschatz als auch grammatikalische Fähigkeiten nehmen nach dem dritten Lebensjahr kontinuierlich zu (Papoušek, 2012). Die grundlegende Grammatik haben Kinder im Alter von vier Jahren erworben (Szagun, 2007).

1.4.1.3. Erfassung der Sprachentwicklung

Der Bewertung sprachlicher Fähigkeiten wird bei der Erfassung des Entwicklungsstandes des Kindes besondere Bedeutung beigemessen, da sie als Entwicklungsindikator angesehen werden und Voraussetzung für den Erwerb weiterer Kompetenzen sind (Ettrich, 2000). Die Sprachkompetenz steht im engen Zusammenhang mit kognitiven und sozialen Fähigkeiten (Grimm, 2003). Der Entwicklungsstand des aktiven und passiven Wortschatzes wird zur Überprüfung des Sprachverständnisses und der Sprachproduktion herangezogen. Diese beiden Eigenschaften ermöglichen eine Beurteilung der sprachlichen Fähigkeiten von Kleinkindern (Ettrich, 2000).

Um die sprachliche Entwicklung des Kindes einschätzen zu können, kann die verbale Interaktion während einer unstrukturierten oder standardisierten Untersuchungssituation beobachtet werden. Eine interessant gestaltete Spiel- oder Bilderbuchsituation regt das Kind zu besonders vielen sprachlichen Äußerungen an. Diese Einschätzung ist allerdings von der Erfahrung der Diagnostikerin/des Diagnostikers abhängig und kann die Objektivität beeinträchtigen. Eine Sprachbeobachtung während der Untersuchungssituation kann leicht zu Fehleinschätzungen führen, da jüngere Kinder mit fremden Personen, insbesondere in einer

ungewohnten Situation, den sprachlichen Kontakt verweigern. Eine alternative Möglichkeit ist eine Spontansprachanalyse, die jedoch sehr aufwändig ist und spezifische Fachkenntnisse verlangt. Dennoch spiegeln sie oftmals nicht das tatsächliche Entwicklungsniveau des Kindes wider, da nur kurze Zeitintervalle erhoben und bewertet werden. Standardisierte Sprachtests ermöglichen eine quantitative Beurteilung des Sprachentwicklungsstandes eines Kindes. Trotz Testgütekriterien zeigen Untersuchungen der Zuverlässigkeit der Ergebnisse, dass Tests zur Bewertung der sprachlichen Entwicklung unscharf und fehlerbehaftet sind. Therapieimplikationen und Interventionsmaßnahmen sollten daher nicht nur auf Basis der Sprachtestergebnisse getroffen werden, sondern weitere Informationen über das Kind einbeziehen. Unstandardisierte Sprachtests sind in der sprachtherapeutischen Praxis am weitesten verbreitet. Aufgrund der fehlenden Testgütekriterien und Normwerte ist die Zuverlässigkeit der gewonnenen Ergebnisse jedoch nur beschränkt (von Suchodoletz, 2012).

1.4.2. Entwicklung der Aufmerksamkeit

Die Aufmerksamkeit ist die Fähigkeit eines Kindes sich zu orientieren, die Aufmerksamkeit zu verlagern und sie auf Geschehnisse, Objekte oder Aufgaben fokussieren zu können (van de Weijer-Bergsma, Wijnroks, & Jongmans, 2008). Viele verschiedene Prozesse und Funktionen sind dabei beteiligt (Heubrock & Petermann, 2001). Erste Messungen der Aufmerksamkeitsfähigkeit sind im Zuge der Erfassung der Entwicklung kognitiver Fähigkeiten, wie dem Gedächtnis und der Habituation, entstanden (van de Weijer-Bergsma et al., 2008).

Mit zunehmenden Alter können Kinder Informationen schneller aufnehmen und ihre Aufmerksamkeit besser handhaben. Während Säuglinge zu Beginn etwa drei oder vier Minuten für die Habituation (Gewöhnung) neuer visueller Reize benötigen, brauchen vier oder fünf Monate alte Kinder nur mehr fünf bis zehn Sekunden. Die lange Zeit, die ein Säugling benötigt um Reize zu erfassen, liegt in der Ursache begründet, dass er seine Aufmerksamkeit nur schwer von interessanten Reizen abziehen kann (Berk, 2005). Junge Kinder sind in einer komplexen Umgebung mit vielen Reizen konfrontiert. Sie müssen selektieren, um sich einzelnen Aufgaben aufmerksam widmen und ihre Aufmerksamkeit auf wichtige Dinge lenken zu können. Sie müssen Durchhaltevermögen erlernen um Handlungen und Aufgaben, trotz Hindernissen und Ablenkungen, erledigen zu können (van de Weijer-Bergsma et al., 2008).

Bereits mit drei bis acht Monaten können Babys der Blickrichtung von Erwachsenen folgen, und das Kind lernt seine Aufmerksamkeit der Aufmerksamkeitsrichtung des Erwachsenen

anzupassen (Rauh, 2008). Im Alter von vier bis sechs Monaten wird die Aufmerksamkeit der Kinder flexibler, und das Kind lernt die Aufmerksamkeit von einem Reiz auf den anderen zu richten (Berk, 2005). Die Aufmerksamkeit des Kindes kann nun zwischen Objekt und Person wechseln. Im Alter von neun Monaten werden die Handlungen zunehmend willentlich gesteuert (Rauh, 2008).

Zwischen dem 18. und 24. Lebensmonat verändert sich das Aufmerksamkeitsverhalten aufgrund der Entwicklung im frontalen Kortex grundlegend. Während dieser Zeit zeigen Kinder eine weitere Zunahme der Aufmerksamkeit während des Spielens oder bei komplexen visuellen Reizen, wie beispielsweise dem Fernsehen. Sie können ihre Aufmerksamkeit beim Spielen mit unterschiedlichen Spielmaterialien fokussieren und ablenkende Reize ausblenden. Die Kinder erlangen weitere wichtige Fähigkeiten und können Handlungen zunehmend planen und ihre Aufmerksamkeit eigenständig steuern (van de Weijer-Bergsma et al., 2008). Im Kleinkindalter sind sie zunehmend fähig zu zweckgerichteten Verhaltensweisen, und die andauernde Aufmerksamkeit verbessert sich. Insbesondere beim Spielen mit Spielzeugen, bei zielgerichteten Handlungen, beispielsweise beim Bauen eines Turmes, muss die Aufmerksamkeit erhalten bleiben um das Ziel zu erreichen (Berk, 2005). Im zweiten Lebensjahr wenden sich Kinder auch Reizen zu, die keinen hohen Aufforderungscharakter besitzen, und zeigen längere Aufmerksamkeitsspannen. Aufgrund der zielgerichteten Aufmerksamkeit können sie geplante Handlungen zu Ende führen (Ruff & Capozzoli, 2003).

Von Leistungsmotivation kann man bei zweijährigen Kindern noch nicht sprechen. Erst im Übergang vom Kleinkind- zum Vorschulalter kann leistungsmotiviertes Handeln beobachtet werden. Leistungsmotivierte Aktivität lässt sich daran erkennen, dass das Kind stolz auf Erfolge ist oder mit Scham auf Misserfolge reagiert. Kleinkinder orientieren sich nur an physikalischen Effekten und freuen sich, wenn diese durch ihr eigenes Handeln hervorgerufen werden. Ebenso reagieren sie mit Ärger, wenn die gewünschte Wirkung ausbleibt, ohne aber die Situation zu bewerten. Erst ab dem dritten Lebensjahr erkennen Kinder, dass Handlungen von den Bezugspersonen wertgeschätzt werden (Holodynski, 2006). Eine Studie konnte zeigen, dass Kinder im Alter von drei Jahren erstes autonomes leistungsorientiertes Handeln zeigen (Stipek, Recchia, McClintic, & Lewis, 1992). Holodynski (2006) ist der Ansicht, dass leistungsmotiviertes Handeln im Vorschulalter kontextabhängig ist und nur in Anwesenheit von Bezugspersonen gezeigt wird.

Das Merkmal *Ausdauer* wird auch häufig im Zusammenhang mit Leistungsmotivation thematisiert, sagt jedoch mehr über die Motivation des Kindes aus (Holodynski, 2006).

1.4.2.1. Erfassung der Aufmerksamkeitsfähigkeit

Beurteilungen der Aufmerksamkeitsfähigkeit eines Kindes dürfen nicht anhand einer einmaligen Beobachtung erfolgen, da sich Beeinträchtigungen meist nicht bei anregenden, neuen Aufgaben oder in strukturierten Situationen, wie es in herkömmlichen Testsituationen der Fall ist, gezeigt werden. Durch die besondere Zuwendung einer erwachsenen Person kann die Aufmerksamkeitsleistung des Kindes überschätzt werden (Renner & Irblich, 2009). Bei einer psychologischen Messung geht es nicht darum einen momentanen Zustand zu beobachten, sondern situationsüberdauernde Merkmale zu erfassen (Kubinger, 2009). Der diagnostische Prozess gestaltet sich somit in mehreren Stufen und bedient sich verschiedener Methoden, die ergänzend zum Einsatz kommen, wie Exploration, Verhaltensbeobachtung und der Einsatz testpsychologischer Verfahren (Heubrock & Petermann, 2001).

Bei der Aufmerksamkeitsfähigkeit handelt es sich um eine psychische Funktion, die sich nur indirekt über Verhaltensweisen ausdrückt und daher schwer messbar ist. Standardisierte Tests stoßen bei der Erfassung solcher psychischen Eigenschaften auf ihre Grenzen (Quaiser-Pohl, 2010b).

Bereits diese kurze Diskussion zur Erfassung der Sprachenwicklung und der Aufmerksamkeitsfähigkeit lässt die Komplexität der Entwicklungsdiagnostik erahnen. Die Testkonstruktion valider Verfahren ist mit einem hohen zeitlichen und finanziellen Aufwand verbunden. Die wichtigsten Testgütekriterien, die die Qualität eines Verfahrens bestimmen, werden im Folgenden erläutert.

1.5. Qualitätsanforderungen an Entwicklungstests

Entwicklungstests müssen methodischen Anforderungen genügen, die in diesem Abschnitt diskutiert werden. Die Qualität und Auswahl psychologischer Verfahren wird durch die Erfüllung der Testgütekriterien bestimmt. In der klassischen Testtheorie spricht man von drei Hauptgütekriterien, die in diesem Kapitel überblicksmäßig dargestellt werden. Vorrangig befasst sich der Diskurs mit der Reliabilität sowie weiteren wichtigen Kennwerten, wie Itemschwierigkeit und Trennschärfe, da sich die vorliegende Studie mit der testtheoretischen Überprüfung dieser Kriterien auseinandersetzt.

1.5.1. Objektivität

Ein Verfahren erfüllt das Kriterium der *Objektivität*, wenn die Durchführung, Auswertung und Interpretation der Ergebnisse unabhängig von der jeweiligen Testleiterin/vom jeweiligem Testleiter sind. Somit werden die Ergebnisse ausschließlich durch die Leistung der Testperson bestimmt (Renner, 2009). Dieses Gütekriterium stellt die Basis für die Standardisierung eines Verfahrens dar (Esser & Petermann, 2010).

Um die *Durchführungsobjektivität* zu gewährleisten, sind unter anderem wörtliche Testinstruktionen, genaue Anweisungen für die Materialvorgabe, Regeln für den Umgang mit Fragen oder schwierigen Untersuchungsbedingungen und Instruktionen für Hilfestellungen notwendig. Genaue Kriterien und Lösungsbeispiele für die Auswertung der Testergebnisse sind erforderlich, um den Ansprüchen der *Auswertungsobjektivität* zu entsprechen. *Interpretationsobjektivität* ist gegeben, wenn unterschiedliche Diagnostikerinnen/Diagnostiker dieselben Schlussfolgerungen aus den Ergebnissen ziehen (Esser & Petermann, 2010; Renner, 2009).

In Abschnitt 1.2 wurden die besonderen Anforderungen an die psychologische Diagnostik von Kleinkindern diskutiert. Den Ausführungen ist zu entnehmen, dass die *Durchführungsobjektivität* bei Testungen mit jungen Kindern nicht entsprechend gewährleistet werden kann.

1.5.2. Validität

Die *Validität* oder *Gültigkeit* gilt als wichtigstes Testgütekriterium und gibt an, in welchem Umfang das Verfahren die Eigenschaft, die es zu messen vorgibt, auch tatsächlich misst. Die Überprüfung der Validität gilt als besonders schwierig. Es werden drei Validitätsarten unterschieden, die *Inhaltsvalidität*, die *Konstruktvalidität* und die *Kriteriumsvalidität* (Kubinger, 2009). Ein Testverfahren gilt als inhaltlich valide, wenn es das Konstrukt präzise abbildet und idealerweise selbst das optimale Kriterium für dieses Merkmal darstellt. Die *Konstruktvalidität* ist erfüllt, wenn die gemessene Eigenschaft mit einem theoretischen Modell übereinstimmt. Die *Kriteriumsvalidität* zielt darauf ab, einen empirischen Zusammenhang zwischen dem Verhalten während der Testsituation und einem Kriterium außerhalb der Testsituation festzustellen (Esser & Petermann, 2010).

Die fehlende theoretische Basis eines Entwicklungstests und die mangelnde empirische Verankerung eines Entwicklungskontinuums sind die zentralen Probleme bei der Beurteilung

der Validität (Ettrich, 2000). Untersuchungen zum Vergleich von Testergebnissen eines Entwicklungstests mit den Resultaten anderer entwicklungsdiagnostischer Verfahren sind im Säuglings- und Kleinkindalter kaum zu finden (Sarimski, 2009).

1.5.3. Reliabilität

Die *Reliabilität* eines Testverfahrens, auch als *Zuverlässigkeit* bezeichnet, beschreibt den Grad der Genauigkeit, mit der ein Test ein bestimmtes psychologisches Merkmal misst (Kubinger, 2009). Reliabilitätskoeffizienten können einen Wert zwischen null und eins annehmen. Ein Reliabilitätskoeffizient von eins bedeutet, dass ein Test exakt und ohne Messfehler misst, unabhängig davon was er zu messen beansprucht (Esser & Petermann, 2010). Zur erforderlichen Höhe der Reliabilitätswerte gibt es unterschiedlichen Ansichten. Fisseni (2004) beurteilt Reliabilitäten ab .80 als mittel und ab .90 als hoch. Ein Reliabilitätskoeffizient von .70 wird als akzeptabel angesehen (Moosbrugger & Kelava, 2012).

Es existieren vier unterschiedliche methodische Zugänge zur Reliabilitätsbestimmung eines Testverfahrens, nämlich die *Paralleltest-Reliabilität*, die *Splithalf-Methode*, die *Retest-Reliabilität* und die *Innere Konsistenz*.

Die *Paralleltest-Reliabilität* wird durch die Testwiederholung mit einer Parallelförmigkeit bestimmt. Im Bereich der Entwicklungsdiagnostik eignet sich diese Methode nicht, da kaum parallele Versionen eines Tests vorliegen (Esser & Petermann, 2010). Kubinger (2009) gibt zu bedenken, dass diese Methode lediglich Aussagen darüber zulässt, wie gut die Konstruktion des Paralleltests gelungen ist.

Zur Erhebung der *Retest-Reliabilität* wird ein und derselbe Test zu zwei verschiedenen Messzeitpunkten vorgegeben. Dabei wird die Korrelation zwischen den beiden Testergebnissen für die einzelnen Funktionsgruppen ermittelt. Die Höhe des Korrelationswertes ist abhängig vom Zeitintervall der beiden Messungen (Moosbrugger & Kelava, 2012). In der Regel erfolgt die Wiederholung der Messung nach zwei bis drei Monaten (Esser & Petermann, 2010). Übungs- und Erinnerungseffekte sowie die Veränderung des gemessenen Merkmals können sich auf den Reliabilitätskoeffizienten auswirken (Moosbrugger & Kelava, 2012). Die Erhebung der Retest-Reliabilität ist bei Entwicklungstests insofern problematisch, als im Kindesalter Lerneffekte und Entwicklungsfortschritte innerhalb kurzer Zeit zu erwarten sind. Man kann also davon ausgehen, dass Kinder zum späteren Messzeitpunkt bessere Testleistungen erzielen. Daher wird in entwicklungsdiagnostischen

Verfahren oftmals auf Angaben und Untersuchungen zur Retest-Reliabilität verzichtet (Petermann et al., 2005).

Für eine Erhebung der Retest-Reliabilität spricht allerdings, dass die augenblickliche Verfassung des Kindes, insbesondere bei der Überprüfung der sprachlichen Fertigkeiten, das Testergebnis maßgeblich beeinflussen kann. Die Bereitschaft zur Mitarbeit und die Sprechfreude können beeinträchtigt werden, wenn sich das Kind bei der Testung nicht wohl fühlt. Das Ergebnis gibt dann kaum Auskunft über sein tatsächliches Entwicklungsniveau. Eine zweimalige Vorgabe ermöglicht die Bestimmung der Retest-Reliabilität und somit die Überprüfung der Genauigkeit des ersten Ergebnisses. Die Resultate sollten zu beiden Zeitpunkten weitgehend übereinstimmen (von Suchodoletz, 2012). Grundsätzlich können Einschätzungen zur Reliabilität, die auf einer einmaligen Messung basieren, die Zuverlässigkeit überschätzen, da dabei situative oder kindbezogene Einflüsse gering gehalten werden (Renner, 2009).

Die häufigste Methode, die zur Reliabilitätseinschätzung herangezogen wird, ist die Berechnung der internen Konsistenzen nach Cronbach Alpha. Dabei wird jedes Item als separater Testteil betrachtet. Diese Methode wird der *Splithalf-Methode* vorgezogen, bei der ein Test in zwei parallele Teile zerlegt und die Korrelation berechnet wird (Moosbrugger & Kelava, 2012). Auch diese beiden Methoden sind in der Entwicklungsdiagnostik nur bedingt einsetzbar, da einzelne Skalen oftmals aus nur wenigen Items bestehen. Eine Grundvoraussetzung zur Durchführung einer Konsistenzanalyse ist die Homogenität der Skalen, die bei Entwicklungstests nicht immer gegeben ist (Moosbrugger & Kelava, 2012; Renner, 2009).

Reliabilitätseinschätzungen sind nicht allgemein gültig, sondern stichprobenabhängig. Je größer eine Stichprobe ist, umso zuverlässiger sind die Reliabilitätswerte (Renner, 2009). Laut Kubinger (2003) können gültige Aussagen erst ab einer Stichprobe von 100 bis 150 Personen getroffen werden.

Darüber hinaus zählen zu den Kriterien der klassischen Testtheorie die Itemtrennschärfe, Itemschwierigkeit und Itemhomogenität, die zur testtheoretischen Absicherung eines Verfahrens notwendig sind (Esser & Petermann, 2010).

Die Itemanalyse ist ein wichtiger Bestandteil der Testkonstruktion, um die psychometrische Qualität der einzelnen Items und somit des gesamten Tests zu bestimmen. Itemeigenschaften

werden als Kennwerte erhoben und anhand von Qualitätsstandards beurteilt (Bortz & Döring, 2006). Laut Lienert (1989) kann die Bestimmung der Itemschwierigkeit im Zusammenhang mit dem Lebensalter Auskunft über die Validität des Verfahrens geben.

Die Itemschwierigkeit, also wie leicht oder schwer ein Item ist, wird für jedes einzelne Item eines Tests berechnet und gibt die Lösungswahrscheinlichkeit an. Eine Itemschwierigkeit von null bedeutet, dass das Item am schwersten ist und von keiner Person gelöst wird. Extrem schwierige oder leichte Items sind wenig informativ und sollten daher ausgeschlossen werden. Mittlere Itemschwierigkeiten im Bereich zwischen .20 und .80 werden bevorzugt (Bortz & Döring, 2006). Für junge Kinder wird die Itemschwierigkeit hoch sein, während sie für ältere Kinder ab einem bestimmten Alter niedriger wird (Ettrich, 2000).

Die Trennschärfe ist der zweite wichtige Kennwert der Itemanalyse. Der Trennschärfeindex gibt Auskunft darüber, ob ein Item zwischen Personen mit hoher und niedriger Merkmalsausprägung differenziert. Eine Trennschärfe mit einem positiven Wert zwischen .30 und .50 wird mittelmäßig beurteilt und Werte über .50 als hoch (Moosbrugger & Kelava, 2012). Laut Field (2009) sollen Items mit einer Itemtrennschärfe unter .30 ausgeschlossen werden. Ebenso sollen Items, die eine negative Trennschärfe aufweisen, nicht in ein Verfahren aufgenommen werden (Moosbrugger & Kelava, 2012).

Die hohen Standards zur Erfüllung der Gütekriterien konnten bei entwicklungsdiagnostischen Verfahren noch nicht zufriedenstellend umgesetzt werden (Renner, 2009). Es sind kaum Instrumente vorhanden, die den hohen methodischen Standards genügen. Die gegenständliche Untersuchung verfolgt das Ziel, zur psychometrischen Qualität eines neu entwickelten Verfahrens beizutragen.

2. Kontext und Ziel der vorliegenden Masterarbeit

Ziel dieser Arbeit ist die Anwendung und testtheoretische Überprüfung des neu entwickelten spielbasierten Itempools zur Erfassung der Entwicklung zweijähriger Kinder (Kuchler et al., 2011). Der Itempool wurde von Kastner-Koller und Deimann in Zusammenarbeit mit den Diplomandinnen Kuchler und Sapper (2011) entwickelt. Im Zuge weiterer Forschungen wurde der Itempool adaptiert und ersten testtheoretischen Analysen unterzogen (Birngruber, 2012; Fuchs-Garderer, 2012; Putzer, 2013; Sindelar, 2013). Von Punz (2013) und Kronberger (2013) erfolgten die letzten Modifikationen der Aufgaben zur Überprüfung der einzelnen Funktionsbereiche. Schließlich führten Tscherne (2014) und Lizst (2014) die letzten testtheoretischen Untersuchungen durch.

Das Verfahren soll erstmals der gleichen Stichprobe im Abstand von vier bis sechs Wochen erneut vorgegeben werden, um Aussagen über die Retest-Reliabilität zu erlangen. Weiters sollen Unterschiede zwischen den beiden Messzeitpunkten analysiert werden.

Der Itempool soll abermals einer Item- und Skalenanalyse unterzogen werden, um die bereits vorhandenen Daten zu erweitern und die Aussagekraft der Ergebnisse zu erhöhen. Der Diskurs beschränkt sich auf die Darstellung und Diskussion der Ergebnisse der Entwicklungsbereiche „Sprache“ und „Aufmerksamkeit“. Für die Untersuchung der anderen Entwicklungsbereiche sei auf Pozniak (in Vorbereitung) Prillinger (in Vorbereitung) und Stark (in Vorbereitung) verwiesen.

Abschließend soll eine Diskussion über die praktische Eignung des Verfahrens stattfinden.

3. Fragestellungen

Die Hauptfragestellung dieser Arbeit ist, ob eine ausreichende Retest-Reliabilität des Itempools gegeben ist.

Zweitens soll beantwortet werden, ob Unterschiede zwischen der ersten Vorgabe und der Testwiederholung bestehen.

Zur Absicherung testtheoretischer Ansprüche soll die dritte Frage ermitteln, ob die Reliabilitäten zufriedenstellend sind und die Ergebnisse der Vorstudie von Liszt (2014) repliziert werden können.

Zuletzt sollen die Itemschwierigkeiten und Itemtrennschärfen dahingehend geprüft werden, ob die Werte innerhalb des empfohlenen Wertebereichs liegen.

Die Überprüfung wird für beide Testzeitpunkte durchgeführt und mit den Ergebnissen der Vorstudie von Liszt (2014) verglichen.

4. Methode

4.1. Vorgehen

Die Datenerhebung erstreckte sich über den Zeitraum von Dezember 2014 bis April 2015. Alle Testungen wurden im Kleinkindertestraum der Fakultät für Psychologie der Universität Wien (Arbeitsbereich Klinische und Gesundheitspsychologie) durchgeführt. Die Datenerhebung erfolgte in einem Team, bestehend aus vier Personen, wobei die Tests zugunsten der Qualitätssicherung immer in Zweiergruppen vorgegeben wurden.

Die Termine wurden, unter Berücksichtigung des Tagesrhythmus des Kindes, individuell mit den Eltern vereinbart und dauerten durchschnittlich 90 bis 120 Minuten. Bei der Terminvereinbarung erhielten die Bezugspersonen genaue Informationen über Zweck, Ablauf und Dauer der Studie und wurden über die Notwendigkeit einer Videoaufzeichnung der Testungen aufgeklärt.

Bei den Testungen war neben der Anwesenheit des Kindes, der Testleiterin und der Beobachterin auch die einer Bezugsperson vorgesehen (Kuchler et al., 2011). Die Testleiterin spielte mit dem Kind, während die Beobachterin sich eher im Hintergrund hielt und am Protokollbogen die einzelnen Items kodierte. Zusätzlich wurde jede Testung auf Video aufgezeichnet. Bevor mit der Testung des Kindes begonnen wurde, erhielten die Eltern Instruktionen, wie sie sich während der Spielsituation verhalten sollten. Sie wurden gebeten, möglichst wenig Eigeninitiative zu zeigen und abseits zu sitzen. Zu Beginn konnten sie die Spielmaterialien gemeinsam mit dem Kind explorieren und auf Wunsch des Kindes in die Testsituation mit eingebunden werden oder das Kind zur Lösung von Aufgaben animieren (Kuchler et al., 2011). Der Bezugsperson wurden ein Elternfragebogen zum Ausfüllen (siehe Anhang D) und eine Einverständniserklärung für die Videoaufzeichnung (siehe Anhang B) zur Unterschrift vorgelegt. Beim ersten Testtermin wurde der Termin für die zweite Testung fixiert. Der Abstand zwischen den beiden Terminen sollte vier bis maximal sechs Wochen betragen. Um im Falle krankheitsbedingter Ausfälle genügend Spielraum für Terminverschiebungen zu behalten, wurde der zweite Termin bereits nach vier Wochen angesetzt. Die Testwiederholung eines jeden Kindes wurde möglichst zur selben Uhrzeit wie die erste Testung durchgeführt. Die Rekrutierung erfolgte über einen Aushang in Kindergärten, Krippen und Spielgruppen in Wien und Umgebung (siehe Anhang A). Ein Teil der Stichprobe umfasste Personen aus dem Bekanntenkreis.

4.2. Setting

Um eine korrekte Messung der Reliabilität zu ermöglichen sind standardisierte Rahmenbedingungen notwendig (Moosbrugger & Kelava, 2012). Alle Kinder wurden im Spielraum der Fakultät für Psychologie der Universität Wien (Arbeitsbereich Klinische und Gesundheitspsychologie) mit den gleichen Materialien getestet. Die Testung und Beobachtung erfolgte zu beiden Zeitpunkten von denselben Personen. Im Raum befanden sich ein Kindertisch mit Kindersesseln, ein Kaufmannsladen mit Tieren, eine Puppenküche, drei Stufen, eine Spielekiste sowie eine Sitzgelegenheit für Beobachterin und Bezugsperson. Die für die Testung benötigten Materialien wurden bei jeder Testung an derselben Stelle im Zimmer platziert. Der Teddy, der zur Kontaktaufnahme mit dem Kind eingesetzt wurde, saß beim Tisch. Dort lagen Stifte und Papier zum Zeichnen bereit. Am Regal oberhalb des Tisches, außerhalb der Sichtweite des Kindes, standen die Perlen, Bausteine und die Materialien zur Gedächtnisüberprüfung bereit. In einer Ecke des Raumes gab es eine Spielekiste mit den Teppichen, Verkleidungsmaterialien und dem Ball. Auf den Stufen lag das Bilderbuch und neben dem Kaufmannsladen das Puzzle. Die Kamera wurde vor dem Fenster positioniert und von der Beobachterin bedient. Alle übrigen Spielmaterialien, die sich im Raum befanden und nicht Teil der Testmaterialien waren, wurden abgedeckt, um eine Ablenkung des Kindes zu vermeiden.

4.3. Stichprobe

Die Stichprobe bestand ursprünglich aus 24 Kindern aus Wien und Umgebung. Drei der Kinder kamen nur zum ersten Testzeitpunkt, aber nicht zum zweiten Termin, und mussten somit aus der Studie ausgeschlossen werden. Die endgültige Stichprobe umfasste damit 21 Kinder.

Zum Zeitpunkt der ersten Erhebung waren die Kinder zwischen 24 und 34 Monaten alt, mit einem Durchschnittsalter von 28.52 Monaten ($SD = 3.11$). Die Geschlechterverteilung war mit 7 Buben (33%) und 14 Mädchen (67%) nicht gleich verteilt (siehe Tabelle 1).

Tabelle 1: Kreuztabelle - Alter und Geschlecht

		Alter in Monaten								Gesamtsumme
		24	25	26	27	28	29	32	34	
Geschlecht	männlich	1	0	1	2	1	0	1	1	7
	weiblich	1	1	2	0	3	4	1	2	14
Gesamtsumme		2	1	3	2	4	4	2	3	21

Alle Kinder wohnten gemeinsam mit Vater und Mutter in einem Haushalt. Insgesamt besuchten 14 Kinder zum Zeitpunkt der ersten Testung den Kindergarten (67%). Außerdem nahmen 11 Kinder (52%) an Sport- oder Freizeitkursen teil. Von den 21 teilnehmenden Kindern waren 13 Einzelkinder (62%), die restlichen 8 Kinder hatten ein bis drei Geschwister.

Nur ein Kind hatte Polnisch als Erstsprache, alle anderen Deutsch (95%). Insgesamt wuchsen sechs der Kinder mehrsprachig auf (29%), vier davon lernten zwei Sprachen und zwei Kinder drei Sprachen.

Alle Elternteile gaben an, dass keine gesundheitlichen Probleme beim Kind bestünden. Bei zwei Kindern lag eine Risikoschwangerschaft vor (10%).

Das Alter der Mütter lag zwischen 25 und 46 Jahren, mit einem Durchschnittsalter von 34.20 Jahren (SD = 6.29). Die Väter waren zwischen 26 und 52 Jahren alt, der Durchschnitt lag bei 37.45 Jahren, wobei ein Elternpaar keine Angaben zum Alter machte.

Im Elternfragebogen wurde überdies die höchste abgeschlossene Ausbildung der Mutter (siehe Tabelle 2) und des Vaters (siehe Tabelle 3) erhoben. Einen Universitätsabschluss hatten 71 Prozent der Mütter und 57 Prozent der Väter.

Tabelle 2: Höchste abgeschlossene Ausbildung der Mütter

	Häufigkeit	Prozent
Lehre/Fachschule	2	10
Matura/Fachhochschule	4	19
Universitätsabschluss	15	71
Gesamtsumme	21	100

Tabelle 3: Höchste abgeschlossene Ausbildung der Väter

	Häufigkeit	Prozent
Lehre/Fachschule	5	24
Matura/Fachhochschule	4	19
Universitätsabschluss	12	57
Gesamtsumme	21	100

4.4. Erhebungsinstrument

Das spielbasierte Verfahren besteht aus einem Itempool zur Erfassung der Entwicklung zweijähriger Kinder. Dieser wurde von Kuchler und Sapper in Zusammenarbeit mit Deimann und Kastner-Koller (2011) entwickelt und von Kronberger (2013) und Punz (2013) letztmals überarbeitet. Das spielbasierte Verfahren ist als allgemeines Entwicklungsverfahren konzipiert und enthält daher Aufgaben zu folgenden Funktionsbereichen:

- Motorische Entwicklung (Grobmotorik, Feinmotorik)
- Sprachliche Entwicklung (Wortschatz, grammatikalische Fähigkeiten)
- Sozial-emotionale Entwicklung (Kind-Bezugsperson-Beziehung, Selbstständigkeit)
- Kognitive Entwicklung (Theory of Mind, numerisches Wissen, Zählen)
- Visuelle Wahrnehmung (Formdifferenzierung, Größendifferenzierung, aktive Farbdifferenzierung, passive Farbdifferenzierung, Stabilität des Farbkonzepts)
- Gedächtnis (phonologisches Gedächtnis, visuelles Gedächtnis)
- Aufmerksamkeit

Das Grundelement der Spielsituation bildet die Puppenküche, in der die einzelnen Items beiläufig beobachtet bzw. von der Testleiterin/vom Testleiter eingebaut werden. Zusätzliche Items werden anhand der Materialien aus einer Spielebox erhoben. Die Vorgabe der einzelnen Items erfolgt individuell angepasst an das jeweilige Kind. Auf die Bedürfnisse und Wünsche des Kindes kann so optimal eingegangen werden. Für den Beginn der Testung wird im Manual jedoch das Spiel in der Puppenküche empfohlen. Der Teddy wird dem Kind vorgestellt, und es wird eingeladen ihm etwas zu kochen, da er Hunger habe. Der Bär kann während der gesamten Spielsituation zur Kontaktaufnahme und Motivation des Kindes eingesetzt werden. Bei der Exploration der Materialien kann die Bezugsperson mit dem Kind gemeinsam die Materialien erkunden. Zwischen dem Spiel in der Puppenküche und den Aufgaben aus der Spielebox kann flexibel gewechselt werden.

Ein Protokollbogen (siehe Anhang C) dient der Kodierung des beobachteten Verhaltens, wobei jeweils das höchste gezeigte Verhalten beurteilt wird. Das Antwortformat der meisten Items ist dichotom. Der Bogen beinhaltet auch Zeitangaben, die erst im Nachhinein mittels Videoaufzeichnung ermittelt werden. Überdies gibt es einen Elternfragebogen, der neben soziodemografischen Daten und Angaben zur Entwicklung des Kindes auch Fragen zur Selbstständigkeit, zum Sozial- und Spielverhalten, zum Wortschatz und zu grammatikalischen Fähigkeiten enthält (siehe Anhang D).

Im Folgenden wird auf die Erfassung der Items zum Entwicklungsbereich der Sprache und der Aufmerksamkeit näher eingegangen. Für die restlichen Funktionsbereiche sei auf Pozniak (in Vorbereitung), Prillinger (in Vorbereitung) und Stark (in Vorbereitung) verwiesen.

4.4.1. Erfassung der sprachlichen Entwicklung

Eine allgemeine Einschätzung der sprachlichen Entwicklung des Kindes erfolgt durch das Zusammenfassen der Bereiche „Wortschatz“ und „Grammatik“.

4.4.1.1. Erfassung des Wortschatzes

Die Skala „Wortschatz“ beinhaltet die zwei Untertests „Wortschatz passiv“ und „Wortschatz aktiv“. Für die Erfassung des passiven und aktiven Wortschatzes werden das Bilderbuch (Theory-of-Mind-Buch) und Spielobst und -gemüse benötigt. Einige Items zum aktiven Wortschatz erfolgen während des Spiels in der Puppenküche, beim Füttern der Tiere. Die restlichen Items werden bei der Bilderbuchbetrachtung vorgegeben. Der Zeitpunkt der Vorgabe hängt von der Bereitschaft des Kindes ab.

Beim passiven Wortschatz soll das Kind Gegenstände, die im Bilderbuch abgebildet sind, suchen. Das Kind wird beispielsweise gefragt: „Findest du die Badewanne?“ Insgesamt werden acht Items vorgegeben.

Das Kind erhält für jede richtige Reaktion einen Punkt, unabhängig davon, ob es den Gegenstand beim Buchlesen oder zu einem anderen Zeitpunkt gezeigt hat. Eine falsche Reaktion wird mit null Punkten kodiert.

Zur Erfassung des aktiven Wortschatzes wird neben dem Bilderbuch auch das Spielobst und -gemüse herangezogen. Das Kind wird nach den Obst- und Gemüseamen, den Zimmern im Bilderbuch und nach verschiedenen Gegenständen gefragt, die es benennen soll. Der Untertest besteht aus 17 Items.

Die richtigen Begriffe werden mit eins kodiert. Falsche Aussprachen werden als richtig gewertet, sofern erkennbar ist, dass das Kind das richtige Wort meint. Wie beim passiven Wortschatz werden richtige Antworten auch zu einem späteren Zeitpunkt gewertet. Es wird empfohlen, alle Aussagen des Kindes zu protokollieren. Antworten in einer anderen Sprache zählen ebenfalls als richtig, sofern die Bezugsperson die Antwort bestätigt.

Die Bezugsperson kann insofern mithelfen, als sie das Kind nach den Obst- und Gemüseamen fragt, wenn das Kind der Testleiterin/dem Testleiter nicht antwortet. Auch das Buch kann mit

der Bezugsperson betrachtet werden, vorausgesetzt die Items werden so vorgelesen, wie sie im Buch stehen.

4.4.1.2. Erfassung der grammatikalischen Fähigkeiten

Die Erfassung der grammatikalischen Fähigkeiten des Kindes erfolgt während der gesamten Spielsituation, ohne zusätzliche Materialien. Das Gesprochene des Kindes wird festgehalten, und die grammatikalischen Fertigkeiten werden anschließend anhand der Aussagen bewertet und im Protokollbogen kodiert. Insgesamt werden 17 Items erhoben, wobei auch Teilpunkte vergeben werden.

4.4.2. Erfassung der Aufmerksamkeit

Die Erfassung der Aufmerksamkeitsleistung erfolgt durch die Beobachtung des Verhaltens des Kindes während des Bauens mit den Bausteinen, des Zeichnens, des Perlenauffädelns und der Bilderbuchbetrachtung des Theory-of-mind-Buches. Es werden keine zusätzlichen Materialien verwendet, sondern lediglich jene, die für die Durchführung der jeweiligen Items erforderlich sind. Die Aufmerksamkeitsleistung des Kindes wird nicht während der Spielzeit bewertet, sondern auf Basis der Videoaufzeichnungen. Das Kind erhält einen Punkt, wenn es bei der Aufgabe die ganze Zeit über aufmerksam war. War das Kind unaufmerksam und unterbrach die Tätigkeit, ließ sich aber wieder zurückholen, wird mit 0.5 Punkten kodiert. Null Punkte erhält das Kind, wenn es unaufmerksam war und sich nicht mehr zur eigentlichen Aufgabe zurückführen ließ. Beim Bilderbuch wird zusätzlich die Seitenanzahl angegeben, die das Kind aufmerksam betrachtet hat. Für jeden Bereich wird die Dauer, in der sich das Kind aufmerksam der Aufgabe widmete, in Sekunden angegeben. Außerdem wird die Zeit bis zur ersten Unterbrechung vermerkt. Diese Angaben werden nur qualitativ erhoben und gehen nicht in die Berechnungen ein.

5. Ergebnisdarstellung

Die statistische Auswertung erfolgte mittels SPSS 22. Nach der Darstellung der Ergebnisse deskriptiver Analysen und fehlender Werte zu beiden Testzeitpunkten, werden die Ergebnisse der Hauptfragestellung zur Retest-Reliabilität dargestellt. Im Anschluss erfolgt der Vergleich der beiden Messzeitpunkte. Zum Schluss erfolgt die Analyse auf Skalen- und Itemebene für beide Erhebungszeitpunkte.

Vor den Berechnungen erfolgte die Überprüfung der Normalverteilung mittels standardisierter Schiefe (siehe Anhang E). Bei kleinen Stichproben entspricht ein Wert unter 1.96 ($z < \pm 1.96$) der Normalverteilung (Field, 2009). Es wurde zweiseitig getestet und das Signifikanzniveau auf $p \leq .05$ festgelegt.

5.1. Ergebnisse der deskriptiven Angaben

Aufgrund der Videoaufzeichnung der Testungen konnten Zeitangaben ermittelt werden, die deskriptiv analysiert wurden. Durchschnittlich dauerte eine Testung 107 Minuten ($M = 107$; $SD = 15.90$). Die Gesamtdauer wird vom Betreten bis zum Verlassen des Testraumes bemessen. Die längste Testung nahm 133 Minuten und die kürzeste 79 Minuten in Anspruch. Beim zweiten Durchgang betrug die durchschnittliche Testdauer 103 Minuten ($M = 102.57$; $SD = 12.43$), die längste dauerte 131 Minuten und die kürzeste 80 Minuten.

Da die Normalverteilung der Differenz der gesamten Testdauer nicht gegeben ist (siehe Anhang E.1), wurde ein parameterfreies Verfahren durchgeführt. Der Zusammenhang zwischen der Gesamtdauer beim ersten und zweiten Termin wurde mittels Spearman-Rangkorrelation berechnet. Es besteht ein mäßiger Zusammenhang, mit einem $r_s = .553$, $p = .009$. Mittels Wilcoxon-Vorzeichen-Rangtest kann ein signifikantes Ergebnis ermittelt werden ($p = .028$). Die Gesamtdauer des Tests war zum zweiten Zeitpunkt signifikant kürzer als beim ersten.

Ebenso wurde die Anlaufzeit gemessen. Das ist die Zeit, die das Kind vor der ersten Vorgabe eines Items zur Exploration der Spielmaterialien benötigte. Die Kinder brauchten unterschiedlich lange für die Exploration, von weniger als eine Minute bis maximal zehn Minuten. Die durchschnittliche Anlaufzeit betrug beim ersten Durchgang drei Minuten ($M = 3.05$; $SD = 2.06$). Die Anlaufzeiten beim zweiten Durchgang waren sehr ähnlich, mit einer längsten Dauer von 11 Minuten. Durchschnittlich benötigten die Kinder bei der

Testwiederholung ebenfalls drei Minuten bis zur eigentlichen Testdurchführung ($M = 3.05$; $SD = 2.50$).

Als Pausen werden jene Phasen bezeichnet, in denen sich das Kind nicht mit den für die Testung vorgesehenen Materialien beschäftigt, sondern beispielsweise isst, trinkt, auf das WC geht oder mit eigenem Spielzeug spielt. Von den 21 Kindern benötigten beim ersten Termin acht Kinder keine Pause (38%), acht Kinder nahmen eine Pause in Anspruch (38%), vier Kinder benötigten zwei Unterbrechungen (19%) und ein Kind drei (5%). Durchschnittlich dauerte eine Pause vier Minuten ($M = 4.40$; $SD = 5.30$). Bei der Testwiederholung benötigten zehn Kinder keine (48%), fünf Kinder eine (24%) und sechs Kinder zwei Unterbrechungen (29%). Die durchschnittliche Pausendauer betrug fünf Minuten ($M = 5.00$; $SD = 6.94$).

Die Zeit, die zur Verabschiedung benötigt wurde, variierte von 5 bis 40 Minuten beim ersten Testdurchgang. Durchschnittlich dauerte die Verabschiedung 13 Minuten ($M = 13.24$; $SD = 8.65$). Bei den zweiten Testterminen dauerte die Verabschiedung zwischen 5 und 35 Minuten, mit einer Durchschnittszeit von 15 Minuten ($M = 14.67$; $SD = 7.02$).

Die durchschnittliche Beschäftigungsdauer mit den Materialien, also die Gesamtdauer minus Anlaufzeit und Verabschiedung, betrug bei den ersten Testterminen zwischen 62 und 107 Minuten. Durchschnittlich dauerte die Materialbeschäftigung 86 Minuten ($M = 86.29$; $SD = 13.29$). Bei der Testwiederholung variierte die Dauer der Materialbeschäftigung von 69 bis 94 Minuten, mit einem Durchschnittswert von 80 Minuten ($M = 80.05$; $SD = 6.83$).

Hinsichtlich der Beschäftigung mit den Materialien zeigt sich beim t-Test mit einer Prüfgröße von $t(20) = 2.150$, $p = .044$ ein signifikantes Ergebnis. Die Kinder beschäftigten sich beim zweiten Termin signifikant kürzer mit den Materialien. Ein Zusammenhang zwischen den beiden Messzeitpunkten besteht nicht ($r = .256$, $p = .263$).

Bezüglich Anlaufzeit, Pausendauer und Verabschiedungszeit konnten keine signifikanten Unterschiede oder Zusammenhänge festgestellt werden (siehe Anhang F).

5.2. Analyse fehlender Werte

Fehlende Werte ergaben sich bei der Erfassung der Sprachentwicklung ausschließlich bei der Betrachtung des Bilderbuches. Es kam öfters vor, dass die Zweijährigen im Bilderbuch vorzeitig umblättern und Items aufgrund der Unaufmerksamkeit des Kindes nicht vorgegeben wurden. In Tabelle 4 sind die Häufigkeiten und Prozentanteile fehlender Werte der einzelnen Items für beide Testzeitpunkte dargestellt. Kritisch anzumerken ist, dass bei der Erfassung sowohl des passiven als auch aktiven Wortschatzes zum Teil nicht ersichtlich war, ob das Kind die Lösung verweigerte oder die Lösung nicht wusste. Die dargestellten fehlenden Werte beziehen sich daher nur auf jene Items, bei denen das Kind augenscheinlich unaufmerksam war oder Items gar nicht vorgegeben wurden. Meistens wurden bei beiden Testterminen dieselben Items verweigert. Auffallend waren die Verweigerungen bei der Befragung nach den Zimmern. Es hatte den Anschein, dass viele Kinder die Instruktion nicht verstanden. Die meisten Kinder zeigten bei der Frage, in welchem Zimmer sich die Menschen befinden, auf die Personen am Bild und kommentierten dies mit „da“. Beim Untertest „Grammatik“ können keine fehlenden Werte entstehen, da nur das Geäußerte der Kinder kodiert wird.

Tabelle 4: Häufigkeiten fehlender Messwerte zu T1 und T2

Item	Testzeitpunkt 1		Testzeitpunkt 2	
	Häufigkeit	Prozent	Häufigkeit	Prozent
<i>Wortschatz aktiv</i>				
Küche benennen	2	10	4	19
Torte/Käse	1	5	1	5
Wohnzimmer benennen	3	14	2	10
Kinderzimmer benennen	3	14	4	19
Schuh	1	5	1	5
Schlafzimmer benennen	5	24	5	24
Kasten	3	14	2	10
Tasche	4	19	5	24
Badezimmer benennen	2	10	6	29
Wasser	3	14	4	19
Lampe			1	5
<i>Wortschatz passiv</i>				

Fernseher	1	5		
Tür	1	5		
Gitterbett	1	5		
Fenster	2	10		
Bild	4	19	3	14
Badewanne	1	5	2	10

Beim Untertest zur Erfassung der Aufmerksamkeitsleistung des Kindes gab es bei beiden Testzeitpunkten nur einen fehlenden Wert, da jeweils ein Kind die Aufgabe „Perlen fädeln“ verweigerte. Somit konnte bei diesem Item keine Aussage über die Aufmerksamkeitsfähigkeit des Kindes getroffen werden.

Um Datenverlust zu vermeiden wurden alle fehlenden Werte für die weiteren Berechnungen als „nicht gelöst“ (0) kodiert.

5.3. Retest-Reliabilitäten der beiden Messzeitpunkte

Die zweimalige Durchführung des spielbasierten Verfahrens im Abstand von vier bis sechs Wochen erlaubt die Berechnung der Retest-Reliabilität. Hierfür wurde die Korrelation zwischen den beiden Testergebnissen für die Skalen „Sprache“ und „Aufmerksamkeit“ sowie die Subskalen „Wortschatz“ und „Grammatik“ ermittelt. Zur Überprüfung der Messwertdifferenzen auf Normalverteilung für die beiden Skalen sowie deren Subskalen wurde die Methode der standardisierten Schiefe durchgeführt (Field, 2009). Die Messwertdifferenzen aller Skalen sind normalverteilt (siehe 0; Anhang E.3), sodass die Berechnungen mittels Pearson Korrelation durchgeführt werden konnten.

5.3.1.1. Retest-Reliabilität der Sprachskala

Die gesamte Sprachskala weist eine akzeptable Retest-Reliabilität mit einem Wert von $r_{tt} = .791$ auf ($p < .001$). Die Retest-Reliabilität der Subskala „Wortschatz“ zeigt ebenfalls einen akzeptablen Wert von $r_{tt} = .774$ ($p < .001$). Die Subskala „Grammatik“ fällt etwas geringer aus, liegt aber mit einem Wert von $r_{tt} = .726$ ($p < .001$) ebenfalls im zufriedenstellenden Bereich.

5.3.1.2. Retest-Reliabilität der Aufmerksamkeitsskala

Die Retest-Reliabilität der Aufmerksamkeitsskala fällt mit einem Wert von $r_{tt} = .457$ ($p = .037$) nicht zufriedenstellend aus.

5.4. Vergleich der beiden Messzeitpunkte

Um Veränderungen zwischen dem ersten Testzeitpunkt (T1) und der Testwiederholung (T2) aufzuzeigen, wurden t-Tests für verbundene Stichproben berechnet.

Die Prüfgröße des t-Tests für die gesamte Sprachskala fällt mit $t(20) = 3.150$, $p = .005$ signifikant aus. Es kann eine Veränderung der sprachlichen Fähigkeiten zwischen Testzeitpunkt eins und zwei angenommen werden. Die Kinder erzielten beim zweiten Testzeitpunkt in allen Subskalen höhere Werte (siehe Tabelle 5).

Tabelle 5: Mittelwertunterschiede der Skala „Sprache“ T1 und T2

	Test- zeitpunkt	Mittelwert	N	Standard- abweichung	Standardfehler Mittelwert	p-Wert
Passiver Wortschatz	1	5.714	21	1.848	0.403	
	2	6.095	21	2.047	0.447	.321
Aktiver Wortschatz	1	6.381	21	3.598	0.785	
	2	7.667	21	3.941	0.860	.016
Wortschatz	1	12.095	21	4.206	0.918	
	2	13.762	21	4.939	1.078	.025
Grammatik	1	7.199	21	4.641	1.013	
	2	9.238	21	4.840	1.057	.015
Sprachskala	1	19.294	21	7.778	1.697	
	2	23.000	21	8.681	1.894	.005

Der t-Test weist mit einer Prüfgröße von $t(20) = 1.017$, $p = .321$ keinen signifikanten Unterschied bei der Subskala „Wortschatz passiv“ auf. Es kann keine statistisch signifikante Veränderung beim passiven Wortschatz festgestellt werden. Mit einer Prüfgröße von $t(20) = 2.631$; $p = .016$ zeigt sich ein signifikanter Unterschied beim „Wortschatz aktiv“. Bei der Testwiederholung erzielten die Kinder signifikant bessere Leistungen beim aktiven Wortschatz. Insgesamt fällt die Prüfgröße beim Untertest „Wortschatz“ signifikant aus ($t(20) = 2.423$, $p = .025$). Ein Leistungszuwachs beim Wortschatz zwischen den beiden Testzeitpunkten kann angenommen werden. Bei der Subskala „Grammatik“ zeigt sich ebenfalls ein signifikantes Ergebnis ($t(20) = 2.661$; $p = .015$). Die Kinder erzielten signifikant bessere Leistungen hinsichtlich der grammatikalischen Fähigkeiten beim zweiten Testzeitpunkt.

Die Ergebnisse der Mittelwertunterschiede der Skala „Aufmerksamkeit“ sind in Tabelle 6 dargestellt. Mit einer Prüfgröße von $t(20) = 1.04$, $p = .311$ zeigt sich kein signifikanter Unterschied zwischen erstem und zweitem Testzeitpunkt.

Tabelle 6: Mittelwertunterschiede der Skala „Aufmerksamkeit“ T1 und T2

	Test-zeitpunkt	Mittelwert	N	Standard-abweichung	Standardfehler Mittelwert	p-Wert
Aufmerk-samkeit	2	5.524	21	6.284	0.137	
	1	5.357	21	0.763	0.167	.311

5.5. Skalen- und Itemanalyse

Die Skalen „Sprache“ und „Aufmerksamkeit“ wurden einer Skalen- und Itemanalyse unterzogen. Beim Fähigkeitsbereich der Sprache wurden zusätzlich die Subskalen „Wortschatz aktiv“ und „Wortschatz passiv“ sowie die Subskala „Grammatik“ analysiert.

Der Koeffizient Alpha nach Cronbach, die korrigierte Itemtrennschärfe und die relative Lösungswahrscheinlichkeit wurden zur Skalen- und Itemanalyse herangezogen.

5.5.1. Fähigkeitsbereich Sprache

Für die gesamte Sprachskala sowie für die Subskalen des aktiven und passiven Wortschatzes und der Grammatik wurde eine Skalen- und Itemanalyse durchgeführt. Um einen Datenverlust zu vermeiden, wurden fehlende Werte auf null gesetzt. Für die gesamte Skala „Sprache“ kann zum ersten Testzeitpunkt ein Alpha-Koeffizient nach Cronbach von $\alpha = .896$ berechnet werden. Insgesamt flossen 40 Items in die Berechnung ein. Acht Items wurden aufgrund einer Varianz von null von der Berechnung ausgeschlossen.

Beim zweiten Testzeitpunkt ergibt sich ein höherer Alpha-Koeffizient nach Cronbach von $\alpha = .915$. Im Unterschied zum ersten Testzeitpunkt wurden nur fünf Items aufgrund ihrer Varianz von null ausgeschlossen.

5.5.1.1. Wortschatz aktiv

Die Subskala „Wortschatz aktiv“ erreicht einen Reliabilitätskoeffizient von $\alpha = .827$ zum ersten Zeitpunkt und einen Reliabilitätskoeffizient von $\alpha = .849$ zum zweiten Termin. Die Items „Wohnzimmer benennen“, „Kinderzimmer benennen“ und „Schlafzimmer benennen“ wurden beim ersten Testzeitpunkt von keinem Kind gelöst und gehen somit nicht in die Konsistenzanalyse ein. Die Itemkennwerte sind in Tabelle 7 dargestellt. Der Großteil der Items

liegt im adäquaten Schwierigkeitsbereich. Lediglich das Item „Küche benennen“ war für die Kinder zu schwierig und liegt unter der empfohlenen Itemschwierigkeit von .20 (Bortz & Döring, 2006). Das Item „Kasten“ liegt mit .19 knapp unter dem empfohlenen Wert, und das Item „Banane“ mit .81 knapp darüber. Bis auf die Items „Torte/Käse“ und „Wasser“ weisen alle Items mittlere bis hohe Trennschärfen auf und sind daher für den Gesamtwert der Skala repräsentativ (siehe Tabelle 7).

Tabelle 7: Itemkennwerte der Subskala „Wortschatz aktiv“ T1

	Itemlösungs- wahrscheinlich- keit (p_i)	Korrigierte Trennschärfe (r_{it})	Korrigierte Trennschärfe nach Selektion
Apfel	.71	.483	.814
Birne	.43	.617	.804
Banane	.81	.646	.805
Zitrone	.29	.639	.803
Karotte	.62	.618	.804
Orange	.29	.363	.823
Küche benennen	.10	.504	.816
Torte/Käse	.52	.243	.832
Lampe	.57	.327	.826
Schuh	.71	.382	.821
Kasten	.19	.525	.812
Tasche	.33	.324	.826
Badezimmer benennen	.29	.603	.806
Wasser	.52	.273	.830

Die Items „Küche benennen“, „Wohnzimmer benennen“, „Kinderzimmer benennen“ und „Schlafzimmer benennen“ stellten sich auch zum zweiten Testzeitpunkt als zu schwierig für die Kinder heraus und konnten nur von 10% der Kinder gelöst werden. Beim zweiten Testlauf gingen alle Werte in die Berechnungen ein. Das Item „Badezimmer benennen“ liegt mit einem Wert von .19 knapp unter dem empfohlenen Wert von .20 und die Items „Banane“ und „Schuh“ mit .81 knapp über dem empfohlenen Wert von .80 (Bortz & Döring, 2006). Die Items der Subskala weisen bis auf die Items „Küche benennen“, „Schuh“ und „Tasche“ mittlere bis hohe Trennschärfen auf und können daher den Gesamtwert gut vorhersagen (Tabelle 8).

Tabelle 8: Itemkennwerte der Subskala „Wortschatz aktiv“ T2

	Itemlösungs- wahrscheinlich- keit (p_i)	Korrigierte Trennschärfe (r_{it})	Korrigierte Trennschärfe nach Selektion
Apfel	.62	.417	.844
Birne	.43	.746	.824
Banane	.81	.587	.835
Zitrone	.24	.361	.846
Karotte	.71	.679	.829
Orange	.48	.568	.835
Küche benennen	.10	.122	.854
Torte/Käse	.76	.382	.845
Wohnzimmer benennen	.10	.476	.842
Lampe	.76	.611	.833
Kinderzimmer benennen	.10	.386	.845
Schuh	.81	.275	.850
Schlafzimmer benennen	.10	.386	.845
Kasten	.29	.551	.836
Tasche	.57	.257	.853
Badezimmer benennen	.19	.504	.839
Wasser	.62	.474	.841

5.5.1.2. Wortschatz passiv

Die interne Konsistenz des Subtests „Wortschatz passiv“ ist mit einem Reliabilitätskoeffizienten von $\alpha = .606$ beim ersten Testzeitpunkt wenig zufriedenstellend. Diese Subskala besteht jedoch insgesamt nur aus acht Items. Beim zweiten Erhebungszeitpunkt zeigt sich ein besseres Cronbach Alpha von $\alpha = .770$.

Beim ersten Zeitpunkt scheinen die Items „Tür“, „Sessel“, „Fenster“ und „Badewanne“ für die Stichprobe eher leicht gewesen zu sein. Mit einem Wert von .81 liegen sie knapp über der bevorzugten Itemschwierigkeit. Auffallend sind die sehr geringen Trennschärfen der Items „Tür“, „Sessel“, „Gitterbett“ und „Badewanne“. Diese tragen nicht zur Repräsentativität des Gesamtwertes bei (siehe Tabelle 9).

Tabelle 9: Itemkennwerte der Subskala „Wortschatz passiv“ T1

	Itemlösungs- wahrscheinlich-keit (p_i)	Korrigierte Trennschärfe (r_{it})	Korrigierte Trennschärfe nach Selektion
Besen	.71	.402	.543

Fernseher	.62	.504	.504
Tür	.81	.265	.584
Sessel	.81	.265	.584
Gitterbett	.48	.151	.625
Fenster	.81	.600	.491
Bild	.67	.411	.539
Badewanne	.81	.093	.670

Das Ergebnis bezüglich der Itemschwierigkeiten konnte beim zweiten Testzeitpunkt bestätigt werden. Das Item „Fenster“ liegt nach wie vor mit einem Wert von .81 knapp über der empfohlenen Itemschwierigkeit (Bortz & Döring, 2006). Die anderen drei wurden von 90% der Kinder gelöst und waren somit zu einfach für die Stichprobe (siehe Tabelle 10).

Die in Tabelle 10 dargestellten Trennschärfen, liegen beim zweiten Termin im mittleren und hohen Bereich. Lediglich das Item „Fernseher“ weist eine niedrige Itemtrennschärfe unter .30 auf.

Tabelle 10: Itemkennwerte der Subskala „Wortschatz passiv“ T2

	Itemlösungs- wahrscheinlichkeit (p_i)	Korrigierte Trennschärfe (r_{it})	Korrigierte Trennschärfe nach Selektion
Besen	.67	.533	.734
Fernseher	.76	.284	.778
Tür	.90	.669	.724
Sessel	.90	.381	.760
Gitterbett	.52	.363	.771
Fenster	.81	.644	.715
Bild	.62	.506	.740
Badewanne	.90	.570	.737

Insgesamt erreicht die Subskala „Wortschatz“, bestehend aus den Aufgaben zum aktiven und passiven Wortschatz, einen Reliabilitätskoeffizienten von $\alpha = .775$ zum ersten und einen Koeffizienten von $\alpha = .850$ zum zweiten Erhebungszeitpunkt. Während beim zweiten Erhebungszeitpunkt alle 25 Items in die Berechnung eingingen, mussten beim ersten Zeitpunkt drei Items ausgeschlossen werden.

5.5.1.3. Grammatik

Die Subskala „Grammatik“ verfügt mit einem Alpha-Koeffizienten von $\alpha = .905$ über eine hohe Reliabilität zum ersten Testzeitpunkt. Die Items „Relativpronomen wo“, „Relativsatz ohne Relativpronomen“, „korrekte Partizipendung“, „Vorkommen von Alternativfragen“, „korrekte Wortstellung Alternativfragen“ wurden von keinem Kind gelöst und weisen somit eine relative Lösungswahrscheinlichkeit von null auf. Diese Items gehen daher nicht in die Berechnungen der Konsistenzanalyse ein.

Die Anzahl an Verbindungswörtern von Haupt- und Nebensatz und die Anzahl an verwendeten Relativpronomen wurden qualitativ erhoben und gehen nicht in die Analyse einher. Ein korrektes Relativpronomen wurde beim ersten Erhebungszeitpunkt lediglich von einem Kind verwendet. Insgesamt drei der Kinder verwendeten Verbindungswörter bei der Bildung von Haupt- und Nebensatz. Ein Kind verwendete beim ersten Testzeitpunkt bereits drei verschiedene Verbindungswörter (und, weil, ob), ein Kind zwei (und, wenn) und das dritte Kind ein Verbindungswort (und).

Beim ersten Zeitpunkt gibt es insgesamt sieben Items, die zu schwer für die Kinder der Stichprobe sind und somit außerhalb der adäquaten Schwierigkeit liegen. Das Item „Vorkommen von Verneinung“ ist zu einfach, bereits 86% der Kinder bildeten eine Verneinung (siehe Tabelle 11).

Tabelle 11: Itemkennwerte der Subskala „Grammatik“ T1

	Itemlösungs- wahrscheinlichkeit (p_i)	Korrigierte Trennschärfe (r_{it})	Korrigierte Trennschärfe nach Selektion
Satzlänge	.79	.680	.899
Subjekt, Prädikat, Objekt	.52	.725	.895
Kombi Haupt-/Nebensatz	.14	.528	.901
Vorkommen von Relativsätzen	.05	.344	.906
korrektes Relativpronomen	.05	.344	.906
korrekte Verbstellung Aussage	.57	.771	.893
Vorkommen eines Perfekts	.14	.480	.903
Perfekt mit „ge-“ gebildet	.14	.480	.903
Hilfsverb bei Perfektbildung	.09	.473	.903
Vorkommen von Verneinungen	.86	.481	.903
korrekte Wortstellung Verneinung	.43	.685	.896
Bildung einer Verneinung	.71	.595	.900
Frageintonation	.57	.453	.905

Vorkommen von Ja-Nein-Fragen	.33	.384	.907
Inversion Subjekt/Objekt	.19	.684	.897
Vorkommen von W-Fragen	.57	.733	.895
Inversion Subjekt/Objekt	.52	.737	.894
Unterscheidung W-Fragen	.50	.536	.902

Die Trennschärfen beim ersten Zeitpunkt liegen alle im mittleren oder hohen Bereich, was für eine gute Repräsentativität des Gesamtwertes spricht (siehe Tabelle 11).

Beim zweiten Erhebungszeitpunkt fällt der Reliabilitätskoeffizient mit einem Wert von $\alpha = .898$ etwas geringer als beim ersten Testdurchgang aus. Bis auf das Item „korrekte Partizipendung“ wurden dieselben Items wie beim ersten Durchgang ausgeschlossen, da sie von keinem Kind gelöst werden konnten. Überdies ging das Item „Vorkommen von Verneinungen“ nicht mehr in die Analyse ein, da bereits alle Kinder eine Verneinung im Sprachgebrauch verwendeten und das Item somit eine Lösungswahrscheinlichkeit von 1 aufwies. Lediglich vier Werte liegen unterhalb des adäquaten Schwierigkeitsbereichs und wurden von nur wenigen Kindern gelöst. Die Items „Satzlänge“ und „Bildung Verneinung“ waren zu einfach für die Kinder. 90% der Stichprobe verwendeten Äußerungen mit zumindest vier Wörtern und 86% der Kinder bildeten eine Verneinung mit „nicht“ (siehe Tabelle 12).

Beim zweiten Erhebungszeitpunkt wurde von drei Kindern ein korrektes Relativpronomen verwendet. Insgesamt benutzten wieder dieselben drei Kinder wie im ersten Durchgang Verbindungswörter. Ein Kind gebrauchte vier Verbindungswörter (und, weil, dass, wenn), ein Kind drei (und, weil, wenn) und ein Kind abermals ein Verbindungswort (und).

Bis auf das Item „Bildung Verneinung“ liegen alle Trennschärfen im mittleren oder hohen Bereich (siehe Tabelle 12).

Tabelle 12: Itemkennwerte der Subskala „Grammatik“ T2

	Itemlösungs- wahrscheinlich keit (p_i)	Korrigierte Trennschärfe (r_{it})	Korrigierte Trennschärfe nach Selektion
Satzlänge	.90	.350	.898
Subjekt, Prädikat, Objekt	.62	.497	.894
Kombi Haupt-/Nebensatz	.38	.556	.892
Vorkommen von Relativsätzen	.14	.486	.894
korrektes Relativpronomen	.14	.486	.894
korrekte Verbstellung Aussage	.67	.493	.894

Vorkommen eines Perfekts	.33	.609	.890
Perfekt mit „ge-“ gebildet	.29	.639	.889
korrekte Partizipendung	.19	.506	.894
Hilfsverb bei Perfektbildung	.19	.576	.892
korrekte Wortstellung Verneinung	.57	.429	.897
Bildung Verneinung	.86	.217	.900
Frageintonation	.67	.521	.893
Vorkommen Ja-Nein-Fragen	.38	.623	.890
Inversion Subjekt/Objekt	.33	.770	.885
Vorkommen von W-Fragen	.57	.600	.891
Inversion Subjekt/Objekt	.48	.634	.889
Unterscheidung W-Fragen	.52	.694	.888

5.5.2. Fähigkeitsbereich Aufmerksamkeit

Die Skala zur Erfassung der Aufmerksamkeitsfähigkeit des Kindes besteht aus nur sechs Items. Beim ersten Erhebungszeitpunkt beträgt die Reliabilität $\alpha = .536$. Beim zweiten Termin ist die interne Konsistenz der Aufmerksamkeitsskala mit einem Reliabilitätskoeffizienten von $\alpha = .281$ noch weniger zufriedenstellend. Während beim ersten Termin fünf Items außerhalb des adäquaten Schwierigkeitsbereichs liegen (siehe Tabelle 13), fallen bei der zweiten Erhebung alle Items mit einer Itemschwierigkeit von über .80 aus und sind somit für die Kinder zu einfach (siehe Tabelle 14).

Die Trennschärfen der Items „Aufmerksamkeit Puzzle“ und „Aufmerksamkeit ToM“ liegen beim ersten Erhebungszeitpunkt nicht im zufriedenstellenden Bereich (siehe Tabelle 13).

Tabelle 13: Itemkennwerte der Skala „Aufmerksamkeit“ T1

	Itemlösungs- wahrscheinlichkeit (p)	Korrigierte Trennschärfe (r_{it})	Korrigierte Trennschärfe nach Selektion
Aufmerksamkeit Zeichnen	.93	.684	.334
Aufmerksamkeit Turmbau	.98	.517	.461
Aufmerksamkeit Perlen	.79	.289	.538
Aufmerksamkeit Puzzle	.90	.071	.597
Aufmerksamkeit ToM	.81	.175	.545
ToM Buchseiten	.95	.533	.458

Beim zweiten Termin waren alle Kinder beim Turmbauen durchwegs aufmerksam, somit hat das Item „Aufmerksamkeit Turmbau“ eine Lösungswahrscheinlichkeit von eins und wurde von

der Analyse ausgeschlossen. Die Items „Aufmerksamkeit Zeichnen“ und „Aufmerksamkeit Perlen“ weisen eine negative Trennschärfe unter .20 auf (siehe Tabelle 14).

Tabelle 14: Itemkennwerte der Skala „Aufmerksamkeit“ T2

	Itemlösungs- wahrscheinlich- keit (p)	Korrigierte Trennschärfe (r_{it})	Korrigierte Trennschärfe nach Selektion
Aufmerksamkeit Zeichnen	.91	-.162	.490
Aufmerksamkeit Perlen	.95	-.157	.450
Aufmerksamkeit Puzzle	.91	.433	-.175
Aufmerksamkeit ToM	.81	.423	-.148
ToM Buchseiten	.95	.327	.167

6. Diskussion

In dieser Arbeit wurden eingangs die Retest-Reliabilität sowie Unterschiede zwischen den Ergebnissen der beiden Messzeitpunkte des vorgegebenen Itempools untersucht. Im Anschluss erfolgte eine Skalen- und Itemanalyse zur testtheoretischen Absicherung des Verfahrens.

Der spielbasierte Entwicklungstest weist bei der gesamten Sprachskala eine zufriedenstellende Retest-Reliabilität auf ($r_{tt} = .791$; $p < .001$). Die Subskalen „Wortschatz“ und „Grammatik“ liegen mit Werten über $r_{tt} = .70$ ebenso im akzeptablen Bereich. Die mit dem Verfahren erhobenen sprachlichen Fertigkeiten können, trotz interindividueller Entwicklungsdynamik, über einen Zeitraum von vier bis sechs Wochen relativ zuverlässig abgebildet werden. Die Kinder entwickeln ihre Fähigkeiten während dieser Zeitspanne zwar weiter, aber ihre Entwicklungsschritte scheinen innerhalb der gegenständlichen Stichprobe weitgehend gleich zu sein. Die Kinder erzielen bei der Messwiederholung in allen Subskalen zur Überprüfung der Sprachentwicklung höhere Werte. Die Ergebnisse beim Wortschatz, bei den grammatikalischen Fähigkeiten sowie bei der gesamten Sprachskala sind beim zweiten Messzeitpunkt signifikant höher. Lediglich beim passiven Wortschatz verändert sich das Leistungsniveau statistisch nicht nachweislich.

Die Werte der Konsistenzanalyse der gesamten Sprachskala sowie der Untertests „Wortschatz“ und „Grammatik“ liegen weitgehend über $r = .80$, sodass eine ausreichende instrumentelle Messgenauigkeit angenommen werden kann. Die Analyse zeigt, dass in der gesamten Sprachskala sowie in den meisten Untertests die gewonnenen Reliabilitätskoeffizienten bei der Testwiederholung höher sind. Dies rechtfertigt die Annahme, dass der Messfehler bei der zweiten Testung geringer ausfällt. Dieses Ergebnis könnte zustande gekommen sein, weil Störeinflüsse bei der Durchführung der zweiten Testungen geringer gehalten wurden. Da die Kinder bereits mit dem Testraum und der Spielsituation vertraut waren, konnten kindbezogene Einflüsse eher vermieden werden. Überdies ermöglichte die Testwiederholung der Testleiterin verstärkt auf Bedürfnisse des Kindes zu reagieren und mit Empfindungen umzugehen. Die Testleiterin konnte somit besser auf das Kind eingehen und eher die bestmögliche Leistung des Kindes abrufen. Das Ergebnis könnte aber auch darauf zurückgeführt werden, dass womöglich eine stärkere Heterogenität der

wahren Werte gegeben war. Der Reliabilitätskoeffizient fällt bei heterogenen Gruppen höher aus (Schermelleh-Engel & Werner, 2012).

Im Abschnitt 1.5.3 wurde das Problem des Übungs- und Lerneffektes bei Testwiederholungen thematisiert. Möglicherweise können die Resultate dadurch erklärt werden. Dadurch, dass den Kindern die Lösungen nicht genannt und die Aufgaben im spielerischen Kontext gestellt werden, ist eher unwahrscheinlich, dass bei den Kindern Lerneffekte entstehen. Es ist daher möglich, das Verfahren mehrfach beim selben Kind anzuwenden, zum Beispiel im Falle einer Verlaufskontrolle. Dennoch könnten sich solche Effekte aufgrund der Anwesenheit der Bezugsperson einstellen und sich negativ auf das Testergebnis auswirken. Trotz des Hinweises zu Beginn der Testung, dass die Kinder noch nicht alle Aufgaben lösen müssen, hatten manche Eltern das Bedürfnis, die Leistungen des Kindes zu rechtfertigen. Es kann daher nicht ausgeschlossen werden, dass sie mit dem Kind zwischen den beiden Testterminen einzelne Aufgaben übten, wodurch die Ergebnisse verzerrt worden sein könnten.

Die Ergebnisse sprechen jedenfalls für eine zweimalige Testdurchführung, da Beurteilungen der Reliabilität, aufgrund einer einmaligen Messung, die Zuverlässigkeit der ermittelten Werte überschätzen können (Renner, 2009). Um Übungs- und Lerneffekte zu vermeiden empfiehlt sich, einen gemeinsamen Spieltermin mit Bezugsperson und Kind als erstes Kennenlernen vor der eigentlichen Vorgabe des Entwicklungstests, zu vereinbaren. So kann das Kind mit der Situation, sowie mit der Umgebung und der Testleiterin/dem Testleiter vertraut werden, wodurch Störeinflüsse bei der eigentlichen Testung vermindert werden können. Dies erscheint auch notwendig zu sein, da gerade jüngere Kinder bei einer Testung den sprachlichen Kontakt mit fremden Personen verweigern (von Suchodoletz, 2012).

Die gewonnenen Konsistenzkoeffizienten der Sprachskala stimmen weitgehend mit jenen der Vorstudie von Liszt (2014) überein. Die Reliabilitäten der Vorstudie fallen mit Werten $r > .80$ und $r > .90$ hoch aus und können, insbesondere bei der Testwiederholung, repliziert werden (Liszt, 2014).

Die Itemanalyse des Untertests „Wortschatz aktiv“ ergibt, dass die Items die Kriterienanforderungen ($.20 \leq p_i \leq .80$; $r_{it} \geq .30$) weitgehend erfüllen. Analog zur Überprüfung beim ersten Testzeitpunkt erfolgte dieselbe Analyse zum zweiten Erhebungszeitpunkt. Während drei Items („Wohnzimmer, Kinderzimmer und Schlafzimmer benennen“) beim ersten Testzeitpunkt von keinem Kind gelöst und somit von der Analyse ausgeschlossen

wurden, konnten diese Aufgaben bei der Testwiederholung bereits von 10% der Kinder gelöst werden. Die Beobachtung, dass lösungsschwere Items zum zweiten Testzeitpunkt im Vergleich zum ersten einfacher bewältigt werden konnten, zeigt den gelungenen Aufbau des Verfahrens. Da angenommen werden kann, dass sich Kinder in dieser Altersspanne innerhalb von vier bis sechs Wochen weiterentwickeln, spricht es für die Itemkonstruktion des Entwicklungstests, dass Kinder zum zweiten Zeitpunkt besser abschneiden. Laut Ettrich (2000) ist die Itemschwierigkeit bei jungen Kindern hoch, während sie bei älteren zunehmend niedriger wird. Die Ergebnisse dieses Untertests stimmen weitgehend mit der Vorstudie von Liszt (2014) überein und sind damit mit großer Wahrscheinlichkeit gültig. Die Analyse der, von den Kindern am häufigsten verweigerten Aufgaben zeigen, dass hauptsächlich Items mit einer niedrigen Itemschwierigkeit verweigert wurden. Diese Items sollten in Zukunft ausgeschlossen werden. Damit würde die Bilderbuchbetrachtung kürzer, was die Motivation der Kinder steigern kann.

Die Itemkennwerte zur Schwierigkeit des Untertests „Wortschatz passiv“ zeigen bei beiden Testdurchgängen, dass diese Items für die Stichprobe eher zu einfach sind. Die Hälfte der zur Erfassung des passiven Wortschatzes herangezogenen Items wurde beim ersten Testtermin von 81% der Kinder gelöst. Bei der Messwiederholung konnten drei Items bereits von 90% der Kinder gemeistert werden. Auch in der Stichprobe von Liszt (2014) zeigte sich, dass einige Items zu einfach für die Kinder waren. Erwartungsgemäß leidet die interne Konsistenz dieser Skala (T1: $\alpha = .606$; T2: $\alpha = .770$) aufgrund der linksschiefen Verteilung.

Auffallend bei diesem Untertest sind die unterschiedlichen Resultate im Hinblick auf die Trennschärfe. Vier Items tragen zum ersten Testzeitpunkt nicht zur Repräsentativität des Gesamtwertes bei. Die Ergebnisse der Trennschärfe unterscheiden sich deutlich von der vorhergehenden Untersuchung von Liszt (2014). Bis auf das Item „Fernseher“ zeigten sich in dieser Studie mittlere bis hohe Trennschärfe bei allen übrigen Items. Dieses Ergebnis kann in der vorliegenden Untersuchung nur bei der Testwiederholung repliziert werden.

Aufgrund der geringen Anzahl von Items bei diesem Untertest ist es nicht sinnvoll Items auszuschließen. Da durch die Ergebnisse sichtbar wird, dass die Messgenauigkeit eines Tests bei gegebenem Itemsatz vom Fähigkeitsniveau der untersuchten Personen abhängig ist (Bortz & Döring, 2006), wird eine erneute Überprüfung mit einer größeren Stichprobe empfohlen. Kubinger (2003) empfiehlt eine Stichprobe mit einer Größe von 100 bis 150 Personen.

Die Analyse der Trennschärfen der Subskala „Grammatik“ zeigt bei beiden Messzeitpunkten mittlere bis hohe Werte, was für eine gute Repräsentativität des Gesamtwertes spricht. Dies stimmt mit den Ergebnissen von Liszt (2014) überein. Einige Items liegen jedoch außerhalb des adäquaten Schwierigkeitsbereiches. Zum Teil handelt es sich um jene, die auch in der Vorstudie zu einfach oder zu schwer für die Stichprobe waren (Liszt, 2014). Aufgrund der hohen Variabilität in der Sprachentwicklung (von Suchodoletz, 2012), die sich auch in den zum Teil sehr unterschiedlichen Ergebnissen der Vorstudien widerspiegelt, und der Tatsache, dass die grammatikalischen Fähigkeiten im Verfahren ohne zusätzlichen Aufwand erhoben werden, erscheint es vorerst nicht notwendig zu schwere Items auszuschließen. Dies ist auch vor dem Hintergrund zu sehen, dass die Sprachentwicklung von Kindern oft nicht kontinuierlich, sondern sprunghaft und unregelmäßig verläuft (von Suchodoletz, 2012). In der vorliegenden Stichprobe waren 86% der Kinder beim ersten Testdurchgang, und alle Kinder bei der Testwiederholung, in der Lage eine Verneinung zu bilden. Diese Ergebnisse entsprechen jenen der Vorstudien von Punz (2013) und Liszt (2014). Dieses Item stellt sich als zu einfach für die Altersgruppe heraus und kann in Zukunft ausgeschlossen werden. Das Ergebnis wird durch wissenschaftliche Untersuchungen bestätigt, wonach Kleinkinder im dritten Lebensjahr bereits in der Lage sind Verneinungen zu bilden (Szagun, 2007).

Zur praktischen Anwendung der Aufgaben zur Erfassung des sprachlichen Entwicklungsstandes des Kindes sei angemerkt, dass die Aufgaben sehr gut durchführbar und für die Kinder interessant waren. Es bedarf jedoch großes Einfühlungsvermögen, seitens der Testleiterin/dem Testleiter, insbesondere bei schüchternen Kindern. Sprachliche Aufgaben sollten erst gestellt werden, wenn das Kind genügend Vertrauen gefasst und bereits mit der Untersucherin/dem Untersucher zu kommunizieren begonnen hat. Da das höchste gezeigte Verhalten der Kinder für die Kodierung beurteilt wird, ist eine hohe Konzentration der Diagnostikerin/des Diagnostikers erforderlich. Der Umstand, dass eine Beobachterin das Gesprochene des Kindes mitprotokollierte, vereinfachte die Erhebung erheblich. Diese Aufzeichnungen waren besonders für die Beurteilung der grammatikalischen Fähigkeiten sehr wertvoll und ersparten das nachträgliche Kodieren anhand der Videoaufzeichnungen. Da es in der Praxis kaum realisierbar ist, dass eine Diagnostik von zwei Personen durchgeführt wird, ist eine Aufzeichnung dennoch von Nutzen. Die nachträgliche Analyse der Aufzeichnung ist allerdings zeitaufwändig und beeinträchtigt die Ökonomie des Verfahrens. Eine Schwierigkeit stellt die Beurteilung dar, ob Aufgaben vom Kind verweigert oder nicht gelöst werden

konnten. Die wiederholte Vorgabe einzelner Items zu verschiedenen Zeitpunkten während der Spielsituation wird daher empfohlen. Untersuchungen bestätigten die Notwendigkeit der wiederholten Vorgabe (Petermann et al., 2005; Sarimski, 2009). Die Itemanzahl im Buch sollte minimiert werden (Ausschluss der Items zur Benennung der Räume), um die Motivation der Kinder durchgängig aufrecht zu erhalten. Die von von Suchodoletz (2012) getätigte Aussage, dass interessant gestaltete Spiel- und Bilderbuchsituationen besonders viele sprachliche Äußerungen anregen, konnte während der Testungen bestätigt werden. Die Kinder fanden großes Gefallen an der Fütterung der Tiere und dem Betrachten des Bilderbuches.

Für das Ziel einer allgemeinen Entwicklungsdiagnostik, nämlich das gesamte Verhalten grob zu erfassen (Deimann & Kastner-Koller, 2011a), ist die Skala „Sprache“ des spielbasierten Verfahrens sehr gut geeignet.

Aufgrund der Tatsache, dass bei der gesamten Sprachskala signifikante Mittelwertunterschiede bei der ersten und zweiten Erhebung bestehen, erscheint die noch ausständige Normierung des Verfahrens als anspruchsvolle Aufgabe. Da sich die Fähigkeiten der Kinder bereits in einem Zeitraum von vier bis sechs Wochen weiterentwickelten, sind Normtabellen mit geringem Abstand erforderlich, um das Entwicklungsprofil eines Kindes entsprechend abbilden zu können. Altersnormen im Abstand von einem Monat sind anzudenken. Geht man davon aus, dass Kleinkinder täglich fünf neue Wörter erlernen (Anglin, 1993), sind kurze Abstände der Normen wichtig, um den Entwicklungsstand des Kindes adäquat beurteilen zu können.

Die Retest-Reliabilität der Aufmerksamkeitsskala fällt nicht zufriedenstellend aus ($r_{tt} = .457$; $p = .037$). Es zeigen sich keine signifikanten Veränderungen zwischen der ersten Testvorgabe und der Messwiederholung. Bei beiden Zeitpunkten sind die Cronbach-Alpha-Werte zu niedrig, um von einem zufriedenstellenden Testergebnis sprechen zu können (T1: $\alpha = .536$; T2: $\alpha = .281$). Die niedrigen internen Konsistenzen beim ersten und zweiten Erhebungszeitpunkt sind für die niedrige Retest-Reliabilität verantwortlich. Die relativ geringen Reliabilitäten sind vor dem Hintergrund zu sehen, dass der Untertest nur über wenige Items verfügt.

Die Items der Skala „Aufmerksamkeit“ genügen nicht den Anforderungen der Testkonstruktion. Beim ersten Erhebungszeitpunkt liegen fünf der sechs Items außerhalb des adäquaten Schwierigkeitsbereichs ($p_i > .80$), und beim zweiten Testzeitpunkt alle Items. Einige

Items weisen eine unzureichende Trennschärfe auf, woraus eine niedrige interne Konsistenz resultiert. Eine hohe Trennschärfe ist aber notwendig, um Items zu einer Skala zu verrechnen, und ist somit eine wichtige Voraussetzung für die Reliabilität der Messung. Die negativen Trennschärfen der Skala sprechen dafür, dass die Items das Merkmal nicht adäquat abbilden. Durch eine mögliche Selektion würde die Skala aber nur mehr aus wenigen Items bestehen, weshalb davon Abstand genommen wurde.

Kritisch anzumerken ist, dass der Untertest die Entwicklungsdimension „Aufmerksamkeit“ womöglich nicht adäquat abbildet und das Verfahren somit in dieser Hinsicht nicht valide ist. Es bedarf dringend weiterer testtheoretischer Analysen. Seit der Überarbeitung des Itempools von Kronberger und Punz (2013) erfolgte keine Skalen- und Itemanalyse für diesen Funktionsbereich. Die Ergebnisse der vorliegenden Arbeit können daher mit keiner Vorstudie verglichen werden. Die Reliabilitätsanalyse von Putzer (2013) ergab einen zufriedenstellenden Wert von $\alpha = .83$. Der Itempool wurde aber seit damals modifiziert und die Itemanzahl verringert. Somit ist ein Vergleich schwer möglich.

Der Umstand, dass die Vorgabe der Items keiner festgelegten Reihenfolge folgt, sondern individuell an das Kind angepasst wird, kann bei der Überprüfung der Aufmerksamkeit zu Problemen führen. Da die Testungen mit einer durchschnittlichen Dauer von 107 bzw. 103 Minuten sehr viel Zeit in Anspruch nahmen, kann es zwischenzeitlich zu Leistungsschwankungen, besonders bei der Aufmerksamkeitsleistung, kommen. Gerade am Ende der Testungen nimmt die Aufmerksamkeit der Kinder ab. Bei der Itemvorgabe muss dieser Aspekt berücksichtigt werden und erfordert das Einfühlungsvermögen der Testleiterin/des Testleiters. Vorausgeschickt sei, dass die Beurteilung der Aufmerksamkeitsfähigkeit des Kindes während der Testung stark von der Fähigkeit der Untersucherin/des Untersuchers abhängt, das Kind zu motivieren. Das Konzept spielbasierter Entwicklungsverfahren sieht vor, das Spielinteresse des Kindes während der Testsituation zu berücksichtigen (Renner, 2009) und flexibel auf die Bedürfnisse der Kinder einzugehen, um so das Optimum an Leistung zu erfassen (Dykeman, 2008). Im Manual des Erhebungsinstrumentes ist beispielsweise für den Fall, dass das Kind unruhig wird, vorgesehen, dass Bewegungsaufgaben zur Erfassung der motorischen Entwicklung vorzugeben sind (Kuchler et al., 2011). Der Einsatz des Verfahrens erfordert daher psychologisches Geschick und Flexibilität. Die einzelnen Aufgaben müssen von der Diagnostikerin/dem Diagnostiker gut beherrscht werden.

Problematisch bei den Items zur Beurteilung der Aufmerksamkeit ist, dass die Aufgaben unterschiedlich viel Zeit in Anspruch nehmen und diese Zeiten auch von Kind zu Kind stark variieren können. Laut Manual (Kuchler et al., 2011) erhält das Kind einen Punkt, wenn es die ganze Zeit über aufmerksam war bis alle Aufgaben vorgegeben wurden oder bis das Kind angibt fertig zu sein. Die Problematik soll anhand eines Beispiels erläutert werden: Ein Kind stapelt drei Steine auf einen Turm, gibt an fertig zu sein und geht nach wenigen Sekunden zur nächsten Aufgabe über. Es erhält einen Punkt, weil es die Aufgabe ohne Unterbrechung beendet hat. Baut jedoch ein Kind über einen längeren Zeitraum aufwendigere Konstruktionen, unterbricht das Bauen und kehrt später wieder zur Aufgabe zurück, erhält das Kind 0.5 Punkte, obwohl es sich über einen längeren Zeitraum dem Bauen widmete.

Zudem unterscheiden sich die Aufgaben hinsichtlich ihrer Zielvorgabe. Während beim Puzzle, bei der Bilderbuchbetrachtung oder beim Auffädeln der Perlen ein konkretes Ziel verfolgt wird um zu einer Lösung zu gelangen (Vervollständigen des Bildes beim Puzzle, vollständiges Betrachten des Buches und Auffädeln aller Perlen), ist dies beim Bauen oder Zeichnen nicht der Fall. Mit solchen Items könnte beispielsweise eher das Durchhaltevermögen des Kindes erfasst werden. Zudem spielt womöglich die Fähigkeit der Kinder, die einzelnen Aufgaben zu lösen, eine entscheidende Rolle. Kinder, die feinmotorisch nicht sehr geschickt sind, sich beim Auffädeln der Perlen plagen und viel Zeit benötigen, beenden diese Aufgabe auch eher vorzeitig oder unterbrechen sie. Eine Überprüfung der Itemhomogenität sollte durchgeführt werden. Moosbrugger und Kelava (2012) folgend sollten Items mit negativer Trennschärfe ausgeschlossen werden.

Die Testungen dauerten bei der Testwiederholung signifikant kürzer als beim ersten Durchgang. Ein mäßiger Zusammenhang zwischen beiden Testzeitpunkten konnte festgestellt werden. Die Kinder beschäftigten sich bei der erneuten Vorgabe des Verfahrens signifikant kürzer mit den Materialien. Ein Zusammenhang zwischen den beiden Testzeitpunkten besteht nicht. Aus der Dauer der Materialbeschäftigung beim ersten Termin lassen sich also keine Vorhersagen für den zweiten ableiten. Der Umstand, dass sich die Kinder an das gemeinsame Spielen erinnerten und von sich aus alle Spiele erneut probieren wollten, vermag die kürzere Dauer des Tests und der Materialbeschäftigung erklären. Die Kinder wussten wohl, was auf sie zukommt, sie waren bereits mit der Umgebung vertraut und mussten zum Teil weniger motiviert werden. Zudem kannte die Testleiterin das Kind bereits und konnte so besser auf

dessen Bedürfnisse eingehen. Dies spricht, wie bereits erwähnt, für die Idee, einen Termin zum Kennenlernen der eigentlichen Testung vorausgehen zu lassen.

Hinsichtlich der Dauer von Pausen, der Anlaufzeit und Verabschiedungsdauer lassen sich keine Zusammenhänge oder Veränderungen zwischen den beiden Messzeitpunkten feststellen.

Die vorliegende Studie weist einige Limitationen auf, die im Folgenden diskutiert werden. Die erste Einschränkung betrifft die Größe der Stichprobe. Es wurden bloß 21 Kinder getestet. Die geringe Anzahl der Getesteten führt zu einer Verminderung der Aussagekraft der Ergebnisse. Um die Aussagekraft dieser und der Vorstudien zu bewerten, müsste eine testtheoretische Überprüfung aller bereits erhobenen Daten erfolgen. Unabhängig davon sollte der Itempool mit einer ausreichend großen Stichprobe erneut evaluiert werden.

Zweitens besteht eine Limitation hinsichtlich der Zusammensetzung der Stichprobe. Der Großteil der Eltern der getesteten Zweijährigen besitzt einen universitären Abschluss. Dies könnte das Ergebnis der Studie verfälscht haben, da keine Repräsentativität für die Gesamtbevölkerung gegeben ist. Forschungsergebnisse zeigen, dass das Bildungsniveau der Eltern Einfluss auf Sprachentwicklung hat (Marjanovič-Umek, Fekonja, Podlesek, & Kranjc, 2011). Weitere soziodemografische Einflussfaktoren, wie zum Beispiel die Anzahl und Stellung der Geschwister, sollten untersucht werden (Rowe, 2012).

Überdies erfolgte die Auswahl der Probanden aus dem geografischen Bereich von Wien und Umgebung. Die ausgewählte Stichprobe ist damit auch in dieser Hinsicht nicht auf die Gesamtbevölkerung generalisierbar.

Kritisch anzumerken ist, dass zu beiden Messzeitpunkten teilweise unterschiedliche Items aus den Berechnungen der Konsistenzanalyse ausgeschlossen wurden. Dies wirft das Problem der Symmetrie auf, wonach bei beiden Messzeitpunkten auch dieselben Items in die Berechnungen einfließen sollten.

Aufgrund der kleinen Stichprobe sind Schlüsse hinsichtlich Messgenauigkeit und Zuverlässigkeit nur begrenzt möglich. Laut Kubinger (2003) ist dazu eine Stichprobe von 100 bis 150 Probanden notwendig. Reliabilitätsanalysen sind überdies stark stichprobenabhängig. Bei der vorliegenden Stichprobe handelte es sich eher um eine homogene Gruppe, wodurch Reliabilitätseinschätzungen unter Umständen schlechter ausfallen als sie tatsächlich sind.

Hinsichtlich der praktischen Durchführbarkeit des spielbasierten Entwicklungstests gibt es einige positive Eigenschaften, die hervorzuheben sind. Durch die flexible Vorgabe der einzelnen Items konnte optimal auf die Bedürfnisse und den Gemütszustand des Kindes eingegangen werden. Die Motivation des Kindes konnte dadurch durchgehend aufrechterhalten werden.

Obwohl es in dieser Altersspanne sehr schwer für Kinder ist aufmerksam zu bleiben und sich längere Zeit einer Aufgabe zu widmen (Petermann et al., 2005), gelang es den Zweijährigen sehr gut. Der Großteil der Kinder war durchwegs aufmerksam, was für das Testverfahren spricht. Die Spielmaterialien, insbesondere das Spielen in der Puppenküche, weckten das Interesse der Kleinkinder. Durch die Exploration der Puppenküche konnten auch schüchterne Kinder angesprochen und zum Mitmachen animiert werden. Der Teddybär eignete sich zusätzlich zur Kontaktaufnahme und war besonders hilfreich zur Lenkung der Aufmerksamkeit des Kindes auf andere Items. Beispielsweise wurde das Kind zur Durchführung einer Aufgabe animiert, indem vorgegeben wurde, dass der Teddy so gerne ein Buch lesen oder zeichnen wolle. Da man bei Kleinkindern noch nicht von einer entsprechenden Arbeitshaltung und Leistungsmotivation ausgehen kann (Irblich & Renner, 2009), ist einerseits der Anreiz der Spielmaterialien wichtig, andererseits aber auch die Kompetenz der Diagnostikerin/des Diagnostikers zur Motivierung des Kindes. Da immer das höchst gezeigte Verhalten des Kindes kodiert wird, obliegt es der Testleiterin/dem Testleiter, wie oft Items wiederholt vorgegeben werden und wie sie/er in der Lage ist, das Kind zum Lösen der Aufgaben zu animieren. Die Diagnostik stellt hohe Anforderungen an die Untersucherin/den Untersucher (Esser & Petermann, 2010). Es bestätigen sich die Aussagen von Irblich und Renner (2009), dass eine Testung mit Kleinkindern viel Erfahrung mit der Altersgruppe und hohes psychologisches Geschick verlangt.

Durch den spielbasierten Zugang und den hohen Aufforderungscharakter der Materialien verlor die Diagnostik den Charakter einer Testsituation. Die Eltern empfanden die Testungen sehr angenehm und stressfrei. Diesen Aspekt hoben auch Gagnon und Nagle (2004) als Vorteil einer spielbasierten Diagnostik hervor.

Die Testungen mit den Zweijährigen zeigten, dass die lange Spieldauer für die Kinder zumutbar ist, da viele nach Beendigung weiterspielen wollten. Dies belegen auch die zum Teil hohen Verabschiedungszeiten. Die Dauer der Diagnostik ist ausschlaggebend um das Optimum an

Leistung der Kinder zu ermöglichen. Das Spielen in der Puppenküche hat sich als Aufwärmphase sehr bewährt. Dennoch wird die Durchführung der Diagnostik an zwei getrennten Terminen empfohlen oder es sollte zumindest ein gesonderter Termin zum Kennenlernen vorhergehen.

Um das Entwicklungsprofil eines Kindes erstellen zu können, müssen neben der Überprüfung der einzelnen Fähigkeitsbereiche auch kindbezogene Merkmale berücksichtigt werden (Macha & Petermann, 2008). Das gemeinsame Spielen in einer angenehmen Atmosphäre ermöglicht es der Diagnostikerin/dem Diagnostiker eine Vielzahl von Informationen durch die Verhaltensbeobachtung zu gewinnen. Beispielsweise können Aussagen über das kindliche Temperament, die Frustrationstoleranz und Verhaltensweisen beim Spielen getroffen werden. Insbesondere für Kinder mit Entwicklungsstörungen bedarf es solcher Einschätzungen für die weitere Therapieplanung (Sarimski, 2009). Das erprobte Verfahren ermöglicht, wertvolle Schlüsselkompetenzen des Kindes zu erfassen, die mit standardisierten Verfahren nicht beobachtbar sind. Auch der Umgang und die Interaktion mit der Bezugsperson können ausschlaggebend und richtungsweisend für das diagnostische Urteil sein. So lässt sich beispielsweise beobachten wie die Bezugsperson reagiert, wenn das Kind Aufgaben nicht löst. Diese Beobachtungen sind in diesem Setting sehr gut möglich, was ebenfalls für das spielbasierte Konzept spricht.

Die Anwesenheit der Bezugsperson bot dem Kind ausreichend Sicherheit, um sich auf das Spiel mit der fremden Person einzulassen. Im Sinne einer förderungsorientierten Diagnostik ist positiv anzumerken, dass Eltern durch die Beobachtung der Spielsituation Hinweise erhalten, wie sie selbst ihr Kind fördern können.

Die zusätzliche Beobachterin sowie die Videoaufzeichnung mit der Kamera wirkten auf viele Kinder befremdlich. Insbesondere gehemmte Kinder waren durch die zusätzliche fremde Person eingeschüchtert. Ein positiver Aspekt dieser Konstellation ist, dass die Beobachterin während der gesamten Testung protokollierte und die Kodierung am Protokollbogen vornahm. Somit konnte sich die Testleiterin völlig der Spielsituation widmen. Dieser Umstand erleichterte die Testdurchführung erheblich.

Die von Kelly-Vance et al. (2002) festgestellten Vorteile spielbasierter Verfahren konnten durch diese Studie bestätigt werden. Dabei sei angemerkt, dass die Testungen von Kelly-Vance et al. (2002) in der natürlichen Umgebung des Kindes stattgefunden haben und bei der

gegenständlichen Studie in einem Testraum, der speziell für die Bedürfnisse der Kinder eingerichtet wurde. Es konnte beobachtet werden, dass die Kinder sich in der Testumgebung wohl fühlten und die Materialien höchst motivierend wirkten. Der flexible Ablauf des Verfahrens half dabei ungemein auf unvorhergesehene Verhaltensweisen der Kinder einzugehen und die Testung trotzdem fortführen zu können. Die außergewöhnlich lange Testdauer von bis zu zwei Stunden trat gegenüber der Freude der Kinder am Spielen mit den Testmaterialien in den Hintergrund und beeinflusste damit den Testablauf nicht negativ. Die Gesamtheit der soeben dargestellten Umstände während der Testung vermochten das volle Leistungspotential des Kindes zu entfesseln.

Erste testtheoretische Bewertungen des Verfahrens wurden durchgeführt. Es bedarf jedenfalls weiterer Forschungen, um eine vollständige Absicherung der Testgütekriterien zu erlangen.

Abschließend sei noch eindringlich darauf hingewiesen, dass die Durchführung spielbasierter Verfahren verglichen mit herkömmlichen Ansätzen, noch höhere Anforderungen an die Testleiterin/den Testleiter stellt. Insbesondere das Einlassen auf die Spielsituation und die kindgerechte Kommunikation erfordern umfassende Kenntnisse des kindlichen Verhaltens und Wissen über die Entwicklung im Kindesalter.

Literaturverzeichnis

- Acton, B. V., Biggs, W. S. G., Creighton, D. E., Penner, K. A. H., Switzer, H. N., Thomas, J. H. P., ... Robertson, C. M. T. (2011). Overestimating Neurodevelopment Using the Bayley-III After Early Complex Cardiac Surgery. *PEDIATRICS*, 128(4), e794–e800. doi:10.1542/peds.2011-0331
- Aktas, M., & Doil, H. (2009). Diagnostik der Sprachentwicklung. In D. Irblich & G. Renner (Eds.), *Diagnostik in der Klinischen Kinderpsychologie. Die ersten sieben Lebensjahre* (pp. 152 – 166). Göttingen: Hogrefe.
- Albers, C. A., & Grieve, A. J. (2007). Test Review: Bayley, N. (2006). Bayley Scales of Infant and Toddler Development- Third Edition. San Antonio, TX: Harcourt Assessment. *Journal of Psychoeducational Assessment*, 25(2), 180–190. doi:10.1177/0734282906297199
- Anglin, J. M. (1993). Vocabulary development : a morphological analysis. *Monographs of the Society for Research in Child Development*, 58(10).
- 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.
- Bagnato, S. J., & Neisworth, J. T. (1994). A National Study of the Social and Treatment „Invalidity“ of Intelligence Testing for Early Intervention. *School Psychology Quarterly*, 9, 81 – 102.
- Bayley, N. (2006). *Bayley Scales of Infant and Toddler Development* (3rd ed.). San Antonio: PsychCorp.
- Bayley, N. (2015). *Bayley Scales of Infant and Toddler Development - Third Edition*. (G. Reuner & J. Rosenkranz, Trans.). Göttingen: Hogrefe.
- Berk, L. E. (2005). *Entwicklungspsychologie* (3rd ed.). München: Pearson Studium.
- 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*. Wien.
- Bortz, J., & Döring, N. (2006). *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler* (4th ed.). Heidelberg: Springer Berlin Heidelberg.
- Deegener, G. (2009). Eltern und Erzieher als Informationsquelle. In D. Irblich & G. Renner (Eds.), *Diagnostik in der Klinischen Psychologie. Die ersten sieben Lebensjahre* (pp. 86 – 96). Göttingen: Hogrefe.
- Deimann, P., & Kastner-Koller, U. (2007). Entwicklungsdiagnostik Developmental Assessment. In M. Hasselhorn & W. Schneider (Eds.), *Handbuch der Entwicklungspsychologie* (pp. 558–569). Göttingen: Hogrefe.
- Deimann, P., & Kastner-Koller, U. (2011a). Entwicklungstests. In L. Hornike, M. Amelang, & M. Kersting (Eds.), *Leistungs-, Intelligenz- und Verhaltensdiagnostik* (pp. 275–302). Göttingen: Hogrefe.

- Deimann, P., & Kastner-Koller, U. (2011b). Maternal evaluations of young children's developmental status: A comparison of clinic- and non-clinic-groups. *Psychological Test and Assessment Modeling*, 53(2), 214–227.
- Deimann, P., Kastner-Koller, U., Benka, M., Kainz, S., & Schmidt, H. (2005). Mütter als Entwicklungsdiagnostikerinnen. *Zeitschrift Für Entwicklungspsychologie Und Pädagogische Psychologie*, 37(3), 122–134. doi:10.1026/0049-8637.37.3.122
- Dykeman, B. F. (2008). Play-Based Neuropsychological Assessment of Toddlers. *Journal of Instructional Psychology*, 35(4), 405 – 408.
- Esser, G., & Petermann, F. (2010). *Entwicklungsdiagnostik*. Göttingen: Hogrefe.
- Ettrich, K. U. (2000). *Eintwicklungsdiagnostik im Vorschulalter*. Göttingen: Vandenhoeck & Ruprecht.
- Field, A. (2009). *Discovering Statistics using SPSS*. London: SAGE.
- Fisseni, H.-J. (2004). *Lehrbuch der psychologischen Diagnostik. Mit Hinweisen zur Intervention*. Göttingen: Hogrefe.
- Fuchs-Garderer, M. (2012). *Die Überarbeitung und testtheoretische Überprüfung eines spielbasierten Itempools zur Erfassung der Entwicklung im dritten Lebensjahr unter besonderer Berücksichtigung der Bereiche Motorik, Sprache und sozioemotionale Entwicklung*. Wien.
- Gagnon, S. G., & Nagle, R. J. (2004). Relationships between peer interactive play and social competence in at-risk preschool children. *Psychology in the Schools*, 41(2), 173–189. doi:10.1002/pits.10120
- Grimm, H. (2003). *Störungen der Sprachentwicklung* (2nd ed.). Göttingen: Hogrefe.
- Grimm, H., & Aktas, M. (2002). Entwicklungstests im Vorschulalter: Berücksichtigung ihrer Nützlichkeit durch praktisch tätige Psychologen. *Frühförderung Interdisziplinär*, 21, 163 – 177.
- Grob, A. (2008). Intelligenz- und Entwicklungsdiagnostik im deutschen Sprachraum. *Kindheit und Entwicklung*, 17(4), 232–242. doi:10.1026/0942-5403.17.4.232
- Hasselhorn, M., & Margraf-Stiksrud, J. (2012). TBS-TK Rezension Entwicklungstest für Kinder von 6 Monaten bis 6 Jahren (ET 6-6). *Psychologische Rundschau*, 63(2), 141–143. doi:10.1026/0033-3042/a000120
- Heubrock, D., & Petermann, F. (2001). *Aufmerksamkeitsdiagnostik*. Göttingen: Hogrefe.
- Holodyski, M. (2006). Die Entwicklung der Leistungsmotivation im Vorschulalter. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 38(1), 2–17. doi:10.1026/0049-8637.38.1.2
- Irblich, D., & Renner, G. (2009). Wie untersucht man Kinder? In D. Irblich & G. Renner (Eds.), *Diagnostik in der Klinischen Psychologie. Die ersten sieben Lebensjahre* (pp. 21 – 33). Göttingen: Hogrefe.

- Kastner-Koller, U., & Deimann, P. (2009). Beobachtung und Befragung von Kindern. In D. Irlich & G. Renner (Eds.), *Diagnostik in der Klinischen Psychologie. Die ersten sieben Lebensjahre* (pp. 97 – 107). Göttingen: Hogrefe.
- Kastner-Koller, U., & Deimann, P. (2012). *WET Wiener Entwicklungstest* (3. ed.). Göttingen: Hogrefe.
- 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., & Glover, K. G. (2002). The Use of Play Assessment to Evaluate the Cognitive Skills of Two-and Three-Year-Old Children. *School Psychology International*, 23(2), 169–185. doi:10.1177/0143034302023002909
- Kronberger, V. M. (2013). *Die Überarbeitung und testtheoretische Überprüfung eines spielbasierten Itempools zur Erfassung der Entwicklung zweijähriger Kinder unter besonderer Berücksichtigung der visuellen Wahrnehmung*. Wien.
- Kubinger, K. D. (2003). *Schlüsselbegriffe der psychologischen Diagnostik*. Weinheim: Beltz.
- Kubinger, K. D. (2009). *Psychologische Diagnostik* (2. überarb.). Göttingen: Hogrefe.
- Kuchler, M., Sapper, E., Kastner-Koller, U., & Deimann, P. (2011). *Manual zum Itempool zur spielbasierten Erfassung der Entwicklung Zweijähriger*. Universität Wien: unveröffentlichtes Manuskript.
- Lienert, G. A. (1989). *Testaufbau und Testanalyse* (4th ed.). München: Psychologie-Verl.-Union.
- Liszt, S. (2014). *Überprüfung der Testgütekriterien eines spielbasierten Inventars für Zwei- bis Dreijährige unter besonderer Berücksichtigung der Bereiche: Motorik, Sprache und sozial-emotionale Entwicklung*. Wien.
- Macha, T., & Petermann, F. (2008). Entwicklungsdiagnostik. In F. Petermann (Ed.), *Lehrbuch der klinischen Kinderpsychologie* (6th ed., pp. 117 – 133). Göttingen: Hogrefe.
- Marjanovič-Umek, L., Fekonja, U., Podlessek, A., & Kranjc, S. (2011). Assessing toddler language competence: agreement of parents' and preschool teachers' assessments. *European Early Childhood Education Research Journal*, 19(1), 21–43. doi:10.1080/1350293X.2011.548957
- Michaelis, R., Berger, R., Nennstiel-Ratzel, U., & Krägeloh-Mann, I. (2013). Validierte und teilvalidierte Grenzsteine der Entwicklung. *Monatsschrift Kinderheilkunde*, 161(10), 898–910. doi:10.1007/s00112-012-2751-0
- Moosbrugger, H., & Kelava, A. (2012). Qualitätsanforderungen an einen psychologischen Test (Testgütekriterien). In H. Moosbrugger & A. Kelava (Eds.), *Testtheorie und Fragebogenkonstruktion* (2nd ed., pp. 7 – 27). Berlin, Heidelberg: Springer.
- Nelson-Gray, R. O. (2003). Treatment Utility of Psychological Assessment. *Psychological Assessment*, 15(4), 521–531. doi:10.1037/1040-3590.15.4.521
- Papoušek, M. (2012). Kommunikation und Sprachentwicklung im ersten Lebensjahr. In M. Cierpka (Ed.), *Frühe Kindheit 0 - 3. Beratung und Psychotherapie für Eltern mit Säuglingen und Kleinkindern*. Berlin, Heidelberg: Springer.

- Pauen, S., Frey, B., & Ganser, L. (2012). Entwicklungspsychologie in den ersten drei Lebensjahre. In M. Cierpka (Ed.), *Frühe Kindheit 0 - 3. Beratung und Psychotherapie für Eltern mit Säuglingen und Kleinkindern* (pp. 21 – 35). Berlin, Heidelberg: Springer.
- Petermann, F., & Macha, T. (2003). Strategien in der testgestützten allgemeinen Entwicklungsdiagnostik. *Monatsschrift 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). *ET 6-6 R Entwicklungstest für Kinder von 6 Monaten bis 6 Jahren - Revision*. Göttingen: Hogrefe.
- Petermann, F., Proske, A., & Macha, T. (2005). Allgemeine Entwicklungsdiagnostik Validität von Entwicklungstests. *Kindheit Und Entwicklung*, 14(3), 150–162.
- Piaget, J. (1951). *Play, dreams, and imitation in childhood*. New York: Norton.
- Pozniak, M. (in Vorbereitung). *Die Retest-Reliabilität eines spielbasierten Entwicklungstests für zweijährige Kinder unter besonderer Berücksichtigung des Gedächtnisses und der visuellen Wahrnehmung*. Wien.
- Prillinger, K. (in Vorbereitung). *Retest-Reliabilität eines spielbasierten Entwicklungstests für zweijährige Kinder unter besonderer Berücksichtigung der kognitiven Entwicklung*. Wien.
- Punz, S. A. (2013). *Die Überarbeitung und testtheoretische Überprüfung eines spielbasierten Itempools zur Erfassung der Entwicklung im dritten Lebensjahr unter besonderer Berücksichtigung der kognitiven Entwicklung*. Wien.
- Putzer, S. (2013). *Die Entwicklung und Erprobung eines Itempools zur spielbasierten Erfassung der Entwicklung Zweijähriger mit besonderer Berücksichtigung der Grobmotorik*. Wien.
- Quaiser-Pohl, C. (2010a). Was ist Entwicklungsdiagnostik? In C. Quaiser-Pohl & H. Rindermann (Eds.), *Entwicklungsdiagnostik* (pp. 18 – 25). München: Ernst Reinhardt Verlag.
- Quaiser-Pohl, C. (2010b). Wie teste ich Kinder? Zur Gestaltung der diagnostischen Situation. In C. Quaiser-Pohl & H. Rindermann (Eds.), *Entwicklungsdiagnostik* (pp. 57 – 68). München: Ernst Reinhardt Verlag.
- Rauh, H. (2008). Vorgeburtliche Entwicklung und frühe Kindeheit. In R. Oerter & L. Montada (Eds.), *Entwicklungspsychologie* (6th ed., pp. 149 – 224). Weinheim, Basel: Beltz Verlag.
- Renner, G. (2009). Testpsychologische Diagnostik bei Kindern. In D. Irblich & G. Renner (Eds.), *Diagnostik in der Klinischen Psychologie. Die ersten sieben Lebensjahre* (pp. 73 – 86). Göttingen: Hogrefe.
- Renner, G., & Irblich, D. (2009). Diagnostik bei Aufmerksamkeitsstörungen. In D. Irblich & G. Renner (Eds.), *Diagnostik in der Klinischen Kinderpsychologie. Die ersten sieben Lebensjahre* (pp. 245 – 259). Göttingen: Hogrefe.

- Reuner, G., & Pietz, J. (2006). Entwicklungsdiagnostik im Säuglings- und Kleinkindalter. *Monatsschrift Kinderheilkunde*, 154(4), 305–313. doi:10.1007/s00112-006-1315-6
- Rowe, M. L. (2012). A Longitudinal Investigation of the Role of Quantity and Quality of Child-Directed Speech in Vocabulary Development. *Child Development*, 83(5), 1762–1774. doi:10.1111/j.1467-8624.2012.01805.x
- Ruff, H. A., & Capozzoli, M. C. (2003). Development of attention and distractibility in the first 4 years of life. *Developmental Psychology*, 39(5), 877–890. doi:10.1037/0012-1649.39.5.877
- Sarimski, K. (2009). Entwicklungsdiagnostik. In D. Irblich & G. Renner (Eds.), *Diagnostik in der Klinischen Psychologie. Die ersten sieben Lebensjahre* (pp. 123 – 135). Göttingen: Hogrefe.
- Schermelleh-Engel, K., & Werner, C. S. (2012). Methoden der Reliabilitätsbestimmung. In H. Moosbrugger & A. Kelava (Eds.), *Testtheorie und Fragebogenkonstruktion* (2nd ed., pp. 119 – 143). Berlin, Heidelberg: Springer.
- Sindelar, I. (2013). *Die Entwicklung und Erprobung eines Itempools zur spielbasierten Erfassung der Entwicklung Zweijähriger unter besonderer Berücksichtigung der Feinmotorik und der Visumotorik*. Wien.
- Stark, M. (in Vorbereitung). *Retest-Reliabilität eines spielbasierten Entwicklungstests für zweijährige Kinder unter besonderer Berücksichtigung der Motorik und der sozial-emotionalen Entwicklung*. Wien.
- Stipek, D., Recchia, S., McClintic, S., & Lewis, M. (1992). Self-Evaluation in Young Children. *Monographs of the Society for Research in Child Development*, 57(1), i. doi:10.2307/1166190
- Szagun, G. (2007). *Das Wunder des Spracherwerbs*. Weinheim, Basel: Beltz.
- Tscherne, N. (2014). *Überprüfung der Testgütekriterien eines spielbasierten Inventars für zwei- bis dreijährige Kinder unter besonderer Berücksichtigung der kognitiven Entwicklung, des Gedächtnisses und der visuellen Wahrnehmung*. Wien.
- Van de Weijer-Bergsma, E., Wijnroks, L., & Jongmans, M. J. (2008). Attention development in infants and preschool children born preterm: A review. *Infant Behavior and Development*, 31(3), 333–351. doi:10.1016/j.infbeh.2007.12.003
- Von Suchodoletz, W. (2012). *Früherkennungs von Sprachentwicklungsstörungen*. Stuttgart: Kohlhammer.
- Willinger, U., & Eisenwort, B. (2005). Mothers' estimates of their children with disorders of language development. *Behavioral Medicine*, 31(3), 117 – 124.

Tabellenverzeichnis

TABELLE 1: KREUZTABELLE - ALTER UND GESCHLECHT.....	28
TABELLE 2: HÖCHSTE ABGESCHLOSSENE AUSBILDUNG DER MÜTTER	29
TABELLE 3: HÖCHSTE ABGESCHLOSSENE AUSBILDUNG DER VÄTER.....	29
TABELLE 4: HÄUFIGKEITEN FEHLENDER MESSWERTE ZU T1 UND T2	35
TABELLE 5: MITTELWERTUNTERSCHIEDE DER SKALA „SPRACHE“ T1 UND T2	37
TABELLE 6: MITTELWERTUNTERSCHIEDE DER SKALA „AUFMERKSAMKEIT“ T1 UND T2	38
TABELLE 7: ITEMKENNWERTE DER SUBSKALA „WORTSCHATZ AKTIV“ T1.....	39
TABELLE 8: ITEMKENNWERTE DER SUBSKALA „WORTSCHATZ AKTIV“ T2.....	40
TABELLE 9: ITEMKENNWERTE DER SUBSKALA „WORTSCHATZ PASSIV“ T1	40
TABELLE 10: ITEMKENNWERTE DER SUBSKALA „WORTSCHATZ PASSIV“ T2.....	41
TABELLE 11: ITEMKENNWERTE DER SUBSKALA „GRAMMATIK“ T1	42
TABELLE 12: ITEMKENNWERTE DER SUBSKALA „GRAMMATIK“ T2	43
TABELLE 13: ITEMKENNWERTE DER SKALA „AUFMERKSAMKEIT“ T1	44
TABELLE 14: ITEMKENNWERTE DER SKALA „AUFMERKSAMKEIT“ T2	45

Anhang

Anhang A. 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 540 92 52

Maria Pozniak, BSc

0650 990 64 12

Karin Prillinger, BSc

0664 158 29 30

Magdalena Stark, BSc

0664 166 22 51

anmeldung-masterarbeit@gmx.at

Anhang B. Einverständniserklärung

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

Anhang C. Protokollbogen

✓ Gelöst

v. verweigert

X nicht gelöst

- nicht vorgegeben

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

KIND-BEZUGSPERSON-BEZIEHUNG WORTSCHATZ PUPPENKÜCHE (OBST)

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

DESKRIPTIVE ANGABEN

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

Visuelle Wahrnehmung PUPPENKÜCHE (Obst sortieren nach Farbe)

Stabilität Farbkonzept Min 3				
Rot				
Grün				
Gelb				
orange				

Feinmotorik PUPPENKÜCHE (0/1 + Anzahl aufschreiben)

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

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

Protokollbogen Jungwirth, Pozniak, Prillinger & Stark

GEDÄCHTNIS - PUPPENKÜCHE

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

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

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			
Hose anziehen	Schuhe anziehen geschafft			
Hut aufsetzen	Versucht selbstständig			
	Sucht Hilfe			
	Desinteresse			
Hut aufsetzen	In Hosenbein reinschlüpfen			
	Hose hochziehen			
T-Shirt anziehen	Versucht selbstständig			
	Sucht Hilfe			
	Desinteresse			
T-Shirt anziehen	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	Küche benennen			
	schmutzige Hände			
Torte/Käse (aktiv)	Torte/Käse (aktiv)			
	Besen passiv			
Wohnzimmer benennen	Wohnzimmer benennen			
	Hungrig			
Lampe (aktiv)	Lampe (aktiv)			
	Fernseher (passiv)			
	Tür (passiv)			
Kinderzimmer benennen	Kinderzimmer benennen			
	Schlafen			
	Pflaster			
Schuh (aktiv)	Schuh (aktiv)			
	Stuhl/Sessel (passiv)			
	Gitterbett (passiv)			
Schlafzimmer benennen	Schlafzimmer benennen			
	durstig			
Kasten (aktiv)	Kasten (aktiv)			
	Tasche (aktiv)			
	Fenster (passiv)			
Bild (passiv)	Bild (passiv)			
Badezimmer (benennen)	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)

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

GROBMOTORIK STEHEN, HÜPFEN, LAUFEN (1/0)

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

FEINMOTORIK & FARBDIFFERENZIERUNG & VISUMOTORIK (1/0)

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

VISUMOTORIK – Formen

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

Skizze von Turm

VISUMOTORIK - TISCHAUFGABEN

Perlen fädeln (min 2)	Anzahl Perlen		
	Vorgezeigt		
	Aufmerksamkeit (0/0,5/1)		
	Dauer bis zur ersten Unterbrechung		
Perlen Farben passiv – GIB MIR	Ließ sich zurückholen		
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			
	• mit korrektem Relativpronomen			
	• mit wo			
	• ohne Relativpronomen			
Verben	Stellung im Aussagesatz			
	Endstellung			
	Verbstellung korrekt			
	Partizip Perfekt (kommt vor)			
Verneinung	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		

Protokollbogen Jungwirth, Pozniak, Prillinger & Stark

Protokoll der vom Kind getätigten Aussagen:

Anhang D. 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:				☐ Ja ☐ Nein
Mutter:				☐ Ja ☐ Nein

Geschwister:

Name	Alter	Geschlecht	gemeinsamer Haushalt
		☐ M ☐ W	☐ Ja ☐ Nein
		☐ M ☐ W	☐ Ja ☐ Nein
		☐ M ☐ W	☐ Ja ☐ Nein
		☐ M ☐ W	☐ Ja ☐ Nein
		☐ M ☐ W	☐ Ja ☐ 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 tagüber 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“).

- | | | |
|------------------------------------|----------------------------------|-------------------------------------|
| ☐ schmutzig | ☐ Auto | ☐ Fenster |
| ☐ Torte | ☐ Apfel | ☐ Bild |
| ☐ Käse | ☐ Birne | ☐ Badezimmer |
| ☐ Besen | ☐ Banane | ☐ Wasser |
| ☐ hungrig | ☐ Zitrone | ☐ Badewanne |
| ☐ Lampe | ☐ Karotte | ☐ Katze |
| ☐ Fernseher | ☐ Orange | ☐ Besteck |
| ☐ Tür | ☐ Sessel | ☐ Schlaf |

- | | | |
|-----------------------------------|---------------------------------------|----------------------------------|
| ☐ schlafen | ☐ Gitterbett | ☐ Fahrrad |
| ☐ Pflaster | ☐ Schlafzimmer | ☐ fliegen |
| ☐ Bub | ☐ durstig | ☐ Elefant |
| ☐ Knie | ☐ Mädchen | ☐ Hase |
| ☐ Mann | ☐ Kasten | ☐ Eis |
| ☐ Schuh | ☐ Tasche | ☐ Hund |

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

Vielen Dank für Ihre Mitarbeit!

Anhang E. Überprüfung der Normalverteilung

Anhang E.1. Prüfung der Messwertdifferenzen deskriptiver Angaben

	Schiefe	Standardfehler	z-Wert
Testdauer	-1.618	0.501	-3.234
Anlaufzeit	0.559	0.501	1.116
Pausendauer	-1.280	0.501	-2.555
Verabschiedungszeit	-0.367	0.501	-0.733
Materialbeschäftigung	-0.332	0.501	-0.663

Anhang E.2. Prüfung der Messwertdifferenzen der Skala „Sprache“

	Schiefe	Standardfehler	z-Wert
Wortschatz passiv	-0.202	0.501	-0.403
Wortschatz aktiv	0.401	0.501	0.800
Wortschatz	0.075	0.501	0.150
Grammatik	0.195	0.501	0.389
Sprache	-0.246	0.501	-0.491

Anhang E.3. Prüfung der Messwertdifferenzen der Skala „Aufmerksamkeit“

	Schiefe	Standardfehler	z-Wert
Aufmerksamkeit	0.923	0.501	1.842

Anhang F. Zusammenhang und Unterschiede der Anlauf-, Verabschiedungszeit und Pausendauer

Anhang F.1. t-Test für abhängige Stichproben

	t	df	Sig.
Anlaufzeit	0.000	20	1.000
Verabschiedungszeit	-0.632	20	.535

Anhang F.2. Spearman – Rangkorrelation

	N	r _s	Sig.
Pausendauer	21	.224	.329

Anhang F.3. Pearson – Korrelation

	N	r	Sig.
Anlaufzeit	21	.660	.001
Verabschiedungsdauer	21	.138	.551

Anhang F.4. Wilcoxon – Vorzeichen – Rangtest

	Pausendauer T1 und Pausendauer T2
U	-.363
Asymp. Sig. (2-seitig)	.717

Anhang G. Curriculum Vitae

ANGABEN ZUR PERSON

Name	Christina Jungwirth, BSc
Adresse	3002 Purkersdorf
Telefon	+43 664 5409252
E-Mail	office@prokids.at
Web	http://www.prokids.at
Staatsangehörigkeit	Österreich
Geburtsdatum	20. Dezember 1985

BERUFSERFAHRUNG

Seit Oktober 2014	Personalvertreterin an der Bildungsanstalt für Kindergartenpädagogik im Sacré Coeur , Pressbaum
Juni 2014 bis Oktober 2014	Psychologisches Pflichtpraktikum in der Kinder- und Jugendpsychologischen Praxis von Dr. Hannelore Koch , Wien
Seit September 2011	Lehrerin an der Bildungsanstalt und am Kolleg für Kindergartenpädagogik im Sacré Coeur , Pressbaum Derzeitiger Unterricht der Gegenstände Kindergartenpraxis, Hortpraxis, Didaktik der Horterziehung, Sonder-, Heilpädagogik und Inklusive Pädagogik, Waldpädagogik
Seit Juli 2009	Selbstständige freiberufliche Tätigkeit als Motopädagogin und Waldpädagogin Veranstaltung von Waldwochen und von Waldmottotagen mit motopädagogischem Schwerpunkt
Seit Juli 2008	Selbstständige freiberufliche Tätigkeit als Legasthenie- und Dyskalkulietrainerin Legastheniepädagogische Austestung und Training von Kindern mit Lese-, Rechtschreib- und Rechenschwäche sowie Projekte zur Leseförderung
April 2007 bis September 2011	Übungshortleiterin im Übungshort Sacré Coeur , Pressbaum Leitung einer Integrationsgruppe mit Kindern im Alter von sechs bis elf Jahren
September 2005 bis April 2007	Horterzieherin bei Erzdiözese Wien Leitung einer Familien-Hortgruppe für Kinder im Alter von acht bis fünfzehn Jahren
Juli 2004 bis August 2004	Animatorin im Feriencamp (Öko- und Kreativcamp in Lackenhof) beim Niederösterreichischen Kinderrettungswerk

AUSBILDUNG

Seit Oktober 2013	Masterstudium der Psychologie an der Universität Wien mit dem Schwerpunkt Entwicklung, Gesundheit und Förderung
Oktober 2010 bis Juli 2013	Bachelorstudium der Psychologie an der Universität Wien mit dem Erweiterungscurriculum „Grundlagen der Heilpädagogik und Inklusiven Pädagogik“, Abschluss als Bachelor of Science (BSc)
September 2010 bis April 2011	Ausbildung zur Waldpädagogin durch den Verein Waldpädagogik Österreich
September 2009 bis April 2010	Ausbildung zur Praxis- und Didaktiklehrerin an der Pädagogischen Hochschule Wien
April 2009 bis April 2010	Ausbildung zur Dipl. Dyskalkulietrainerin durch den EÖDL – Erster Österreichischer Dachverband für Legasthenie
August 2008 bis Oktober 2009	Ausbildung zur Motopädagogin durch die VALEO – Psychomotorische Entwicklungsbegleitung GmbH
Oktober 2007 bis Juni 2008	Ausbildung zur Dipl. Legasthietrainerin durch den EÖDL – Erster Österreichischer Dachverband für Legasthenie
September 2000 bis Juni 2005	Reife- und Diplomprüfung an der Bildungsanstalt für Kindergartenpädagogik der Schulschwestern , Amstetten Zusatzausbildung Hortpädagogik

FÄHIGKEITEN UND KOMPETENZEN

Muttersprache	Deutsch
Fremdsprache	Englisch
Psychologische Spezialisierung	▪ Kinder und Jugendliche ▪ Diagnostik und Intervention ▪ Lernpsychologie, Lerntherapie ▪ Entwicklungspsychologie ▪ Spielbasierte Entwicklungsverfahren für Kleinkinder
Sonstige Kompetenzen	▪ Ersthilfekurs ▪ Outdoor Ersthilfekurs ▪ Kleinkinder-Notfallkurs ▪ Helferschein