

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17

Zusammenhänge zwischen den Bayley Scales III und dem Beobachtungsverfahren MONDEY: Eine Validitätsanalyse

Groß, Rika¹
Traum, Jasmin¹
Heilig, Lena¹
Pauen, Sabina¹
Peykarjou, Stefanie^{1,2}

¹Affiliation: Psychologisches Institut, Universität Heidelberg
²Affiliation: Charlotte-Fresenius Hochschule, University of Psychology, Wiesbaden
Kontakt: stefanie.peykarjou@charlotte-fresenius-uni.de

Zusammenfassung: *Theoretischer Hintergrund:* Alltagsnahe Entwicklungsdiagnostik durch Beobachtung kann eine wichtige Ergänzung zu standardisierten Entwicklungstest darstellen. *Fragestellung:* Die vorliegende Arbeit untersucht Zusammenhänge zwischen dem standardisierten Entwicklungstest Bayley-III und dem Beobachtungsinstrument MONDEY hinsichtlich konvergenter Validität. Der Fokus liegt auf den Entwicklungsbereichen Denken und Sprache bei Kindern im Alter von 21 ($n_1 = 33$, $w = 14$) und 26 ($n_2 = 30$, $w = 15$) Monaten.

Methode: Die elterlichen Einschätzungen der Entwicklung wurden mittels MONDEY erhoben und mit den standardisierten Testergebnissen der Bayley-III in Zusammenhang gesetzt. Die Übereinstimmung der Beurteilungen innerhalb der Entwicklungsbereiche wurden korrelativ untersucht. Zusätzlich wurden explorative Analysen zu korrelativen Zusammenhängen über Entwicklungsbereiche hinweg durchgeführt. *Ergebnisse:* Die Ergebnisse zeigen hohe Korrelationen zwischen den Sprachskalen ($r = .75$, $p < .001$) sowie moderate Zusammenhänge zwischen den Kognitionsskalen ($r = .50$, $p < .001$) beider Instrumente. Die Skaleninterkorrelationen fallen bei den Bayley-III höher als in MONDEY. Diskussion und Schlussfolgerung: Die Befunde unterstützen die Annahme, dass MONDEY eine valide Ergänzung zur standardisierten Entwicklungsdiagnostik darstellen kann. Zudem scheint die Trennschärfe der MONDEY-Skalen für die Bereiche Kognition und Sprache höher zu sein als die der Bayley-III. Somit lassen sich Hinweise auf konvergente, aber auch diskriminante Validität der Instrumente finden.

Schlüsselwörter: Entwicklungsdiagnostik, Frühe Kindheit, Bayley-III, MONDEY

Relations between Bayley Scales III and the observation tool MONDEY: An analysis of validity

Abstract: *Theoretical background:* Developmental diagnostics via everyday observations and questionnaires can be an important addition to standardised tests, provided there are valid instruments. *Objective:* This study explores relations between the German versions of the Bayley Scales of Infant and Toddler Development III (Bayley-III) and the parent observation tool „Milestones of Normal Development in Early Years” (MONDEY) concerning convergent validity, focusing on the cognition and language scales among children aged 21 ($n_1 = 33$, $f = 14$) and 26 ($n_2 = 30$, $f = 15$) months. *Method:* The degree of agreement between parental

1 evaluations using MONDEY and standardized Bayley-III test results was evaluated. Agreement
2 among assessments was calculated via correlation analyses, and additional explorative analyses
3 were employed to assess correlational relations between areas of development. *Results:*
4 Findings reveal strong correlations between the language scales ($r = .75, p < .001$) and moderate
5 relations between the cognition scales ($r = .50, p < .001$) of both instruments. In addition,
6 relations across areas of development seem to be stronger in the Bayley-III assessment than for
7 the MONDEY scale. *Discussion and Conclusion:* These findings support the notion that
8 MONDEY can serve as a valid complement to standardized developmental diagnostics. In
9 addition, the statistical power of the MONDEY scales for discriminating the areas of cognition
10 and language appears to be higher than that of the Bayley-III. Thus, this study provides evidence
11 for both convergent and discriminant validity of the analyzed instruments.

12 **Keywords:** developmental diagnostics, early childhood, Bayley-III, MONDEY

Zusammenhänge zwischen den Bayley Scales III und dem Beobachtungsverfahren MONDEY: Eine Validitätsanalyse

„Die gegenwärtige Entwicklungspsychologie [ist] zu einem großen Teil die Wissenschaft fremdartigen Verhaltens von Kindern in fremden Situationen mit fremden Erwachsenen in kürzestmöglichen Zeitabschnitten“ (Bronfenbrenner, 1978, S. 33).

Entwicklungsdiagnostik im Säuglings- und Kleinkindalter hat zum grundsätzlichen Ziel, Aussagen über den Entwicklungsstand eines Kindes zu treffen, indem systematisch entwicklungsbezogene Daten erfasst werden (Frey, 2012). Dazu wird das gegebene Verhalten eines Kindes mit dem einer Normstichprobe verglichen und so auf der entsprechenden Entwicklungsvariable eingeordnet (Esser & Petermann, 2010). Besonders im Bereich der Frühprävention und -intervention hat die Entwicklungsdiagnostik damit in den letzten Jahren an Bedeutung gewonnen. Laut der zu Beginn des Artikels zitierten Beobachtung Bronfenbrenners liegt eine zentrale Schwierigkeit entwicklungspsychologischer Diagnostik darin, dass standardisierte, alltagsferne diagnostische Situationen zwar eine objektive Beurteilung von kindlichen Kompetenzen ermöglichen, aber nur einen Teil alltagsbedeutsamer Fähigkeiten widerspiegeln. Eine alltagsnahe Entwicklungsdiagnostik anhand von Beobachtungen sowie Befragungen kann daher eine wichtige Ergänzung darstellen, um effektiv valide und übertragbare Erkenntnisse über den Entwicklungsstand von Kindern zu sammeln. Standardisierte Entwicklungsdiagnostik bedeutet einen hohen Aufwand und stellt eine Ausnahmesituation für das Kind sowie die Familie dar. In der Frühpädagogik ist daher eine diagnostische Entwicklungsbeobachtung für alle betreuten Kinder gesetzlich vorgeschrieben (JKM & KMK, 2004). Auch Hagmann-von Arx, Meyer und Grob (2008) betonen, dass Entwicklungstests nur eine Informationsquelle zur Beurteilung des kindlichen Entwicklungsstandes sind: Erst die Integration von Beobachtungen verschiedener Parteien ermöglicht eine valide Einschätzung (Burgess et al., 2025).

Dabei sollten zunächst Zusammenhänge zwischen standardisierten Entwicklungstests und alltagsnaher Entwicklungsbeobachtung untersucht werden, um Rückschlüsse hinsichtlich kon-

1 kurrenter und diskriminanter Validität beider Erhebungsverfahren ziehen zu können. Das Elekt-
 2 ronische Supplement (ESM) 1 bietet eine Übersicht über bisherige Befunde zu Zusammenhän-
 3 gen elternbasierter Entwicklungsdiagnostik und standardisierter Entwicklungstests in den Be-
 4 reichen Denken, Sprache (allgemein, rezeptiv und expressiv) sowie gesamte Entwicklung. Ins-
 5 gesamt zeigt sich hier, dass der Zusammenhang zwischen Elternfragebögen und standardisier-
 6 ten Entwicklungstests moderat bis groß ist, wobei Eltern- und Testurteile zur Erfassung der
 7 sprachlichen Entwicklung tendenziell stärker übereinstimmen als Urteile zur kognitiven Ent-
 8 wicklung (*siehe ESM 1*). Innerhalb der sprachlichen Domäne scheinen Testskalen, die die ex-
 9 pressive Sprachentwicklung erfassen, höher mit Elternurteilen zusammenzuhängen als
 10 Testskalen zur rezeptiven Sprachentwicklung, was darauf zurückzuführen sein könnte, dass El-
 11 ternfragebögen meist ausschließlich das expressive Sprachverhalten abfragen (Perra et al.,
 12 2015; Reuner & Rosenkranz, 2014; Sachse & Von Suchodoletz, 2008). Einige Beobachtungs-
 13 instrumente für Eltern erfassen dennoch auch das rezeptive Sprachverständnis, indem die Be-
 14 obachter_innen beurteilen, ob ein Kind beispielsweise auf sprachliche Benennungen oder Auf-
 15 forderungen ohne Zeigegesten reagiert (z. B. Pauen, 2011).

16 Die vorliegende Arbeit beschäftigt sich mit der Frage, inwiefern elterliche Einschätzungen
 17 des kindlichen Entwicklungsstands mithilfe des Beobachtungsinstruments “Milestones of
 18 Normal Development in Early Years” (MONDEY; Pauen, 2011) und des standardisierten
 19 Entwicklungstests Bayley-III (Bayley, 2006; deutsche Version Reuner & Rosenkranz, 2014)
 20 übereinstimmen. Dazu wurden die Entwicklungsbereiche Denken und Sprache ausgewählt,
 21 welche eine hohe Relevanz für die psychologische Frühdiagnostik aufweisen. Auf die Durch-
 22 führung und Analyse der Motorik-Skalen wurde aus Rücksicht auf den zeitlichen Aufwand für
 23 die teilnehmenden Familien verzichtet. Der vorliegende Beitrag konzentriert sich auf die für
 24 die psychologische Diagnostik hoch bedeutsamen Bereiche Sprache und Kognition. Es wur-
 25 den zwei Stichproben mit Kindern im Alter von 21 und 26 Monaten getestet, ein Altersbe-
 26 reich, in dem große Fortschritte in der Denk- und Sprachentwicklung auftreten (z. B. Lohaus,
 27 Vierhaus & Lemola, 2024; Reuner & Pietz, 2006) und in welchem Sprachdiagnostik empfoh-
 28 len wird (Sansavini et al., 2021).

29 Quantitativ-diagnostische Beobachtungsverfahren, wie MONDEY (Pauen, 2011), gehen
 30 davon aus, dass sich kindliche Kompetenzniveaus anhand von Entwicklungsnormen erfassen
 31 und vergleichen lassen. Hierbei wird das kindliche Verhalten anhand von standardisierten
 32 Merkmalslisten und konkreten Verhaltensbeschreibungen für verschiedene Entwicklungsbe-

1 reiche von einer Bezugsperson beurteilt (Viernickel, 2014). MONDEY-Meilensteine beschrei-
 2 ben alltagsrelevantes Verhalten der Kinder in standardisierter Form. Die Bezugsperson ent-
 3 scheidet, ob das Kind die beschriebene Verhaltensweise schon mindestens zweimal in der be-
 4 schriebenen Form gezeigt hat. Damit geht MONDEY weniger von einer bestimmten Theorie
 5 der Entwicklung als vielmehr von der Frage aus, welche Kompetenzen das Kind im Alltag
 6 zeigt (Frey, 2012). Standardisierte Entwicklungstests folgen meist keinem einheitlichen, kon-
 7 sistenten und theoretischen Fundament (Esser & Petermann, 2010). Auch die Bayley-III ba-
 8 siert nicht auf einer speziellen Entwicklungstheorie, sondern versucht, aktuelle entwicklungs-
 9 psychologische Erkenntnisse und praxisgerechte Anwendung gemäß psychologischer Test-
 10 standards zu vereinen. Die Konstruktion entspricht der einer Stufenleiter, ordnet Aufgaben
 11 also in ansteigender Schwierigkeit an (Reuner & Rosenkranz, 2014).

12 Aufgrund vielfältiger Befunde, welche eine grundlegende psychometrische Güte der
 13 Bayley-III indiziert (z. B. Bayley, 2006; Burgess et al., 2025, Deroma et al., 2013; Hammerl
 14 et al., 2025, Lopes da Silva et al., 2024; Macha & Petermann, 2015; Orioli, Johnston, Krebs,
 15 Francisco & Carvalho, 2022; Reuner & Rosenkranz, 2014) sowie des breiten Einsatzes in un-
 16 terschiedlichsten Kontexten, wurde dieser Test in der vorliegenden Studie zur Validierung der
 17 MONDEY-Skalen gewählt. Gleichzeitig ist davon auszugehen, dass bei der Entwicklungsbe-
 18 urteilung durch MONDEY aufgrund der alltagsnahen, auf einen längeren Beobachtungszeit-
 19 raum bezogenen Datengrundlage des Instruments wertvolle Informationen einfließen können,
 20 welche in einer standardisierten Entwicklungsdiagnostik anhand der Bayley-III nicht verfüg-
 21 bar sind. Hindernisse wie Schüchternheit, Zeitdruck oder Verweigerung treten genauso wenig
 22 auf wie Fragen der Generalisierbarkeit oder zu enge Normvorgaben (Perra et al., 2015; Bel-
 23 ler & Beller, 2010; Esser & Petermann, 2010). Es handelt sich bei der vorliegenden Arbeit
 24 somit nicht um eine rein konvergente Validierungsstudie, bei welcher eine höchstmögliche
 25 Korrelation zwischen beiden Instrumenten angestrebt wird. Vielmehr können Unterschiede in
 26 den Ergebnissen von MONDEY und Bayley-III und dadurch eine eingeschränkte Korrelation
 27 auch auf valide Differenzen einer Entwicklungsbeurteilung durch Erziehungsberechtigte im
 28 Vergleich zu einer standardisierten Testung hinweisen (Macha, Proske & Petermann, 2005).

29 Die Analyse der zu vergleichenden Instrumente auf Item-Ebene stützt die Vermutung, dass
 30 ein signifikanter Zusammenhang zwischen MONDEY und Bayley-III zu erwarten ist, welcher
 31 aufgrund spezifischer Anteile beider Instrumente jedoch eher im mittleren Bereich liegen
 32 sollte. Die Items zur Erfassung der sprachlichen Entwicklung weisen mehr Ähnlichkeiten zwi-
 33 schen den beiden Instrumenten auf als im kognitiven Bereich. Zudem wird die Sprache im

Bayley-III zum Teil aufgrund von Beobachtungen des Sprachverhaltens während der gesamten Testung, *nicht bloß* anhand von festen Aufgaben, eingeschätzt, was der Beobachtungsmethode von MONDEY näherkommt und die Items ähnlicher macht. Der kognitive Entwicklungsstand wird dagegen im Bayley-III vorwiegend durch gezielte Aufgaben ermittelt, *wobei es schwierig sein kann, valide Items zur Testung kognitiver Fähigkeiten zu entwickeln* (Weiers et al., 2025), während MONDEY auch hier ausschließlich auf Beobachtungen im Alltag basiert. Dies könnte zu einem höheren Zusammenhang für die Sprachskalen von MONDEY und Bayley-III als für die Kognitionsskalen führen.

Auf Basis der vorliegenden Literatur (*siehe ESM 1*), der vorangegangenen Erläuterungen zu konvergenten und diskriminanten Aspekten der zu vergleichenden Instrumente sowie des Vergleichs der beiden Instrumente auf Item-Ebene wurden folgende Hypothesen postuliert: Es wird eine signifikante Korrelation mittlerer Größe ($.30 \leq r < .50$) zwischen der Kognitionsskala des Entwicklungstests Bayley-III und dem Entwicklungsbereich Denken von MONDEY erwartet. Weiterhin wird zwischen der Sprachskala des Entwicklungstests Bayley-III und der MONDEY-Sprachskala ein hoher korrelativer Zusammenhang erwartet ($.50 \leq r \leq .80$).

An Entwicklungstests und insbesondere an den Bayley Scales wird häufig kritisiert, dass einzelne Entwicklungsbereiche nicht klar voneinander abgegrenzt erfasst werden (Anderson & Burnett, 2017; Reuner, Fields, Wittke, Löffprich & Pietz, 2013). Vor allem die kognitive Entwicklung werde nicht isoliert, sondern unter Einbezug sprachlicher und motorischer Fähigkeiten gemessen (z. B. Morgan, Honan, Allsop, Novak & Badawi, 2019; Sherwell, Reid, Reddihough, Wrennall, Ong & Stargatt, 2014). Laut Kipping und Kolleg_innen (2024) sind die kognitiven Scores prädiktiv für motorische Leistungen im Schulalter. Viele kognitive Items des Bayley-III setzen sprachliches Verständnis oder motorische Ausführung voraus. Zudem können auch soziale Entwicklung und Selbstregulation die Testergebnisse beeinflussen, bspw. wenn Kinder in der Testsituation gehemmt reagieren (Perra et al., 2015). Die MONDEY-Meilensteine im Bereich Denken scheinen die genannte Problematik zu vermeiden, da sie auf Alltagsbeobachtungen basieren und nicht direkt auf sprachliche oder feinmotorische Fähigkeiten beruhen (Pauen, 2011). Inwiefern die Skalenspezifität für MONDEY höher ausfällt als für den Bayley-III, wurde in dieser Studie explorativ untersucht.

Methoden

Stichprobe

Für die Studie wurden insgesamt $N = 63$ Kinder (weiblich $n = 29$) über die Kartei an Studienteilnahmen interessierter Familien des Lehrstuhls für Entwicklungs- und Biologische Psychologie der Universität Heidelberg rekrutiert. Eine Aufwandsentschädigung erfolgte nicht. Die Stichprobe setzt sich aus zwei Kohorten zusammen: Kinder im Alter von 21 Monaten ($n_1 = 33$, weiblich $n = 14$) und 26 Monaten ($n_2 = 30$, weiblich $n = 15$). In der ersten Kohorte erstreckt sich das Alter der Kinder von 20 Monaten und 22 Tagen bis 23 Monate und 00 Tagen ($M = 21.49$, $SD = .62$). Das Alter der Kinder der zweiten Kohorte zum Zeitpunkt der ersten Testung reicht von 25 Monaten und 23 Tagen bis 26 Monate und 29 Tage ($M = 26.12$, $SD = .18$). Die Erhebung wurde im Zeitraum von November 2019 bis September 2022 durchgeführt. Sie basiert auf verschiedenen Qualifikationsarbeiten bzw. Praktika-Tätigkeiten Studierender und musste aufgrund der Corona-Pandemie für mehrere Monate unterbrochen werden. Die finale Datenauswertung erfolgte aus internen Gründen erst 2024. Eine A-priori-Poweranalyse ergab bei einem erwarteten mittleren Effekt von $\rho = .3$ und einem einseitigen Test mit $\alpha = .05$ für eine hohe Testpower von $1-\beta = .8$ eine benötigte Stichprobengröße von $N = 64$. Mit der final erzielten Stichprobengröße von $N = 63$ kann laut Post-hoc-Analyse eine Testpower von $1-\beta = .74$ erreicht werden (berechnet mit G*Power; Faul, Erdfelder, Lang & Buchner, 2007).

Der Großteil der Eltern gab an, über einen Hochschulabschluss (Mutter 54 %, Vater 57 %) bzw. Fachhochschulabschluss (Mutter 16 %, Vater 14 %) zu verfügen. 19.0% der Mütter und 15.9% der Väter gaben das Abitur, 9.5% der Mütter und Väter die Mittlere Reife und nur 1 Vater (1.6%) den Hauptschulabschluss als höchsten Schulabschluss an. Die Entwicklungsdokumentation via MONDEY wurde größtenteils durch die Mütter der Kinder vorgenommen (92.1%).

Instrumente

Zur Erfassung des kindlichen Entwicklungsstandes wurde zum einen das Entwicklungsbeobachtungsinstrument MONDEY 0-3 (Milestones of Normal Development in Early Years) eingesetzt (Pauen, Heilig, Danner, Haffner, Tettenborn & Ross, 2012b). MONDEY 0-3 besteht insgesamt aus 111 Meilensteinen, die acht Entwicklungsbereichen zugeordnet sind

(Grobmotorik, Feinmotorik, Wahrnehmung, Denken, Sprache, Soziale Beziehungen, Selbstregulation, Gefühle). Die Meilensteine werden in der Regel von normal entwickelten Kindern im Laufe der ersten drei Lebensjahre erreicht. Um eine möglichst genaue Beurteilung zu gewährleisten, gibt es zu jedem Meilenstein eine Definition sowie Kriterien, unter welchen Umständen der Meilenstein als erreicht gilt. *Sie sind im Alltag in Form von Verhaltensweisen oder Fähigkeiten gut beobachtbar. Die Beurteilung erfolgt dabei durch Erziehungsberechtigte oder pädagogische Fachkräfte.* Vorliegende Studien zeigen eine sehr gute Konstruktvalidität (Pauen et al., 2012b) sowie eine gute *interne Konsistenz von $\alpha = .97$ (Pauen & Heilig, 2012) und Inter-Rater-Reliabilitäten zwischen zwei Fachkräften von $\kappa = .77$ (Entwicklungsbereich Sprache) bis $\kappa = .49$ (Entwicklungsbereich Gefühle)* (Heilig & Pauen, 2013). Als Screening ermöglicht MONDEY eine erste Orientierung über den Entwicklungsstand eines Kindes, welcher ggf. im Rahmen einer ausführlicheren Diagnostik objektiviert und differenziert werden muss (Lohaus et al., 2024). Die Denk- und Sprachentwicklung werden in MONDEY über zwei getrennte Skalen und insgesamt 24 Meilensteine abgedeckt. Die MONDEY Skala „Denken“ beinhaltet acht Meilensteine aus den Bereichen „darstellen und symbolisieren“, „räumlich ordnen“ und „planen“. Die MONDEY Skala „Sprache“ besteht aus 16 Meilensteinen zum „Wort- und Satzverständnis“, zur „Laut-, Silben-, Wort- und Satzproduktion“ sowie zur Verwendung grammatikalisch bedeutsamer Strukturen, wie etwa Personalpronomen. *Es werden also sowohl rezeptive als auch produktive Aspekte der Sprachentwicklung erfasst.*

Als Vergleichsinstrument wurde der standardisierte Entwicklungstest Bayley-III genutzt (Bayley Scales of Infant and Toddler Development 3, deutsche Version; Reuner & Rosenkranz, 2014). Dieser besteht aus drei Skalen: Kognitionsskala, Sprachskala (Untertests Sprache Rezeptiv und Sprache Expressiv) sowie Motorik-Skala (Untertests Grob- und Feinmotorik). *Die Kognitionsskala beinhaltet 91 Aufgaben zu Themen wie Objektpermanenz, Spielverhalten, Zahlenkonzepten, Puzzeln oder Gruppenbildung. Im Untertest Sprache Rezeptiv wird das Sprachverständnis des Kindes in 49 Aufgaben geprüft, die expressiven Sprachkenntnisse werden mit 46 Items erfasst. Die hohe Aufgabendichte bedingt eine lange Durchlaufdauer, die laut Autor_innen je nach Alter des Kindes 50 bis 90 Minuten für die gesamte Testung beträgt. MONDEY erfasst die Bereiche Sprache und Kognition somit mit deutlich weniger Items als Bayley-III.* Der Fokus der vorliegenden Studie liegt auf den psychologisch primär relevanten Bereichen der Denk- und Sprachentwicklung, daher wurden die Kognitionssowie Sprachskala (rezeptiv und expressiv) durchgeführt. *Der deutsche Bayley-III wurde von Reuner und Rosenkranz (2014) an insgesamt $N = 1009$ Kindern normiert, sodass aus den Rohwerten standardisierte Skalenwerte ($M = 100$, $SD = 15$) und dazugehörige Prozentränge*

ermittelt werden können. Laut Macha und Petermann (2015) liegen durch das standardisierte Material, die vorgegebenen Instruktionen und Testreihenfolge sowie die Normwerttabellen eine hohe Durchführungs-, Auswertungs- und Interpretationsobjektivität vor. *Obwohl im Rahmen dieser Studie nur die Kognitions- und Sprachskala durchgeführt wurden, bleiben Testeigenschaften wie standardisiertes Material und Instruktionen erhalten. Auch die für die Durchführungsobjektivität wichtige feste Testreihenfolge wird dadurch nicht gestört, da ausschließlich die Reihenfolge innerhalb der Skalen eingehalten werden muss, während die Reihenfolge der Skalen flexibel angepasst werden kann (Reuner & Rosenkranz, 2014).* Als gegeben können ebenfalls die Inhalts- sowie die Konstruktvalidität betrachtet werden (Reuner & Rosenkranz, 2014).

Durchführung

Die Erhebung des Bayley-III fand *im Einzelsetting* in einem ruhigen Testraum statt. Aufgrund der Länge des Tests war bei 27 Kindern ein zweiter Termin erforderlich, um die Testung abschließen zu können. *Die Zeitangabe von 50-90 Minuten für alle Skalen des Bayley-III sahen wir in der Praxis nicht bestätigt. Mögliche Gründe hierfür könnten die wie im Ergebnisteil berichteten oft überdurchschnittlichen Testergebnisse der untersuchten Kinder sein, welche sich im Bayley-III durch die Konstruktion als Stufenleiter direkt auf die Durchführungsdauer auswirken.* Das Intervall zwischen den beiden Testzeitpunkten wurde dabei so gering wie möglich gehalten und betrug im Durchschnitt auf 13,96 Tage ($SD = 10$; $min = 4$, $max = 50$). Bei acht Kindern konnte nur die kognitive Skala erhoben werden, *da der vereinbarte zweite Termin in den Beginn der Coronapandemie fiel und somit nicht stattfinden konnte.* Somit liegen vollständige Datensätze von $N = 55$ Kindern vor ($n_1 = 30$; $n_2 = 25$). Um Dropouts zu vermeiden, flossen in die Analysen zu den kognitiven Skalen die Daten der gesamten Stichprobe ($N = 63$) ein.

Nach Abschluss der Bayley-III Testung wurde mit den Eltern ein Termin für ein Telefongespräch vereinbart, um sie in die Entwicklungsdokumentation MONDEY einzuführen. Während des Telefonats gaben die Eltern *über das Online-Dokumentationstool des Instruments („MONDEYOnline“)* ein, welche Meilensteine das Kind bereits erreicht hatte. Dieser Termin wurde möglichst bald nach Abschluss der Bayley-III Testung vereinbart, um den zeitlichen Abstand zwischen beiden Messungen möglichst gering zu halten ($M = 16.05$ Tage; $SD = 12.01$; $min = 0$, $max = 46$).

Analysen

Alle Analysen basieren auf den Skalenrohwerten. Diese setzen sich bei MONDEY 0-3 aus der Summe der erreichten Meilensteine in jedem einzelnen Bereich zusammen. *Der MONDEY-0-3 Gesamtscore bildet sich aus der skalenübergreifenden Summe aller gekonnten Meilensteine.* Der Rohwert der Kognitionsskala des Bayley-III errechnet sich aus der Summe der gekonnten Items. Für die Sprachskala des Bayley-III wurden die Rohwerte der beiden Untertests Sprache Rezeptiv und Expressiv summiert. Die Daten wurden mithilfe der Statistiksoftware SPSS Statistics 28 ausgewertet. Alle Analysen wurden sowohl für die Gesamtstichprobe als auch für jede Subgruppe (21 Monate, 26 Monate) getrennt durchgeführt. Alle Variablen wurden auf Normalverteilung überprüft. Die Bayley-III-Skalen waren normalverteilt, die MONDEY-0-3 Skalenwerte zeigten sich überwiegend deutlich linksschief und nicht normalverteilt. *Die Mittelwerte und Standardabweichungen aller für die Analysen relevanten Variablen sowie Dispersions- und Verteilungsmaße wie Schiefe und Kurtosis werden für die Gesamtstichprobe und getrennt nach Kohorten in ESM 2 dargestellt. Aufgrund der Verteilungen wurden non-parametrische Rangkorrelationen nach Spearman berechnet (s. Tabelle 1). Die Signifikanztests erfolgten einseitig mit einem α -Niveau von .05.* Eine Ausreißeranalyse mittels Boxplots ergab pro Gruppe drei Ausreißer. In einer Kontrollanalyse wurden diese Versuchspersonen ausgeschlossen, was jedoch zu keiner bedeutsamen Veränderung der Ergebnisse führte. Ausgehend von diesem Befund werden nachfolgend sämtliche Analysen unter Berücksichtigung des vollständigen Datensatzes berichtet.

Ergebnisse

In beiden Kohorten erreichen die Kinder auf beiden Instrumenten hohe Werte. In der Gruppe der 26 Monate alten Kinder haben sowohl im MONDEY 0-3 Entwicklungsbereich Denken als auch im Entwicklungsbereich Sprache über die Hälfte der Kinder nach Einschätzung ihrer Eltern bereits alle Meilensteine erreicht ($n = 21$ Kinder im Bereich Denken, $n = 16$ im Bereich Sprache), *es zeigen sich somit Deckeneffekte.* In der Gruppe der 21 Monate alten Kinder war die Varianz der Einschätzungen in MONDEY 0-3 deutlich höher. Hier erreichten nur $n = 10$ Kinder alle Meilensteine im Bereich Denken und nur $n = 2$ Kinder alle Meilensteine im Bereich Sprache. Auch im Bayley-III schnitten alle Kinder sehr gut ab. Es zeigen

sich gemittelte Skalenwerte *der Kognitionsskala im oberen Durchschnittsbereich im Vergleich zur Bayley-III-Normstichprobe* (normiert auf $M = 100$, $SD = 15$), sowohl in der Gruppe der 21 Monate alten Kinder ($M = 110.91$, $SD = 17.16$) als auch in der Gruppe der 26 Monate alten Kinder ($M = 107.33$, $SD = 15.53$). Auch die Skalenwerte der Sprachskala von $M = 112.07$ ($SD = 20.13$) in der jüngeren Gruppe liegen im oberen Durchschnittsbereich und sind mit $M = 121.12$ ($SD = 20.52$) in der älteren Gruppe sogar überdurchschnittlich. Dies schränkt die Varianz zur Berechnung von Korrelationen ein und kann zu einer Unterschätzung der tatsächlichen Zusammenhänge führen.

Korrelation der Kognitionsskalen

In der Gesamtstichprobe ergab sich zwischen den Rohwerten der Kognitionsskala des Bayley-III und dem MONDEY 0-3 Entwicklungsbereich Denken eine Korrelation von $r = .50$, $p < .001$ ($N = 63$; siehe Tabelle 1). Für die Kohorte der 21 Monate alten Kinder wurde zwischen den Rohwerten der Kognitionsskala des Bayley-III und des MONDEY 0-3 Entwicklungsbereichs Denken eine Korrelation von $r = .25$, $p = .20$ ($n1 = 33$; siehe ESM 3 Tabelle 1a) gefunden. In der Kohorte der 26 Monate alten Kinder ergab sich eine Korrelation von $r = .44$, $p < .05$ ($n2 = 30$; siehe ESM 3, Tabelle 1b). Somit kann die Annahme, dass die Kognitionsskalen des Bayley-III und des MONDEY 0-3 einen Zusammenhang mittlerer Größe ($r = .30 - .50$) aufweisen, für die Gesamtstichprobe und die Gruppe der älteren Kinder bestätigt werden. Für die Kohorte der 21 Monate alten Kinder dagegen fällt der Zusammenhang zwischen den Kognitionsskalen beider Instrumente geringer aus als postuliert.

Korrelation der Sprachskalen

In der Gesamtstichprobe ergab sich zwischen der Sprachskala des Bayley-III und dem MONDEY Entwicklungsbereich Sprache eine Korrelation von $r = .75$, $p < .001$ ($N = 55$; siehe Tabelle 1). Auch in der Kohorte der 21 Monate alten Kinder zeigte sich zwischen den Sprachskalen eine statistisch signifikante Korrelation von $r = .62$ ($p < .001$; $n1 = 30$; siehe ESM 3, Tabelle 1a). Bei den 26 Monate alten Kindern konnte eine Korrelation von $r = .59$, $p < .001$ ($n2 = 25$; siehe ESM 3, Tabelle 1b) zwischen den Sprachskalen beider Instrumente gefunden werden. Folglich stützen die vorliegenden Daten die zweite Hypothese, welche von einem hohen Zusammenhang ($r = .50 - .80$) zwischen den Bayley-III und MONDEY-0-3 Sprachskalen ausgeht, sowohl für die Gesamtstichprobe als auch für die beiden Teilstichproben.

((Tabelle 1 hier einfügen))

Explorative Zusatzanalysen

Eine Visualisierung der berechneten Korrelationen ist in Abbildung 1 zu finden.

((Abbildung 1 hier einfügen))

Zwischen der Kognitionsskala des Bayley-III und dem allgemeinen Entwicklungsstand nach MONDEY 0-3 wurde für die Gesamtstichprobe eine Korrelation von $r = .63, p < .01$ gefunden (26 Monate alte Kinder $r = .52, p < .01$, 21 Monate alte Kinder $r = .50, p < .01$). Die Korrelation zwischen der Kognitionsskala des Bayley-III und dem MONDEY-Entwicklungsbereich Denken fiel, wie oben berichtet, deutlich kleiner aus, ($z = 2.96, p < .01$). *Somit korreliert die Bayley Kognitionsskala enger mit dem MONDEY Gesamtscore als mit der Denkskala (siehe Angaben zu „Fragestellung 1“ in Abbildung 1).*

Zwischen der Kognitions- und der Sprachskala des Bayley-III wurde in der Gesamtstichprobe eine Korrelation von $r = .75, p < .$ (26 Monate $r = .75, p < .$, 21 Monate $r = .50, p < .$) berechnet. Bei MONDEY korrelierten die beiden Skalen zu $r = .63, p < .$ miteinander (26 Monate alte Kinder $r = .70, p < .$, 21 Monate alte Kinder $r = .38, p < .$). Der Unterschied der beiden Korrelationen in der Gesamtstichprobe ist signifikant ($z = 2.47, p = .02$), *was darauf hinweist, dass MONDEY die Entwicklungsbereiche Denken und Sprache trennschärfer erfassen kann als der Bayley III.*

Diskussion

Die vorliegende Studie untersuchte die Zusammenhänge zwischen MONDEY 0-3 (Pauen, 2011), einem Eltern-Rating von Entwicklungsmeilensteinen, und dem Bayley-III (Bayley, 2006), einem standardisierten Entwicklungstest, in einer Stichprobe von 21 und 26 Monate alten Kindern. In der Gesamtstichprobe korrelierten die Sprachskalen hoch miteinander ($r = .73$), was wie erwartet auf eine hohe konvergente Validität beider Instrumente hindeutet. Zwischen den Testergebnissen der Kognitionsskalen zeigte sich wie erwartet eine Korrelation im mittleren Bereich ($r = .50$), welche auf eine mittlere konvergente Validität schließen lässt. Explorative Analysen unterstützen die Annahme, dass MONDEY 0-3 eine trennschärfere Erfassung der Kognition ermöglicht als die Bayley-III-Skalen.

Die Sprachskalen von MONDEY 0-3 und des Bayley-III weisen inhaltliche Überlappungen auf, weshalb eine hohe Korrelation erwartet wurde. Dies konnte in allen Analysen bestätigt werden (Gesamtgruppe $r = .75$; 21 Monate alte Kinder $r = .62$; 26 Monate alte Kinder $r = .59$). Unterschiede zwischen den Skalen lassen sich vermutlich auf methodische Differenzen zwischen beiden Instrumenten (Elterneinschätzung vs. Entwicklungstest) sowie einen vergleichsweise höheren Anteil an Items zum rezeptiven Sprachverständnis im Bayley-III erklären.

Die Kognitionsskalen beider Instrumente weisen trotz gleichem Entwicklungsfokus weniger inhaltliche Überschneidungen in den Items auf. *Wie erwartet* wurden Korrelationen im mittleren Bereich gemessen (Gesamtstichprobe $r = .50$; 21 Monate alte Kinder $r = .25$ (ns); 26 Monate alte Kinder $r = .44$), *was vermutlich auf die Berücksichtigung elterliche Einschätzungen basierend auf alltäglichen Erfahrungen in MONDEY 0-3 zurückzuführen ist, welche in einer klassischen Entwicklungstestung nicht genutzt werden.*

In den Ergebnissen zeigt sich deutlich, dass beide Instrumente zwar konvergente Validität, aber auch diskriminante Anteile aufweisen. Als Beobachtungs- und Dokumentationsinstrument ist die Datengrundlage von MONDEY 0-3 das natürliche Verhalten eines Kleinkindes in dessen Alltag, welches von Eltern anhand möglichst konkret formulierter Anhaltspunkte eingeschätzt werden soll. Subjektive Einschätzungen sowie Vorannahmen oder Wunschvorstellungen können hierbei die Dokumentation verzerren. Als standardisierter Entwicklungstest basieren die Ergebnisse des Bayley-III auf klar vorgegebenen Aufgaben und folgen festen Bewertungsregeln. Damit ist der Test objektiver, aber auch situationsspezifischer.

Theoretisch lässt sich vermuten, dass die Unterschiede zwischen den Kognitionsskalen teilweise auf mangelnde Trennschärfe der Bayley-III Skalen zurückzuführen sein könnten. Es scheint wahrscheinlich, dass die kognitive Entwicklung in der Testung nicht isoliert, sondern mit Anteilen sprachlicher, motorischer und sozialer Fähigkeiten erfasst wird (z. B. Morgan et al., 2019; Sherwell et al., 2014). Die Items der MONDEY 0-3 Denkskala können im Alltag hingegen vermutlich besser *unabhängig von* anderen Entwicklungsbereichen *untersucht* werden. Zusatzanalysen bestätigen diese Überlegungen: *Die Korrelation zwischen der Bayley-III Kognitionsskala und dem MONDEY 0-3 -Gesamtentwicklungsscore fällt signifikant höher aus als die Korrelation mit der Subskala Denken.*

Darüber hinaus fallen die Skaleninterkorrelationen zwischen Sprache und Denken bei den Bayley-III Scales höher aus als bei MONDEY, ein weiterer Hinweis darauf, dass MONDEY

1 die Entwicklungsbereiche Denken und Sprache trennschärfer erfasst als der Bayley-III. *Dieses*
 2 *Ergebnis unterstützt die Vermutung, dass in die Kognitionsskala des Bayley-III neben Denkfä-*
 3 *higkeiten auch weitere Entwicklungskomponenten einfließen, da die kognitive Entwicklung*
 4 *nicht isoliert, sondern unter Voraussetzung sprachlicher, motorischer und sozialer Fähigkei-*
 5 *ten zum Bearbeiten der Testaufgaben erfasst wird (z.B. Morgan et al., 2019; Sherwell et al.,*
 6 *2014), was bei der reinen Alltagsbeobachtung durch MONDEY weniger zutrifft.*

7 Limitationen der vorliegenden Studie liegen in der kleinen Stichprobe, geringen Altersran-
 8 ges, den Deckeneffekten in den MONDEY 0-3 Skalen und der fehlenden Repräsentativität
 9 der Stichprobe. *Es zeigt sich ein deutlich überdurchschnittlicher Akademikeranteil unter den*
 10 *Eltern. Dies könnte zu einer Unterschätzung der Zusammenhänge zwischen Bayley-III Maßen*
 11 *und MONDEY-Scores geführt haben. Für die vollständige Beantwortung diagnostischer*
 12 *Fragestellungen wäre es hilfreich, auch Daten von Kindern aus Elternhäusern mit gerin-*
 13 *gerem Bildungshintergrund zu erheben. MONDEY 0–3 könnte insbesondere in Kitas helfen,*
 14 *Entwicklungsrisiken über alle Bevölkerungsgruppen hinweg frühzeitig zu erkennen. Auch*
 15 *wenn die MONDEY Meilensteine gemeinsam mit Fachkräften und Eltern formuliert wurden,*
 16 *sollten zukünftige Studien gezielt prüfen, ob die Items sprachlich so formuliert sind, dass sie*
 17 *auch für Eltern mit geringerem Bildungsstand gut verständlich sind. Für das Abschneiden in*
 18 *entwicklungspsychologischen Testungen ist der Bildungsstand und das Einkommen der Eltern*
 19 *hochrelevant, vermutlich unter anderem durch ein kognitiv stimulierenderes häusliches Um-*
 20 *feld und positivere Erziehungspraktiken wie Responsivität und Wärme (Linver, Brooks-Gunn*
 21 *& Kohen, 2002; Rakesh, Lee, Gaikwad & McLaughlin, 2024; Thomas & Coecke, 2023). Dies*
 22 *könnte die Deckeneffekte in der vorliegenden Stichprobe erklären. Unabhängige Replikati-*
 23 *onsstudien zur Untersuchung der Validität mit repräsentativen Samples sind daher wün-*
 24 *schenswert.*

25 Einerseits ist eine Überschätzung von Effektstärken bei kleinen Stichproben besonders
 26 wahrscheinlich (Fritz, Morris & Richler, 2012), andererseits führt eine Einschränkung der Va-
 27 rianz (z. B. durch selektive Stichproben) zu einer Unterschätzung von Effektstärken für Zu-
 28 sammenhänge zwischen Variablen. Beide Effekte könnten die vorliegenden korrelativen Be-
 29 funde dieser Studie gegenläufig beeinflusst haben.

30 *Eine weitere Limitation betrifft den Studienaufbau: Die Bayley-III-Testung fand vor der*
 31 *MONDEY 0–3-Beurteilung statt, bei welcher ein Elternteil anwesend war. So kann nicht aus-*
 32 *geschlossen werden, dass die Beobachtung während der Testung die elterliche Einschätzung*
 33 *beeinflusst hat. Eine direkte Übertragung des Testungsgeschehens in die MONDEY 0-3-Werte*

1 *ist jedoch unwahrscheinlich, da die dort definierten Meilensteine in mindestens zwei unab-*
2 *hängigen Situationen und Tagen beobachtet worden sein müssen, um als erfüllt zu gelten. Zur*
3 *Kontrolle dieses Aspekts könnte jedoch in weiterführenden Studien die Elternbeurteilung vor*
4 *der Testung oder durch den nicht-anwesenden Elternteil erfolgen.*

5 Gerade in den ersten Lebensjahren können situative und soziale Aspekte einer Testung die
6 Ergebnisse stark beeinflussen, wodurch die Reliabilität und Validität frühkindlicher Diagnos-
7 tik potentiell eingeschränkt werden. Der Bayley-III ist psychometrisch umfassend untersucht,
8 weist ausreichende Gütekriterien auf und bildet einen zentralen Pfeiler der diagnostischen Ab-
9 klärung bei Verdacht auf Entwicklungsverzögerung. Die externe Validität der standardisierten
10 Diagnostik könnte jedoch durch Einbezug standardisierter Entwicklungsbeobachtung gestärkt
11 werden. Als Beobachtungsinstrument für Eltern und pädagogische Fachkräfte bietet MON-
12 DEY 0-3 eine Antwort auf die Kritik Bronfenbrenners (1978, S.33), Entwicklungspsycholo-
13 gie sei „zu einem großen Teil die Wissenschaft fremdartigen Verhaltens von Kindern in frem-
14 den Situationen mit fremden Erwachsenen in kürzestmöglichen Zeitabschnitten“. *Das bedeu-*
15 *tet nicht, dass eine Entwicklungsbeurteilung über MONDEY 0-3 eine standardisierte Entwick-*
16 *lungstestung gänzlich ersetzen kann bzw. sollte, da diese Beurteilung nicht durch hierfür qua-*
17 *lifizierte Fachkräfte erfolgt. Dies macht eine Diagnosestellung unzulässig. Leistungen eines*
18 *Kindes in einer Laborsituation und im Alltag sind nicht unbedingt deckungsgleich. Aus die-*
19 *sem Grund scheint es ratsam, in der Praxis ein kombiniertes Vorgehen zu wählen, um kindli-*
20 *che Kompetenzen aus möglichst unterschiedlichen Blickwinkeln erfassen zu können.*

Literatur

- Anderson, P. J. & Burnett, A. (2017). Assessing developmental delay in early childhood –
concerns with the Bayley-III scales. *The Clinical Neuropsychologist*, 31(2), 371-381.
DOI: 10.1080/13854046.2016.1216518
- Bayley, N. (2006). *Bayley Scales of Infant and Toddler Development: Technical Manual* (3.
ed.). Psychological Corporation.
- Beller, K. & Beller, S. (2010). *Kuno Beller's Entwicklungstabelle* (9. Aufl.). Freie Universität
Berlin.
- Blaggan, S., Guy, A., Boyle, E. M., Spata, E., Manktelow, B. N., Wolke, D. et al. (2014). A
parent questionnaire for developmental screening in infants born late and moderately
preterm. *Pediatrics*, 134(1), 55-62. <https://doi.org/10.1542/peds.2014-0266>
- Bornstein, M. H. & Haynes, O. M. (1998). Vocabulary competence in early childhood: Meas-
urement, latent construct, and predictive validity. *Child Development*, 69(3), 654–671.
<https://doi.org/10.2307/1132196>
- Bronfenbrenner, U. (1978). Ansätze zu einer experimentellen Ökologie menschlicher Ent-
wicklung. In R. Oerter (Hrsg.), *Entwicklung als lebenslanger Prozess – Aspekte und
Perspektiven* (S. 33-65). Hamburg: Hoffmann Campe.
- Burgess, A., Luke, C., Jackman, M., Wotherspoon, J., Whittingham, K., Benfer, K. et al.
(2025). Clinical utility and psychometric properties of tools for early detection of de-
velopmental concerns and disability in young children: A scoping review. *Develop-
mental Medicine & Child Neurology*, 67, 286-306.
<https://doi.org/10.1111/dmcn.16076>
- Deroma, L., Bin, M., Tognin, V., Rosolen, V., Valent, F., Barbone, F. & Carrozzi, M. (2013).
Concordanza interesaminatore del test Bayley III in una coorte di bambini del Nord-
Est Italia [Interrater reliability of the Bayley III test in the Italian Northern-Adriatic
Cohort II]. *Epidemiologia e prevenzione*, 37(4-5), 297-302.
- Esser, G. & Petermann, F. (2010). *Entwicklungsdiagnostik*. Göttingen: Hogrefe.
- Faul, F., Erdfelder, E., Lang, A.-G. & Buchner, A. (2007). GPower 3: A flexible statistical
power analysis program for the social, behavioral, and biomedical sciences. *Behavior
Research Methods*, 39(2), 175–191. <https://doi.org/10.3758/BF03193146>
- Frey, B. (2012). Entwicklungsdiagnostik. In M. Cierpka (Hrsg.), *Frühe Kindheit 0-3. Bera-
tung und Psychotherapie für Eltern mit Säuglingen und Kleinkindern* (S. 457–466).
Berlin, Heidelberg: Springer.

- 1 Fritz, C. O., Morris, P. E. & Richler, J. J. (2012). Effect size estimates: current use, calcula-
2 tions, and interpretation. *Journal of Experimental Psychology: General*, 141(1), 2-18.
3 <https://doi.org/10.1037/a0024338>
- 4 Hagmann-von Arx, P., Meyer, C. S. & Grob, A. (2008). Intelligenz- und Entwicklungsdiag-
5 nostik im deutschen Sprachraum. *Kindheit und Entwicklung*, 17(4), 232–242.
6 <https://doi.org/10.1026/0942-5403.17.4.232>
- 7 Hammerl, M., Zimmermann, M., Posod, A., Pupp Peglow, U., Höch, M., Griesmaier, E. et al.
8 (2025), *Comparative analysis of developmental outcomes in very preterm infants:*
9 *BSID-II versus Bayley-III German norms*. *PLOS ONE*, 20(1), E0318263.
10 <https://doi.org/10.1371/journal.pone.0318263>
- 11 Heilig, L. & Pauen, S. (2013). Wie wirkt sich die Beobachterrolle auf die Beurteilungen früh-
12 kindlicher Entwicklung aus? Ein Vergleich der MONDEY-Dokumentation von Eltern
13 und pädagogischen Fachkräften. *Frühe Bildung*, 2(3), 144-151.
14 <https://doi.org/10.1026/2191-9186/a000101>
- 15 JMK & KMK, Jugendminister- und Kultusministerkonferenz. (2004). *Gemeinsamer Rahmen*
16 *der Länder für die frühe Bildung in Kindertageseinrichtungen*. Fassung vom
17 06.05.2021 (JKM) und 24.03.2022 (KMK). Verfügbar unter
18 [https://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2004/2004_06_03-](https://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2004/2004_06_03-Fruehe-Bildung-Kindertageseinrichtungen.pdf)
19 [Fruehe-Bildung-Kindertageseinrichtungen.pdf](https://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2004/2004_06_03-Fruehe-Bildung-Kindertageseinrichtungen.pdf)
- 20 Kipping, S. M., Kiess, W., Ludwig, J., Meigen, C. & Poulai, T. (2024). *Are the results of the*
21 *Bayley Scales of Infant and Toddler Development (Third Edition) predictive for later*
22 *motor skills and school performance?* *Children*, 11(12), 1486.
23 <https://doi.org/10.3390/children11121486>
- 24 Linver, M. R., Brooks-Gunn, J. & Kohen, D. E. (2002). *Family processes as pathways from*
25 *in-come to young children's development*. *Developmental Psychology*, 38(5), 719–734.
26 <https://doi.org/10.1037/0012-1649.38.5.719>
- 27 Lohaus, A., Vierhaus, M. & Lemola, S. (2024). *Entwicklungspsychologie des Kindes- und Ju-*
28 *gendalters für Bachelor (5. Auflage)*. Berlin, Heidelberg: Springer Verlag.
- 29 Lopes da Silva, T. C., dos Santos, G. A. A., dos Santos, L. M., Madaschi, V., Saigh Jurdi, A. P,
30 & dos Santos, D. N. (2024). *Reliability of the AYLEY Scale: Assessing Children Af-*
31 *ected by Zika Virus*. *Psicologia: Teoria e Prática*, 26(2), ePTPPA14444.
32 <https://doi.org/10.5935/1980-6906/ePTPPA14444.en>

- Macha, T. & Petermann, F. (2015). *Bayley Scales of Infant and Toddler Development, Third Edition – Deutsche Fassung*. *Zeitschrift Für Psychiatrie, Psychologie und Psychotherapie*, 63(2), 139–143. <https://doi.org/10.1024/1661-4747/a000232>
- Macha, T., Proske, A. & Petermann, F. (2005). *Validität von Entwicklungstests*. *Kindheit und Entwicklung*, 14(3), 150–162. <https://doi.org/10.1026/0942-5403.14.3.150>
- Martin, A. J., Darlow, B. A., Salt, A., Hague, W., Sebastian, L., McNeill, N. & Tarnow-Mordi, W. (2013). Performance of the Parent Report of Children's Abilities-Revised (PARCA-R) versus the Bayley Scales of Infant Development III. *Archives of Disease in Childhood*, 98(12), 955–958. <https://doi.org/10.1136/archdischild-2012-303288>
- Morgan, C., Honan, I., Allsop, A., Novak, I. & Badawi, N. (2019). Psychometric properties of assessments of cognition in infants with cerebral palsy or motor impairment: A systematic review. *Journal of Pediatric Psychology*, 44(2), 238–252. <https://doi.org/10.1093/jpepsy/jsy068>
- Orioli, P.A., Johnston, C., Krebs, V.L.J., Franciso, R.P.V. & Carvalho, W.B. (2022). *Assessment of child development by the Bayley III Scale: A systematic review*. *Clinical Case Reports: Open Access*, 5(1), 205. DOI:10.46527/2582-5038.205
- Pauen, S. (2011). *Vom Baby zum Kleinkind. Entwicklungstagebuch zur Beobachtung und Begleitung in den ersten Lebensjahren. MONDEY – Milestones of normal development in early years*. Heidelberg: Springer. <https://doi.org/10.1007/978-3-8274-2780-9>
- Pauen, S., Frey, B. & Ganser, L. (2012a). Entwicklungspsychologie in den ersten drei Lebensjahren. In M. Cierpka (Hrsg.), *Frühe Kindheit 0-3. Beratung und Psychotherapie für Eltern mit Säuglingen und Kleinkindern* (S. 22–38). Berlin, Heidelberg: Springer.
- Pauen, S. & Heilig, L. (2012, 07–09 Juni). *Milestones of Normal Development in Early Years. Testing objectivity, reliability, and construct validity of a new standardized inventory to monitor development from 0–3* [Poster-Präsentation], XVIII Biennial International Conference of Infant Studies, Minneapolis, Minnesota.
- Pauen, S., Heilig, L., Danner, D., Haffner, J., Tettenborn, A. & Ross, J. (2012b). Milestones of Normal Development in Early Years (MONDEY). Konzeption und Überprüfung eines Programms zur Beobachtung und Dokumentation der frühkindlichen Entwicklung von 0-3 Jahren. *Frühe Bildung*, 1(2), 64-70. <https://doi.org/10.1026/2191-9186/a000032>

- Perra, O., McGowan, J. E., Grunau, R. E., Doran, J. B., Craig, S., Johnston, L., Jenkins, J., Holmes, V. A. & Alderdice, F. A. (2015). Parent ratings of child cognition and language compared with Bayley-III in preterm 3-year-olds. *Early Human Development*, 91(3), 211–216. <https://doi.org/10.1016/j.earlhumdev.2015.01.009>
- Peyton, C., Msall, M., Wroblewski, K., Rogers, E. E., Kohn, M. & Glass, H. C. (2020). Concurrent validity of the Warner Initial Developmental Evaluation of Adaptive and Functional Skills and the Bayley Scales of Infant and Toddler Development (Third Edition). *Developmental Medicine & Child Neurology*, 63, 349-354. DOI: 10.1111/dmcn.14737
- Pitchik, H. O., Tofail, F., Akter, F. Shoab, A. K. M. Sultana, J., Huda, T. M. N. et al. (2023). Concurrent validity of the ages and stages questionnaire inventory and the Bayley Scales of Infant and Toddler Development in rural Bangladesh. *BMC Pediatr*, 23(93). doi: 10.1186/s12887-022-03800-6
- Rakesh, D., Lee, P. A., Gaikwad, A. & McLaughlin, K. A. (2024). Associations of socioeconomic status with cognitive function, language ability, and academic achievement in youth: a systematic review of mechanisms and protective factors. *Journal of Child Psychology and Psychiatry*, 0(0), 1-23. doi:10.1111/jcpp.14082
- Reuner, G., Fields, A. C., Wittke, A. Löprrich, M. & Pietz, J. (2013). Comparison of the developmental tests Bayley-III and Bayley-II in 7-month-old infants born preterm. *European Journal of Pediatrics*, 172, 393-400. DOI: 10.1007/s00431-012-1902-6
- Reuner, G. & Pietz, J. (2006). Entwicklungsdiagnostik im Säuglings- und Kleinkindalter. *Monatsschrift Kinderheilkunde*, 154(4), 305–313. <https://doi.org/10.1007/s00112-006-1315-6>
- Reuner, G. & Rosenkranz, J. (2014). *Bayley-Scales of Infant and Toddler Development, Third Edition – Deutsche Version: Technisches Manual zu den Bayley III Skalen und dem Bayley III Screening Test*. Frankfurt am Main: Pearson Assessment.
- Sachse, S. & Von Suchodoletz, W. (2008). Early identification of language delay by direct language assessment or parent report? *Journal of Developmental and Behavioral Pediatrics*, 29(1), 34–41. <https://doi.org/10.1097/DBP.0b013e318146902a>
- Sansavani, A., Favilla, M. E., Guasti, M. T., Marini, A., Millepiedi, S., Di Martino, M. P et al. (2021). Developmental language disorder: Early predictors, age for the diagnosis, and diagnostic tools. A scoping review. *Brain sciences*, 11(5), 654. DOI: 10.3390/brainsci11050654

- Schonhaut, L., Armijo, I., Schönstedt, M., Alvarez, J. & Cordero, M. (2013). Validity of the ages and stages questionnaires in term and preterm infants. *Pediatrics*, 131(5), 1468-1474. <https://doi.org/10.1542/peds.2012-3313>
- Sherwell, S., Reid, S. M., Reddihough, D. S., Wrennall, J., Ong, B. & Stargatt, R. (2014). Measuring intellectual ability in children with cerebral palsy: Can we do better? *Research in Developmental Disabilities*, 35(10), 2558–2567. <https://doi.org/10.1016/j.ridd.2014.06.019>
- Simpson, S., Eadie, T., Khoo, S. T., Titmuss, A., Maple-Brown, L. J. Thompson, R. et al. (2021). The ASQ-TRAK: Validating a culturally adapted developmental screening tool for Australian Aboriginal children. *Early Human Development*, 163, 105481. <https://doi.org/10.1016/j.earlhumdev.2021.105481>
- Thomas, M. S. C. & Coecke, S. (2023). Associations between Socioeconomic Status, Cognition, and Brain Structure: Evaluating Potential Causal Pathways Through Mechanistic Models of Development. *Cognitive Science*, 47, e13217. DOI: 10.1111/cogs.13217
- Viernickel, S. (2014). Beobachtung und Dokumentation. In R. Braches-Chyrek, C. Röhner, H. Sünger & M. Hopf (Hrsg.), *Handbuch frühe Kindheit* (S. 491–502). Opladen, Berlin. Toronto: Babara Budrich.
- Weiers, H., Ghattaura, S., Scerif, G., Sella, F., Simms, V., Xenidou-Dervou, I. et al. (2025). Methods Used to Assess Early Mathematical Skills in 2-Year-Olds: A Review of the Literature. *Infant and Child Development*, 34, :e70001. <https://doi.org/10.1002/icd.70001>

Tabelle 1.

Spearman-Rho-Korrelationen der Outcomevariablen für die Gesamtstichprobe

Stichprobe	Variablen	Korrelation (<i>r</i>) <i>P</i> -Wert(<i>p</i>) Stichprobengröße (<i>N</i>)				
		KogB	SprB	KogM	SprM	AllgM
Gesamt	KogB	1.00				
	SprB	.75** (<.001) <i>N</i> =55	1.00			
	KogM	.50** (<.001) <i>N</i> =63	.52** (<.001) <i>N</i> =55	1.00		
	SprM	.55** (<.001) <i>N</i> =63	.75** (<.001) <i>N</i> =55	.63** (<.001) <i>N</i> =63	1.00	
	AllgM	.63** (<.001) <i>N</i> =63	.67** (<.001) <i>N</i> =55	.77** (<.001) <i>N</i> =63	.88** (<.001) <i>N</i> =63	1.00

Anmerkungen. Kognition Bayley (KogB), Sprache Bayley (SprB), Kognition MONDEY (KogM), Sprache MONDEY (SprM) und MONDEY Allgemeiner Entwicklungsstand (AllgM). ***p* < .01.