

Methodenbericht

Längsschnitterhebung im Projekt Sprachbildung und -entwicklung im Kita-Alltag in Nordrhein-Westfalen (SEIKA-NRW)

Das übergeordnete Ziel der Längsschnitterhebung im Verbundprojekt SEIKA-NRW war die Untersuchung der Sprachentwicklung von Kindern im Elementarbereich in Nordrhein-Westfalen. Von besonderem Interesse war die Identifikation von Faktoren, die sich positiv auf die Sprachentwicklung auswirken. Insbesondere wurden Daten erhoben, die Rückschlüsse darüber erlauben, ob sich die Bereitstellung von zwei zusätzlichen finanziellen Förderlinien durch das Familienministerium positiv auf die pädagogische Qualität der Betreuung in den Kitas und so auch auf die Entwicklung der Sprachkompetenzen der Kinder auswirkt. Die Verbundpartner im Projekt waren das Deutsche Jugendinstitut in München, die Technische Universität (TU) Dortmund und die Ruhr-Universität Bochum. Im Rahmen der Längsschnitterhebung hat die Projektgruppe an der Ruhr-Universität Bochum in drei Erhebungszeiträumen (jeweils im Jahr 2016, 2017 und 2018) Daten erhoben, um die Sprachentwicklung von Kindern, die eine Kita besucht haben, zu erfassen. Die Eltern von 626 Kindern im Alter von 2;6 bis 3;11 Jahren stimmten im Jahr 2016 der Teilnahme ihres Kindes an der Längsschnitterhebung zu. Die Kinder waren auf 175 Gruppen in 92 Kitas aufgeteilt. Zusätzlich wurden im zweiten und dritten Erhebungszeitraum weitere Kinder in die Stichprobe aufgenommen, die dem Alter nach in die Stichprobe gehörten, jedoch erst später in die Kita gekommen waren. Im zweiten Erhebungszeitraum wurden 347 Kinder und im dritten Erhebungszeitraum 74 Kinder neu aufgenommen. Für alle Kinder liegen informierte Einwilligungen der Eltern vor. Neben den Daten zu den Sprachkompetenzen der Kinder wurden Daten von den Müttern sowie von den pädagogischen Fachkräften erhoben. Die hier beschriebenen Methoden beziehen sich ausschließlich auf die Erhebungen im SEIKA-NRW Projekt, die verantwortlich an der Ruhr-Universität durchgeführt wurden.

Datenerhebungsinstitut:	Interdisziplinäres Zentrum für Familienforschung an der Ruhr-Universität Bochum
Projektleitung:	Prof. Birgit Leyendecker
ProjektmitarbeiterInnen:	Dr. Alexandru Agache, Dr. Lilly-Marlen Bihler, Dr. Jessica Willard, Dr. Katharina Kohl, Katharina Heuser, M.Sc., Stefanie Madsen, M.Sc., Angelika Seubert, M.Sc.
Administration des Projekts:	Yasemin Akar, B.A., Helga Pistorius
Fördernde Institution:	Ministerium für Kinder, Familie, Flüchtlinge und Integration des Landes NRW (ehemals: Ministerium für Familie, Kinder, Jugend, Kultur und Sport)

1. Hintergrund des Projekts SEIKA-NRW

Gute Deutschkenntnisse sind unbestritten eine zentrale Voraussetzung für den Schulerfolg von Kindern in unserem Land. Dies bedeutet, dass Kinder bis zum Eintritt in die Schule gute semantisch-lexikalische sowie morphologisch-syntaktische Basisqualifikationen erworben haben sollten, die es ihnen erlauben, sich sprachlich gut auszudrücken und Sprache entschlüsseln zu können. Weiterhin gilt die phonologische Bewusstheit, also die Unterscheidung und Produktion von Sprachlauten, Silben und Wörtern als eine wichtige Voraussetzung für den späteren Schriftspracherwerb in der Schule (Weinert & Grimm, 2018). Hinzu kommt, dass der Erwerb von Sprache eine Basis für die Entwicklung von höheren kognitiven Funktionen bildet und deswegen einen nicht zu unterschätzenden Einfluss auf die kognitiven Fähigkeiten von Kindern hat. Viele – oder gar die meisten – Lernprozesse sind sprachlich vermittelt. Für eine aktive Teilhabe an unserer Gesellschaft ist das Beherrschen der (Bildungs-) Sprache eine entscheidende Schlüsselkompetenz – nur wer die Sprache versteht und sich selber sprachlich und schriftlich ausdrücken kann, hat eine Chance in der Schule zurechtzukommen und eine berufliche Ausbildung zu erwerben, die seinen oder ihren Fähigkeiten und Interessen entspricht. Wenn sprachliche Basiskompetenzen nicht in der frühen Kindheit erworben werden, sind lebenslange Benachteiligungen eine wahrscheinliche Konsequenz.

Der sich hieraus ergebende Bildungsauftrag an die Elementarpädagogik hat in den vergangenen Jahren noch einmal besondere Aufmerksamkeit bekommen und die Feststellung und Förderung kindlicher Sprachkompetenzen ist in den Mittelpunkt des pädagogischen Alltags gerückt. Sprachbildung soll, wie das Ministerium für Familie, Kinder, Jugend, Kultur und Sport des Landes Nordrhein-Westfalen (NRW; heute: Ministerium für Kinder, Familie, Flüchtlinge und Integration, MKFFI) 2014 in der Präambel zu seiner Broschüre „Alltagsintegrierte Sprachbildung und Beobachtung im Elementarbereich – Grundlagen für NRW“ schrieb, „sowohl an der Lebenserfahrung als auch an den individuellen Lebenslagen der Kinder orientiert und integriert im pädagogischen Alltag“ stattfinden. Während Sprachförderung früher als additiv verstanden wurde und in kleinen, zeitlich begrenzten und isolierten Einheiten stattfand, wurde 2014 explizit das Ziel formuliert, dass Sprachbildung den gesamten pädagogischen Alltag durchdringen soll. Um dies zu erreichen, hat damals das Land NRW, neben Informationsveranstaltungen für die pädagogischen Fachkräfte zu den Konzepten der alltagsintegrierten Sprachförderung, mehrere eng miteinander verzahnte und auf Nachhaltigkeit angelegte Maßnahmen eingeleitet. So wurden u.a. zwei Förderlinien etabliert, bei denen Kitas zusätzliche Finanzmittel zur Verfügung gestellt wurden. Ein Ziel des Projektes SEIKA-NRW war es die Wirksamkeit der beiden Förderlinien zu evaluieren.

Eine Förderlinie war vorgesehen für Kitas mit einem hohen Anteil an Kindern für die ein hoher zusätzlicher Sprachförderbedarf postuliert wurde (Sprachförder-Kitas). Die andere Förderlinie war vorgesehen für Kitas mit einem hohen Anteil an Kindern für die ein besonderer Unterstützungsbedarf in Bildungsprozessen postuliert wurde (plusKITAs). Sprachförder-Kitas wurden jährlich mit mindestens 5000 Euro gefördert, plusKITAs jährlich mit mindestens 25.000 Euro. Die finanziellen Fördermittel standen in der Regel ab dem

01.08.2014 für die Dauer von 5 Kita-Jahren zur Verfügung. Insgesamt betrachtet erhielt die Mehrheit der Kitas keine zusätzliche Förderung (57%, *Kontrollgruppe*). Die zweitgrößte Gruppe erhielt ausschließlich Zuschüsse aus der Sprachförderlinie (25%, *Treat1*). Die drittgrößte Gruppe erhielt zum Zuschuss aus der Sprachförderlinie zusätzlich noch den Zuschuss für plusKITAs (14%, *Treat2*). Zuletzt gab es noch eine geringe Anzahl an Kitas, die nur den Zuschuss für plusKITAs erhalten haben, die aber nicht Gegenstand der Evaluation sind.

Das Land NRW hat die beiden Förderlinien basierend auf Kriterien für die Vergabe der Zuschüsse laut dem Kinderbildungsgesetz (KiBiz) an die Jugendämter in NRW verteilt. Die Kriterien umfassten (1) den Anteil der Kinder unter sieben Jahren in Familien mit Leistungsbezug nach dem SGB II an der landesweiten Gesamtzahl der Kinder unter sieben Jahren in Familien mit SGB-II-Leistungsbezug (Kriterium für beide Förderlinien) und (2) die Anzahl der Kinder im Jugendamtsbezirk in Kindertageseinrichtungen, in deren Familien vorrangig nicht Deutsch gesprochen wird (zusätzliches Kriterium für die Auswahl von Sprachförder-Kitas; §§ 21a - 21b KiBiz). Die Jugendämter wiederum verteilten die Fördermittel basierend auf sozialräumlichen Kriterien, die nicht näher im Gesetz spezifiziert wurden.

Die übergeordneten Ziele des Projektes SEIKA-NRW waren die Evaluation der in NRW unternommenen Anstrengungen zur Implementierung der alltagsintegrierten Sprachbildung im Elementarbereich und die Erforschung der Wirksamkeit früher Sprachbildung. Hierzu wurden drei Module für das Verbundprojekt SEIKA-NRW konzipiert. Die Studienpartner sind das Deutsche Jugendinstitut in München gemeinsam mit der TU Dortmund sowie der Ruhr-Universität Bochum.

Die Projektgruppe am Deutschen Jugendinstitut in München übernahm verantwortlich das Modul 1 „Der Weg des Geldes“, in dem primär die Verteilung der zusätzlichen Ressourcen durch die Jugendämter, Träger und die Kita-Leitungen untersucht wurde.

Die Projektgruppe an der Ruhr-Universität Bochum übernahm die Verantwortung für die Module 2 und 3. Ziel von Modul 2 war die Feststellung des Sprachförderbedarfs im Elementarbereich mit einer Querschnittsstichprobe von Kindern im Alter von 2 ½ bis 6 Jahren. Ziel von Modul 3 war die Untersuchung der Entwicklung der sprachlichen Kompetenzen der Kinder im Längsschnitt sowie die Identifizierung der Faktoren, die sich positiv auf die Sprachbildung auswirken. Von besonderer Relevanz war hier die Frage, ob die beiden Förderlinien, die durch das Land NRW zur Verfügung gestellt wurden, positive Auswirkungen auf die pädagogische Qualität der Betreuung in den Kitas und auf die Entwicklung der Sprachkompetenzen der Kinder haben.

Mithilfe des hier zur Verfügung gestellten Längsschnittdatensatz aus Modul 3 lassen sich insbesondere außerfamiliäre und familiäre Faktoren identifizieren, die mit der Sprachentwicklung der Kinder zusammenhängen.

2. Vorbereitung der Module 2 und 3 des Projektes

2.1 Ethikkommission

Zu Beginn des Projekts wurde für die Erhebung der Daten in den Modulen 2 und 3 ein Antrag bei der Ethikkommission der Fakultät für Psychologie der Ruhr-Universität Bochum gestellt, der als ethisch unbedenklich votiert wurde.

2.2 Anwendung eines Propensity Score Matching Verfahrens für die Stichprobenauswahl

Eine übergeordnete Fragestellung des Projektes SEIKA-NRW fokussierte u.a. die Frage nach der Wirksamkeit von zwei Arten der finanziellen Förderung von Kitas (*Treat1* und *Treat2*; vgl. Abschnitt 1). Die Herausforderung der Evaluation der Wirksamkeit lag darin, dass die finanziellen Mittel aufgrund von sozialpolitischen und ethischen Gründen nicht randomisiert zugewiesen werden konnten. Es gab eine Vielzahl von Variablen, die sowohl mit der Zuweisung der beiden Förderlinien als auch mit den Merkmalen der Kitas und den interessierenden Sprachoutcomes der Kinder konfundiert sein konnten. Zum Beispiel war der mittlere Anteil an 3-6-Jährigen mit nicht deutscher Familiensprache an allen 3-6-Jährigen in der Kita je nach Förderlinie unterschiedlich hoch (*Kontrollgruppe*: $M = 12\%$, *Treat1*: $M = 28\%$, *Treat2*: $M = 50\%$). Es war anzunehmen, dass dieser und weitere Unterschiede die Entwicklung der Kinder in den Kitas der jeweiligen Förderlinien verschieden beeinflussen.

Durch die Anwendung eines Propensity Score Matching Verfahrens zur Stichprobenauswahl sollte schon vor der Datenerhebung sichergestellt werden, dass eine Stichprobe von Kita-Drillingen (*Kontrollgruppe*, *Treat1*, *Treat2*) gewonnen wurde, in der sich die drei Gruppen von Kitas auf einer Vielzahl von relevanten sozialräumlichen und strukturellen Merkmalen nicht bedeutsam voneinander unterscheiden haben. Für die Anwendung des Verfahrens wurden Variablen aus vier Datenquellen genutzt:

- 1) Administrative KiBiz.web-Daten von 8906 Einrichtungen aus dem Kita-Jahr 2014/2015. Der Rohdatensatz wurde von unseren Projektpartnern an der TU Dortmund aufbereitet (Alexander Sommer).
- 2) Sozialräumliche microm-Daten (Datenaufbereitung: TU Dortmund, Alexander Sommer).
- 3) Dokumentenanalyse der TU Dortmund (Nora Jehles): Kodierungen der angewandten Kriterien der Jugendämter zur Vergabe der beiden Förderlinien.
- 4) Informationen darüber, ob eine Kita mit hoher Wahrscheinlichkeit ab Januar 2016 im Bundesprogramm *Sprach-Kitas* gefördert wird. Informationen zum Bundesprogramm siehe <https://sprach-kitas.fruehe-chancen.de/>.

Aus diesen vier Datenquellen wurden, basierend auf theoretischen und empirischen Überlegungen, 45 Variablen für die Anwendung des Propensity Score Matching Verfahrens ausgewählt. Diese Variablen bildeten ein breites Spektrum an sozialräumlichen und strukturellen Kita-Merkmalen ab, z.B. die Arbeitslosenquote und den Anteil zugewanderter Familien im nahen Umkreis der Kita sowie den Personal-Kind-Schlüssel und den Anteil von Kindern mit einer 45-Stunden-Betreuung gegenüber einer 25- und 35-Stunden-Betreuung pro Woche in der Kita.

Insgesamt bestanden im Daten-Gesamtpool für etwa 5% der Kitas fehlende Werte. Die fehlenden Werte wurden mit einem Expectation-Maximum-Likelihood (EM) Verfahren in zwei Schritten imputiert. Zunächst wurden die KiBiz.web-Daten auf der Ebene der Jugendämter aggregiert und die fehlenden Kriterien auf Jugendamtsebene geschätzt. Im zweiten Schritt wurden die vollständigen Kriterien auf Jugendamtsebene den KiBiz.web-Daten hinzugespielt und so die restlichen fehlenden Daten auf Kita-Ebene geschätzt.

Das Propensity Score Matching Verfahren wurde mithilfe des Pakets *TriMatch* in R durchgeführt (Bryer, 2013). Das Verfahren basiert auf drei logistischen Modellen mit je 45 Kovariaten. Anhand dieser Modelle wurden für jede Kita die Wahrscheinlichkeiten berechnet, dass sie nicht gefördert, einfach oder doppelt gefördert wird. Einrichtungen, die ähnliche Wahrscheinlichkeiten aufwiesen, aber aus unterschiedlichen Gruppen stammten (je 1x *Kontrollgruppe*, 1x *Treat1* und 1x *Treat2*), wurden zu einem Drilling zusammengefasst. Es konnten 292 Drillings identifiziert werden, in denen sich die drei Gruppen von Kitas auf den 45 Kriterien stark ähneln. In keinem Drilling stammten mehrere Kitas aus demselben Jugendamt. Durch die Anwendung des Propensity Score Matching Verfahrens zur Stichprobenauswahl konnte sichergestellt werden, dass sich die drei Gruppen auf einer Vielzahl von Variablen nicht bedeutsam voneinander unterscheiden. Hierdurch verfügt das Projekt über ein quasi-experimentelles Design. Eine detaillierte Beschreibung des Propensity Score Matching Verfahrens, das hier eingesetzt wurde, findet sich in Bihler, Agache, Willard, Kohl und Leyendecker (2020/ im Druck).

Im Folgenden wird dargestellt wie die Stichprobe für das Modul 2 des Projektes SEIKA-NRW ausgewählt wurde. Es ist wichtig diesen Auswahlprozess zu verstehen, um anschließend nachvollziehen zu können wie die Stichprobe für das Modul 3 (Längsschnitterhebung, Thema dieses Methodenberichts) ausgewählt wurde.

2.3 Gewinnung einer Stichprobe für das Projekt SEIKA-NRW

Ziel war es etwa 33 Drillings (99 Kitas) für die Teilnahme am Projekt SEIKA-NRW zu gewinnen. Anfang Dezember 2015 wurden über das Familienministerium in NRW zunächst Anschreiben mit Informationen zum Projekt an die Spitzenverbände von 150 Kindertageseinrichtungen verschickt. Darin wurden die Spitzenverbände gebeten, die in ihrem Verband befindlichen Träger der 150 ausgewählten Einrichtungen anhand eines beigelegten Briefes und Flyers über das Projekt zu informieren. Nach einem Zeitraum von ca. 5 Wochen wurden die Träger der entsprechenden Kitas ebenfalls angeschrieben und gebeten, die Einrichtungen darüber zu informieren, dass diese für die Studie ausgewählt wurden. Wiederum ca. 2 Wochen später haben wir die einzelnen Kitas zunächst schriftlich und ca. 1 Woche nach Erhalt des Briefes auch telefonisch kontaktiert und erfragt, ob sie an der Studie teilnehmen. Wir konnten 27 der 50 Drillings für die Teilnahme gewinnen. In den restlichen 23 Drillings stimmten nicht alle Einrichtungsleitungen der Teilnahme zu. Zu Beginn wurden Drillings ausgeschlossen, sobald nur eine Einrichtung aus dem Drilling keine Bereitschaft zur Teilnahme zeigte. Im Laufe der Stichprobengewinnung wurde dazu übergegangen, zu untersuchen, ob sich auf Basis der geschätzten Propensity Scores der Einrichtungen, die gerne an der Studie teilnehmen wollten, neue Drillings bilden lassen. Das Matching per Hand gelang uns für fünf Drillings (15 Einrichtungen), wodurch wir eine

Stichprobe von insgesamt 32 Drillingen (96 Einrichtungen) realisieren konnten. Eine Einrichtung zog ihre Teilnahmebereitschaft kurz vor der Datenerhebung aus organisatorischen Gründen zurück, wodurch sich die effektive Stichprobe auf 95 Einrichtungen reduzierte. Die Stichprobe setzte sich wie folgt zusammen:

- 32 Einrichtungen der *Kontrollgruppe*
- 31 Einrichtungen, die *Treat1* erhalten haben
- 32 Einrichtungen, die *Treat2* erhalten haben

Auf 44 der 45 Variablen, die ins Propensity Score Matching Verfahren einbezogen wurden, unterschieden sich die drei Gruppen von Kitas nicht statistisch bedeutsam voneinander.

Zum Zeitpunkt der Datenerhebung besuchten im Mittel 67 Kinder die Einrichtungen mit einem Minimum von 36 und einem Maximum von 122 Kindern. Die Anzahl der Gruppen pro Einrichtung lag zwischen 2 und 6.

Die Datenerhebung startete etwa 18 Monate nach Beginn der Förderung und ist daher methodisch als ex-post-Evaluation anzusehen.

2.4 Vergleichbarkeit der teilnehmenden Kitas mit allen Kitas in NRW

Die teilnehmenden Kitas waren über ganz NRW verteilt (siehe Abbildung 1 rechts). Ein Abgleich der teilnehmenden Kitas mit den restlichen Kitas in NRW ergab, dass sich die teilnehmenden Kitas auf vielen wichtigen sozialräumlichen und strukturellen Merkmalen im Mittel nicht von den restlichen Kitas in NRW unterschieden haben.

Während die Gesamtheit der an der Studie SEIKA-NRW teilnehmenden Kitas weitestgehend die Gesamtheit der Kitas in NRW abgebildet hat, unterschieden sich die ausgewählten Kitas innerhalb einer Förderlinie von der Gesamtheit der Kitas in NRW innerhalb der jeweiligen Förderlinie. Das war zwangsläufig der Fall, da es vor der Stichprobenauswahl große

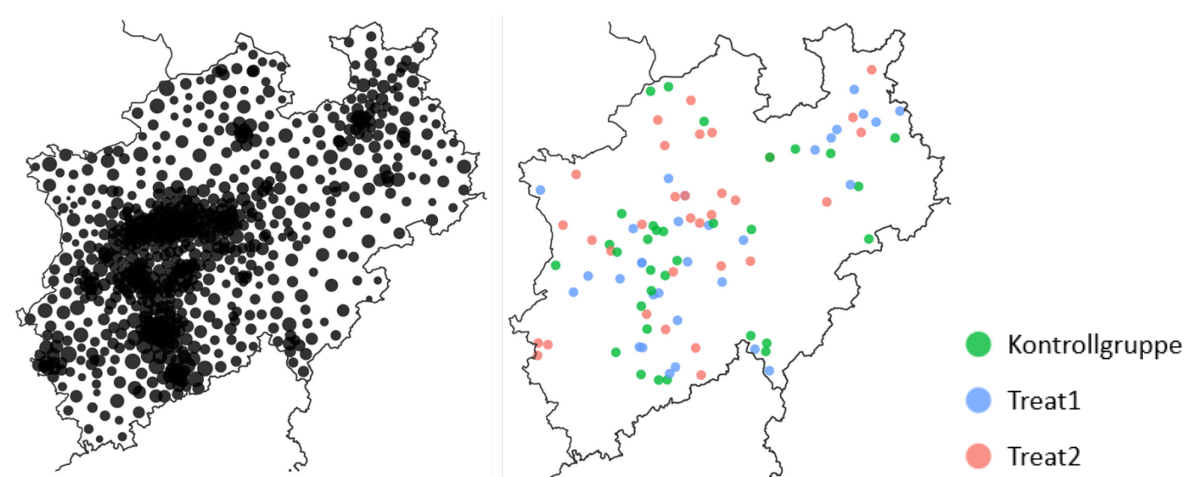


Abbildung 1: Links: Verteilung der Grundgesamtheit der 8906 Kitas in NRW, deren Daten im Rahmen des Propensity Score Matching Verfahrens genutzt wurden. Je größer die Punkte sind, desto höher ist die Dichte an Kindertageseinrichtungen in dieser Region. **Rechts:** Verteilung der 95 Einrichtungen aus der SEIKA-Stichprobe nach der jeweiligen Förderlinie.

Unterschiede in den Mittelwerten zwischen den Förderlinien gab. Ein Beispiel soll dies verdeutlichen. Die mittlere Ausprägung des Merkmals „mittlerer Anteil an 3-6-Jährigen mit nicht deutscher Familiensprache an allen 3-6-Jährigen in der Kita“ unterschied sich vor dem Matching stark zwischen den Förderlinien (Kontrollgruppe: M = 12%, *Treat1*: M = 28%, *Treat2*: M = 50%). Beiläufig erwähnt, spricht dies für die Sinnhaftigkeit der Verteilung der Mittel aus beiden Förderlinien. Innerhalb der gesamten Stichprobe von Kitas im Projekt SEIKA-NRW hat sich die Ausprägung des Merkmals aufgrund des Matchings nicht mehr signifikant unterschieden (Kontrollgruppe: M = 22%, *Treat1*: M = 25%, *Treat2*: M = 24%). Die Kitas innerhalb der Förderlinien weisen also einen ähnlichen mittleren Anteil an 3-6-Jährigen mit nicht deutscher Familiensprache an allen 3-6-Jährigen in der Kita auf. Auf Grundlage des vorangegangenen Matching Verfahrens können Aussagen darüber getätigt werden, inwieweit die Förderlinien bei ansonsten gleichen Rahmenbedingungen zu einer Veränderung der Sprachkompetenzen der Kinder geführt haben.

2.5 Auswahl der teilnehmenden Gruppen

Ziel war es pro Einrichtung zwei Gruppen für die Teilnahme an der Studie zu gewinnen. Die teilnehmenden Kitas gaben Auskunft über die Anzahl und die konzeptionelle Ausrichtung der Gruppen. Falls vorhanden, wurden zwei altersgemischte Gruppen zur Teilnahme an der Studie ausgewählt. Altersgemischte Gruppen wurden ausgewählt um die Wahrscheinlichkeit eines Gruppenwechsels der teilnehmenden Kinder im Rahmen der Längsschnitterhebung zu minimieren. Verfügte die Einrichtung nur über eine altersgemischte Gruppe, so wurde als zweite Gruppe, die Gruppe ausgewählt, die die höchste Zahl an 2 ½ - 4-Jährigen aufwies. Verfügte die Einrichtung nur über altersspezifische Gruppen, so wurde individuell entschieden.

Insgesamt haben 192 Gruppen (je 2 Gruppen in 93 Kita, je 3 Gruppen in 2 Kitas) am Projekt SEIKA-NRW teilgenommen. Die durchschnittliche Gruppengröße lag bei 21 Kindern, wobei die Größe zwischen 10 und 35 variierte. Die Unterschiede in den Gruppengrößen sind darauf zurückzuführen, dass sowohl kleinere Krippen-Gruppen als auch größere altersgemischte und altersspezifische Gruppen in der Stichprobe vertreten waren. Die 192 Gruppen wurden von insgesamt 4010 Kindern besucht. Zu diesen Kindern wurden vor der eigentlichen Datenerhebung basale Informationen in einem Gruppenfragebogen eingeholt, u.a. Geschlecht, Geburtsdatum und vorrangige Familiensprache(n). Diese Daten dienen u.a. der Auswahl der Stichprobe der teilnehmenden Kinder. Die Gruppenfragebögen wurden passwortgeschützt als Excel-Dateien verschickt. Die Angaben aus den Gruppenfragebögen sind im SEIKA-Querschnittsdatensatz, der ebenfalls beim IQB veröffentlicht ist, enthalten:

Agache, A., Bihler, L.-M., Willard, J. & Leyendecker, B. (2020). *Sprachbildung und -entwicklung im Kita-Alltag in Nordrhein-Westfalen (SEIKA-NRW)* (Version 1) [Datensatz].

Berlin: IQB – Institut zur Qualitätsentwicklung im Bildungswesen.

http://doi.org/10.5159/IQB_SEIKA-NRW_v1

Diese Angaben können mit dem Längsschnittdatensatz über die Einrichtungs-ID verknüpft werden, z.B. um Kompositionsvariablen für die Gruppen zu bilden.

2.6 Auswahl der teilnehmenden Kinder für Modul 2 (Querschnittserhebung)

Die teilnehmenden Einrichtungen erhielten vor der Querschnitterhebung (Modul 2) im Jahr 2016 ein Anschreiben an alle Eltern, das die Eltern über das Projekt informierte. Dieses Anschreiben wurde für zugewanderte Familien in sechs Sprachen übersetzt (arabisch, bulgarisch, englisch, polnisch, russisch und türkisch). Die Eltern wurden gebeten, der Teilnahme ihres Kindes an der Querschnitterhebung zu widersprechen, wenn sie nicht mit der Teilnahme ihres Kindes einverstanden waren. Zusätzlich wurden von den meisten Einrichtungen Aushänge angebracht, die die Eltern kurz vor der Datenerhebung erneut an die Studie erinnerten. In vier Einrichtungen wurde auf Wunsch der Einrichtungsleitung aktiv die Zustimmung der Eltern zur Teilnahme eingeholt.

Zur Zielstichprobe der Querschnittserhebung zählten Kinder im Alter von 2;6 bis 6;11 Jahren. In diesem Alter waren 3868 der 4010 Kinder in den Gruppen. Von diesen Kindern wurden 113 Kinder, die nachweislich einen erhöhten Förderbedarf aufwiesen sowie 85 Kinder, deren Eltern der Teilnahme widersprochen haben, von der Studie ausgeschlossen. Die verbleibenden 3670 Kinder wurden innerhalb ihrer Gruppen nach Altersgruppen aufgeteilt und dann innerhalb ihrer Altersgruppen in eine randomisierte Reihenfolge gebracht. Anschließend wurden die Kinder unter Berücksichtigung eines Oversamplings von Kindern im Alter von 2;6-3;11 Jahren in diesen Reihenfolgen getestet. Das setzte voraus, dass die Kinder kein Ausschlusskriterium erfüllt haben (z.B. an allen Testtagen abwesend/krank waren, zwischenzeitlich von der Kita abgemeldet wurden oder schon ein Geschwisterkind unter 4 Jahre getestet wurde; Details siehe Tabelle 1).

Tabelle 1. Teilnahmestatus der Kinder in der Querschnitterhebung ($N = 4010$).

Wurde das Kind getestet? Wenn nicht, warum nicht?	Nennungen
ja, getestet	2319
ja, aber Daten entfernt (aus diversen Gründen)	24
nein, Kind wurde im Rahmen der Zufallsauswahl nicht ausgewählt	883
nein, Kind war an allen Testtagen abwesend oder krank	477
nein, Kind hat Deutschland verlassen	2
nein, Eltern haben der Teilnahme ihres Kindes widersprochen	85
nein, Kind hat nachweislich einen erhöhten Förderbedarf	117
nein, es wurde schon ein Geschwisterkind unter 4 Jahre getestet (Vermeidung einer zusätzlichen Ebene der Nestung)	28
nein, Kind weist unzureichende Deutschkenntnisse auf um Testinstruktionen zu verstehen	4
nein, Kind möchte nicht teilnehmen	42
nein, Kind wurde zwischenzeitlich von der Kita abgemeldet	29

2.7 Auswahl der teilnehmenden Kinder für Modul 3 (Längsschnitterhebung)

Die Eltern, deren Kinder im Alter von 2;6 bis 3;11 Jahren an der Querschnitterhebung teilgenommen hatten, wurden gebeten, zusätzlich der Teilnahme ihres Kindes an der Längsschnitterhebung zuzustimmen. Von insgesamt 968 Ziel-Kindern in der genannten Altersgruppe liegt für 626 Kinder (65%)¹ mindestens ein weiterer Datenpunkt im Rahmen der Längsschnitterhebung vor. Während der folgenden Erhebungen in den Jahren 2017 und 2018 wurden zusätzlich Kinder aufgenommen, die dem Alter nach in die Stichprobe gehörten, aber erst nach dem letzten Erhebungszeitpunkt in die Kita gekommen waren. Bis zum zweiten Erhebungszeitpunkt waren 537 Zielkinder im entsprechenden Alter neu in die teilnehmenden Kitas gekommen. Die Eltern von 358 dieser Kinder stimmten der Teilnahme ihres Kindes an der Studie zu (67%)². Von diesen 358 Kindern liegt für 347 Kinder (97%) mindestens ein Datenpunkt vor. In Welle 3 waren 141 Zielkinder neu in die Kita-Gruppen gekommen. Die Eltern von 74 dieser Kinder stimmten der Teilnahme zu (52%). Für alle 74 Kinder liegt mindestens ein Datenpunkt vor.

Kinder, die eines der Ausschlusskriterien aus Tabelle 1 erfüllten, z.B. einen nachweislich erhöhten Förderbedarf, wurden auch in W2 und W3 nicht in die Stichprobe aufgenommen.

Die gesamte Stichprobe der Längsschnitterhebung umfasst 1047 Kinder. Der Anteil von Jungen in der Stichprobe liegt bei 51%. Das mittlere Alter der Kinder liegt zum ersten Erhebungszeitpunkt bei 40 Monaten ($SD = 5$), zum zweiten Erhebungszeitpunkt bei 49 Monaten ($SD = 6$) und zum dritten Erhebungszeitpunkt bei 60 Monaten ($SD = 6$). Der Anteil von Kindern mit einer anderen Familiensprache als Deutsch liegt laut Mutterangabe (Telefoninterview) je nach Erhebungszeitraum zwischen 31% und 33%. Für mehr als die Hälfte der Kinder in der Stichprobe (552 Kinder) liegt in jedem Erhebungszeitraum mindestens ein Datenpunkt vor. Es ist wichtig zu wissen, dass 139 dieser 552 Kinder die Kita-Gruppe während der Projektlaufzeit mindestens einmal gewechselt haben. Das kann anhand der Gruppen-ID für den jeweiligen Erhebungszeitraum nachvollzogen werden. Die Anzahl von Kindern mit mind. einem Datenpunkt zu den drei Erhebungszeiträumen ist in Abbildung 2 dargestellt.

¹Bei vier Kindern stellte sich heraus, dass das Geburtsdatum im Gruppenfragebogen falsch angegeben worden war und sich die Kinder außerhalb der gewünschten Altersgruppe befanden. Ein Kind in der Stichprobe war erst 29 Monate alt und drei Kinder waren 48 Monate alt. Diese Kinder wurden trotzdem in der Stichprobe behalten.

² Bei drei Kindern deutet die Angabe im Gruppenfragebogen darauf hin, dass die Kinder die Kita schon im ersten Erhebungszeitraum besucht haben. Das Kind mit der $kID_2 = 040114$ wurde versehentlich aufgenommen. Es war im ersten Erhebungszeitraum an allen Testtagen abwesend. Die anderen beiden Kinder wurden neu in den Gruppenfragebogen eingetragen (kID_2 : 066230 und 066317).

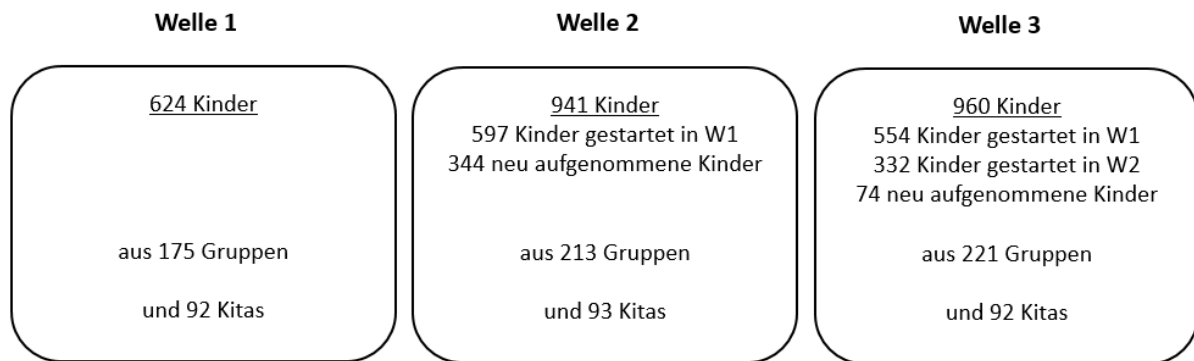


Abbildung 2. Anzahl der Kinder mit mind. einem Datenpunkt pro Erhebungszeitraum (Testabbrüche nicht mitgezählt) sowie Anzahl der Gruppen und Kitas, die von den Kindern besucht wurden.

3. Erhebungsinstrumente

Die Instrumente zur Datenerhebung für das Projekt SEIKA-NRW wurden so gewählt, dass Informationen von multiplen Datenquellen auf unterschiedlichen Ebenen (Kind, Gruppe und Einrichtung) und unter Verwendung verschiedener Methoden erfasst wurden (standardisierte Tests, Beobachtungsverfahren, Fragebögen und Interviews). Für die Auswahl geeigneter Instrumente wurden als Experte für Diagnostik Professor Dr. Wolfgang Lenhard (Universität Würzburg) und als Expertin für Sprachentwicklung und Mehrsprachigkeit Professorin Dr. Annick De Houwer (Universität Erfurt) hinzugezogen.

3.1 Standardisierte Tests

Anhand standardisierter Tests wurden mehrere Indikatoren der Sprachentwicklung im Deutschen erhoben. Zu allen drei Erhebungszeitpunkten wurde mit den Kindern zuerst der rezeptive Wortschatztest *Peabody Picture Vocabulary Test 4* durchgeführt (PPVT-4; Dunn & Dunn, 2007; Lenhard, Lenhard, Segerer, & Suggate, 2015). Der rezeptive Wortschatz korreliert moderat bis hoch mit Maßen expressiver und rezeptiver gesprochener Sprache und kann daher als Indikator für den Gesamtsprachstand betrachtet werden (Lenhard et al., 2015).

Anschließend wurde zu jedem Erhebungszeitpunkt der Subtest *Figuren legen* aus dem Hannover-Wechsler Intelligenztest für das Vorschulalter-III (HAWIVA III; Ricken, Fritz, Schuck, & Preuß, 2007) durchgeführt. Dieser Test dient als Indikator für die kognitive Entwicklung der Kinder. Der Test stellt ein nonverbales Intelligenzmaß dar und ist somit auch für Kinder mit geringen Sprachkenntnissen im Deutschen geeignet.

Darauf folgend wurden zu jedem Erhebungszeitpunkt zwei Subtests aus dem *Sprachentwicklungstest für 3–5-Jährige Kinder* (SETK 3-5; Grimm, 2001) durchgeführt. Erstens der Subtest *phonologisches Arbeitsgedächtnis für Nichtwörter*, der einen guten Prädiktor für die Wortschatzentwicklung darstellt und diagnostisch aufschlussreich ist (Armon-Lotem, De Jong, & Meir, 2015; Weinert, Ebert, Lockl, & Kuger, 2012), zweitens der Subtest *morphologische Regelbildung*, der die expressiven morphologischen Fähigkeiten der Kinder erfasst.

Ab dem zweiten Erhebungszeitpunkt wurde mit allen Kindern ab 4 Jahren zusätzlich der Subtest *Satzgedächtnis* aus dem SETK 3-5 (Grimm, 2001) durchgeführt. Dieser Subtest fragt sowohl das lexikalisch-semantische als auch das grammatikalische Wissen und die phonologischen Fähigkeiten ab. Die Ergebnisse in diesem Test sind sowohl prädiktiv für die frühe Lese- und Rechtschreibkompetenz, als auch für das frühe Leseverständnis der Kinder (Ebert & Weinert, 2013; Von Goldammer, Mähler, Bockmann, & Hasselhorn, 2010).

Der SETK 3-5 wurde bereits in früheren Studien erfolgreich mit Kindern aus zugewanderten Familien eingesetzt (Caspar & Leyendecker, 2011; Dubowy, Ebert, von Maurice, & Weinert, 2008).

Alle angewandten Tests sind für Kinder ab 3 Jahren konzipiert. Vor der Durchführung der Testungen in den Zieleinrichtungen wurden die Testverfahren in der Kindertagesstätte „Unikids“ des Akademischen Förderungswerkes der Ruhr-Universität Bochum mit 13 Kindern im Alter von 2;6 bis 6;0 Jahren pilotiert. Dabei wurde festgestellt, dass es möglich ist Kinder ab einem Alter von 2;6 Jahren zu testen, wenn das auf eine spielerische Art und Weise geschieht. Wir haben dazu u.a. für alle Testungen in der ersten Erhebungswelle und in den folgenden Erhebungswellen nach Ermessen des Testleiters eine Papageien-Handpuppe mit dem Namen *Seika* genutzt, die die Instruktionen vorgesprochen hat.

3.2 Beobachtung der Interaktionsqualität (CLASS)

Geschulte und zertifizierte Beobachter haben die Interaktionsqualität zwischen pädagogischen Fachkräften und Kindern einmal pro Erhebungszeitraum anhand des Beobachtungsverfahrens *Classroom Assessment Scoring System Pre-K* (CLASS Pre-K; Pianta, La Paro, & Hamre, 2008) erfasst. CLASS erlaubt eine objektive und reliable Einschätzung von Interaktionsmerkmalen, die mit der kindlichen Sprachentwicklung in Zusammenhang stehen (Mashburn et al., 2008).

In Gruppen mit teilnehmenden Kindern wurden Beobachtungen durchgeführt. Das waren 161 Beobachtungen im ersten Erhebungszeitraum (2016), 186 Beobachtungen im zweiten Erhebungszeitraum (2017) und 170 Beobachtungen im dritten Erhebungszeitraum (2018). Zum zweiten und dritten Erhebungszeitraum wurden i.d.R. Gruppen beobachtet, die weiterhin von mindestens einem Kind aus einem vorherigen Erhebungszeitraum besucht wurden. Gruppen wurden i.d.R. neu in die Stichprobe der CLASS-Beobachtungen aufgenommen, wenn sie von mindestens zwei neu in die Stichprobe aufgenommenen Kindern besucht wurden.

Es wurden 4 Zyklen je etwa 20 Minuten pro Gruppe beobachtet³. Die Beobachtungen fanden überwiegend vormittags statt. Dabei wurden insbesondere Bring-Situationen sowie Stuhlkreise beobachtet. Pro Zyklus wurde möglichst eine pädagogische Fachkraft beobachtet. Manchmal verließ die pädagogische Fachkraft während der Beobachtung für

³ Hiervon gibt es drei Ausnahmen. Im zweiten Erhebungszeitraum wurden in der Gruppe mit der ID 151003023 nur zwei Zyklen beobachtet (Fehler lag beim Beobachter). Zum dritten Erhebungszeitraum wurden in den Gruppen mit den IDs 090004021 und 442005123 nur je drei Zyklen beobachtet (bei der erstgenannten Gruppe lag der Fehler beim Beobachter, bei der zweitgenannten Gruppe waren alle Kinder für längere Zeit unbeaufsichtigt, weshalb kein weiterer Zyklus mehr beobachtet werden konnte).

längere Zeit den Raum und eine andere pädagogische Fachkraft kümmerte sich um die Kinder. In diesen Fällen wurde die pädagogische Fachkraft, die die Aufsicht übernommen hatte, weiter beobachtet und ein gewichteter Wert für das Rating gebildet. Im Laufe der vier Zyklen sollten alle anwesenden Fachkräfte der Gruppe beobachtet werden. Pädagogische Fachkräfte, die für eine begrenzte Zeit organisatorische Aufgaben übernommen hatten, wurden nicht beobachtet. „Aushilfserzieher“, die die Kinder nicht oder kaum kannten, wurden nicht beobachtet.

In Einrichtungen mit offenem Konzept wurden insgesamt nur 4 Zyklen beobachtet. Die Einschätzung darüber, ob ein offenes Konzept vorliegt, wurde von unseren geschulten Beobachtern vorgenommen. Die Einschätzung der Beobachter kann von der Angabe der Einrichtungsleitung im Einrichtungsfragebogen abweichen.

Spezifische Angebote für kleine Gruppen von Kindern sollten nur beobachtet werden, wenn sie gruppenintern durchgeführt wurden. Beobachtungen im Außenbereich, insbesondere von gemischten Gruppen, sollten vermieden werden. Schlafenszeiten wurden in keinem Fall beobachtet.

3.3 Fragebögen und Telefoninterviews

In allen drei Erhebungszeiträumen wurden Informationen von verschiedenen Informanten anhand von Fragebögen und Telefoninterviews eingeholt. Dies geschah im ersten Erhebungszeitraum ausschließlich nach der Datenerhebung in den Einrichtungen. Das gilt für Mutterfragebogen, Telefoninterview und Erzieher-Kind-Fragebogen größtenteils auch für den zweiten und dritten Erhebungszeitraum. Für die Angaben im Einrichtungsleitungs- und Gruppenleitungsfragebogen gilt, dass sie teilweise auch vor der Datenerhebung in den Einrichtungen eingeholt wurden. Der zeitliche Abstand zwischen der Testung und der Bearbeitung des Fragebogens bzw. der Durchführung des Telefoninterviews lässt sich den Variablen aus dem Datenhandbuch entnehmen, die mit dem Namen *dmt* beginnen.

3.3.1 Einrichtungsleitungen

Die Einrichtungsleitungen wurden gebeten, Auskünfte über strukturelle Merkmale ihrer Einrichtung (z.B. Personal, Anteil Kinder mit Migrationshintergrund) sowie über Maßnahmen zur Sprachförderung (z.B. Fortbildungen, Sprachförderprogramme) zu erteilen. Im ersten Erhebungszeitraum wurden diese Informationen in einem Online-Fragebogen erfasst. Da der Online-Fragebogen von einigen Einrichtungsleitungen nicht ausgefüllt wurde und darin wichtige strukturelle Informationen wie die Art der finanziellen Förderung abgefragt wurden, wurden die wichtigsten fehlenden Angaben in einem sehr kurzen Telefoninterview nacherhoben. Ab dem zweiten Erhebungszeitraum wurde der Online-Fragebogen komplett durch ein erheblich gekürztes Telefoninterview ersetzt (vgl. Abbildung 3). Dadurch konnten höhere Teilnahmen-Quoten erzielt werden.

Während der ersten Erhebung stellte sich heraus, dass einige Fragen für Einrichtungen mit einer offenen konzeptionellen Ausrichtung schwer zu beantworten waren oder nicht auf sie zutrafen. Zum Beispiel konnte die Frage „Wie viele pädagogische Fachkräfte arbeiten in Gruppe 1?“ von diesen Einrichtungen oft nicht beantwortet werden. Einrichtungen mit

einem offenen Konzept weisen die Kinder zwar häufig formell „Stammgruppen“ zu, aber haben im Alltag keine festen Gruppenstrukturen. Daher wurden solche „schwierigen“ Fragen ab dem zweiten Erhebungszeitraum in angepasster Form und in einem eigenen Interview für Einrichtungen mit einem offenen Konzept gestellt. Die Unterschiede in den Fragebögen lassen sich dem Datenhandbuch bzw. den Fragebogen-Dokumenten entnehmen. Zum Teil hat sich die konzeptionelle Ausrichtung der Einrichtungen über die drei Erhebungszeiträume verändert.

3.3.2 Gruppenleitungsfragebogen

Die Gruppenleitungen wurden gebeten, Auskünfte über strukturelle Merkmale der Gruppe sowie über Maßnahmen zur Sprachförderung zu erteilen. Darüber hinaus sollten die Gruppenleitungen Angaben zum Alltag in der Gruppe tätigen. Der Fragebogen sollte vorzugsweise von der jeweiligen Gruppenleitung ausgefüllt werden. Teilweise wurden die Angaben jedoch auch von anderen pädagogischen Fachkräften in der jeweiligen (Stamm-) Gruppe getätigt. In der ersten Erhebungswelle wurden alle Informationen online erfasst. Auf mehrfachen Wunsch wurden die Online-Fragebögen ab dem zweiten Erhebungszeitraum durch Papier-Fragebögen ersetzt (vgl. Abbildung 3).

Wie auch beim Einrichtungsfragebogen zeigte sich, dass die Gruppenleitungen, die in Einrichtungen mit einer offenen konzeptionellen Ausrichtung arbeiteten, einen Teil der Fragen im Gruppenleitungsfragebogen nicht oder nur schwer beantworten konnten. Daher wurde ab dem zweiten Erhebungszeitpunkt ebenfalls ein separater Fragebogen für Gruppenleitungen aus Einrichtungen mit einer offenen konzeptionellen Ausrichtung erstellt.

3.3.3 Erzieher-Kind-Fragebogen

Für jedes Kind, das an der Längsschnitterhebung teilnahm, sollte ein Erzieher-Kind-Fragebogen ausgefüllt werden. Diese Angaben wurden durch eine pädagogische Fachkraft aus der Gruppe, die das Kind besucht, getätigt. In der ersten Erhebungswelle wurden diese Informationen online erfasst. Auf mehrfachen Wunsch wurden die Online-Fragebögen ab dem zweiten Erhebungszeitraum durch Papier-Fragebögen ersetzt (vgl. Abbildung 3).

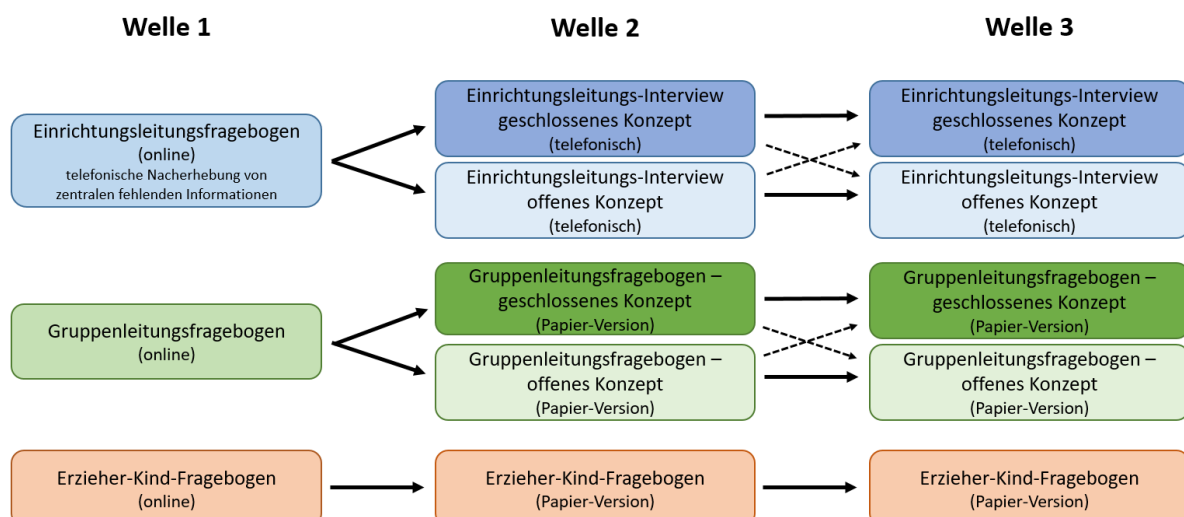


Abbildung 3. Befragung der pädagogischen Fachkräfte.

3.3.4 Telefoninterview mit der Mutter

Die leiblichen Mütter der Kinder wurden nach jeder Datenerhebung gebeten, an einem Telefoninterview teilzunehmen. Das Telefoninterview wurde mit einer anderen Person durchgeführt, wenn ein Kind zum Zeitpunkt der Datenerhebung nicht mit seiner leiblichen Mutter zusammenlebte. In diesem Fall wurde die primäre Bezugsperson des Kindes interviewt, z.B. der alleinerziehende Vater, die Stiefmutter oder die Oma. Anhand der Variablen *tver_x* (x bitte durch den jeweiligen Erhebungszeitraum ersetzen) im Datensatz lässt sich nachvollziehen, mit welcher Person das Telefoninterview durchgeführt wurde.

Im Rahmen des Interviews wurden u.a. soziodemografische Merkmale sowie Informationen zum Sprachverhalten in der Familie erfasst. Die Antworten wurden von den MitarbeiterInnen, die das Interview durchführte, direkt in einen Online-Fragebogen eingegeben und gespeichert.

Mit Müttern, die schon einmal an einem Interview, dem sogenannten Erstinterview (ETI), teilgenommen haben, wurde zu den folgenden Erhebungszeitpunkten ein Zweit- bzw. Dritrinterview durchgeführt (ZTI bzw. DTI). Dieses ZTI bzw. DTI ist eine gekürzte Version des ETIs, in dem unveränderliche Merkmale wie das Geburtsdatum nicht erneut erfragt wurden. Auch Merkmale, die mit einer hohen Wahrscheinlichkeit stabil bleiben, z.B. die Bildung der Mutter, wurden nicht erneut erfasst. Davon abgewichen wurde, wenn die Mutter über eine Filterfrage angegeben hat, dass es seit dem letzten Interview eine Veränderung bezüglich des Merkmals gegeben hat (vgl. Abbildung 4).

Trotz großer Sorgfalt wurde in neun Fälle zweimal ein ETI durchgeführt. Wir raten in diesen Fällen dazu, bei den betroffenen Merkmalen die Angaben aus dem ersten ETI zu verwenden, da das zweite ETI als ZTI/DTI durchgeführt werden sollte. Weiter gibt es acht Fälle in denen nur Daten aus einem ZTI/DTI und keine Daten aus einem ETI vorliegen.

Für Merkmale, die unveränderlich sind oder mit hoher Wahrscheinlichkeit stabil bleiben und daher i.d.R. nur einmalig erfasst wurden, wurde – falls sinnvoll – eine wellenübergreifende Variable gebildet. Das trifft u.a. auf die Erfassung des Geburtsgewichts (*tkgw*) und der Geburtsgröße (*tkgg*) des Zielkindes zu. Eine wellenübergreifende Variable ist daran erkennbar, dass sie am Ende nicht den Zusatz *_1*, *_2* bzw. *_3* enthält (dieser steht bei den anderen Variablen für den Erhebungszeitraum).

Da es unser Ziel war, auch Familien mit anderen Familiensprachen als Deutsch einzubeziehen, wurden die Fragebögen in zahlreiche Sprachen übersetzt und die Telefoninterviews in insgesamt 17 unterschiedlichen Sprachen angeboten.

3.3.5 Mutterfragebogen

Darüber hinaus wurden die leiblichen Mütter gebeten, einen kurzen Online-Fragebogen auszufüllen. Es wurden u.a. Fragen zum aktuellen Entwicklungsstand des Kindes im deutschen Wortschatz und in der deutschen Grammatik (SBE-3-KT, Kademann, Tippelt & von Suchodoletz, 2009) sowie zu einer Reihe von familiären Einflussfaktoren, die für die Sprachentwicklung eines Kindes bedeutsam sein können, gestellt. Der Fragebogen wurde

pro Kind nur einmal während der gesamten Projektlaufzeit ausgefüllt (vgl. Abbildung 4). In der Regel geschah das zu dem Erhebungszeitraum, zu dem das Kind in die Studie aufgenommen wurde. Anhand der Variablen *bwelle* lässt sich entnehmen, wann das Kind in die Studie aufgenommen wurde. Es gibt 27 Fälle, in denen der Mutterfragebogen versehentlich zweimal ausgefüllt wurde.

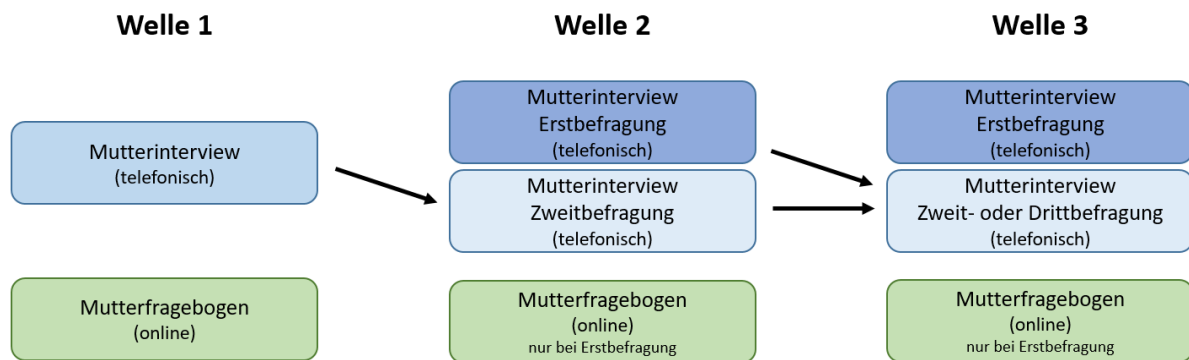


Abbildung 4. Befragung der Mütter.

4. Schulung der MitarbeiterInnen

4.1 Testung der Kinder

In einer 16-stündigen Schulung über zwei Tage wurden im Dezember 2015 die ersten neun studentischen MitarbeiterInnen (ausschließlich weiblich) von den ProjektmitarbeiterInnen für die Durchführung der o.g. Testungen der Kinder in den Einrichtungen geschult. Inhalte der Schulung waren neben den allgemeinen Informationen zum Projekt auch die testtheoretischen Zusammenhänge, die Übung konkreter Testsituationen sowie die Erläuterungen allgemeiner Verhaltensregeln im Umgang mit Kindern und Fachkräften während der Testsituation.

Zwei ProjektmitarbeiterInnen, die an der Schulung der studentischen MitarbeiterInnen beteiligt waren, haben die Datenerhebung unterstützt und die studentischen MitarbeiterInnen bei ihren ersten Testungen begleitet und supervidiert.

Aufgrund des hohen Bedarfs wurden im April 2016 fünf weitere Datenerheberinnen in einer zweitägigen Schulung unter Beteiligung der bisherigen Datenerheberinnen mit den gleichen Inhalten für die Kindertestungen instruiert. Außerdem erfolgte eine Nachschulung aller Datenerheberinnen auf Grund zwischenzeitlich aufgetretener Probleme oder Unklarheiten in der Testsituation.

Im Dezember 2016 wurden elf der Datenerheberinnen, die schon an der ersten Erhebung teilgenommen hatten, für die zweite Datenerhebung vorbereitet. Wissen wurde aufgefrischt und die Durchführung des neuen Tests „Satzgedächtnis“ erläutert und geübt. Weiterhin wurden acht neue studentische MitarbeiterInnen für die Datenerhebung geschult.

Im Dezember 2017 wurde die dritte Datenerhebung vorbereitet. Dreizehn Datenerheberinnen erhielten eine Nachschulung und 7 studentische Mitarbeiterinnen wurden neu für die Datenerhebungen geschult.

Bevor die Datenerheberinnen selbständig Testungen durchführen durften, durchliefen alle mindestens sowohl eine Hospitation als auch eine Supervision durch eine Projektmitarbeiterin in einer Einrichtung. Außerdem wurde von allen Datenerheberinnen ein erweitertes polizeiliches Führungszeugnis angefordert. In regelmäßigen Abständen erfolgten Nachschulungen aller Datenerheberinnen um zwischenzeitlich aufgetretene Probleme oder Unklarheiten in der Testsituation zu besprechen und zu beseitigen.

4.2 Beobachtung der Fachkraft-Kind Interaktionsqualität

Im Dezember 2015 wurden in einem zweitägigen In-House-Seminar durch eine Mitarbeiterin des amerikanischen Unternehmens Teachstone 5 studentische MitarbeiterInnen und 6 ProjektmitarbeiterInnen des Standorts Bochum offiziell in der Anwendung des standardisierten Beobachtungsverfahrens „Classroom Assessment Scoring System Pre-K“ (CLASS Pre-K) geschult. Anschließend hatten die studentischen MitarbeiterInnen und ProjektmitarbeiterInnen acht Wochen Zeit, um eine Online Zertifizierung zu bestehen, so dass sie auch tatsächlich für die Beobachtungen eingesetzt werden konnten. Dazu mussten fünf Videos, die in amerikanischen Kitas aufgenommen wurden, betrachtet und bewertet werden. Auf einer siebenstufigen Skala mussten 80% der Bewertungen und mindestens 40% der Bewertungen pro Dimension nicht mehr als einen Punkt von sogenannten Master Codes abweichen. Alle studentische MitarbeiterInnen und ProjektmitarbeiterInnen bestanden den Test mit einer Reliabilität von 84%.

Im Januar 2017 wurden zehn weitere studentische MitarbeiterInnen und eine Projektmitarbeiterin neu in der Anwendung des Beobachtungsverfahrens CLASS Pre-K geschult. Die Schulung wurde mit Erlaubnis von Teachstone von Frau Dr. Claudia Wirts vom Staatsinstitut für Frühpädagogik durchgeführt. Die studentischen MitarbeiterInnen mussten die Online Zertifizierung ebenfalls bestehen.

Alle CLASS-Beobachter haben die Zertifizierung im jährlichen Abstand wiederholt. Sowohl für den zweiten als auch für den dritten Erhebungszeitraum lag die Reliabilität der Re-Zertifizierung bei 89%.

Manche Mitarbeiterinnen führten sowohl CLASS-Beobachtungen als auch Datenerhebungen durch. Andere MitarbeiterInnen waren nur für die CLASS-Beobachtung oder für die Datenerhebung geschult.

4.3 Telefoninterviews mit den Müttern

Im Mai 2016 fand die erste Schulung zum Thema „Durchführung der telefonischen Mutterinterviews“ mit allen Datenerheberinnen statt. Hier wurde ein besonderer Schwerpunkt auf interkulturelle Sensibilität und Standardisierung gelegt. Nach- und Neuschulungen erfolgten im März 2017 für den zweiten Erhebungszeitraum und im Dezember 2017 für den dritten Erhebungszeitraum.

5. Ablauf der Datenerhebungen in den Einrichtungen

In den Einrichtungen wurden jeweils von Februar bis Oktober 2016 (Welle 1), von Januar bis Oktober 2017 (Welle 2) und von Januar bis September 2018 (Welle 3) die Sprachkompetenzen der Kinder erfasst. Für jeden Erhebungszeitraum wurden mit jeder Einrichtung zwei bis drei Termine an möglichst aufeinanderfolgenden Tagen vereinbart. Es wurden ein bis drei Datenerheberinnen und ein bis zwei CLASS-Beobachter in die Einrichtungen geschickt.

Während des zweiten Erhebungszeitraums kam es zu insgesamt 10 Nachtestungen im November und im Dezember 2017. Diese Nachtestungen wurden mit Kindern durchgeführt, die schon im letzten Jahr an der Studie teilgenommen hatten, aber in diesem Jahr an allen vereinbarten Testtagen abwesend waren.

Nach der Ankunft in der Einrichtung besprachen die Datenerheberinnen das allgemeine Vorgehen mit den involvierten pädagogischen Fachkräften. Die anwesenden pädagogischen Fachkräfte wurden gebeten nicht in den Testablauf einzugreifen. Die Testung fand in einem möglichst ruhigen, separaten Raum statt. Während der Testung im ersten Erhebungszeitraum war in der Regel eine pädagogische Fachkraft der jeweiligen Gruppe anwesend. Ab dem zweiten Erhebungszeitraum wurde von der pädagogischen Fachkraft bzw. auch oft von den Kindern selbst entschieden, ob die Testung durch eine pädagogische Fachkraft begleitet wird.

Es wurde darauf geachtet, dass Kinder, die eine Brille benötigen, diese auch getragen haben, ggf. wurde die Testung um einen Tag verschoben.

Der Ablauf der Testung war stets gleich. Begonnen wurde mit der Begrüßung der Kinder und einer Einführung in die Testsituation (in Welle 1 mit der Papageien-Handpuppe *Seika*). Danach folgte die Durchführung des PPVT-4. Der PPVT-4 ist ein adaptives Verfahren, das sich bei der Auswahl der zur Bearbeitung vorgelegten Items am Antwortverhalten der Kinder orientiert. Da das Verfahren hohe Ansprüche an die Testdurchführung stellt, wurde eine computerbasierte Version am Tablet verwendet. Für jedes Kind wurde eine separate Datei mit den Testergebnissen erzeugt. Anschließend wurde mit den Kindern der Subtest Figuren legen des HAWIVA III sowie die beiden Subtests PGN und MR des SETK 3-5 durchgeführt. Ab der zweiten Erhebungswelle wurde für Kinder ab 4 Jahren zusätzlich der Subtest SG aus dem SETK 3-5 angewandt. Die drei bzw. vier Testverfahren wurden ohne Tablet durchgeführt. Die Ergebnisse wurden für jedes Kind auf einem Protokollbogen notiert. Nach Abschluss der Testung bekam jedes Kind eine Forscherurkunde ausgehändigt.

In einigen Fällen mussten einzelne Testverfahren oder gesamte Testungen abgebrochen werden, z.B. wenn ein Kind die Aufgabenstellung nicht verstand oder nicht zur Teilnahme zu motivieren war. Aufschluss über solche Fälle geben testspezifische Variablen (Näheres siehe Variablenhandbuch, Stichwort „Abbruch“). Weiterhin gibt es testspezifische Variablen, die beschreiben, ob es Zweifel an der Belastbarkeit der Testergebnisse gibt, z.B. aufgrund unerwünschter Hilfestellungen einer anwesenden pädagogischen Fachkraft oder aufgrund von technischen Problemen mit dem Tablet bei der Durchführung des PPVT-4 (Stichwort,

„bedenkliches Testergebnis“). Im Falle eines Abbruchs oder von schwerwiegenden Bedenken wurde der Gesamt-Score des jeweiligen Tests auf NA (Missing) gesetzt.

Parallel zu den Testungen wurde an einem Tag das Beobachtungsverfahren CLASS Pre-K durchgeführt. In Kitas mit einem offenen Konzept wurde die Beobachtung nur einmal pro Erhebungszeitraum für die gesamte Einrichtung durchgeführt, da sich die Kinder keiner festen Gruppe zuordnen ließen. Der CLASS-Beobachtungsbogen wurde in den Kitas ausgefüllt und die Ergebnisse wurden zu einem späteren Zeitpunkt in ein Computerprogramm eingegeben.

Die Dauer eines Besuchs richtete sich nach den Öffnungszeiten der Einrichtungen. Meistens erfolgte die Datenerhebung vormittags bis in die frühen Mittagsstunden.

6. Stichprobenpflege

6.1 Einrichtungen

Im ersten Erhebungszeitraum erhielt jede Einrichtung für ihre Teilnahme 100 € in Form eines Gutscheins sowie kleinere Geschenke wie SEIKA-Blöcke, -Kugelschreiber und -Luftballons. Jede teilnehmende Gruppe erhielt nach der Datenerhebung eine Papageien-Handpuppe. Zusätzlich erhielt jede Einrichtung je Kind im Alter von 2;6 bis 3;11 Jahren, das an der Längsschnitterhebung teilnahm, einen 30 €-Gutschein.

Im zweiten und dritten Erhebungszeitraum erhielt jede Einrichtung 15€ für jedes Kind, das an der Längsschnitt-Studie teilnahm. Zusätzlich erhielt jede Einrichtung im zweiten Erhebungszeitraum je eine Einkaufstasche mit SEIKA-Logo, eine Teebox, ein Papageienbuch für jede teilnehmende Gruppe und Spielzeug für die Kinder. Im dritten Erhebungszeitraum erhielt jede Einrichtung einen Stoffbeutel mit SEIKA-Logo, eine Packung Glückstee, Spielzeug für die Kinder und für jede teilnehmende Gruppe ein weiteres Buch.

6.2 Mütter

Im ersten Erhebungszeitraum konnten alle Mütter, die sowohl am telefonischen Mutterinterview teilgenommen als auch den Online-Fragebogen ausgefüllt hatten, zwischen einer Barüberweisung oder einem Gutschein im Wert von 15 € auswählen. Darüber hinaus erhielt jede Mutter ein 5€-Los der Deutschen Klassenlotterie zur Teilnahme an sieben Ziehungen.

Im zweiten und dritten Erhebungszeitraum konnten alle Mütter, die am Telefoninterview teilgenommen hatten, zwischen verschiedenen Gutscheinen im Wert von 20 € wählen.

6.3 Kinder

In jedem Erhebungszeitraum erhielt jedes getestete Kind eine Teilnehmerurkunde, die es selbst mit einem Papageien-Stempel abstempeln durfte.

Jeweils im Dezember 2016, 2017 und 2018 wurde den Kindern eine Postkarte mit Aufklebern zugeschickt. Mit den Postkarten bedankten wir uns jeweils für die Teilnahme im jeweiligen

Jahr und erinnerten im Jahr 2016 und 2017 gleichzeitig an die Studienteilnahme im nächsten Jahr.

7. Datenaufbereitung

Für die Zusammenführung der Daten wurden IDs für die Kinder, die Einrichtungen, die Gruppen und die CLASS-Beobachtungen vergeben. Die Zusammenführung und Aufbereitung der Daten wurde hauptsächlich in den Statistikprogrammen R und SPSS durchgeführt.

Vor den vereinbarten Terminen in den Einrichtungen schickten wir Excel-Gruppenfragebögen an die Einrichtungsleitungen. Diese erhielten wir ausgefüllt per Email zurück (passwortgeschützt). Die Dateien wurden abgespeichert, in R eingelesen und zusammengefügt.

Im Rahmen der Testungen mit dem PPVT-4 wurde für jedes Kind eine separate Datei mit den Testergebnissen erstellt. Diese Dateien wurden in R eingelesen und zusammengefügt.

Für die Eingabe der Testergebnisse, die auf den Protokollbögen notiert waren (Figuren legen, PGN, MR und SG) wurde für jede Welle eine Eingabemaske in C# mit Windows Forms erstellt. Diese Masken wurden so konfiguriert, dass es bei der Eingabe eines ungültigen Wertes zu einer Warnung kam und die Eingabe nicht gespeichert werden konnte. Durch dieses Vorgehen wurden Eingabefehler minimiert. Die eingegebenen Daten wurden als Text-Dateien ausgegeben. Die Dateien wurden in R eingelesen und zusammengefügt.

Die Eingabe der CLASS-Beobachtungsbögen erfolgte für jeden Erhebungszeitraum in einer separaten SPSS-Datei. Die eingegebenen Daten wurden auf ungültige Werte überprüft.

Im ersten Erhebungszeitraum erfolgte die Beantwortung der Fragen im Einrichtungs-, Gruppenleitungs- und Erzieher-Kind-Fragebogen jeweils online in einem Umfrageportal. Im Umfrageportal wurde jeweils eine SPSS-Datei für jeden der drei Fragebögen ausgegeben. Ab der zweiten Erhebungswelle wurde der Einrichtungsleitungsfragebogen durch ein Telefoninterview ersetzt. Die am Telefon getätigten Angaben wurden von den Interviewerinnen direkt in einen Online-Fragebogen im Umfrageportal eingetragen und gespeichert. Die Gruppenleitungs- und Erzieher-Kind-Fragebogen wurden ab der zweiten Erhebungswelle durch Papier-Fragebögen ersetzt. Die ausgefüllten Fragebögen wurden später von studentischen MitarbeiterInnen in das Umfrageportal eingegeben. Die Daten aus den drei Fragebögen wurden in R zusammengefügt und entsprechend umstrukturiert.

Alle separat aufbereiteten Dateien wurden in R zusammengefügt. Dichotome Variablen mit nein/ja-Antwortformat wurden einheitlich kodiert (0 = nein, 1 = ja). Neue Variablen wurden gebildet (näheres siehe Variablenhandbuch). Negativ formulierte Items wurden rekodiert und erhielten die Endung „r“ im Variablennamen.

8. Hinweise zur Nutzung der Daten

8.1 Berücksichtigung der Mehrebenen-Struktur

Für die Erhebung der Daten wurde ein Mehrebenen-Stichproben-Design realisiert.

Zur Querschnittserhebung in der ersten Welle waren 4010 Kinder genestet in 192 Gruppen. Die Gruppen waren in 95 Kitas genestet und 3 Förderlinien zugehörig.

Zur Längsschnitt-Stichprobe zählen insgesamt 1047 Kinder aus insgesamt 245 verschiedenen Gruppen. Die Gruppen waren in 94 Einrichtungen genestet. Die CLASS-Beobachtungseinheiten stellen eine alternative Ebene zur Gruppenebene dar. In Tabelle 2 ist die Schachtelung je Erhebungszeitpunkt dargestellt.

Es ist davon auszugehen, dass sich die Mitglieder innerhalb einer Ebenen-Einheit ähnlich sind. Zum Beispiel ist zu erwarten, dass der Sprachstand eines einzelnen Kindes mit dem Sprachstand der anderen Kinder in der Gruppe bzw. Kita zusammenhängt. Das Ignorieren der Abhängigkeiten kann zu verzerrten Standardfehlern führen. Daher sollten die verschiedenen Ebenen berücksichtigt werden, wenn die vorliegenden Daten analysiert werden. Das kann durch die Anwendung eines Mehrebenen-Modells oder zumindest durch die Korrektur von Standardfehlern für die Nestung geschehen.

Tabelle 2. Anzahl ineinander geschachtelter Untersuchungseinheiten im Längsschnitt.

	Welle 1			Welle 2			Welle 3		
	<i>M</i>	<i>Min</i>	<i>Max</i>	<i>M</i>	<i>Min</i>	<i>Max</i>	<i>M</i>	<i>Min</i>	<i>Max</i>
Gruppen in Einrichtungen	1.90	1	3	2.29	1	4	2.40	1	5
CLASS-Beobachtungseinheiten in Einrichtungen	1.75	1	3	2.00	1	4	1.85	1	4
Kinder in Kita-Gruppen	3.58	1	9	4.46	1	13	4.46	1	13
Kinder in CLASS-Beobachtungseinheiten	3.89	1	15	5.08	1	18	5.62	1	21

8.2 Variablenhandbuch

Das Variablenhandbuch enthält eine Beschreibung aller Variablen im Datensatz. Die Reihenfolge der Variablen im Handbuch entspricht der Reihenfolge der Variablen im Datensatz. Die Variablennamen sind eindeutig identifizierbar und erlauben eine Zuordnung zu dem Instrument, aus dem sie stammen. Die Beschreibung der Variablen umfasst Informationen zur Frageformulierung, zu den Variablenwerten, zur Berechnungsweise (im Falle von neu gebildeten Variablen), zur Quelle sowie weiteren wichtigen Anmerkungen.

9. Projektbezogene Publikationen:

- Agache, A., Bihler, L.-M., Kohl, K., Willard, J., & Leyendecker, B. (2019). *Introducing the network analysis approach to unpack the multi-causal system of early child-care quality*. Unveröffentlichtes Manuskript, Fakultät für Psychologie, Ruhr-Universität Bochum.
- Agache, A., Bihler, L.-M., Willard, J., & Leyendecker, B. (2020). *Sprachbildung und -entwicklung im Kita-Alltag in Nordrhein-Westfalen (SEIKA-NRW)* (Version 1) [Datensatz]. Berlin: IQB – Institut zur Qualitätsentwicklung im Bildungswesen.
http://doi.org/10.5159/IQB_SEIKA-NRW_v1
- Bihler, L., Agache, A., Schneller, K., Willard, J.A. & Leyendecker, B. (2018). Expressive morphological skills of dual language learning and monolingual German children: Exploring links to duration of preschool attendance, classroom quality and classroom composition. *Frontiers in Psychology, Developmental Psychology*.
<https://www.frontiersin.org/articles/10.3389/fpsyg.2018.00212/full>
- Bihler, L., Agache, A., & Leyendecker, B. (2018). Factor analysis of the Classroom Assessment Scoring System replicates the three domain structure and reveals no support for the bifactorial model in German preschools. *Frontiers in Psychology, Educational Psychology*. <https://doi.org/10.3389/fpsyg.2018.01232>
- Kohl, K., Willard, J., Agache, A., Bihler, L.-M., & Leyendecker, B. (2019). Classroom quality, classroom composition, and age at entry: Experiences in early childhood education and care and single and dual language learners' vocabulary. *AERA Open*.
<https://doi.org/10.1177/2332858419832513>
- Bihler, L.-M., Agache, A., Willard, J.A., Kohl, K. & Leyendecker, B. (2020). Propensity score matching as a procedure for sample selection. Illustrating the utility for selecting three groups. In K. Blatter, K. Groth & M. Hasselhorn (Eds.), *Evidenzbasierte Überprüfung von Sprachförderkonzepten*, Edition ZfE, Band 6. Springer VS, Wiesbaden (im Druck).
- Kohl, K., Bihler, L., Willard, J., Agache, A., & Leyendecker, B. (2020) Linking quantity and quality of early childhood education and care to children's socio-emotional adjustment: A German cross-sectional study. *Early Education and Development*, 31, 177-199.
doi:10.1080/10409289.2019.1650543
- Willard, J.A., Agache, A., Bihler, L.-M., Kohl, K., & Leyendecker, B. (in Begutachtung) Language development of dual language learners and lower SES children in early childhood education: The role of quality, composition and early access. *Child Development*.
- Willard, J.A., Kohl, K., Bihler, L.-M., Agache, A., & Leyendecker, B. (im Druck). Family Literacy Activities and their Interplay with Family and Preschool Language Environments: Links to Dual Language Learners' German Vocabulary. *Early Education and Development*.

10. Literaturverzeichnis

- Armon-Lotem, S., De Jong, J., & Meir, N. (2015). *Assessing multilingual children. Disentangling bilingualism from language impairment*. Bristol: Multilingual Matters.
- Bryer, J. (2013). *TriMatch: An R package for propensity score matching of non-binary treatments*. Retrieved from <https://pdfs.semanticscholar.org/716d/cc948d6481c362c0c7b9c2d9cd16fb31be52.pdf>
- Caspar, U., & Leyendecker, B. (2011). Deutsch als Zweitsprache. Die Sprachentwicklung türkischstämmiger Vorschulkinder in Deutschland. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 43(3), 118–132. <https://doi.org/10.1026/0049-8637/a000046>
- Dubowy, M., Ebert, S., von Maurice, J., & Weinert, S. (2008). Sprachlich-kognitive Kompetenzen beim Eintritt in den Kindergarten: Ein Vergleich von Kindern mit und ohne Migrationshintergrund. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 40(3), 124–134. <https://doi.org/10.1026/0049-8637.40.3.124>
- Dunn, L. M., & Dunn, D. M. (2007). *Peabody Picture Vocabulary Test (4th ed.)*. (N. Pearson, Ed.). Minneapolis.
- Ebert, S., & Weinert, S. (2013). Predicting reading literacy in primary school: The contribution of various language indicators in preschool. In M. Pfost, C. Artelt, & S. Weinert (Eds.), *The development of reading literacy from early childhood to adolescence* (pp. 93–149). Bamberg: University of Bamberg Press. <https://doi.org/https://doi.org/10.20378/irbo-51696>
- Grimm, H. (2001). *SETK 3-5. Sprachentwicklungstest für drei- bis fünfjährige Kinder. Diagnose von Sprachverarbeitungsfähigkeiten und auditiven Gedächtnisleistungen*. Göttingen: Hogrefe.
- Lenhard, A., Lenhard, W., Segerer, R., & Suggate, S. (2015). *Peabody Picture Vocabulary Test (4. Ausgabe)*. Frankfurt: Pearson.
- Mashburn, A. J., Pianta, R. C., Hamre, B. K., Downer, J. T., Barbarin, O. A., Bryant, D., ... Howes, C. (2008). Measures of classroom quality in prekindergarten and children's development of academic, language, and social skills. *Child Development*, 79(3), 732–749. <https://doi.org/10.1111/j.1467-8624.2008.01154.x>
- Pianta, R. C., La Paro, K. M., & Hamre, B. K. (2008). *Classroom Assessment Scoring System. Manual, Pre-K*. Baltimore: Brookes Publishing Co.
- Ricken, G., Fritz, A., Schuck, K. D., & Preuß, U. (Eds.). (2007). *HAWIVA -III, Hannover-Wechsler Intelligenztest für das Vorschulalter-III. Manual zur Durchführung und Auswertung*. Göttingen: Huber.
- Von Goldammer, A., Mähler, C., Bockmann, A. K., & Hasselhorn, M. (2010). Vorhersage früher schriftsprachleistungen aus vorschulischen kompetenzen der sprache und der phonologischen informationsverarbeitung. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 42(1), 48–56. <https://doi.org/10.1026/0049-8637/a000005>
- Weinert, S., Ebert, S., Lockl, K., & Kuger, S. (2012). Disparitäten im Wortschatzerwerb: Zum

Einfluss des Arbeitsgedächtnisses und der Anregungsqualität in Kindergarten und Familie auf den Erwerb lexikalischen Wissens. *Unterrichtswissenschaft*, 40(1), 4–25.

Weinert, S., & Grimm, H. (2018). Sprachentwicklung. In W. Schneider & U. Lindenberger (Eds.), *Entwicklungspsychologie* (8., überar., pp. 445–470). Weinheim: Beltz/PVU.