

# Langfristige Wirkungen des Fortbildungsansatzes „Mit Kindern im Gespräch“ zur Sprachförderung in Kindertagesstätten

Gisela Kammermeyer<sup>1</sup>, Anja Leber<sup>1</sup>, Astrid Metz<sup>1</sup>,  
Susanna Roux<sup>2</sup>, Beate Biskup-Ackermann<sup>1</sup>, Eva Fondel<sup>3</sup>

<sup>1</sup> Universität Koblenz-Landau, Campus Landau

<sup>2</sup> Pädagogische Hochschule Weingarten

<sup>3</sup> Evaluation, Methoden- und Forschungsberatung (Statistik) Landau

*Zusammenfassung:* Die langfristigen Wirkungen von Fortbildungen werden nur selten untersucht; dies gilt auch für den Elementarbereich. In einer randomisierten experimentellen Prä-Post-Follow-up-Studie im Feld werden zwei Fortbildungsansätze zur Sprachförderung in Kindertagesstätten verglichen. Zur Erfassung der Wirkungen der pädagogischen Fachkräfte ( $N=50$ ) wird das internationale anerkannte *Classroom Assessment Scoring System* (CLASS) von Pianta, La Paro und Hamre (2008) eingesetzt. Varianzanalytische Auswertungen zeigen, dass sich diejenigen pädagogischen Fachkräfte, die mit dem inhaltlich auf Sprachförderstrategien und methodisch auf situiertes Lernen fokussierten Fortbildungsansatz „Mit Kindern im Gespräch“ (Kammermeyer et al., 2017) weitergebildet wurden, ein Jahr nach Fortbildungsende signifikant stärker in der Anregungsqualität (*instructional support*) verbessert haben als diejenigen, die mit dem inhaltlich breit angelegten Ansatz „Sprache – Schlüssel zur Welt“ (Katholische Erwachsenenbildung Rheinland-Pfalz, 2009) fortgebildet wurden. Es gibt einen signifikanten Unterschied zwischen den Fortbildungsgruppen im Anteil derjenigen, die ein Jahr nach der Fortbildung einen als bedeutsam angenommenen Schwellenwert übersteigen.

**Schlüsselbegriffe:** Fortbildung, Sprachförderung, Kindertagesstätte

## *Long-term Impact of the Language Professional Development Program “Mit Kindern im Gespräch (Talking with Children)” in German Preschools*

*Summary:* Long-term impacts of in-service trainings are seldom examined; this is also true for early childhood education. In a randomized experimental pre-post-follow-up-study in the field, two approaches for in-service trainings were compared. To measure the effects, the standardized *Classroom Assessment Scoring System* (CLASS, Pianta, La Paro & Hamre, 2008) was used. One year after the end of the in-service training variance-analytical evaluations showed that preschool teachers, who have been trained with the methodically and content-focused approach “Mit Kindern im Gespräch” (Kammermeyer et al., 2017) indicated significant greater gains in instructional support relative to teachers in the control group with the broadly based approach “Sprache – Schlüssel zur Welt” (Katholische Erwachsenenbildung Rheinland-Pfalz, 2009) and a significant difference between the two in-service-trainings in exceeding a threshold considered to be meaningful.

**Keywords:** Professional development, language support, preschool

Wissenschaft und Praxis wurden durch die Evaluationsergebnisse aufgeschreckt, wonach aufwendige und teure Sprachförderprogramme nicht die erhofften Wirkungen hatten (Roos,

Polotzek & Schöler, 2010). Möglicherweise liegen die fehlenden Effekte jedoch weniger an der Konzeption der Sprachförderprogramme als an den Fortbildungen, die zu deren Vermitt-

lung eingesetzt wurden. Der Transfer von Forschungserkenntnissen über Fortbildung in die Praxis ist ein sehr komplexer und anspruchsvoller Prozess (Lipowsky, 2010), der auch von der methodischen Gestaltung der Fortbildung (z. B. Berücksichtigung des Anwendungskontexts durch Methoden des situierten Lernens) abhängt. Von hoher praktischer Relevanz ist deshalb zum einen die Frage, ob die neu konzipierte Fortbildung zur Verbesserung der Anregungsqualität „Mit Kindern im Gespräch“ größere Wirkungen erzielt als die etablierte „Sprache – Schlüssel zur Welt“. Aufgrund des Zusammenhangs der Anregungsqualität mit der sprachlichen und kognitiven Entwicklung von Kindern (Mashburn et al., 2008) ist die Frage bedeutsam, inwieweit es durch die Gestaltung einer Fortbildung gelingt, die Qualität des pädagogischen Handelns einer Fachkraft langfristig so zu steigern, dass Verbesserungen in der Entwicklung von Kindern erwartet werden können.

### *Theoretischer Hintergrund*

Wenn es um Wirkungen von Fortbildungen geht, stellt sich zunächst die Frage nach dem Kriterium, an dem diese gemessen werden sollen. Weit verbreitet ist die Orientierung an dem aus der beruflichen Weiterbildung stammenden Vier-Ebenen-Modell von Kirkpatrick (1979), in dem bei den Teilnehmenden unterschieden wird zwischen (a) deren *Reaktion* (z. B. Zufriedenheit), (b) deren *Lernergebnissen* (z. B. Veränderung von Wissen und Orientierungen), (c) deren *Verhaltensänderungen* (z. B. Qualität und Quantität von Interaktionen) sowie (d) den *Resultaten dieses Verhaltens* (z. B. der Entwicklung von Kindern). Die dritte Ebene wird nach Lipowsky (2010, S. 58) häufig ausgespart, weil von Veränderungen auf *Kinderebene* (Ebene 4) auf potenziell stattgefundene Veränderungen auf der *Unterrichtsebene* (Ebene 3) rückgeschlossen wird. Wenn jedoch, wie in der vorliegenden Studie, Verhaltensänderungen der Erzieherinnen (Ebene 3) gemessen werden, so können aufgrund bestehender

Erkenntnisse über die Zusammenhänge mit Kind-Outcomes (vgl. Schwellenwerte nach Burchinal, Vandergrift, Pianta & Mashburn, 2010) auch erste Folgerungen auf die Kindebene (Ebene 4) abgeleitet werden.

Ein Modell zur Beschreibung des Wegs von der Fortbildung der pädagogischen Fachkraft bis zu den Auswirkungen bei den Kindern stammt von Fukkink und Lont (2007). Danach führt ein Weg von (a) der Fortbildung über (b) Wissen und Orientierungen zum (c) Handeln und von dort dann (d) auf die Entwicklung von Kindern. Einen zusätzlichen direkten Weg von der Fortbildung zum pädagogischen Handeln über das Identifizieren von gelungenen Situationen gibt es nach dem Zweiweg-Modell von Hamre et al. (2012).

### *Forschungsstand*

Der Weg von der Fortbildung zu den Auswirkungen bei den Kindern ist zwar weit, wenn es jedoch gelingt, durch eine Fortbildung die Anregungsqualität zu erhöhen, dann kann aufgrund der Erkenntnisse von Mashburn et al. (2008), in ihrer Studie mit 2439 Kindern in 671 Einrichtungen in 11 US-Bundesstaaten, von Wirkungen auf die sprachliche und kognitive Entwicklung der Kinder ausgegangen werden. Mittels der mit CLASS erhobenen Anregungsqualität konnte die sprachliche und kognitive Entwicklung signifikant vorhergesagt werden. Der Frage, ob es unterschiedliche Zusammenhänge zwischen kindlicher Entwicklung und der Anregungsqualität in Abhängigkeit von unterschiedlichen Qualitätsniveaus gibt, gingen die Studien zu Schwellenwerten von Burchinal et al. (2010), Burchinal, Vernon-Feagans, Vitiello, Greenberg und Family Life Project Key Investigators (2014) sowie von Burchinal, Zaslow und Tarullo (2016) in den USA nach. Burchinal et al. (2010) kommen aufgrund ihrer Analysen mit einem a priori gesetzten Schwellenwert für die Anregungsqualität von 3.25 auf der siebenstufigen CLASS-Skala zu einem signifikanten Unterschied des Zusammenhangs mit *expres-*

*sive language* in den so voneinander abgetrennten Qualitätsbereichen. Eine Anregungsqualität über 3.25 erwies sich als ein stärkerer Prädiktor als eine unter 3.25. In anderen Studien dieser Forschergruppe wurden für ähnliche Fragestellungen auch A-priori-Schwellenwerte (Anregungsqualität) von 3.00 (Burchinal et al., 2014) bzw. 2.75 (Burchinal et al., 2016) gewählt.

Die Verbesserung der Anregungsqualität ist dringend notwendig, denn in Deutschland ist diese auf einem niedrigen Niveau. Der CLASS-Skalenwert beträgt bei Stuck, Kammermeyer und Roux (2016) 1.61, bei von Suchodoletz, Fäsche, Gunzenhauser und Hamre (2014) 2.47 und bei Wildgruber, Wertfein, Wirts, Kammermeyer und Danay (2016) 2.49. In den USA liegt die Anregungsqualität bei 87 % unter 3.25 (Burchinal et al., 2010).

Es gibt international nur wenige Studien, die langfristige Effekte von Fortbildungsmaßnahmen mit Designs von hoher Qualität untersuchen, wie die Metaanalysen im Elementarbereich zeigen (Egert, 2015; Fukkink & Lont, 2007; Markussen-Brown et al., 2017; Schachter, 2015; Werner, Linting, Vermeer & van Ijzendoorn, 2016; Yoon, Duncan, Lee, Scarloss & Shapley, 2007). Langfristige Wirkungen wurden in keiner der 73 von Schachter (2015) in ihre Literaturanalyse einbezogenen Fortbildungsstudien untersucht. In der Metaanalyse von Yoon et al. (2007), in der Fortbildungsstudien zwischen 1986 und 2006 analysiert wurden, fanden sich unter den 1343 Studien nur 9, die über ein quasi-experimentelles Design oder ein randomisiertes Kontrollgruppendesign verfügten. Hinzu kommt, dass die Verwendung von standardisierten Beobachtungsinstrumenten zur Messung der Wirkung von Fortbildungen auf das pädagogische Handeln eher die Ausnahme als die Regel darstellt. Im Literaturüberblick zu Fortbildungen im Elementarbereich von Schachter (2015) wurden von den 73 einbezogenen Studien in 40 % der Fälle von den Forschern selbst entwickelte Erhebungsinstrumente verwendet, nur 15 setzten standardisierte, psychometrisch getestete Verfahren

ein, um Veränderungen zu messen (wie z. B. das Classroom Assessment Scoring System [CLASS] von Pianta, La Paro und Hamre [2008], das von Schachter [2015] als eines der besten derzeit verfügbaren Erhebungsinstrumente bezeichnet wird).

Mit CLASS wurden die Wirkungen von Fortbildungen in mehreren randomisierten Studien gemessen: In der Studie von Wasik und Hindman (2011) zeigten die Lehrpersonen in der Interventionsgruppe nach Ende der Fortbildung eine signifikant höhere Anregungsqualität als die der Vergleichsgruppe in der Größenordnung von ca. einem Punkt auf der siebenstufigen CLASS-Skala.

Signifikante Effekte in der Anregungsqualität wurden auch in der Studie von Hamre et al. (2012) gefunden. Die Differenz zwischen Interventions- und Kontrollgruppe (*business as usual supports*) auf der siebenstufigen Ratingsskala beträgt knapp einen halben Punkt. Die webbasierte Fortbildung „MyTeachingPartner“ (MTP), in der es um die Identifizierung von gelingenden Interaktionssituationen in fremden und eigenen Videosequenzen geht und die Videobeispiele und webbasierte Beratung im Hinblick auf die CLASS-Dimensionen enthält, wurde von Pianta, Mashburn, Downer, Hamre und Justice (2008) mit einer Fortbildung verglichen, die lediglich aus dem Zugang zu einer Website mit Videoclips bestand. Positive Wirkungen auf die Anregungsqualität wurden vor allem bei pädagogischen Fachkräften in Klassen mit einem größeren Anteil an Risikokindern gefunden, die Effekte waren jedoch nur klein, sie bestanden, in CLASS-Skalenwerten übersetzt, aus einem halben Punkt.

Die ausschließlich webbasierte Fortbildung „MyTeachingPartner“ (MTP) wurde in der Studie von Early, Maxwell, Ponder und Pan (2017) mit dem face-to-face Fortbildungskonzept „Making the Most of Classroom Interactions“ (MMCI) verglichen. Bei MMCI lernen die pädagogischen Fachkräfte mithilfe der CLASS effektive Interaktionen zu identifizieren und zu analysieren. Zwischen den Sitzungen haben die pädagogischen Fachkräfte die Aufgabe,

Videos aus der CLASS-Videothek zu analysieren und einzelne Interaktionsformen in ihrer Praxis auszuprobieren. Der Vergleich zwischen der MMCI-Fortbildung, der MTP-Fortbildung und der Kontrollgruppe zeigt, dass es nur zwischen dem MMCI-Treatment und der Kontrollgruppe einen signifikanten Unterschied in der Anregungsqualität gab, in der Höhe von  $\Delta = 0.27$  ( $p = 0.011$ ) auf der siebenstufigen CLASS-Skala.

Auch zur Frage, von welchen Faktoren die Wirkungen einer Fortbildung abhängen, gibt es einschlägige Forschungserkenntnisse (Egert, 2015; Egert, Eckhardt & Fukkink, 2017). Fortbildungen mit einem Fokus auf spezifische Inhalte, die sich auf evidenzbasierte Lehrstrategien beziehen, vergrößern die Effekte im Elementarbereich (Birman, Desimone, Porter & Garet, 2000; Egert, 2015; Fukkink & Lont, 2007). Belegt sind vor allem die Wirkungen von Fortbildungen mit Fokussierung auf die Interaktionsqualität (Egert, 2015). Das methodische Vorgehen spielt ebenfalls eine Rolle. Im Ansatz des Situiereten Lernens (Fölling-Albers, Hartinger & Mörtl-Hafizovic, 2004) werden verschiedene methodische Einzelmerkmale gebündelt, die zur Erzeugung flexibel anwendbaren Wissens beitragen, das auf die eigene Praxis übertragen werden kann (*authentische Lernsituationen, kooperatives Lernen, kognitive Aktivierung, Perspektivwechsel, Berücksichtigung des Anwendungskontexts*). Die Bedeutung des Situiereten Lernens kann auch belegt werden durch die Ergebnisse der Studien zum Fortbildungsprogramm „MyTeachingPartner“, in denen sich die Identifikation von gelingenden Interaktionssituationen als eine besonders geeignete authentische Lernsituation erwies. Sie stellte sich in dieser Studie als bedeutsame Variable zur Erklärung des Fortbildungserfolgs auf alle Dimensionen der Anregungsqualität heraus und bestätigt somit das Zweiwege-Modell von Hamre et al. (2012) mit einem zusätzlichen direkten Weg von der Fortbildung zum pädagogischen Handeln über das Identifizieren von gelungenen Situationen.

In Deutschland wurden die langfristigen Wirkungen einer Fortbildung im Elementarbereich zur Verbesserung der Interaktionsqualität im Vergleich zu einer Kontrollgruppe mit einer Basis-Fortbildung von Beller (2014) analysiert. Die Wirkungen zwei Jahre nach Abschluss der Intervention wurden mit einem selbst entwickelten niedrig-inferenten Erhebungsinstrument gemessen, als abhängige Variable wurde ein eng auf die Intervention abgestimmtes Beobachtungssystem verwendet, das u. a. auch das sprachliche Anregungsniveau beinhaltet. Der Vergleich der Effektstärken zwischen Prä- und Follow-up-Test zeigt nach Beller (2014) eine signifikante Interaktion (Intervention  $\times$  Zeit) mit großem Effekt im Hinblick auf die Skala *Sprachvorbild* und *verbale Austausch anregen*. In den beiden Skalen *Gespräche führen* und *Erzieherin spricht* hingegen gibt es keine signifikanten Interaktionseffekte.

Auch in der randomisierten Studie von Simon und Sachse (2011) wurden die Veränderungen des sprachlichen Verhaltens der pädagogischen Fachkräfte vor und nach einer Fortbildung (mit dem „Heidelberger Trainingsprogramm“ von Buschmann & Jooss, 2011) mit einem selbst entwickelten niedrig-inferenten Kategoriensystem im Hinblick auf die langfristigen Wirkungen (5 Monate nach dem Posttest) untersucht. Beim Redeanteil der Erzieherinnen findet man einen mittleren Effekt zum Posttest und einen großen Effekt zum Follow-up im Unterschied zum Prätest. Bei den Sprachlehrstrategien einschließlich *korrekatives Feedback* gibt es sowohl einen großen Effekt zum Posttest als auch zum Follow-up. Bedeutsam ist hier zudem, dass es vom Posttest zum Follow-up-Zeitpunkt einen Anstieg der Effektstärken gibt.

Die Untersuchung der kurzfristigen Wirkungen (unmittelbar nach Ende der Fortbildung) der Fortbildung „Mit Kindern im Gespräch“ im Vergleich zu „Sprache – Schlüssel zur Welt“ ergab signifikante Interaktionseffekte (Intervention  $\times$  Zeit) in allen drei Dimensionen der Anregungsqualität, einen großen Effekt zu Sprachbildung und Konzeptentwicklung und einen mittleren Effekt zu Feedbackqualität (Kammermeyer et al., 2018).

Zusammenfassend kann festgehalten werden, dass es in den internationalen Fortbildungsstudien, die sich auf die Verbesserung der Interaktionsqualität beziehen und diese mit dem international anerkannten, standardisierten hoch-inferenten Classroom Assessment Scoring System (CLASS) von Pianta, La Paro et al. (2008) messen, keinen Follow-up-Messzeitpunkt gibt. In den deutschen Studien hingegen, die langfristige Wirkungen von Fortbildungen zur Interaktionsqualität erfassen und signifikante Effekte zeigen, wird die abhängige Variable mit einem selbst entwickelten niedrig-inferenten Instrument gemessen.

Wie sich die durch Fortbildung verbesserte Anregungsqualität von pädagogischen Fachkräften auch langfristig über ein Jahr hinweg entwickelt, wenn die Wirkungen mit einem hoch-inferenten Instrument gemessen werden, mit dem diese auch international eingeordnet werden können, ist unklar. Die Wirkungen zwischen Post- und Follow-up-Test hängen vermutlich von interventionsabhängigen und -unabhängigen Faktoren ab (Hager & Hasselhorn, 2000), möglich sind ein *Fading-Effekt* (die Anregungsqualität sinkt), ein *Stabilisierungseffekt* (die erreichte Anregungsqualität bleibt erhalten) oder ein *Entwicklungsschub* (die Anregungsqualität erhöht sich).

Für einen Fading-Effekt sprechen zum einen interventionsabhängige Faktoren, z. B. wenn das in der Fortbildung Gelernte nicht nachhaltig genug im Handlungsrepertoire der Teilnehmenden verankert wurde und keine ausreichende Motivation zum Einsatz der Sprachförderstrategien aufgebaut wurde. Zum anderen können interventionsunabhängige Faktoren wie z. B. fehlende Anwendungsmöglichkeiten in der Einrichtung und mangelnde Unterstützung durch Leitung und Team zu einem solchen Effekt führen. Interventionsabhängige Faktoren (z. B. die Durchführung von Auffrischungssitzungen) und interventionsunabhängige Faktoren (z. B. Unterstützung durch neue Einrichtungsleitung) können zu einem Stabilisierungseffekt und einem Entwicklungsschub beitragen.

## Fragestellung

Die vorliegende Studie<sup>1</sup> geht folgenden beiden Fragen nach:

### Fragestellung 1

Gibt es langfristig mittlere Unterschiede in der Anregungsqualität der pädagogischen Fachkräfte je nachdem, ob sie durch die Fortbildung „Mit Kindern im Gespräch“ (Interventionsgruppen) oder „Sprache – Schlüssel zur Welt“ (Vergleichsgruppe) weitergebildet wurden?

Es wird zum einen erwartet, dass die pädagogischen Fachkräfte in der Anregungsqualität nach der Fortbildung (Post- und Follow-up-Test) im Mittel signifikant besser abschneiden als zuvor und dass es zwischen Post- und Follow-up-Test keinen signifikanten Unterschied gibt (Hypothese 1). Außerdem wird angenommen, dass die pädagogischen Fachkräfte, die an der Fortbildung „Mit Kindern im Gespräch“ (Interventionsgruppe) teilgenommen haben, zum Posttest und zum Follow-up-Test über eine höhere Anregungsqualität verfügen als diejenigen, die in der Fortbildung „Sprache – Schlüssel zur Welt“ (Vergleichsgruppe) ihre Qualifikation als Sprachförderkraft erworben haben (Hypothese 2). Des Weiteren wird angenommen, dass sich die Teilnehmenden der Interventionsgruppen in der Anregungsqualität stärker und nachhaltiger verbessern als die in der Vergleichsgruppe (Hypothese 3). Zwischen Post- und Follow-up-Test wird erwartet, dass die Anregungsqualität in den Interventionsgruppen nicht abnimmt (Stabilisierungseffekt oder Entwicklungsschub; Hypothese 4).

Für die erste Hypothese, dass sich die Teilnehmenden aller Fortbildungsgruppen verbessern, spricht die umfangreiche Auseinandersetzung mit dem Thema Sprachförderung. Unterschiede zwischen den Fortbildungsgruppen werden erwartet, weil davon ausgegangen wird, dass der fokussierte Ansatz effektiver für den Erwerb eines ho-

<sup>1</sup> Das Forschungsprojekt wurde von der Deutschen Forschungsgemeinschaft (DFG) gefördert (DFG-KA 16 57/4-1).

hen Anregungsniveaus ist (Hypothese 2). Die stärkere Verbesserung in der Anregungsqualität in „Mit Kindern im Gespräch“ (Hypothese 3) wird vor allem auf deren inhaltliche und methodische Fokussierung zurückgeführt. Für einen Stabilisierungseffekt (Hypothese 4) in den Interventionsgruppen sprechen vor allem die Auffrischungssitzungen, die im Jahr nach der Fortbildung stattfinden. In der Vergleichsgruppe könnte aufgrund der Auffrischungssitzungen die Anregungsqualität ebenfalls stabil bleiben. Da das dort erworbene Fachwissen im Laufe der Zeit trotz Auffrischungssitzungen ggf. eher vergessen wird, kann es jedoch auch zu einem Fading-Effekt kommen.

## Fragestellung 2

Gibt es einen signifikanten Unterschied in der Entwicklung der Anregungsqualität (vom Präbis zum Follow-up-Test) zwischen Interventions- und Vergleichsgruppe bei den Teilnehmenden, die zum Prätest unter dem Schwellenwert von 3.25 in der Anregungsqualität liegen?

Die Antwort auf diese Frage ist für die Einschätzung der praktischen Bedeutsamkeit der erzielten Wirkungen der Fortbildungen relevant. Praktisch bedeutsam ist eine Fortbildung, wenn sich die durch sie erreichten Verbesserungen in einem Bereich oberhalb einer Schwelle befinden, ab der höhere Zusammenhänge mit der Entwicklung der Kinder aufgezeigt werden können als unterhalb dieser Schwelle.

Es wird erwartet, dass sich die Entwicklung der Teilnehmenden, die im Prätest in der Anregungsqualität unter dem Schwellenwert 3.25 liegen, in Interventions- und Kontrollgruppe signifikant unterscheidet: In der Interventionsgruppe verbessern sich signifikant mehr pädagogische Fachkräfte so stark, dass sie zum Follow-up-Test über dem Schwellenwert 3.25 liegen als in der Vergleichsgruppe (Hypothese 5).

Analog zur Begründung von Hypothese 3 wird davon ausgegangen, dass die inhaltliche und methodische Fokussierung dazu führt, dass sich mehr pädagogische Fachkräfte in der Anregungsqualität stark genug verbessern, sodass sie den Schwellenwert von 3.25 überschreiten.

## Methode

### Design

Zur Beantwortung der Fragestellungen wurde eine experimentelle Prä-Post-Follow-up-Interventionsstudie im Feld durchgeführt mit randomisierter Zuordnung der pädagogischen Fachkräfte zu einer Vergleichsgruppe (Gruppe A) und zu zwei Interventionsgruppen (Gruppe B1 und B2). In allen Gruppen fanden zwischen Prä- und Posttest (ca. 10 Monate) neun ganztägige Veranstaltungen im Abstand von ca. einem Monat statt. Zwischen Posttest und Follow-up fanden in einem Zeitraum von 12 Monaten vier halbtägige Auffrischungssitzungen zur Sicherung der Nachhaltigkeit statt. In Tabelle 1 wird die Konzeption der beiden Fortbildungen gegenübergestellt, die in der vorliegenden Studie verglichen werden.

### Fortbildungskonzept

Der Vergleich der beiden Fortbildungskonzepte in Tabelle 1 zeigt, dass diese im Hinblick auf die Zielgruppe, bestimmte Merkmale der Fortbildnerin, das Ziel, ein Zertifikat als Sprachförderkraft zu erwerben, und im Hinblick auf Strukturmerkmale vergleichbar sind. Zum anderen wird erkennbar, dass der zentrale Unterschied darin liegt, dass das Konzept in der Vergleichsgruppe inhaltlich breit angelegt und jenes in den Interventionsgruppen inhaltlich fokussiert ist. Außerdem unterscheiden sich die beiden Konzepte im Hinblick auf die eingesetzte Methode. Im Gegensatz zur Vergleichsgruppe, in der es kein spezifisches methodisches Konzept gibt, werden in den Interventionsgruppen die Fortbildungen nach dem Konzept des Situiereten Lernens durchgeführt.

Grundlage für die Fokussierung auf Sprachförderstrategien in den Interventionsgruppen (Kammermeyer et al., 2017) sind die Entwicklungstheorie von Vygotsky (1987, vgl. Bodrova & Leong, 2007), die Distanzierungstheorie nach Sigel (u. a. 2002), die Analytische Methode und Ankermethode nach Silverman (2007) und die Methode *Text talk* von Beck und McKeown (2001). Eine weitere Basis (v. a. im Hinblick auf die Strukturierung der Anregungsqualität) stellt das Konzept von Pianta, auf dem die CLASS beruht, dar. Die methodische Fokussierung in den Interventionsgruppen basiert auf dem Konzept des Situiereten Lernens (*situated learning*); (Stark & Klauer, 2018), zentrale Methode ist die Videoanalyse (Blomberg, Sherin, Renkl, Glogger & Seidel, 2014; Sherin & van Es, 2009), zuerst mit fremden, dann mit eigenen Videos (Kleinknecht, Schneider & Syring, 2014).

Tab. 1: Vergleich der beiden Fortbildungen

Konzept	„Sprache – Schlüssel zur Welt“ zur Förderung der Sprache (Kath. Erwachsenenbildung Rheinland-Pfalz, 2009)	„Mit Kindern im Gespräch“ zur Förderung von Sprechen und Denken (Kammermeyer et al., 2017)
	A	B1
	Vergleichsgruppe	Interventionsgruppe
	Pädagogische Fachkräfte (v. a. Erzieherinnen und Erzieher)	
Zielgruppe		
Inhalt	<p>Inhaltlich breit angelegt, bezieht sich auf</p> <ol style="list-style-type: none"> <li>(1) Leitideen</li> <li>(2) Rahmenbedingungen und Querschnittsaufgaben der frühkindlichen Pädagogik für die sprachliche Förderung in Kitas</li> <li>(3) Wahrnehmung und Beschreibung kindlicher Sprachentwicklungsprozesse</li> <li>(4) Beobachtung und Dokumentation von Sprache und Sprachentwicklung</li> <li>(5) Methodisch-didaktisches Wissen für die Sprachförderung</li> <li>(6) Konzeption und Durchführung von Sprachförderheiten</li> <li>(7) Dokumentation, Auswertung und Weiterentwicklung von Sprachförderheiten</li> <li>(8) Unter Dreijährige</li> <li>(9) Reflexions- bzw. Präsentationstag</li> </ol>	<p>Inhaltlich fokussiert angelegt, bezieht sich auf</p> <p>I Sprachförderstrategien</p> <ol style="list-style-type: none"> <li>(1) Frage- und Modellierungsstrategien: z. B. offene Fragen stellen, handlungs- und denkbegleitendes Sprechen</li> <li>(2) Strategien zur Konzeptentwicklung, z. B. zum Herstellen von Zusammenhängen anregen, vergleichen</li> <li>(3) Rückmeldestrategien, z. B. korrekatives Feedback, erweitern, infrage stellen, Denken des Kindes sichtbar machen</li> </ol> <p>II Anwendung in Schlüssel-situationen</p> <ol style="list-style-type: none"> <li>(4) Lesesituationen</li> <li>(5) Routinesituationen</li> <li>(6) Gezielte Aktivitäten</li> <li>(7) Symbol- und Rollenspiele</li> <li>(8) Spontane Sprechanlässe</li> <li>(9) Markt der Möglichkeiten – Reflexion eigener Videos</li> </ol>
Ziel	<p>Vermittlung von Fachwissen, fachdidaktischem Wissen und Handlungskompetenzen</p>	<p>Vermittlung von Handlungskompetenzen</p>
	Zertifikat zur Sprachförderkraft	
	Zwei Fortbilderinnen zu gleichen Teilen	
Fortbildnerin	<p>Zwei Fortbildnerinnen mit langjähriger Erfahrung mit dem Fortbildungskonzept, vom kooperierenden Fortbildungsträger als besonders kompetent und erfolgreich empfohlen</p>	<p>Zwei wissenschaftliche Mitarbeiterinnen, die an der Entwicklung des Konzepts beteiligt waren</p>
Methode	<p>Übliche Methoden der Erwachsenenbildung, v. a. Vortrag und Diskussion</p>	<p>Keine Erfahrung</p> <p>Erfahrung mit Gruppe B1</p>
	Situierendes Lernen, z. B. Aufgreifen authentischer Situationen u. a. mithilfe von Videoanalysen	
Struktur	<p>Neun Tage Fortbildung ganztägig in monatlichem Abstand im Laufe eines Kita-Jahres sowie vier halbtägige Auffrischungssitzungen im Jahr nach der Fortbildung</p> <p>Ergänzend zu den Präsenzveranstaltungen gibt es schriftliche Selbstlernmaterialien zu jedem Modul mit Lernzielen und Aufgaben (und deren Lösungen) und Diskussionsanregungen, die von den Teilnehmenden zur Vorbereitung und zur Verarbeitung der Fortbildungssitzungen in Kleingruppen bearbeitet werden sollen. In Kleingruppen eine Förderinheit geplant, durchgeführt und präsentiert.</p>	<p>Zwei einstündige individuelle Coachingsitzungen mit einer eigenen video-graphierten Förder-situation durchgeführt</p>

Bei der Vergleichsgruppe handelt es sich um eine Fortbildung auf der Grundlage eines Curriculums, das seit 2008 in Rheinland-Pfalz verpflichtend für die Qualifizierung von Sprachförderkräften durchgeführt wird.

### Stichprobe

Die Stichprobe wurde rekrutiert aus den pädagogischen Fachkräften, die sich bei den größten regionalen Fortbildungsanbietern in Rheinland-Pfalz für die reguläre Fortbildungsmaßnahme zur Sprachförderkraft angemeldet hatten. Alle pädagogischen Fachkräfte erklärten sich im Vorfeld bereit, sowohl an der regulären als auch an der alternativen *fokussierten* Fortbildungsmaßnahme, die zum gleichen Zertifikat führt, teilzunehmen. Sie wurden dann per Zufall auf zwei Interventionsgruppen und eine Vergleichsgruppe aufgeteilt (vgl. Tab. 2).

Interventions- und Vergleichsgruppe unterscheiden sich nicht wesentlich hinsichtlich zentraler Personenvariablen. Die Schwundrate beträgt insgesamt 12%.

### Variablen und Erhebungsmethoden

#### Abhängige Variable Interaktionsqualität

Zur Erfassung der Interaktionsqualität wurde das Classroom Assessment Scoring System (CLASS) von Pianta, La Paro et al. (2008) eingesetzt. Hierbei handelt es sich um ein Instrument, dessen Struktur nicht nur in den USA, sondern auch in Deutschland sowohl für die additive Sprachförderung (Stuck et al., 2016) als auch für die alltagsintegrierte Förderung (von Suchodoletz et al., 2014) sowohl in der Live-

Beobachtung als auch in der Videobeobachtung belegt wurde. Aus diesem Grund kann davon ausgegangen werden, dass dieses in den USA entwickelte Instrument auch für den Einsatz in Deutschland geeignet ist. Die CLASS wurde zudem ausgewählt, weil es in Deutschland kein Erhebungsinstrument gibt, das in gleicher Weise auf die Interaktionsqualität fokussiert ist und dessen Güte entsprechend belegt ist (vgl. hierzu auch den Überblick über den Stand der Forschung zur Validität der drei am häufigsten eingesetzten Beobachtungsinstrumente von Mayer & Beckh, 2018): Nach Mayer und Beckh (2018) ist die Kindergarten-Einschätzskala KES-R (Tietze, Schuster, Grenner & Roßbach, 2007) im Vergleich zur CLASS sehr breit angelegt und erfasst nur in zwei Skalen die hier interessierende Anregungsqualität, die Caregiver Interaction Scale CIS (Arnett, 1989), differenziert zu wenig im mittleren Bereich der Interaktionsqualität. In Deutschland entwickelte Instrumente, wie die Dortmunder Ratingskala DO-RESI (Fried & Briedigkeit, 2008) und das Beobachtungsinstrument GiNA (Weltzien, 2013) sind nicht annähernd so gut untersucht wie die CLASS und international nicht anschlussfähig. Insofern stellt die CLASS derzeit aus unserer Sicht das beste Erhebungsinstrument zur Erfassung der Interaktions- und Anregungsqualität dar.

Zur theoretischen Begründung der CLASS kann auf das „Teaching through Interactions Framework“ (TTI) von Hamre et al. (2013) verwiesen werden. Für die Struktur der CLASS liegen Ergebnisse aus 4341 Einrichtungen im Elementar- und Primarbereich vor, die zeigen, dass die Erzieherin-Kind- bzw. Lehrkraft-Schüler-Interaktion aus den drei Domänen *Emotionale Unterstützung*, *Organisation der Lernsituation* und *Anregungsqualität* besteht und dass dieses

Tab. 2: Stichprobe

	Vergleichsgruppe A	Interventionsgruppe B1	Interventionsgruppe B2	gesamt
Teilnehmer (Prätest)	18	16	16	50
Teilnehmer (Posttest)	15	16	16	47
Teilnehmer (Follow-up-Test)	12	16	15	43
Geschlecht (Frauen/Männer)	18/0	16/0	15/1	49/1
Qualifikation Erzieherin – Fachschule (geringer als Erzieherin/Erzieherin – Hochschule/fehlend)	14 (1/2/1)	15 (1/0/0)	14 (2/0/0)	43 (3/3/1)
Eigene Mehrsprachigkeit (Anzahl)	1	0	3	4

Anmerkung: Sofern nicht anders angegeben, beziehen sich die Aussagen auf den Zeitpunkt vor der Fortbildung.

Tab. 3: Domänen und Qualitätsdimensionen des Classroom Assessment Scoring Systems (CLASS)

Domäne (domains)	Qualitätsdimension (dimension)	Beschreibung
Emotionale Unterstützung ( <i>Emotional Support</i> )	Positives Klima	Ausmaß der gegenseitigen emotionalen Verbundenheit, Respekt und Äußerungen positiver Gefühle zwischen Erzieherinnen und/oder Kindern
	Negatives Klima	Ausmaß negativer Emotionen, wie z. B. Verärgerung, Feindseligkeit oder Aggression zwischen Erzieherinnen und/oder Kindern
	Sensibilität der Erzieherin	Ausmaß des bewussten Wahrnehmens und Reagierens auf lernbezogene und emotionale Belange der Kinder
	Orientierung am Kind	Ausmaß der Berücksichtigung von Interessen, Motivation und Sichtweisen der Kinder in der sozialen Interaktion oder im Lernprozess
Organisation der Lernsituation ( <i>Classroom Organization</i> )	Verhaltensmanagement	Ausmaß der Effizienz, mit welcher Verhaltensweisen der Kinder gesteuert, verhindert und korrigiert werden
	Produktivität	Ausmaß der effektiven Organisation von Routineabläufen, Aktivitäten und Anleitungen, die gewährleistet, dass ein Maximum der verfügbaren Zeit für Lernaktivitäten genutzt werden kann
	Lern-Arrangement	Ausmaß der Unterstützung bezüglich der Lernaktivitäten sowie des zur Verfügung gestellten stimulierenden Materials, das einen aktiven Miteinbezug der Kinder in den Lernprozess und ein Maximum an Lerngelegenheiten ermöglicht
Anregungsqualität ( <i>Instructional Support</i> )	Sprachbildung	Ausmaß der Unterstützung und Förderung der Sprachfähigkeiten der Kinder
	Konzeptentwicklung	Ausmaß pädagogisch-didaktischer Diskussionen und Aktivitäten zur Unterstützung komplexerer Denkprozesse der Kinder
	Feedbackqualität	Ausmaß der Förderung der Lernprozesse der Kinder durch Antworten, Reaktionen und Kommentare auf kindliche Ideen und Arbeitsprodukte

Dreifaktorenmodell einen besseren Fit aufweist als ein Ein- oder Zweifaktorenmodell (Hamre et al., 2013, Hamre, Hatfield, Pianta & Jamil, 2014). Einen Überblick über die Domänen und Dimensionen der CLASS, die auf einer siebenstufigen Ratingskala gemessen werden, gibt Tabelle 3.

Die Einschätzung der Qualitätsdimensionen mit dem Classroom Assessment Scoring System (CLASS) auf der Grundlage von Videografien wurde von Ratern übernommen, die an einer CLASS-Schulung gemäß der Vorgaben der Entwicklungsfirma Teachstone<sup>2</sup> teilgenommen hatten. Sie verfügten zum Zeitpunkt der Einschätzungen über eine aktuell be-

standene Reliabilitätsprüfung. Die Einschätzungen bezogen sich zu jedem Erhebungszeitpunkt pro pädagogische Fachkraft jeweils auf eine Sprachfördereinheit von 20 Minuten Länge. Das Raten erfolgte verblindet, d. h. die beiden Rater wussten weder, ob es sich um Videos der Prä-, Post- oder Follow-up-Erhebung, noch ob es sich um Videos der Interventions- oder Vergleichsgruppe handelte. 20 % der Videos wurden von zwei Ratern eingeschätzt, bei

<sup>2</sup> <https://teachstone.com/trainings>. letzter Zugriff am 23.7.2019

allen Dimensionen wurde der von den CLASS-Autoren geforderte Standard, bei mindestens 80 % der Ratings höchstens ein Punktwert Abweichung, eingehalten. Insgesamt gab es bei 95,4 % der Ratings Übereinstimmung. Dieser hohe Wert wurde erreicht, weil es für die Rater – wie von Teachstone empfohlen – zusätzliche Kalibrierungssitzungen mit Übungsvideos gab.

### Ablauf der Untersuchung

Vor der Fortbildung (Prätest: September bis November 2014) wurde pro pädagogische Fachkraft eine Fördereinheit videografiert, um die Interaktionsqualität vor der Fortbildung zu erfassen und überprüfen zu können, ob sich die pädagogischen Fachkräfte in Interventions- und Vergleichsgruppe unterscheiden. Am Ende der Fortbildung (Posttest: Mai bis Juli 2015) wurden erneut Videos in Fördersituationen aufgenommen. Der Follow-up-Test fand ein Jahr später statt (Mai bis Juni 2016). Bei den Videoaufnahmen konnte die pädagogische Fachkraft eine Sprachförderaktivität mit durchschnittlich fünf Kindern frei wählen. Die Fortbildungen fanden etwa einmal monatlich statt, die Fortbildungen der Interventionsgruppe B1 und B2 wurden jeweils an zwei aufeinanderfolgenden Tagen durchgeführt, Gruppe B1 jeweils einen Tag vor Gruppe B2. Zwischen Posttest und Follow-up gab es vier halbtägige Auffrischungssitzungen.

### Auswertungsverfahren

Als abhängige Variable wurde der CLASS-Wert der Domäne *Anregungsqualität* verwendet, der sich aus dem Mittelwert der drei Dimensionen *Sprachbildung*, *Konzeptentwicklung* und *Feedbackqualität* zusammensetzt. Die Voraussetzungen für diese Mittelwertbildung waren gegeben, da sich die Dreifaktorenstruktur der CLASS auch in Deutschland (Stuck et al., 2016) empirisch (trotz nicht idealem Fit) als die beste im Vergleich zur Ein- und Zweifaktorenstruktur herausgestellt hat. Zum anderen reiht sich die Studie auf diese Weise in die Tradition der CLASS-Studien ein. Hinzu kommt, dass es sich bei der Domäne *Anregungsqualität* über verschiedene Studien hinweg um die Domäne handelt, die die geringsten Interkorrelationen zu den anderen Domänen aufweist (vgl. Hamre et al., 2014; Hamre et al., 2013; Stuck et al., 2016). In exploratorischen Faktorenanalysen wurde mit dem vorliegenden Da-

tenzatz zudem zu jedem Zeitpunkt eine einfaktorielle Struktur über die drei genannten Dimensionen hinweg mit wenig variierenden Faktorladungen gefunden.

Zur Beantwortung der ersten Fragestellung wurde eine zweifaktorielle  $3 \times 3$  Varianzanalyse mit Messwiederholung (Vergleichsgruppe A vs. Interventionsgruppe B1 vs. Interventionsgruppe B2 und Prätest vs. Posttest vs. Follow-up-Test) mit der abhängigen Variablen *Anregungsqualität (instructional support)* gerechnet. Es wurden geplante Helmert-Kontraste (Gruppe A vs. Gruppe B1/B2; Gruppe B1 vs. Gruppe B2) und ( $t_1$  vs.  $t_2/t_3$ ;  $t_2$  vs.  $t_3$ ) durchgeführt. Zur Kontrolle der gefundenen Unterschiede im Prätest wurde ergänzend eine Kovarianzanalyse durchgeführt.

In der zweiten Fragestellung ging es darum, wie viele pädagogische Fachkräfte ihre *Anregungsqualität* in der Richtung verändern, dass sie vor der Fortbildung unter und ein Jahr nach Ende der Fortbildung über einem bestimmten Schwellenwert liegen. Der Schwellenwert wurde a priori auf 3.25 festgesetzt. Hierbei handelt es sich um den höchsten in der Literatur diskutierten Schwellenwert, für den Unterschiede in der Prädiktionskraft von der *Anregungsqualität* auf die sprachliche Entwicklung des Kindes gefunden wurden (Burchinal et al., 2010). Zur Prüfung der Signifikanz wurde aufgrund der kleinen Zellenbesetzung der exakte Test nach Fisher verwendet.

## Ergebnisse

### Deskriptive Ergebnisse

In Tabelle 4 sind die deskriptiven Ergebnisse der abhängigen Variablen *Anregungsqualität* und ihrer Dimensionen *Sprachbildung*, *Konzeptentwicklung* und *Feedbackqualität*, die mit dem Classroom Assessment Scoring System (CLASS) von Pianta, La Paro et al. (2008) gemessen wurden, zu den drei Messzeitpunkten dargestellt.

Der deskriptive Vergleich zeigt, dass die Vergleichsgruppe A und die Interventionsgruppe B2 mit gleichen Ausgangswerten starten und dass Interventionsgruppe B1 höhere Werte als die beiden anderen Gruppen in der *Anregungsqualität* aufweist. Die Werte zum Follow-up sind bei beiden Interventionsgruppen um mehr als einen Skalenpunkt (auf einer Skala von 1 bis 7) höher als in der Vergleichsgruppe.

Tab. 4: Anregungsqualität – Deskriptive Werte von Vergleichsgruppe A und den beiden Interventionsgruppen B1 und B2

	Vergleichsgruppe (A) M (SD) N = 12			Interventionsgruppe (B1) M (SD) N = 16			Interventionsgruppe (B2) M (SD) N = 15		
	Prätest	Posttest	Follow-up	Prätest	Posttest	Follow-up	Prätest	Posttest	Follow-up
Sprachbildung	2.21 (1.20)	2.46 (1.16)	2.71 (1.05)	3.41 (1.23)	4.31 (1.09)	4.38 (1.01)	2.23 (0.94)	3.90 (1.65)	4.57 (1.29)
Konzeptentwicklung	1.54 (0.72)	1.83 (0.83)	2.21 (0.58)	2.28 (0.68)	3.44 (0.98)	3.03 (0.78)	1.50 (0.63)	3.23 (1.66)	3.47 (1.20)
Feedbackqualität	1.67 (0.98)	2.00 (0.85)	2.17 (1.11)	2.62 (1.02)	3.75 (1.47)	3.63 (1.19)	2.00 (1.07)	3.50 (1.48)	4.00 (1.27)
Anregungsqualität	1.81 (.92)	2.10 (0.87)	2.36 (0.86)	2.77 (.87)	3.83 (1.10)	3.68 (0.88)	1.91 (.78)	3.54 (1.53)	4.01 (1.12)

### Inferenzstatistische Absicherung der Unterschiede im Prätest

Trotz Randomisierung unterscheiden sich die drei Gruppen bereits vor Beginn der Fortbildungen signifikant in der Anregungsqualität ( $F[2,40] = 5.751, p < .05$ ; Posthoc: A vs. B1:  $p < .05$ ; B1 vs. B2:  $p < .05$ ; A vs. B2:  $p > .05$ ).

### Fragestellung 1: Gibt es langfristige mittlere Unterschiede in der Anregungsqualität zwischen den beiden Fortbildungen „Mit Kindern im Gespräch“?

Überprüfung der Voraussetzungen für die messwiederholte Varianzanalyse:

Die Überprüfung der Normalverteilung ergibt für Gruppen und Messzeitpunkte getrennt betrachtet nur eine signifikante Abweichung, und zwar im Prätest-Wert der Interventionsgruppe B2 (Kolmogorov-Smirnov-Test:  $D[15] = .255, p = .01$ ).

Von der Gleichheit der Fehlervarianzen kann nach Levene ausgegangen werden (Prätest:  $F[2,40] = 0.367, p > .05$ ; Posttest:  $F(2;40) = 3.188, p > .05$ ; Follow-up-Test:  $F[2,40] = 1.670, p > .05$ ). Auch die Homogenität der Kovarianzmatrizen gemäß dem Box-Test ist gegeben ( $p > .05$ ). Da eine Verletzung der Voraussetzung der Sphärizität vorliegt, wird eine Huynh-Feldt-Korrektur der Freiheitsgrade vorgenommen (Mauchly-W [2] = .841,  $p < .05, \epsilon = 0.943$ ).

Die messwiederholte Varianzanalyse ergibt einen signifikanten Haupteffekt für den Innersubjektfaktor Zeit mit  $F(1,89, 75.43) = 20.43, p < .001$  und einer Effektstärke von  $\eta^2_p = .34$ . Die a priori geplanten Helmert-Kontraste zeigen an, dass es signifikante Unterschiede zwischen dem Prätest und den beiden Messzeitpunkten nach der Intervention (Posttest und Follow-up-Test) gibt:  $F(1;40) = 52.86, p < .001$  und eine Effektstärke von  $\eta^2_p = .57$ , allerdings gibt es keine signifikanten Unterschiede mehr zwischen Post-Test und Follow-up-Test:  $F(1,40) = 0.74, p > .05$ .

Auch der Haupteffekt, der die verschiedenen Fortbildungen vergleicht, wurde mit  $F(2,40) = 13.52, p < .001$  signifikant und ist mit

einer Effektstärke von  $\eta^2_p = .40$  sogar etwas größer als der Effekt der Zeit. Die ebenfalls durchgeführten Helmert-Kontraste vergleichen hier zunächst die Vergleichsgruppe A mit den beiden Interventionsgruppen (B1/B2) und zeigen eine geschätzte mittlere Differenz von  $\Delta = -1.20$  ( $p < .001$ ). Das heißt, die Vergleichsgruppe wird in Bezug auf die Anregungsqualität im Schnitt gut einen Skalenwert geringer eingeschätzt als die beiden Interventionsgruppen. Die beiden Interventionsgruppen wiederum unterscheiden sich bei einem Kontrastschätzer  $\Delta = 0.27$  ( $p > .05$ ) nicht signifikant voneinander.

Von besonderem Interesse ist nun der Interaktionseffekt, auch dieser wird mit  $F(3,77, 75.43) = 3.38$ ,  $p < .05$  und einer Effektstärke von  $\eta^2_p = .14$  signifikant. Betrachtet man nun die Interaktionskontraste, so zeigt sich das gleiche Bild wie bei dem messwiederholten Faktor: Ein signifikanter Interaktionseffekt für den Vergleich des Prätests mit den Messzeitpunkten nach der Intervention ( $F[2,40] = 7.58$ ,  $p < .05$ ,  $\eta^2_p = .28$ ), allerdings keine signifikanten Interaktionseffekte mehr zwischen Post- und Follow-up-Messung ( $F[2,40] = 0.75$ ,  $p > .05$ ). Dass die Interaktion zwischen Fortbildungsgruppe und Messzeitpunkt signifikant wird, bedeutet, dass sich die pädagogischen Fachkräfte in der neu konzipierten Fortbildung „Mit Kindern im Ge-

spräch“ signifikant stärker verbessert haben als diejenigen in der etablierten „Sprache – Schlüssel zur Welt“.

Die beschriebene Entwicklung der Anregungsqualität der pädagogischen Fachkräfte in den drei Gruppen zeigt Abbildung 1.

Es gibt signifikante mittlere Differenzen von über einem Skalenwert zum Post- und Follow-up-Test, beim Posttest zwischen A und B1 ( $\Delta = 1.74$ ,  $p < .001$ ), zwischen A und B2 ( $\Delta = 1.45$ ,  $p < .05$ ), jedoch nicht zwischen B1 und B2 ( $\Delta = 0.29$ ,  $p > .05$ ). Beim Follow-up-Test finden sich ebenfalls signifikante mittlere Differenzen zwischen A und B1 ( $\Delta = 1.32$ ,  $p < .001$ ), zwischen A und B2 ( $\Delta = 1.65$ ,  $p < .001$ ), jedoch nicht zwischen B1 und B2 ( $\Delta = 0.33$ ,  $p > .05$ ).

Um auszuschließen, dass die gefundenen Effekte auf die nicht funktionierende Randomisierung, also bereits auf Unterschiede vor der Intervention, zurückzuführen sind, wird in einer ergänzenden Kovarianzanalyse für die ungleichen Prätestwerte kontrolliert. Die Voraussetzungen für die Kovarianzanalyse sind für den Posttest als Kriterium nicht, für den Follow-up-Test hinreichend erfüllt:

Die Kovariate Anregungsqualität-Prätest und abhängige Variable Anregungsqualität-Follow-up-Test kovariieren signifikant ( $r = .38$ ,  $p < .05$ ), die Gleichheit der Fehlervarianzen ist

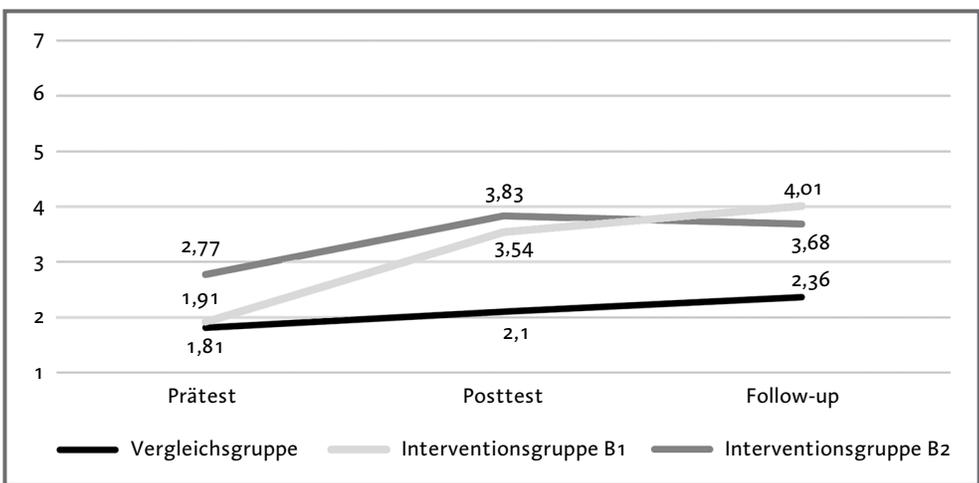


Abb. 1: Entwicklung der Anregungsqualität zwischen Prätest und Follow-up.

gegeben (Levene-Test:  $F(2,40) = 0.80, p > .05$ ) und die Interaktion zwischen Gruppenfaktor und Kovariate ist nicht signifikant unterschiedlich ( $F(2,37) = 0.86, p > .05$ ).

Die Ergebnisse der Kovarianzanalyse zeigen, dass auch über den signifikanten Effekt des Prätests als Kovariate hinaus ( $F[1,39] = 37.78, p < .001, \eta^2_p = .15$ ) Unterschiede zwischen den Gruppen bestehen ( $F[3,39] = 10.48, p < .001, \eta^2_p = .35$ ), und dass dieser Effekt sogar noch größer ist. Erneut wird ein geplanter Helmert-Kontrast für die Gruppenunterschiede gerechnet, welcher anhand des Prätestwerts adjustierte mittlere Differenzen von  $\Delta = -1.25$  mit  $p < .001$  als Kontrastschätzer zwischen A und B1/B2 und  $\Delta = -0.72$  mit  $p > .05$  zwischen B1 und B2 angibt. Damit sind die Ergebnisse der Varianzanalyse auch unter Berücksichtigung der unterschiedlichen Ausgangswerte in den Gruppen bestätigt.

Zusammenfassend kann festgehalten werden, dass die pädagogischen Fachkräfte, wie in Hypothese 1 angenommen, in der Anregungsqualität nach den Fortbildungen (Post- und Follow-up-Test) im Mittel signifikant besser

abschneiden als zuvor. Zwischen Post- und Follow-up-Test kommt es zu einer Stabilisierung, der Unterschied wird nicht mehr signifikant. Ebenfalls wie erwartet (Hypothese 2) gibt es auch signifikante Unterschiede zwischen den beiden Fortbildungen. Die pädagogischen Fachkräfte, die in der Interventionsgruppe fortgebildet wurden, verfügen zum Posttest und zum Follow-up-Test über eine signifikant höhere Anregungsqualität als diejenigen, die in der Vergleichsgruppe ihre Qualifikation als Sprachförderkraft erworben haben. Besonders bedeutsam ist das Ergebnis, dass sich die Teilnehmenden an „Mit Kindern im Gespräch“ zwischen Prä- und Posttest signifikant stärker (um ca. einen CLASS-Skalenwert) in der Anregungsqualität verbessern als die Teilnehmenden an der Fortbildung „Sprache – Schlüssel zur Welt“. Die Teilnehmenden der Interventionsgruppen verbessern sich in der Anregungsqualität stärker als die in der Vergleichsgruppe (Hypothese 3). Zwischen Post- und Follow-up-Test gibt es keinen Entwicklungsschub, sondern einen Stabilisierungseffekt (Hypothese 4).

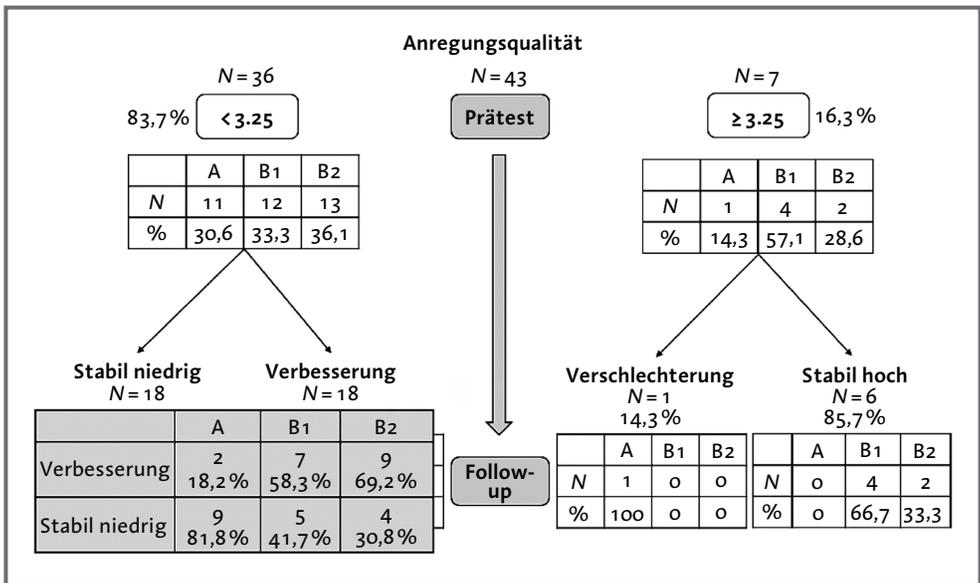


Abb. 2: Entwicklung der Anzahl der pädagogischen Fachkräfte in der Anregungsqualität vom Prätest zum Follow-up in Abhängigkeit vom Schwellenwert 3.25.

*Fragestellung 2: Gibt es einen signifikanten Unterschied in der Entwicklung der Anregungsqualität zwischen Interventions- und Vergleichsgruppe zwischen Prä- und Follow-up-Test bei den Teilnehmenden, die zum Prätest unter dem Schwellenwert von 3.25 in der Anregungsqualität liegen?*

Wie sich die Teilnehmenden in der Anregungsqualität in Abhängigkeit zum Schwellenwert von 3.25 entwickeln, zeigt Abbildung 2.

Vor Beginn der Fortbildung weisen 83 % der Teilnehmenden eine Anregungsqualität unterhalb des Schwellenwertes auf, von dem angenommen wird, dass es einen höheren Zusammenhang zwischen Verhalten der pädagogischen Fachkraft und der Entwicklung des Kindes gibt. Die drei Gruppen unterscheiden sich hierbei nicht. Ein Jahr nach der Fortbildung zum Follow-up haben sich 50 % dieser Teilnehmenden so stark verbessert, dass sie diese Schwelle überwinden. Der Unterschied zwischen den drei Fortbildungsgruppen wird – wie in Hypothese 5 angenommen – signifikant (exakter Test nach Fisher:  $p < .05$ ), mit einem *Cramers V* von .43 liegt die Effektstärke nach Cohen (1988) im mittleren bis großen Bereich.

Nur sieben Teilnehmende liegen bereits zum Prätest über dem Schwellenwert 3.25, sie verteilen sich unterschiedlich auf die drei Gruppen. Die einzige pädagogische Fachkraft, die sich verschlechtert, stammt aus Gruppe A.

## Diskussion

Die Frage, ob es langfristig mittlere Unterschiede in der Anregungsqualität der pädagogischen Fachkräfte gibt, je nachdem, ob sie durch die Fortbildung „Mit Kindern im Gespräch“ (Interventionsgruppen) oder „Sprache – Schlüssel zur Welt“ (Vergleichsgruppe) weitergebildet wurden, kann aufgrund der experimentellen Feldstudie klar beantwortet werden. Nach der Fortbildung erreichten zwar alle fortgebildeten pädagogischen Fachkräfte signifikant höhere Werte in der Anregungsqualität, die Teilneh-

menden in den beiden Interventionsgruppen erzielten jedoch signifikant höhere Werte als die in der Vergleichsgruppe. Wie erwartet, gab es keinen signifikanten Unterschied zwischen den Gruppen B1 und B2. Der fehlende signifikante Unterschied zwischen Post- und Follow-up-Test kann als Stabilisierungseffekt interpretiert werden.

In der Fortbildung „Mit Kindern im Gespräch“ ist es somit gelungen, die nach der Fortbildung erreichte Anregungsqualität auch noch ein Jahr nach der Fortbildung auf einem stabilen Niveau zu halten. Es gab keinen Fading-Effekt, jedoch auch keinen Entwicklungsschub wie in der Studie von Simon und Sachse (2011). Zu dieser Stabilität haben vermutlich vor allem die vier halbtägigen Auffrischungssitzungen im Jahr nach der Fortbildung beigetragen. In den beiden Interventionsgruppen wurden mit den Mittelwerten in der Anregungsqualität in Gruppe B1 ( $M = 3.68$ ) und in Gruppe B2 ( $M = 4.01$ ) Werte auf der siebenstufigen CLASS-Skala erreicht, die nach Burchinal et al. (2014) in Studien bisher kaum vorkommen. Die Differenz der Mittelwerte zwischen den Interventionsgruppen und der behandelten Vergleichsgruppe zum Follow-up-Test lag ein Jahr nach Beendigung der Fortbildung deutlich über einem Skalenpunkt (A vs. B1:  $\Delta = 1.32$ ; A vs. B2:  $\Delta = 1.65$ ), der Zuwachs zwischen Prätest und Follow-up betrug in der Vergleichsgruppe A ( $\Delta = 0.55$ ), in Interventionsgruppe B1 ( $\Delta = 0.91$ ) und in Interventionsgruppe B2 sogar über zwei Skalenpunkte ( $\Delta = 2.10$ ). Einen Hinweis zur Interpretation dieser Zuwächse geben Pianta, Mashburn et al. (2008). Im Diskussionsteil ihrer Arbeit weisen sie unter Bezug auf die Studie von Hamre und Pianta (2005) darauf hin, dass ein Zuwachs von einem Skalenpunkt auf der CLASS-Skala mit einem Gewinn an kindlichen Outcome-Maßen zwischen .3 und .5 Standardabweichungen in einem standardisierten Leistungstest in der ersten Klasse einhergeht. Es wird jedoch in der Studie von Hamre und Pianta (2005) weder dieses Ergebnis explizit benannt noch diskutiert, ob es relevant ist, an welcher Stelle auf der CLASS-Skala diese Veränderung stattfindet,

d. h. beispielsweise ob im Zusammenhang mit dem kindlichen Outcome eine Verbesserung von einem Wert von 1 auf 2 (auf der siebenstufigen Ratingskala der CLASS) mit einer Verbesserung von einem Wert von 3 auf 4 gleichbedeutend ist. Hiermit ist das Thema Schwellenwerte angesprochen, mit dem sich die Studien von Burchinal et al. (2010), Burchinal et al. (2014) sowie von Burchinal et al. (2016) beschäftigen. Die Anregungsqualität erwies sich z. B. in der Studie von Burchinal et al. (2010) über einem CLASS-Skalenwert von 3.25 als ein stärkerer Prädiktor für *expressive language* als ein Wert unter 3.25.

In der vorliegenden Studie gab es signifikant mehr pädagogische Fachkräfte in den Interventionsgruppen B1 und B2 als in der Vergleichsgruppe A, die sich über diesen Schwellenwert – auch langfristig ein Jahr nach Ende der Fortbildung – verbesserten. Dies ist ein besonders vielversprechendes Ergebnis, denn mit dieser Auswertung auf der Ebene des Verhaltens der pädagogischen Fachkräfte können erste Hinweise auf mögliche positive Zusammenhänge mit der sprachlichen Entwicklung abgeleitet werden, was letztendlich das eigentliche Ziel der Entwicklung der Fortbildung „Mit Kindern im Gespräch“ war. Ob es auch tatsächlich Wirkungen auf die sprachliche Entwicklung der Kinder gibt, werden weitere Analysen mit Kinddaten zeigen, die ebenfalls erhoben wurden.

Als Stärke der vorgestellten Studie kann vor allem das experimentelle Design herausgestellt werden, mit der Randomisierung erfüllt es den Goldstandard der Wirkungsforschung. Die Messung der abhängigen Variable mit dem Classroom Assessment Scoring System (CLASS) von Pianta, La Paro et al. (2008), das den Vergleich mit der internationalen Forschung erlaubt, kann als eine weitere Stärke der Studie angesehen werden. Die CLASS ist aus unserer Sicht aufgrund der umfangreichen Studien in den USA und der wenigen Studien in Deutschland derzeit das beste Erhebungsinstrument zur Erfassung der Interaktions- und der Anregungsqualität. Es gibt jedoch auch Bedenken zum Einsatz der CLASS. Mayer und Beckh (2018)

kritisieren, dass die hoch-inferente Struktur der Skala eine differenzierte Messung des latenten Konstrukts suggeriert, dass die Bewertung dann jedoch auf der Ebene der Dimensionen nur in ein oberflächliches, globales Gesamturteil, im Sinne von eher gut versus schlecht, resultiert. Die spezifischen Qualitätsaspekte wie die Angemessenheit werden zu wenig berücksichtigt. Mit dieser Kritik beziehen sich Mayer und Beckh (2018) jedoch weniger auf die CLASS an sich als vielmehr auf das generelle Problem hoch-inferenter Verfahren, da sie einen hohen Grad an Schlussfolgerungen von den Beobachtern erfordern, was die Validität des Verfahrens beeinträchtigen könnte. Dieses Problem ist aus unserer Sicht jedoch bei der CLASS in geringerem Maße als bei anderen hoch-inferenten Beobachtungsverfahren vorhanden, da die Qualität der Schlussfolgerungen durch eine anspruchsvolle Reliabilitätsprüfung gesichert wird. Um dies zu erreichen wurde in der vorliegenden Studie außerdem ein zusätzliches Kalibrierungsverfahren eingesetzt (Leber, Kammermeyer & Roux, 2019). Hinzu kommt, dass in der dargestellten Studie vergleichbare Ergebnisse auch mit einem niedrig-inferenten Verfahren erzielt wurden (Kammermeyer, Metz et al., im Druck).

Da sich die Struktur des Fortbildungskonzepts an der Struktur der CLASS orientiert, könnte auch die Gefahr eines Zirkelschlusses gesehen werden. Dem kann entgegengehalten werden, dass es sich bei der CLASS um ein Instrument mit ausreichender Programmferne handelt, da das Fortbildungskonzept auf der Basis verschiedener theoretischer und empirischer Erkenntnisse entwickelt wurde (Kammermeyer et al., 2017). Hinzu kommt, dass von den pädagogischen Fachkräften ein dreifacher Transfer ihrer erworbenen Kompetenzen geleistet werden muss (Hager & Hasselhorn, 2000). Sie stehen vor der Aufgabe, die gelernten Sprachförderstrategien über einen Zeitraum von einem Jahr aufrechtzuerhalten (*zeitlicher Transfer*) und diese in einer ökologisch validen Situation mit Kindern in der Kita (*Situationstransfer*) im Rahmen einer selbst gestalteten Fördersituation (*Anforderungstransfer*) anzuwenden.

## Limitationen

Bei der Interpretation der Befunde ist einschränkend zu berücksichtigen, dass es sich hier um eine kleine Studie mit nur drei Fortbildungsgruppen handelt, an der insgesamt lediglich 50 pädagogische Fachkräfte teilgenommen haben. Dadurch ist ein Rückschluss auf eine Grundgesamtheit schwierig. Eine größere Stichprobe ist jedoch bei einer so aufwendigen, experimentellen und gleichzeitig ökologisch validen, einjährigen Fortbildung sehr ressourcenintensiv und logistisch anspruchsvoll. Um die vorliegenden vielversprechenden Befunde auf eine breitere empirische Basis zu stellen, wäre es jedoch wünschenswert, hierin zu investieren.

Als weitere Limitation könnte die nicht gezielte Randomisierung bewertet werden. Die ergänzend durchgeführte Kovarianzanalyse zeigte jedoch, dass sich unter Kontrolle der Prätestwerte die gleichen Rückschlüsse ziehen lassen. Ein weiteres Problem lag darin, dass die Anregungsqualität zu jedem Messzeitpunkt nur einmal gemessen wurde. Praetorius (2014) hat in ihrer deutschen Studie zur Messung der Unterrichtsqualität durch Ratings (jedoch nicht mit CLASS) aufgezeigt, dass die Dimension kognitive Aktivierung starken kontextuellen Einflüssen ausgesetzt ist und deshalb über Unterrichtsstunden hinweg nur wenig stabil ist. Dies könnte auch für die Anregungsqualität in Kindertagesstätten zutreffen.

Als weiteres Problem könnte auch gesehen werden, dass die Fortbildungen in der Interventions- und in der Vergleichsgruppe nicht von den gleichen Fortbildnerinnen durchgeführt wurden. Die beiden Interventionsgruppen wurden von zwei Projektmitarbeiterinnen, die Vergleichsgruppe von zwei erfahrenen und als besonders kompetent empfohlenen, regulären Weiterbildnerinnen durchgeführt.

Das Fortbildungskonzept „Mit Kindern im Gespräch“ ist seit 2017 verpflichtendes Landeskonzept für die Qualifizierung von Sprachförderkräften in Rheinland-Pfalz. Den eigentlichen Praxistest besteht das Fortbildungskonzept erst dann, wenn auch reguläre Fortbildnerinnen die gewünschten langfristigen Wirkungen erzielen.

## Literatur

- Arnett, J. (1989). Caregivers in day-care centers: Does training matter? *Journal of Applied Developmental Psychology*, 10, 541–552. [https://dx.doi.org/10.1016/0193-3973\(89\)90026-9](https://dx.doi.org/10.1016/0193-3973(89)90026-9)
- Beck, I. & McKeown, M. G. (2001). Text Talk: Capturing benefits of read-aloud experiences for young children. *The Reading Teacher*, 55, 10–20.
- Beller, S. (2014). *ESLA-Erzieherfortbildung zur sprachlich-interaktiven Anregung: Evaluation langfristiger Effekte bei pädagogischen Fachkräften und Kindern*. Ort: Südwestdeutscher Verlag für Hochschulschriften.
- Birman, B. F., Desimone, L., Porter, A. C. & Garet, M. S. (2000). Designing professional development that works. *Educational Leadership*, 57(8), 28–33.
- Blomberg, G., Sherin, M. G., Renkl, A., Glogger, I. & Seidel, T. (2014). Understanding video as a tool for teacher education: investigating instructional strategies to promote reflection. *Instructional Science*, 42, 443–463. <https://dx.doi.org/10.1007/s11251-013-9281-6>
- Bodrova, E. & Leong, D. J. (2007). *Tools of the mind* (2nd ed.). Englewood, CO: Prentice Hall.
- Burchinal, M., Vandergrift, N., Pianta, R. & Mashburn, A. (2010). Threshold analysis of association between child care quality and child outcomes for low-income children in pre-kindergarten programs. *Early Childhood Research Quarterly*, 25, 166–176. <https://dx.doi.org/10.1016/j.ecresq.2009.10.004>
- Burchinal, M., Vernon-Feagans, L., Vitiello, V., Greenberg, M. & Family Life Project Key Investigators (2014). Thresholds in the association between child care quality and child outcomes in rural preschool children. *Early Childhood Research Quarterly*, 29, 41–51. <https://dx.doi.org/10.1016/j.ecresq.2013.09.004>
- Burchinal, M., Zaslow, M. & Tarullo, L. (Eds.). (2016). *Quality thresholds, features, and dosage in early care and education: secondary data analyses of child outcomes*. Monographs of the Society for Research in Child Development, 81 (Serial No. 81). [http://inid.gse.uci.edu/files/2011/03/MSRCD\\_ECE-Quality-Thresholds-Features-Dosage-final.pdf](http://inid.gse.uci.edu/files/2011/03/MSRCD_ECE-Quality-Thresholds-Features-Dosage-final.pdf), letzter Zugriff am 20.7.2019
- Buschmann, A. & Jooss, B. (2011). Alltagsintegrierte Sprachförderung in der Kinderkrippe. *Verhaltenstherapie und psychosoziale Praxis*, 43, 303–312.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: L. Erlbaum Associates.
- Early, D. M., Maxwell, K. L., Ponder, B. D. & Pan, Y. (2017). Improving teacher-child interactions: A randomized controlled trial of Making the Most of Classroom Interactions and MyTeachingPartner professional development models. *Early Childhood Research Quarterly*, 38, 57–70. <https://dx.doi.org/10.1016/j.ecresq.2016.08.005>
- Egert, F. (2015). *Meta-analysis on the impact of in-service professional development programs for preschool teachers on quality ratings and child outcomes*. Unveröffentlichte Dissertation, Universität Bamberg. Zugriff am 29.1.2018 unter <https://d-nb.info/1081935197/34>
- Egert, F., Eckhardt, A. G. & Fukkink, R. G. (2017). Zentrale Wirkmechanismen von Weiterbildungen zur Qualitätssteigerung in Kindertageseinrichtungen. *Frühe Bildung*, 6, 58–66. <https://dx.doi.org/10.1026/2191-9186/a000309>
- Fölling-Albers, M., Hartinger, A. & Mörtl-Hafizovic, D. (2004). Situierendes Lernen in der Lehrerbildung. *Zeitschrift für Pädagogik*, 50, 727–747.

- Fried, L. & Briedigkeit, E. (2008). *Sprachförderkompetenz: Selbst- und Teamqualifizierung für Erzieherinnen, Fachberatungen und Ausbilder*. Cornelsen Scriptor.
- Fukkink, R. G. & Lont, A. (2007). Does training matter? A meta-analysis and review of caregiver training studies. *Early Childhood Research Quarterly*, 22, 294–311. <https://dx.doi.org/10.1016/j.ecresq.2007.04.005>
- Hager, W. & Hasselhorn, M. (2000). Psychologische Interventionsmaßnahmen: Was sollen sie bewirken können? In W. Hager, J.-L. Patry & H. Brezing (Hrsg.), *Evaluation psychologischer Interventionsmaßnahmen. Standards und Kriterien: Ein Handbuch* (S. 41–85). Göttingen: Hogrefe.
- Hamre, B., Hatfield, B., Pianta, R. & Jamil, F. (2014). Evidence for general and domain-specific elements of teacher-child interactions: Associations with preschool children's development. *Child Development*, 85, 1257–1274. <https://dx.doi.org/10.1111/cdev.12184>
- Hamre, B. K. & Pianta, R. C. (2005). Can instructional and emotional support in the first-grade classroom make a difference for children at risk of school failure? *Child Development*, 76, 949–967. <https://dx.doi.org/10.1111/j.1467-8624.2005.00889.x>
- Hamre, B. K., Pianta, R. C., Burchinal, M., Field, S., LoCassale-Crouch, J., Downer, J. T., ... & Scott-Little, C. (2012). Course on effective teacher-child interactions: Effects on teacher beliefs, knowledge, and observed practice. *American Educational Research Journal*, 49, 88–123. <https://dx.doi.org/10.3102/0002831211434596>
- Hamre, B. K., Pianta, R. C., Downer, J. T., DeCoster, J., Mashburn, A. J., Jones, S. M., ... Brackett, M. A. (2013). Teaching through interactions: Testing a developmental framework of teacher effectiveness in over 4000 classrooms. *The Elementary School Journal*, 113, 461–487. <https://dx.doi.org/10.1086/669616>
- Kammermeyer, G., King, S., Göbel, P., Lämmerhirt, A., Leber, A., Metz, A., ... Roux, S. (2017). *Mit Kindern im Gespräch (Kita). Strategien zur Sprachbildung und Sprachförderung von Kindern in Kindertageseinrichtungen*. Donauwörth: Auer.
- Kammermeyer, G., Metz, A., Leber, A., Roux, S., Biskup-Ackermann, B. & Fondel, E. (im Druck). Wie wirken sich Weiterbildungen auf die Anwendung von Sprachförderstrategien in Kitas aus? *Frühe Bildung*.
- Kammermeyer, G., Roux, S., Metz, A., Leber, A., Biskup-Ackermann, B. & Fondel, E. (2018). Verbessert der Weiterbildungsansatz „Mit Kindern im Gespräch“ die Anregungsqualität in der Kita? *Empirische Pädagogik*, 32, 198–219.
- Katholische Erwachsenenbildung Rheinland-Pfalz – Landesarbeitsgemeinschaft e.V. für den Trägerverbund FIF (Hrsg.). (2009). *Sprache – Schlüssel zur Welt. Materialien zur Qualifizierung von Sprachförderkräften in Rheinland-Pfalz*. Mainz: KEB.
- Kirkpatrick, D. L. (1979). Techniques for evaluating training programs. *Training and Development Journal*, 33, 78–92.
- Kleinknecht, M., Schneider, J. & Syring, M. (2014). Varianten videobasierten Lehrens und Lernens in der Lehrpersonenaus- und -fortbildung: Empirische Befunde und didaktische Empfehlungen zum Einsatz unterschiedlicher Lehr-Lern-Konzepte und Videotypen. *Beiträge zur Lehrerinnen- und Lehrerbildung*, 32, 210–220.
- Leber, A., Kammermeyer, G. & Roux, S. (2019). Sicherung der Qualität von Beobachtungen der ErzieherIn-Kind-Interaktion am Beispiel des Classroom Assessment Scoring Systems (CLASS). Manuskript eingereicht zur Publikation.
- Lipowsky, F. (2010). Lernen im Beruf. In F. Müller, A. Eichenberger, M. Lüders & J. Mayr (Hrsg.), *Lehrerinnen und Lehrer lernen* (S. 51–69). Münster: Waxmann.
- Markussen-Brown, J., Juhl, C. B., Piasta, S. B., Bleses, D., Højen, A. & Justice, L. M. (2017). The effects of language- and literacy-focused professional development on early educators and children: A best-evidence meta-analysis. *Early Childhood Research Quarterly*, 38, 97–115. <https://dx.doi.org/10.1016/j.ecresq.2016.07.002>
- 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, 732–749. <https://dx.doi.org/10.1111/j.1467-8624.2008.01154.x>
- Mayer, D. & Beckh, K. (2018). Erfassung pädagogischer Qualität in Kindertageseinrichtungen. *Frühe Bildung*, 7, 67–76. <https://dx.doi.org/10.1026/2191-9186/a000370>
- Pianta, R. C., LaParo, K. M. & Hamre, B. K. (2008). *Classroom Assessment Scoring System. Manual Pre-K*. Baltimore, MD: Brookes Publishing.
- Pianta, R. C., Mashburn, A. J., Downer, J. T., Hamre, B. K. & Justice, L. (2008). Effects of web-mediated professional development resources on teacher-child interactions in pre-kindergarten classrooms. *Early Childhood Research Quarterly*, 23, 431–451. <https://dx.doi.org/10.1016/j.ecresq.2008.02.001>
- Praetorius, A. K. (2014). *Messung der Unterrichtsqualität durch Ratings*. Münster: Waxmann.
- Roos, J., Polotzek, S. & Schöler, H. (2010). *Evaluationsstudie zur Sprachförderung von Vorschulkindern. Wissenschaftliche Begleitung der Sprachfördermaßnahmen im Programm „Sag' mal was – Sprachförderung für Vorschulkinder“ Der Baden-Württemberg-Stiftung*. Zugriff am 4. 1. 2012 unter <http://www.sagmalwas-bw.de>
- Schachter, R. E. (2015). An analytic study of the professional development research in early childhood education. *Early Education and Development*, 26, 1057–1085. <https://dx.doi.org/10.1080/10409289.2015.1009335>
- Sherin, M. G. & van Es, E. A. (2009). Effects of video club participation on teachers' professional vision. *Journal of Teacher Education*, 60, 20–37. <https://dx.doi.org/10.1177/10022487108328155>
- Sigel, I. E. (2002). The psychological distancing model: A study of the socialization of cognition. *Culture & Psychology*, 8, 189–214. <https://dx.doi.org/10.1177/1354067X02008002438>
- Silverman, R. (2007). A comparison of three methods of vocabulary instruction during read-alouds in kindergarten. *The Elementary School Journal*, 108, 97–113. <https://dx.doi.org/10.1086/525549>
- Simon, S. & Sachse, S. (2011). Sprachförderung in der Kindertagesstätte – Verbessert ein Interaktionstraining das sprachförderliche Verhalten von Erzieherinnen? *Empirische Pädagogik*, 20, 462–480.
- Stark, R. & Klauer, J. (2018). Situiertes Lernen. In D. H. Rost, J. R. Sparfeldt & S. Buch (Hrsg.), *Handwörterbuch Pädagogische Psychologie* (5. Aufl.) (S. 763–771). Weinheim: Beltz.
- Stuck, A., Kammermeyer, G. & Roux, S. (2016). The reliability and structure of the Classroom Assessment Scoring System (CLASS) in German preschools. *European Early Childhood Education Research Journal*, 24, 873–894. <https://dx.doi.org/10.1080/1350293X.2016.1239324>

- Suchodoletz, A. von, Fäsche, A., Gunzenhauser, C. & Hamre, B. K. (2014). A typical morning in preschool: Observations of teacher-child interactions in German preschools. *Early Childhood Research Quarterly*, 29, 509–519. <https://dx.doi.org/10.1016/j.ecresq.2014.05.010>
- Tietze, W., Schuster, K.-M., Grenner, K. & Roßbach, H.-G. (2007). *Kindergarten-Skala (KES-R). Feststellung und Unterstützung pädagogischer Qualität in Kindergärten* (3. Aufl.). Berlin: Cornelsen Scriptor.
- Wasik, B. A. & Hindman, A. H. (2011). Improving vocabulary and pre-literacy skills of at risk-preschoolers through teacher professional development. *Journal of Educational Psychology*, 103, 455–469. <https://dx.doi.org/10.1037/a0023067>
- Weltzien, D. (2013). Erfassung von Interaktionsgelegenheiten im Alltag – erste Ergebnisse der Entwicklung und Überprüfung des Beobachtungsverfahrens GInA. In K. Fröhlich-Gildhoff, I. Nentwig-Gesemann, A. König, U. Stenger & D. Weltzien (Hrsg.), *Forschung in der Frühpädagogik VI. Schwerpunkt: Interaktion zwischen Fachkräften und Kindern* (S. 59–58). Freiburg im Breisgau: FEL.
- Werner, C. D., Linting, M., Vermeer, H. J. & Van Ijzendoorn, M. H. (2016). Do intervention programs in childcare promote the quality of caregiver-child interactions? A meta-analysis of randomized controlled trials. *Prevention Science*, 17, 259–273. <https://dx.doi.org/10.1007/s11121-015-0602-7>
- Wildgruber, A., Wertfein, M., Wirts, C., Kammermeier, M. & Danay, E. (2016). Situative Unterschiede der Interaktionsqualität im Verlauf des Kindergartenalltags. *Frühe Bildung*, 5, 206–213. <https://dx.doi.org/10.1026/2191-9186/a000283>
- Yoon, K. S., Duncan, T., Lee, S. W.-Y., Scarloss, B. & Shapley, K. (2007). *Reviewing the evidence on how teacher professional development affects student achievement* (Issues & Answers Report, REL 2007 – No. 033). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Southwest. Zugriff am 5. 6. 2013 unter [http://ies.ed.gov/ncee/edlabs/regions/southwest/pdf/rel\\_2007033.pdf](http://ies.ed.gov/ncee/edlabs/regions/southwest/pdf/rel_2007033.pdf)
- Prof. Dr. Gisela Kammermeyer**  
**Anja Leber**  
**Astrid Metz**  
**Beate Biskup-Ackermann**  
Universität Koblenz-Landau  
Campus Landau  
Institut für Bildung im Kindes- und Jugendalter  
August-Croissant-Str. 5  
D-76829 Landau  
E-Mail: [kammermeyer@uni-landau.de](mailto:kammermeyer@uni-landau.de)  
[leber@uni-landau.de](mailto:leber@uni-landau.de)  
[metz@uni-landau.de](mailto:metz@uni-landau.de)  
[biskup-ackermann@uni-landau.de](mailto:biskup-ackermann@uni-landau.de)
- Prof. Dr. Susanna Roux**  
Pädagogische Hochschule Weingarten  
Kirchplatz 2  
D-88250 Weingarten  
E-Mail: [roux@ph-weingarten.de](mailto:roux@ph-weingarten.de)
- Eva Fondel**  
Evaluation, Methoden- und Forschungsberatung  
(Statistik)  
Kapuzinergasse 3  
D-76829 Landau  
E-Mail: [beratung@evafondel.de](mailto:beratung@evafondel.de)