

# Metaanalyse

zur Wirkung von Weiterbildungen für pädagogische Fachkräfte  
zur Steigerung der Interaktionsqualität

**Oktober 2018**



Franziska Egert & Verena Dederer

Zitiervorschlag: Egert, F. & Dederer, V. (2018). Metaanalyse zur Wirkung von Weiterbildungen für pädagogische Fachkräfte zur Steigerung der Interaktionsqualität. München: Staatsinstitut für Frühpädagogik. Online verfügbar unter: [www.ifp.bayern.de](http://www.ifp.bayern.de)

Förderkennzeichen:

BiSS-E2: BIS.00.00005.15

Gefördert vom



Bundesministerium  
für Familie, Senioren, Frauen  
und Jugend

## Inhaltsverzeichnis

<b>1</b>	<b>Hintergrund</b>	<b>7</b>
<b>2</b>	<b>Ziele und Fragestellungen</b>	<b>9</b>
<b>3</b>	<b>Methode</b>	<b>9</b>
3.1	Vorgehensweise	9
3.2	Selektionskriterien	10
3.3	Systematische Literaturrecherche	10
3.4	Volltextanalyse	10
3.5	Analyse	15
3.6	Classroom Assessment Scoring System	15
<b>4</b>	<b>Ergebnisse</b>	<b>17</b>
4.1	Ergebnisse zum Literatursuch- und Codierprozess	17
4.2	Ergebnisse des Systematischen Reviews	18
4.3	Ergebnisse der Metaanalyse	21
4.4	Ergebnisse der Subgruppen- und Moderatorenanalysen	24
<b>5</b>	<b>Diskussion</b>	<b>26</b>
	<b>Literatur</b>	<b>30</b>

## Tabellenverzeichnis

Tabelle 1 Gutachterübereinstimmung zum Codierschema (Short Form) und Verzerrungsindex.....	11
Tabelle 2 Gutachterübereinstimmung zur Feincodierung von Weiterbildungsmerkmalen und Forschungsdesign	12
Tabelle 3 Gutachterübereinstimmung zur Neucodierung von Studien und Weiterbildungsmerkmalen .....	15
Tabelle 4 Ergebnisse der elektronischen Suche.....	17
Tabelle 5 Studienübersicht .....	19
Tabelle 6 Übersicht zu Weiterbildungsmerkmale.....	20
Tabelle 7 Überblick zu spezifischen Weiterbildungsdidaktik.....	21
Tabelle 8 Weiterbildungseffekte auf Interaktionsqualität .....	22

## Abbildungsverzeichnis

Abbildung 1 Beschreibung des Classroom Assessment Scoring System Version Pre-K .....	16
Abbildung 2 Flow-Diagramm zur systematischen Literatursuche .....	18
Abbildung 3 Übersicht der einzelnen Effektstärken .....	21
Abbildung 4 Funnel Plot zur Publikationsverzerrung .....	23

## Zusammenfassung

Die Qualität der Fachkraft-Kind-Interaktionen in Kindertageseinrichtungen steht in Zusammenhang mit positiven Lernmöglichkeiten und kindlicher Entwicklung im Vorschulalter (Mashburn et al., 2008). Jedoch ist die Qualität der Fachkraft-Kind-Interaktionen, insbesondere im Bereich der sprachlichen und kognitiven Anregung, oftmals nur mittelmäßig vorzufinden (Pakarinen et al., 2010; von Suchdoletz et al., 2014). In der Politik wurden deshalb viele Anstrengungen unternommen, die pädagogische Qualität von Kindertageseinrichtungen durch Aus- und Weiterbildung von frühpädagogischen Fachkräften zu optimieren. In der Praxis findet sich ein breites Spektrum an Qualifizierungsangeboten für frühpädagogische Fachkräfte, von kurzen Impuls-Workshops bis hin zu berufsbegleitenden und konsekutiven Studiengängen (Baumeister & Griese, 2011). Eine erste Übersicht zu den Effekten von Weiterbildungen mit eingeschlossener Literaturrecherche bis 2011 machte allerdings deutlich, dass nur wenige der Weiterqualifizierungen wissenschaftlich evaluiert wurden (Egert et al., 2017). Die aktuelle Metaanalyse geht unter Einbezug von neuerer nationaler und internationaler Forschungsliteratur bis 2017 der Frage nach, wie effektiv die Qualität von Fachkraft-Kind-Interaktionen im Alltag von Kindertageseinrichtungen durch Weiterbildungen gesteigert werden kann. Es wurden nur Evaluationen eingeschlossen, die die Qualität der Fachkraft-Kind-Interaktionen mittel des Classroom Assessment Scoring Systems (CLASS; Pianta, La Paro, & Hamre, 2008) erfassten. Weiter soll analysiert werden, wie methodische Faktoren und Kontextbedingungen die Wirksamkeit beeinflussen. Zur Beantwortung der praxisrelevanten Fragestellung, wie gute und effiziente Weiterbildungen gestaltet werden sollten, werden die Befunde von unterschiedlichen Weiterbildungsformaten und didaktischen Elementen gegenübergestellt. Die systematische Literaturrecherche für Literatur von 1970 bis 2017 erfolgte in elektronischen Datenbanken (ERIC, PSYCINFO, SOCIndex, Proquest D&T und Dissertation Express) und manuell in renommierten Fachzeitschriften, Reviews und Literaturverzeichnissen relevanter Studien. Ergänzend wurde eine freie Suche in Internetsuchmaschinen durchgeführt. Insgesamt wurden 543 potentielle Referenzen gefunden und die Titel und Abstracts gescreent. Als relevant für die Bearbeitung der Fragestellungen wurden 93 Volltexte analysiert und inhaltlich codiert. Nach Bündelung der Studien und Ausschluss von mangelhaften Untersuchungen fanden sich 15 Studien mit 18 verschiedenen Weiterbildungsmaßnahmen. Der gesamte Suchprozess und die inhaltliche Codierung erfolgte durch zwei unabhängige Gutachter. Die Übereinstimmung war überwiegend zufriedenstellend. Es lagen Daten zur Berechnung von 115 Effektstärken mittels Hedges'  $g$  vor. Die Metaanalyse wurde mit Hilfe eines Random-Effect-Model unter Berücksichtigung der Mehrebenenstruktur durchgeführt. Es ergab sich ein aggregierter Gesamteffekt von  $g'=0,42$  für die gefundenen Weiterbildungseffekte. Die Qualität der Studien wurde als sehr hoch eingeschätzt, da sich keine bedeutsamen Verzerrungen aufgrund von Untersuchungsmerkmalen (u.a. Randomisierung, externe Evaluation, Erfassung, Ausbildung) oder Kontextbedingungen (u.a. Träger und Gruppenzusammensetzung) fanden. Innerhalb der Weiterbildungen ließen sich einige Wirkmechanismen nachweisen. So erzielten Weiterbildungen mit intensiveren Gruppenformaten (Stunden pro Monat) und Maßnahmen mit Selbstreflexion signifikant höhere Werte. Insgesamt setzten die evaluierten Maßnahmen Empfehlungen aus bisherigen Forschungsbefunden in Bezug auf aktives Lernen, Praxisbezug und Reflexion prototypartig um, so dass von einem mittelgroßen Erwartungseffekt von  $g'=0,36 - 0,45$  ausgegangen werden kann. Aus der Metaanalyse geht hervor, dass Weiterbildungen das Potential haben, Fachkraft-Kind-Interaktionen systematisch zu verändern und die Qualität der emotions- und lernunterstützenden Erfahrungen von Kindern in Kindertageseinrichtungen zu steigern.

## Abstract

In early childhood education and care (ECEC), teachers are required to display a wide range of skills in their interactions with young children. Teachers offer young children emotional support and are also responsible for classroom organization and the support of children's cognitive development (Hamre, 2014). However, large scale studies indicate that the quality of cognitive and language stimulating interactions is low in North America and most of the European countries (e.g., Mashburn et al., 2008; Pakarinen et al., 2010; von Suchodoletz et al., 2014). In policy and practice, the importance of ongoing in-service professional development for teachers is widely acknowledged as a key mechanism to improve classroom quality. In practice, a wide spectrum of in-service trainings is available for teachers that range from one-day workshops to year-long trainings. Unfortunately, only a small number of programs have been evaluated (Egert, 2015).

The meta-analysis aggregates the impact of in-service professional development programs on interaction quality measured with the Classroom Assessment Scoring System (CLASS Pre-K; Pianta, La Paro & Hamre, 2008). A systematic search of studies that were published between 1970 and 2017 in electronic data bases (e.g., ERIC, PSYCINFO, SOCIndex, Proquest D&T and Dissertation Express), renowned journals, and reference lists of relevant reviews was conducted. In sum, the systematic search revealed 543 hits and after abstract screening, 93 full texts were coded. All abstracts and full texts were coded by two independent reviewers and agreement was satisfactory for most of the codes. Statistical information was used to estimate effect sizes (Hedges'  $g$ ) with CMA (Comprehensive Meta-Analysis Software V3). A random-effects model that nested effect sizes under treatments was used to aggregate the overall effect. Only codes with a high interrater reliability (ICC or Kappa  $>0.7$ ) were included in moderator analysis or meta-regression.

Overall, 15 studies with 18 in-service treatments were included in the meta-analysis. Based on 115 effect sizes, the overall professional development effect on quality of teacher-child-interaction was small-to-medium range ( $g'=0.42$ ). The quality of experimental studies was homogeneous and at a very high level (e.g., randomization, blind observers, external evaluation). There was no systematic bias based on publication, study quality or contextual features. Further, most of the in-service programs included features of active learning opportunities and on-site support. Programs that offered opportunities for self-reflection and self-evaluation were more effective than in-service approaches without. Further, training intensity matters. We found a linear relation between hours per month in group training formats (e.g., workshops, course) and effect sizes. In sum, the meta-analysis shows that in-service professional development programs have the strength to change the quality of teacher-child-interactions in early childhood education and care.

# 1 Hintergrund

Forschungsergebnisse zeigen, dass die Qualität von Fachkraft-Kind-Interaktionen in Zusammenhang mit der kognitiven, sprachlichen und sozial-emotionalen Entwicklung von Kindern im Vorschulalter steht (Mashburn et al., 2008). Allerdings ist die Qualität in Kindertageseinrichtungen in nordamerikanischen und vielen europäischen Ländern meist nur mittelmäßig (Tietze et al., 2013). Während emotions- und verhaltensunterstützende Interaktionen als gut bewertet werden, befinden sich die sprachlich-kognitiv anregenden Interaktionen in Deutschland im niedrigen Qualitätsbereich (Kammermeyer, Roux & Stuck, 2011; Suchodoletz et al., 2014). Die unzureichende Qualifizierung von pädagogischen Fachkräften wird oftmals als Erklärungsansatz für niedrige Werte herangezogen. Die Ergebnisse zum Zusammenhang der frühpädagogischen Qualität mit der Ausbildung von pädagogischen Fachkräften, insbesondere der Akademisierung, sind eher inkonsistent (Early et al., 2006; 2007) und bieten derzeit keine Hinweise für die systematische Optimierung der Qualität in Kindertageseinrichtungen. Auch viele andere politische Bestrebungen der letzten Jahre wie z.B. die Nationale Qualitätsoffensive oder die Einführung von Bildungs- und Rahmenplänen haben bislang wenig dazu beigetragen, die Qualität zu verbessern (Eckhardt et al., 2013). Demgegenüber zeigen die zusammenfassenden Ergebnisse von Metaanalysen die Potentiale von Weiterbildungen auf (Egert, 2015; Fukkink & Lont, 2007; Markussen-Brown et al., 2017; Werner et al., 2016).

In der Metaanalyse von Fukkink und Lont (2007) fanden sich 17 Interventionsstudien, die zwischen 1980 und 2005 publiziert wurden. Es wurde ein aggregierter Effekt auf Fachkraftniveau von  $ES=0,40$  für die Optimierung des pädagogischen Verhaltens gefunden. Einige Jahre später wurden von Werner und Kollegen (2016) 19 Zeitschriftenartikel (publiziert bis 2013) in eine Metaanalyse eingeschlossen. Hier ließen sich u.a. ein aggregierter Effekt von  $ES=0,44$  für das verbesserte Verhalten der Fachkräfte und ein Effekt von  $ES=0,39$  für höhere Qualitätswerte auf Gruppenebene berechnen. In die Metaanalyse von Markussen-Brown und Kollegen (2017) wurden 25 Weiterbildungsstudien (publiziert bis 2014) eingeschlossen. Es ergab sich ein mittelgroßer Effekt für Weiterbildungen mit dem Fokus auf Sprache, Literacy und Interaktionen von  $ES=0,59$  auf Indikatoren zur Prozessqualität in Gruppen. Die internationale Metaanalyse von Egert, Fukkink und Eckhardt (2018) umfasste 36 Interventionsstudien aus dem Zeitraum von 1970 bis 2011. Hier zeigte sich ein mittelgroßer Effekt von  $ES=0,68$  auf Indikatoren zur Messung der pädagogischen und interaktionsspezifischen Prozessqualität auf Gruppenebene. Die aggregierten Befunde der geschilderten Metaanalysen deuten zwar auf das Potential von Weiterbildungen hin, die Interaktionsqualität in Kindertageseinrichtungen zu steigern. Bei genauer Betrachtung fanden sich jedoch auch einige Primärstudien, bei denen Weiterbildungen keine oder negative Effekte produzierten. Auch die Konfidenzintervalle der aggregierten Effektstärken weisen auf eine gewisse Varianz zwischen den Maßnahmen hin. Diese Streuung der Effektstärken kann durch unterschiedliche Bedingungen beeinflusst sein. So divergiert die Effektivität der Weiterbildungen u.a. durch das verwendete Konstrukt bzw. Instrument zur Erfassung von Qualität in Kindertageseinrichtungen (Egert et al., 2018). Dabei wurden signifikant höhere Effektwerte der Qualitätsindikatoren bei Messungen mit dem Classroom Assessment Scoring System (CLASS; Pianta, La Paro, & Hamre, 2008) erzielt als bei anderen handelsüblichen



Qualitätsratingverfahren (Egert, 2015). Es ist somit ratsam, die Befunde lediglich zu einem spezifischen Qualitätsmaß zu synthetisieren.

Auch Weiterbildungskomponenten beeinflussen den Weiterbildungserfolg. So wurden individuelle Trainingseinheiten, wie z.B. Coaching, als besonders effektive Wirkmechanismen innerhalb der Weiterbildungen identifiziert (Egert, 2015; Werner et al., 2016). Bislang wurde kein direkter linearer Zusammenhang zwischen Qualitätssteigerung und Weiterbildungsdauer (in Monaten) sowie Intensität (Stunden) gefunden. Mehr „Weiterbildung“ führt demnach nicht automatisch zu höherer Wirkung. Die metaanalytischen Ergebnisse deuten eher auf die Bedeutung der Intensität individueller Weiterbildungskomponenten hin (Markussen-Brown et al., 2017). Narrative Analysen von hoch wirksamen Weiterbildungen mit Effektstärken von  $ES > 0,8$  unterstützen diese Annahme und indizieren, dass diese Programme einiges gemeinsam haben. So war eine größere Wirkung zu erwarten, wenn die Maßnahmen (1) aktives und anwendungsbezogenes Lernen durch Übungen im Kita-Kontext (performance-based assignments), (2) individuelle Unterstützung und/oder (3) Videoanalysen als Reflexions-, Feedback- oder Demonstrationsformate beinhalteten (Egert, Eckhardt & Fukkink, 2017). Die Potentiale dieser didaktischen Überlegungen und der Weiterbildungsintensität gilt es weiterführend zu prüfen.

Neben didaktischen Überlegungen werden Aspekte der Studienqualität oftmals als Begründung für unterschiedliche Weiterbildungswirkungen herangezogen. So fanden sich in den beschriebenen Metaanalysen mehrere methodologische Effektmoderatoren, wie beispielsweise Kontrollgruppenbedingung oder systematischer Stichprobenverlust, die die Weiterbildungswirkung systematisch verzerren (Markussen-Brown et al., 2017, Werner et al., 2016). Einen großen Anteil bei der Weiterbildungswirkung hat dabei auch das verwendete Qualitätsratingverfahren. So wurden höhere Wirkungen erzielt, wenn die Gütekriterien in Bezug auf Beobachterübereinstimmung und interne Konsistenz zufriedenstellend waren (Egert, 2015). Auch die Gruppe, die höhere Ausgangswerte im Qualitätsrating besaß (entweder Interventions- oder Kontrollgruppe), erzielte signifikant größere Effektstärken (Egert, Fukkink & Eckhardt, 2018).

In Deutschland hat die Bereitstellung von berufsbezogenen Weiterbildungen in der frühkindlichen Bildung zunehmend an Bedeutung gewonnen. Dies geht einher mit den gewachsenen Ansprüchen und der Komplexität des Arbeitsfeldes der frühkindlichen Bildung. Erworbene Wissensbestände und Fähigkeiten reichen oftmals nicht aus, um die anstehenden Herausforderungen zu meistern und dem Bildungsauftrag gerecht zu werden. Derzeit wird nahezu das volle Spektrum an berufsbegleitenden Weiterqualifizierungen für pädagogische Fachkräfte angeboten. Dieses reicht von kurzen Workshops bis hin zu mehrjährigen Maßnahmen (Baumeister & Griese, 2011). Auch die Bereitschaft, eine Fort- und Weiterbildung zu besuchen, ist bei pädagogischen Fachkräften sehr hoch (Beher & Walter, 2012). Hauptmotiv der hohen Teilnahmequote ist die Erweiterung beruflicher Kompetenzen (ebd.). Der politisch formulierte Bedarf an einer empirisch abgesicherten Praxis lässt deshalb auch die Frage nach der Wirksamkeit von Weiterbildungen im Bereich der Frühpädagogik aufkommen. Obwohl eine Vielzahl an Programmen angeboten wird, sind bislang nur wenige der angebotenen Fort- und Weiterbildungsmaßnahmen in Deutschland, insbesondere zur Steigerung der Qualität in Kindertageseinrichtungen, wissenschaftlich evaluiert (Egert et al., 2017).



## 2 Ziele und Fragestellungen

Ziel der Metaanalyse ist es, den derzeit verfügbaren nationalen und internationalen Forschungsstand zu Weiterbildungseffekten auf die Qualität von Fachkraft-Kind-Interaktionen in Kindertageseinrichtungen zu identifizieren. Dies erfolgt anhand einer systematischen Literaturrecherche und der quantitativen Analyse vorhandener Studienergebnisse. Die Metaanalyse geht primär dem Ziel nach, wie effektiv die Interaktionsqualität in Kindertageseinrichtungen durch Weiterbildungen gesteigert werden kann. Weiter werden unterschiedliche Effektmoderatoren auf Seiten der Weiterbildung und teilnehmenden Personen wie auch in Bezug auf Studienmerkmale untersucht. Dabei werden folgende Punkte bearbeitet:

- (1) *Zunächst wird der zu erwartende Weiterbildungseffekt auf die Interaktionsqualität, gemessen mit dem Classroom Assessment Scoring System CLASS-PreK, berechnet.*
- (2) *Es finden weiterführende Analysen zur Robustheit des erwartbaren Effekts und zum Risiko einer Publikationsverzerrung statt. Auch die Varianz der Studienergebnisse wird untersucht.*
- (3) *Um die Aussagekraft der Ergebnisse und Compliance der Untersuchungen besser einschätzen zu können und Implikationen für Interventionsstudien zu formulieren, wird die Höhe des Effekts in Abhängigkeit zu methodologischen Variablen und Studienqualität untersucht. Dabei werden Variablen wie Randomisierungsebene, Kontrollgruppenbedingung, Erhebungsart (Videorating vs. Live-Beobachtung), Autoren-Allegiance und Stichprobenszusammensetzung näher untersucht.*
- (4) *Zur Beantwortung der praxisrelevanten Fragestellung, von welchen Merkmalen die Wirksamkeit der Weiterbildung abhängt, werden mehrere Subgruppen- und Moderatorenanalysen durchgeführt. Dabei wird näher auf die Weiterbildungsformate und Intensität eingegangen.*
- (5) *Um die Reichweite der Ergebnisse einordnen zu können, werden Kontextmerkmale als potentielle Effektmoderatoren analysiert.*

## 3 Methode

### 3.1 Vorgehensweise

In der BiSS E2 Studie, die vom Bundesministerium für Familie, Senioren, Frauen und Jugend gefördert wird, wurde im Rahmen eines Teilprojekts eine Metaanalyse zur Wirkung von Weiterbildungen für frühpädagogische Fachkräfte auf die Interaktionsqualität durchgeführt. Hauptverantwortliche und Ansprechpartnerin für dieses Teilprojekt ist Dr. Franziska Egert.

Die Bearbeitung der Fragestellungen beinhaltet (1) eine systematische Literatursuche, (2) eine Volltextanalyse und Codierung relevanter Informationen, (3) die Extrahierung statistischer Kennwerte und die Transformation in Effektstärken, (4) eine Metaanalyse der gesammelten Daten sowie (5) Subgruppen- und Moderatorenanalysen.

### 3.2 Selektionskriterien

In die Metaanalyse wurden (1) Interventionsstudien mit Prä- und Postmessungen (u.a. randomisierte und quasi-experimentelle Studien) eingeschlossen, die die Weiterbildungswirkung für pädagogische Fachkräfte in Kindertageseinrichtungen untersuchten. Studien mit in der Tagespflege tätigen Personen wurden ausgeschlossen. Als Outcome-Maß (2) wurden nur externe Qualitätsratings mit dem Classroom Assessment Scoring System (CLASS) zugelassen. Eingeschlossen wurden (3) alle Programme zur Weiterqualifizierung von frühpädagogischen Fachkräften neben dem Beruf, jedoch keine Maßnahmen zum Erwerb von Basisqualifikationen (u.a. Ausbildung, Studium). Berücksichtigt wurde (4) deutsch- und englischsprachige Literatur von 1970 bis 2017. Weiteres Einschlusskriterium war (5) die Bereitstellung statistischer Kennwerte zur Bildung von Effektstärken.

### 3.3 Systematische Literaturrecherche

Die systematische Literaturrecherche umfasste drei Zugänge. (1) Die elektronische Literatursuche erfolgte in den Datenbanken ERIC, PSYCINFO, SOCIndex, ProQuest Dissertation & Thesis und Dissertation Express. In den jeweiligen Datenbanken wurde die folgende Kombination von Schlüsselwörtern verwendet:

((in-service training for teachers) OR (professional development) OR (coaching)) AND ((interaction quality) or (teacher child interaction)) AND (preschool or kindergarten or early childhood education) AND (effect or impact or influence)

(2) Weiter wurde eine manuelle Suche in bisherigen Überblicksartikeln (u.a. Egert, Fukkink, & Eckhardt, 2018; Fukkink & Lont, 2007; Klein & Gomby, 2008; Markussen-Brown et al. 2017; Werner et al. 2016; Zaslow et al., 2010a; 2010b) und relevanten Zeitschriften (u.a. Early Childhood Research Quarterly) durchgeführt. (3) Um an weitere Literaturquellen zu gelangen wurde eine freie Suche in webbasierten Suchmaschinen (u.a. Google, Bing) mit den oben beschriebenen Schlüsselwörtern durchgeführt.

### 3.4 Volltextanalyse

Die Volltextanalyse erfolgte in drei Schritten. (1) Zunächst wurde die Studienqualität, einfache Merkmale der Weiterbildungen und potentielle Verzerrungsbedingungen in der Short Form erfasst. (2) Es erfolgte eine Feincodierung zu Lern- und Trainingstransferfaktoren. (3) In einem weiteren Schritt wurde eine Umformulierung und Neucodierung zu Items aus den ersten beiden Codierschemen durchgeführt, die im ersten Codierprozess nicht reliabel eingeschätzt werden konnten. Alle Texte wurden von zwei unabhängigen Gutachterinnen analysiert<sup>1</sup>. Nicht-Übereinstimmungen wurden durch erneutes gemeinsames Lesen diskutiert und finale Codes determiniert. Die Berechnung der Übereinstimmung wurde mittels Kappa und Intraklassenkorrelationskoeffizient (ICC) berechnet. Nach Wirtz und Casper (2002) bedeutet ein ICC- oder Kappa-Wert über 0.7 eine gute Übereinstimmungs-Reliabilität.

---

<sup>1</sup> Wir danken Julia Gaupp für die Unterstützung bei der Volltextanalyse zur Feincodierung von Studien.

Tabelle 1 Gutachterübereinstimmung zum Codierschema (Short Form) und Verzerrungsindex

<b>Variable der Codierung (Short Form)</b>	<b>Kappa</b>	<b>ICC</b>	<b>k</b>
Publikationstyp (1=publiziert; 2=graue/unveröffentlichte Literatur)	0,98		89
Vergleichskondition (1=KG; 2=keine KG; 3=alternatives Treatment)	0,88		86
Studientyp (1=experimentell; 2=quasi;3=Korrelationsstudie; 4=nicht-experimentell; 5=Single-Subjekt /Multiple-Baseline)	0,81		85
Zuweisung der Probanden (1=randomisiert Fachkraft; 2=randomisiert Kita-Gruppe/Kita; 3=Parallelisiert; 4=Bestehende Gruppen(convenience), 5=nicht zuordenbar)	0,67	0,61	48
Zielgruppe (1=preschool; 2=K-12-Teacher; 3=andere)	0,93		83
n Interventionsgruppe 1		0,94	76
n Kontrollgruppe		0,93	39
n Gesamtstichprobe Kontrast		0,93	84
Qualifizierung (1=Weiterbildung; 2=Ausbildung; 3=sonstige)	1,00		79
Workshop (1=ja)	0,76		79
Kurs/Seminar (1=ja)	0,67		79
Online-Modul (1=ja)	0,82		79
Community of Practice (1=ja)	1,00		79
Individuelle Unterstützung (1=ja)	0,90		79
Unterstützungsmethode (1=Coaching, 2=Beratung; 3=sonstige)	0,68		57
Dauer in Monaten		0,95	41
Intensität in Stunden		1,00	34
CLASS-Score (1=ja)	1,00		86
Erhebungsart (1=Video; 2=Beobachtung; 3=externes Rating durch Leitung/Kollegen; 4=Selbsteinschätzung)	0,88		88
Statistische Kennwerte (1=vorhanden; 0=nicht)	0,88		86
Relevante Studie (1=ja)	0,93		89
<b>Variablen zum Verzerrungsindex</b>	<b>Kappa</b>	<b>ICC</b>	<b>k</b>
Randomisierung/ Zufallsauswahl (1=ja)	0,75		87
Gruppen zu Prätest äquivalent (1=ja)	0,77		21
% Dropout Prä-Post		0,99	18
Umsetzungsqualität der Weiterbildung hoch (1=ja)	0,74		88
Konfundierung mit anderem Treatment (1=ja)	0,90		82
Weiterbildung wurde extern evaluiert (1=ja)	0,69		71
Gute Reliabilitätswerte berichtet (1= $\alpha > .65$ , Kappa oder ICC $> .70$ )	0,65		89
Effektstärken werden berichtet (1=ja)	0,77		89
Setting mit überwiegend Risikokindern (1=ja)	0,61		85
Setting mit Kindern unter 3 Jahren/Krippe (1=ja)	0,85		85
% akademische Fachkräfte in Gesamtstichprobe		1,00	34
% akademische Fachkräfte in Interventionsgruppe		1,00	28
% akademische Fachkräfte in Kontrollgruppe		1,00	22

Das **Codierschema (Short Form)** bestand aus 24 Items. Der erste Teil (9 Items) beinhaltet Informationen zum Codierer sowie zur Studien- und Treatmentnummer und Publikation, um Studien bündeln und identifizieren zu können (Tabelle 1). Dabei wird der Publikationstyp dichotom erfasst (1=publiziert als peer-review Artikel, Buchkapitel oder Buch und 0=graue Literatur wie nicht publizierte

Berichte oder Dissertationsschriften), um eine Publikationsverzerrung berechnen zu können. Durch fünf Items wird das Studiendesign (u.a. Art der experimentellen Studie, Zuweisung der Probanden, Probandentypus, Anzahl der Probanden in Interventions- und Kontrollbedingung) näher analysiert. Eine erste Beschreibung bzw. Codierung der Weiterbildung erfolgt über sechs Items. Dabei wird explizit geprüft, ob es sich um eine Weiterbildung im Beruf stehender pädagogischer Fachkräfte handelt oder um eine Basisqualifikation (Ausbildung, Studium, Kurz-Qualifizierung für Quereinsteiger), die zur Arbeit in Kindertageseinrichtungen befähigt. Mit welchem Instrument die Outcome-Messung erfolgte (1=CLASS-Scores; 0=keine CLASS-Scores bzw. andere Skala) und die Art der Erhebung (CLASS-Rating durch 1=Video, 2=Beobachtung) sowie die Bereitstellung von statistischen Kennwerten wurden in drei Items erfasst. Ein weiteres Item wurde verwendet, um die Relevanz der Studien für die Metaanalyse zu codieren (1=relevant; 0=nicht relevant). Für diesen finalen Code mussten alle oben genannten Selektionskriterien erfüllt sein.

Zur Überprüfung möglicher Verzerrungen der Effektstärken aufgrund von methodologischen Aspekten wurde ergänzend ein **Verzerrungsindex**, bestehend aus zehn Items, codiert. Für eine einfache Codierung wurden überwiegend dichotome Items zu Informationen zur Randomisierung, Äquivalenz der Gruppen, Umsetzungsqualität/Fidelity, Konfundierungen mit anderen Treatments/Manipulationen, externer Evaluation, Bereitstellung von Reliabilitätswerten und statistischen Kennwerten verwendet. Weiter wurden Prozentangaben zum Dropout und der Anzahl der akademisch ausgebildeten Fachkräfte in den jeweiligen Stichproben erhoben.

Das **erweiterte Codierschema (Feincodierung)** beschäftigt sich im Detail mit den Inhalten, der didaktischen Gestaltung und den Trainern der Weiterbildungen. Hiermit wurden nur Studien mit CLASS-Daten analysiert.

*Tabelle 2 Gutachterübereinstimmung zur Feincodierung von Weiterbildungsmerkmalen und Forschungsdesign*

<b>Variable</b>	<b>Kappa</b>	<b>ICC</b>	<b>k</b>
Publikationsjahr		1,00	22
<b>Didaktische Gestaltung /Weiterbildungselemente (Wie)</b>			
Weiterbildung mit individuellen Einheiten (1=ja)	1,00		23
Weiterbildung in organisierten Gruppen außerhalb der Kita (1=ja)	1,00		23
Weiterbildung als in-house Schulung (1=ja)	x		23
Verhaltensdemonstration und -modellierung/ Verhaltensbeobachtung (1=ja)	0,73		23
Rollenspiel (1=ja)	1,00		23
Aktive Übungen/Hausaufgaben für Praxis (1=ja)	0,55		23
Dokumentation der Übungen/Hausaufgaben (1=ja)	0,55		23
Diskussion und Austausch (1=ja)	0,59		23
Planung und Agenda/Einheiten zur Erstellung von Aktionsplänen (1=ja)	0,62		23
Vorträge/rezeptiver Wissensinput (1=ja)	0,47		23

Tabelle 2 (fortgesetzt)

Variable	Kappa	ICC	k
Qualitätsprofile der pädagogischer Arbeit (1=ja)	x		23
Selbstevaluation(1=ja)	0,74		23
Feedback allgemein vorhanden (1=ja)	0,72		23
- mündliches Feedback (1=ja)	0,78		23
- schriftliches Feedback (1=ja)	0,62		23
- Feedback erfolgte zeitnah (1=ja)	0,45		23
- Feedback erfolgt distal/online/telefonisch (1=ja)	1,00		23
- Feedback erfolgte vor Ort (1=ja)	0,47		23
Weiterbildung beinhaltete Videos (1=ja)	0,91		23
- Videoreflection/Selbsteinschätzung des Videos (1=ja)	0,74		23
- Videofeedback von externen (1=ja)	0,83		23
- Es wurden externe (unpersönliche) Videos verwendet (1=ja)	0,64		23
- Es wurden eigenen Videos verwendet (1=ja)	0,62		23
Anzahl der individuellen Trainingsstunden		1,00	23
Anzahl der Trainingsstunden in organisierten Gruppensettings		1,00	14
Training wurde von einer Universität/Hochschule durchgeführt (1=ja)	0,83		21
<b>Inhaltliche Gestaltung von Weiterbildungen (Was)</b>			
Trainingsinhalt (1=problemzentriert, 2=bedürfnisorientiert, 3=zielorientiert, 4=festes Curriculum)	1,00		22
Trainingsziel = Bewusstsein(1=ja)	0,51		23
Trainingsziel = Wissen (1=ja)	0,64		23
Trainingsziel = Job-Performanz (1=ja)	k		23
Trainingsreichweite (1=spezifisches Training, 2=universelles/breiter Fokus)	0,64		21
Verknüpfung von Trainingsinhalt mit Messinstrument (1=nah; 2=keine)	1,00		21
Inhalt: Implementierung von Programmen/Curricula (1=ja)	0,65		23
Inhalt: Software/Tools für Diagnostik/Monitoring kindlicher Entwicklung (1=ja)	x		23
Inhaltlicher Fokus (1=emotionale Unterstützung, 2=Organisation im Kita-Alltag, 3=Lernunterstützung, 4=alle CLASS-Dimensionen, 5=Sprache und Literacy, 6=Mehrsprachigkeit, 7=Pädagogische Qualität allgemein, 8=MINT, 9=andere)	0,83		23
<b>Merkmale des Trainer (Wer)</b>			
Anzahl Trainer/innen		1,00	6
Merkmale der Trainer/innen (1=Autor, 2=zertifizierte/r Trainer/in, 3=Multiplikator, 4=andere)	0,66		14
Qualifikation des Trainer/in (1=akademischer Abschluss, 0=nicht)	1,00		8
<b>Forschungsdesign</b>			
Stichprobe: Alle Fachkräfte sind aus Head Start Einrichtungen (1=ja)	0,83		23
Störvariablen: Kindertageseinrichtung/Gruppe erhält finanzielle Unterstützung (1=ja)	0,65		23
Störvariablen: Fachkraft erhält finanzielle Unterstützung (Stipendien) (1=ja)	0,59		23
Störvariablen: Materiale Unterstützung	0,34		23
Studienqualität: Erheber/innen waren blind (1=ja)	0,65		18

Anmerkung: x=wurde in keiner Studie erwähnt; k= Konstante.

Die Items zum Vorhandensein spezifischer Weiterbildungselemente (wie) und Weiterbildungsinhalte (was) wurden dichotom zum Ankreuzen codiert (1=ja/Vorhandensein eines Element und 0=nicht vorhanden/keine Information). Lediglich das Publikationsjahr, die Anzahl der Stunden für individuelle Weiterbildungseinheiten, die Anzahl der Weiterbildungsstunden in organisierten Gruppeneinheiten sowie die Anzahl der Weiterbildner/innen wurde offen und intervallskaliert angegeben. Die Gestaltung der Weiterbildungsinhalte wurde kategorial erfasst (1=problemzentriert, 2=bedürfnisorientiert, 3=zielorientiert, 4=festes Curriculum). Für die Codierung des inhaltlichen Fokus wurden Kategorien verwendet, die analog zu den unterschiedlichen Aspekten von Interaktionsqualität und deren Messung durch die CLASS sind (1=emotionale Unterstützung, 2=Organisation im Kita-Alltag, 3=Lernunterstützung, 4=alle CLASS-Dimensionen), sowie weitere breite Weiterbildungsinhalte (5=Sprache und Literacy, 6=Mehrsprachigkeit, 7=Pädagogische Qualität allgemein, 8=MINT, 9=andere). Daran anschließend erfolgte eine dichotome Einschätzung, wie gut die Trainingsinhalte mit dem Messinstrument CLASS verknüpft sind (1=nahe Verknüpfung/alignment; 2=keine Verknüpfung). Auch die thematische Bandbreite der Weiterbildung bzw. die Spezifizierung der Weiterbildungsinhalte wurde dichotom codiert (1=spezifisches Training, 2=universelles/breiter Fokus). Weiter wurden Informationen zu den Weiterbildner/innen gesammelt. Dabei wurde kategorial erfasst, wer als Weiterbildner/in in der Studie agierte (1=Autor(en), 2=zertifizierte/r Trainer/in, 3=Multiplikator(en), 4=andere). Auch die Qualifikation des Weiterbildners/der Weiterbildnerin wurde dichotom codiert (1=akademischer Abschluss, 0=keine akademischer Abschluss). Auch Variablen zu Störfaktoren und zur Stichprobe wurden dichotom zum Ankreuzen erfasst, wobei die Bejahung der Frage als 1 und das Verneinen als 0 codiert wurde.

Die Übereinstimmung der Gutachterinnen beim unabhängigen Codieren der Studien war für die meisten Items zufriedenstellend bis sehr gut (Tabellen 1 und 2). Somit zeigt sich, dass nur bei wenigen Kategorien ein großer Klärungsbedarf bestand und die Texte erneut analysiert werden mussten. Dies war überwiegend bei detaillierten Informationen zu didaktischen Elementen der Weiterbildungen nötig. Bei einzelnen Items, die relevant für die Moderatorenanalyse erschienen, aber eine geringere Übereinstimmung (ICC oder Kappa  $<0,7$ ) aufwiesen, wurde eine Neucodierung durchgeführt (Tabelle 3).

So wurden ausgewählte Merkmale der Studienqualität erneut codiert. Der Randomisierungstyp sowie die Verblindung wurden nochmals dichotom erfasst. Bei den vorherrschenden Kontrollgruppenbedingungen besteht das Risiko einer Effektverzerrung, bei der insbesondere Studien mit alternativen Weiterbildungstreatments betroffen sind und geringere Effektstärken aufweisen könnten. Da es sich bei den Studien um Feldzugänge handelt und in einigen Ländern eine Weiterbildungspflicht besteht, ist somit eine unbehandelte (weiterbildungsfreie) Kontrollgruppe nur schwer realisierbar.

Weiterbildungsmerkmale und didaktische Elemente wie Kursformat, Möglichkeit zum Austausch/Diskussion und verpflichtende Praxisübungen wurden erneut dichotom (1=ja, vorhanden; 0=nein/nicht vorhanden) codiert. Es wurde ein erneuter Versuch durchgeführt herauszufinden, welche konkrete individuelle Unterstützungsform (1=Coaching; 2=Mentoring; 3=andere) verwendet wurde. Auch die Anzahl der Trainingsstunden in individuellen Unterstützungsformaten oder Gruppensettings

wurde nochmals erfasst und bei Nichtvorhandensein einer solchen Weiterbildungskomponente die Stundenzahl auf 0 gesetzt.

*Tabelle 3 Gutachterübereinstimmung zur Neucodierung von Studien und Weiterbildungsmerkmalen*

<b>Variable der Neucodierung</b>	<b>Kappa</b>	<b>ICC</b>	<b>K</b>
Randomisierungstyp (1=RCT mit Randomisierung der Fachkräfte; 0=CRT mit Randomisierung auf Gruppen oder Einrichtungsebene)	0,77		18
Verblindung der Erheber_in für Studienkondition (1=blind; 0=nicht blind)	0,77		18
Kontrollgruppenbedingung (1=Business as usual; 2=business as usual mit regulären Weiterbildungen/Unterstützungsangeboten für bestehende Curricula mit gleichem Umfang wie experimentelles Treatment; 3= Alternatives experimentelles Treatment wie z.B. verkürzte Weiterbildung)	0,79		18
Weiterbildung im Kursformat mit mehreren Seminareinheiten (1=ja; 0=nicht vorhanden)	0,87		18
Anzahl der individuellen Trainingsstunden (bei keiner individuellen Unterstützung wurde die Anzahl auf 0 Stunden festgelegt)		0,98	18
Anzahl der Weiterbildungsstunden in Gruppensettings (u.a. Workshops, Kurse)		0,83	18
Möglichkeit zum Austausch und Diskussion (1=ja; 0=nein)	0,26		18
Weiterbildung mit verpflichtenden (standardisierten) Übungen (performance-based assignments) für die Praxisphase (1=ja; 0=nein)	0,37		18
Methode der individuellen Unterstützung (1=Coaching; 2=Mentoring; 3=andere)	0,60		18

### 3.5 Analyse

Die Umwandlung der statistischen Daten in Effektstärken sowie die Berechnung der Metaanalyse und Meta-Regression erfolgte anhand der Software „Comprehensive Meta-Analysis Software Version 3“ (Bornstein, Hedges, Higgins, & Rothstein, 2014). Dabei wurde die Effektstärke Hedges'g verwendet, die als robust für Stichprobengrößen gilt (Hedges & Olkin, 1985). Für die Berechnung des Gesamteffekts wurde ein Random-Effekt-Modell verwendet. Da bei den meisten Untersuchungen Effekte für mehrere CLASS-Items oder CLASS-Dimensionen vorhanden waren, ist von einer hierarchischen Struktur auszugehen. In dem Mehrebenenmodell wurden die Effektstärken unter der Studie bzw. dem Weiterbildungstreatment geschichtet. Um doppelte Informationen und somit eine daraus resultierende Übergewichtung von Einzelbefunden auszuschließen, wurden die Informationen auf kleinster Ebene für die Berechnung verwendet. So gingen beispielsweise die einzelnen CLASS-Items anstelle des aggregierten Werts auf Domain-Ebene (Emotionale Unterstützung/Emotional Support, Organisation des Kita-Alltags/Classroom Organization, Lernunterstützung/Instructional Support) in die Berechnungen ein. Gleichsam verhielt es sich mit dem CLASS-Gesamtwert. Bei der Berechnung des Gesamtwertes wurden die Untersuchungen nach Varianz gewichtet.

### 3.6 Classroom Assessment Scoring System

Das Classroom Assessment Scoring System ist ein in den USA entwickeltes, international eingesetztes Qualitätsmessverfahren zur Erfassung verschiedener Aspekte von Fachkraft-Kind-Interaktionen. Die Erhebung mit der CLASS ist eine strukturierte und standardisierte Beobachtung von Alltagssituationen, die entweder live im Gruppenraum oder als Videoring erfolgt. In der Regel werden



mehrere (ca. vier bis fünf) 20-minütige Cycles an einem Vormittag erhoben. Für unterschiedliche Altersgruppen liegen eigene Versionen des Instruments vor. Die Version CLASS Pre-K ist für die Beobachtung mit dem Altersschwerpunkt 3-6 Jahre (Kindergartenalter) konzipiert. Auf einer 7-stufigen Skala wird die Qualität der Fachkraft-Kind-Interaktionen zu verschiedenen Aspekten beleuchtet. Ein Wert von 1 bis 2 wird als niedrige, 3 bis 5 als mittlere und, 6 bis 7 als hohe Qualität eingeschätzt. Die handelsübliche Version CLASS Pre-K verfügt über die folgenden 10 Items:

- Positives Klima (Positive Climate)
- Negatives Klima (Negative Climate)
- Feinfühligkeit (Teacher Sensitivity)
- Orientierung am Kind (Regard for Student Perspective)
- Verhaltensmanagement (Behavior Management)
- Beschäftigungsgrad der Kinder (Productivity)
- Lernarrangement (Instructional Learning Formats)
- Kognitive Anregung (Concept Development)
- Feedbackqualität (Quality of Feedback)
- Unterstützung sprachlichen Lernens (Language Modeling).

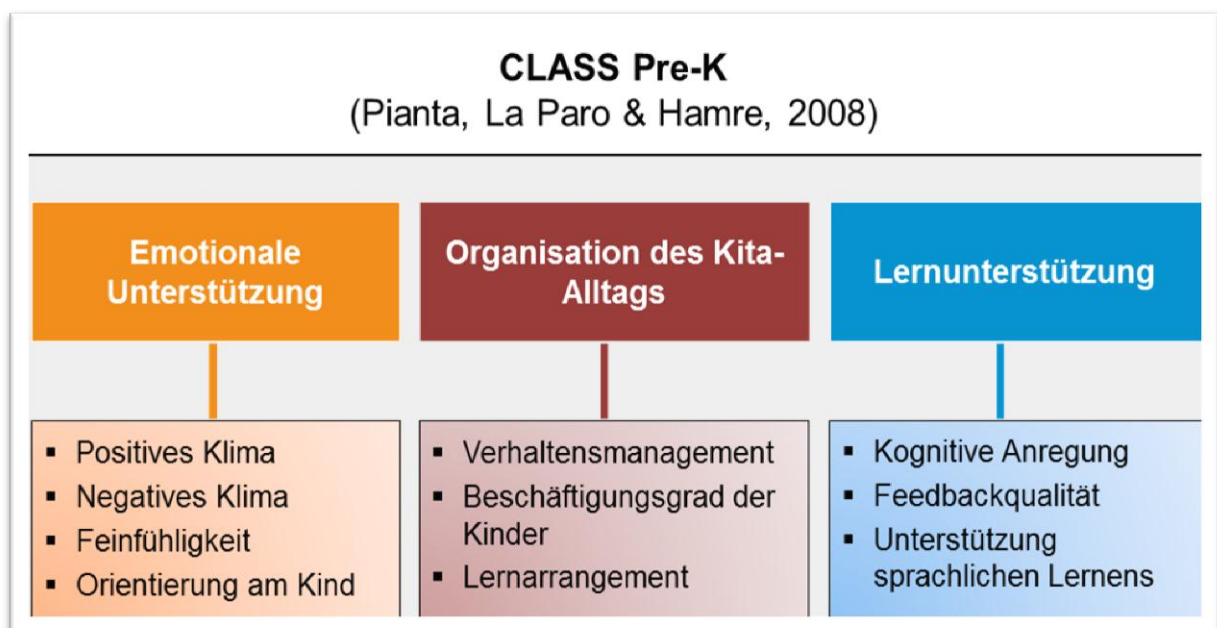


Abbildung 1 Beschreibung des Classroom Assessment Scoring System Version Pre-K

Zu Forschungszwecken wurde ein weiteres Item zum Thema Literacy entwickelt, das derzeit keinem Qualitätsbereich zugeordnet ist (vgl. Justice, Mashburn, Hamre & Pianta, 2008). In der handelsüblichen Version CLASS Pre-K wird die Interaktionsqualität in den Kindergartengruppen in drei Qualitätsbereiche (Domänen) unterteilt: (1) Emotionale Unterstützung (Emotional Support), (2) Organisation des Kita-Alltags (Classroom Organization) und (3) Lernunterstützung (Instructional Support). Die Aufteilung der Items auf die drei Qualitätsbereiche sind in Abbildung 1 ersichtlich.

## 4 Ergebnisse

Im Folgenden werden zunächst die Ergebnisse des systematischen Literatursuchprozesses dargestellt und die relevanten Studien sowie deren Effektstärken präsentiert. Weiter wurden Subgruppen- und Moderatorenanalysen durchgeführt, um herauszufinden von welchen Aspekten der Weiterbildungserfolg abhängt.

### 4.1 Ergebnisse zum Literatursuch- und Codierprozess

Die elektronische Suche in den unterschiedlichen elektronischen Datenbanken ergab 511 Treffer (Tabelle 4). Nach dem Titel und Abstract-Screening und nach Überprüfung von Duplikaten wurden in der elektronischen Suche 87 Publikationen als potentiell relevant für die Fragestellung identifiziert.

*Tabelle 4 Ergebnisse der elektronischen Suche*

	Treffer	relevante Treffer Codierer 1	relevante Treffer Codierer 2	<b>relevante Treffer gesamt</b>
PSYCINFO	117	35	37	<b>35</b>
SOCINFO	18	0	0	<b>0</b>
ERIC	153	41	35	<b>33</b>
ProQUEST D&T	143	11	11	<b>12</b>
DISSEXRESS	80	5	7	<b>7</b>
Gesamt:	511			<b>87</b>

In einer freien Suche im Internet mit den identischen Schlüsselwörtern, die auch in der elektronischen Suche verwendet wurden, sowie in den Literaturverzeichnissen anderer Metaanalysen und Reviews zu Weiterbildungseffekten (u.a. Egert, 2015; Fukkink & Lont, 2007; Markussen-Brown et al., 2017; Werner et al., 2016) wurde nach neuen, noch nicht identifizierten Studien gesucht (Abbildung 2). Dabei wurden weitere 32 Referenzen identifiziert und dem Abstract-Screening unterzogen. Insgesamt wurden die Volltexte von 93 Studien anhand eines Codierschemas (Short Form) und des Verzerrungsindex analysiert. Nach Bündelung der Studien und der strengen Begutachtung waren 15 Studien mit 18 verschiedenen Weiterbildungsmaßnahmen für die Metaanalyse geeignet. Alle Studien verfügen über eine Interventions- und Vergleichsgruppe.

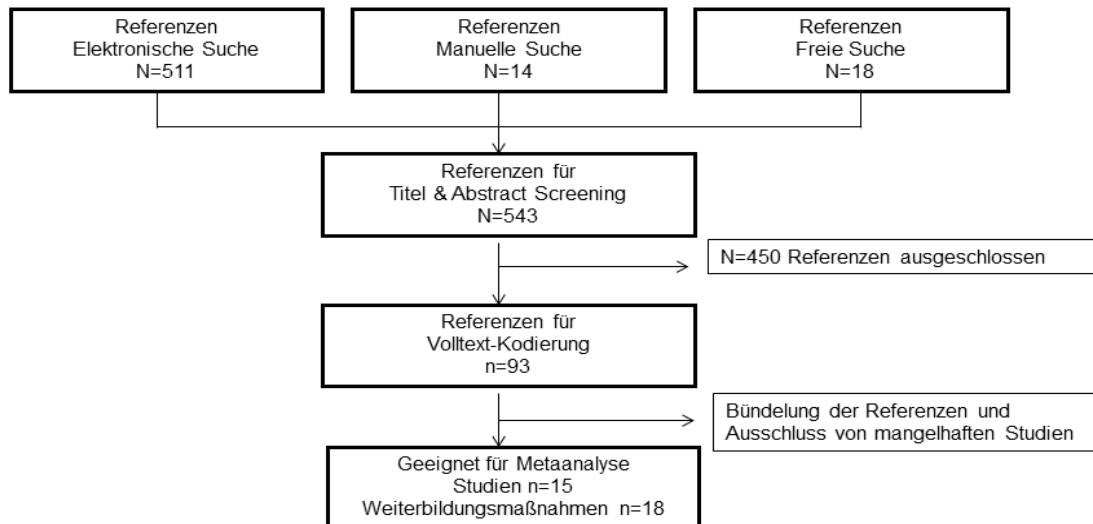


Abbildung 2 Flow-Diagramm zur systematischen Literatursuche

## 4.2 Ergebnisse des Systematischen Reviews

### Beschreibung der Studienqualität der Primärstudien

Der größte Teil (n=12) der 15 Studien wurde in wissenschaftlichen Zeitschriften publiziert. Weitere vier Untersuchungen waren als Abschlussberichte oder unveröffentlichte Doktorarbeiten erhältlich (vgl. Tabelle 5). Alle Untersuchungen wurden in den USA durchgeführt, mit Ausnahme der Studie von Yoshikawa et al. (2015), die in Chile stattfand. Insgesamt wurden 6 Studien (mit 9 Maßnahmen) als externe Evaluation durchgeführt, bei 7 Studien waren die Autoren gleichzeitig für die Weiterbildung und Evaluation zuständig und bei 2 Studien waren keine Informationen diesbezüglich zu finden. Randomisiert-kontrollierte Studien gelten als Goldstandard für Wirksamkeitsüberprüfungen (Campbell Collaboration, 2014). Die Zuweisung der Probanden zur Interventions- oder Kontrollgruppe erfolgte bei allen Studien per Zufallsprinzip. Bei 7 Studien wurden die einzelnen Fachkräfte randomisiert, bei den anderen 8 Untersuchungen erfolgte eine randomisierte Zuordnung auf Gruppen- oder Einrichtungsebene. Durch die zufällige Zuordnung zu Interventions- und Kontrollgruppe sollen potentielle Einflussbedingungen der Probanden und Kontexte gleichmäßig verteilt werden und somit eine Vergleichbarkeit der Bedingungen sichergestellt sein (Higgins & Altman, 2008; Higgins & Green, 2008). Mit einer Randomisierung kann einer Ergebnisverzerrung entgegen gewirkt werden, die während des Selektionsprozesses entsteht und als Selektionsbias bezeichnet wird. In den Studien wurden die experimentellen Gruppen unterschiedlichen Vergleichsbedingungen gegenübergestellt. Dabei wurden 11-mal business as usual Vergleichsgruppen gegenübergestellt. Bei 2 weiteren business as usual Bedingungen erhielten die Probanden der Vergleichsgruppen in ähnlichem Stundenumfang wie die Interventionsgruppe die reguläre Unterstützung/Weiterbildung für das vorherrschende Curriculum. Bei vier Untersuchungen wurde ein alternatives Treatment in Form einer verkürzten Version der Weiterbildung oder eines Zugangs zu einer Online-Plattform gewählt. Die

beteiligten Stichproben variierten deutlich von 25% bis 100% in ihrem Anteil an pädagogischen Fachkräften mit Ausbildung auf akademischen Niveau (Bachelor oder Master). Bei drei Studien wurde der Ausbildungsgrad der pädagogischen Fachkräfte nicht näher beschrieben. Bei den identifizierten Untersuchungen handelt es zumeist um Pilotstudien mit kleinen Stichproben unter 200 Probanden. Für die Gesamtstichprobe wurde die Anzahl der Fachkräfte zu Beginn der Studie codiert. Die Zahl der Probanden, die in die Analysen einging, kann deshalb abweichen.

*Tabelle 5 Studienübersicht*

Studie	Publikation	Evaluation	Design	KG Typ	Akad. Ausbildung in %	IG (n)	KG (n)	Gesamt (n)
Barnett et al., 2008	publiziert	extern	RCT	BAU+C		14	22	36
Breffni, 2011	publiziert	k.A.	RCT	BAU		9	8	17
Domitrovich et al., 2009	publiziert	involviert	CRT	BAU+C	32	44	44	88
Early, Pan, & Maxwell, 2014 (1)	beides	extern	RCT	BAU	91	175	160	335
Early, Pan, & Maxwell, 2014 (2)	beides	extern	RCT	BAU	91	151	160	311
Hamre et al., 2012; 2010	publiziert	involviert	RCT	BAU	61	223	217	440
Hemmeter et al., 2016	publiziert	involviert	RCT	BAU	100	20	20	40
Morris et al., 2010	grau	extern	CRT	BAU	98	26	25	51
Morris et al., 2014 (1)	grau	extern	CRT	BAU	62	77	77	154
Morris et al., 2014 (2)	grau	extern	CRT	BAU	62	77	77	154
Morris et al., 2014 (3)	grau	extern	CRT	BAU	62	76	77	153
Pianta et al., 2008	publiziert	involviert	CRT	AT	100	82	91	173
Pronchenko-Jain, 2012	grau	extern	RCT	AT	50	12	12	24
Raver et al., 2008	publiziert	k.A.	CRT	BAU	25			87
Wasik & Hindman, 2011	publiziert	involviert	CRT	BAU	63	19	11	40
Whittaker et al., 2015	publiziert	involviert	CRT	BAU	100	14	11	25
Yoshikawa et al., 2015	publiziert	extern	CRT	AT		120	93	213
Zan & Donegan-Ritter, 2014	publiziert	involviert	RCT	BAU	63	38	22	60

Anmerkung: RCT=Randomized Controlled Trial mit Zufallsziehung der Fachkräfte; CRT=Cluster Randomized Trial mit Zufallsauswahl auf Gruppen oder Einrichtungsebene; BAU=Business as usual Bedingung; BAU+C=Business as usual mit Curriculum-Training; AT=Alternatives Treatment (Weiterbildung).

### **Beschreibung der Weiterbildungsmerkmale**

Die Feincodierung gab Aufschluss über Weiterbildungsmerkmale. Insgesamt wurde für 11 Weiterbildungen eine gute Umsetzungsqualität (Fidelity) berichtet (Tabelle 6). Die Dokumentation und Prüfung der Umsetzungsqualität erfolgte über Videoanalysen, Protokolle, Tagebücher und Teilnehmerlisten. Für die anderen sechs Weiterbildungen lag keine detaillierte Beschreibung vor. Es ist unklar, ob die Weiterbildungen so umgesetzt wurden wie es geplant war. In der Regel wurde eine Kombination aus unterschiedlichen Weiterbildungsformaten wie Workshops, Seminaren/Kursen, Online-Plattformen und individuellen Unterstützungsangeboten angeboten. Bei den meisten Maßnahmen (n=15) erhielten die Teilnehmer eine individuelle Unterstützung, u.a. in Form von Coaching, Mentoring, Supervision oder Technical Assistance. Innerhalb der individuellen Unterstützungsmethoden wurde mehrfach das My Teaching Partner (MTP) Coachingkonzept evaluiert. Dabei handelt es sich um ein Distance Coachingformat, bei dem die

Weiterbildungsteilnehmer Videos ihrer pädagogischen Praxis im Kita-Alltag einsenden und dann anschließend mit einem Coach gemeinsam online besprechen. Lediglich bei Breffni (2011) wurde die Verwendung eines traditionellen Weiterbildungsformats in Form eines Kurses ohne Kombination mit anderen Weiterbildungsformaten evaluiert. Bei zwei weiteren Untersuchungen wurden die traditionellen Weiterbildungselemente mit Online-Angeboten ergänzt (Early, Pan & Maxwell, 2014 (1); Whittaker et al., 2015). Die Dauer und Intensität der Weiterbildung in Monaten wurde nur von einem Teil der Studien genau dokumentiert. Es fanden sich kurze Maßnahmen von 2 Monaten bis längerfristige Qualifizierungen von bis zu 18 Monaten. Auch die Intensität variierte stark. So wurden bis zu 162 Stunden individuelle Unterstützung und bis zu 48 Weiterbildungsstunden im Gruppensetting angeboten.

*Tabelle 6 Übersicht zu Weiterbildungsmerkmalen*

Studie	Gute Umsetzung	WS	Kurs	Online	Individuelle Unterstützung	Dauer Monate	Ind. Std.	Gruppe Std.
Barnett et al., 2008		x	x		x	9	18	41
Breffni, 2011			x			2	0	32
Domitrovich et al., 2009		x			x	9	144	32
Early, Pan, & Maxwell, 2014 (1)	ja		x	x		5	0	25
Early, Pan, & Maxwell, 2014 (2)				x	x	8		0
Hamre et al., 2012; 2010	ja		x		x	18	0	42
Hemmeter et al., 2016	ja	x			x	4	35,2	19,5
Morris et al., 2010		x	x		x	10	162	30
Morris et al., 2014 (1)	ja		x		x	8	51,1	48
Morris et al., 2014 (2)	ja		x		x	6	52,6	32
Morris et al., 2014 (3)	ja		x		x	8	50,8	40
Pianta et al., 2008	ja			x	x			
Pronchenko-Jain, 2012	ja	x	x		x	6	18,5	33
Raver et al., 2008	ja		x		x	9	82	30
Wasik & Hindman, 2011	ja	x			x	9		43
Whittaker et al., 2015			x	x			0	25,5
Yoshikawa et al., 2015	ja		x		x	6		48
Zan & Donegan-Ritter, 2014			x		x	8	12,33	12

Anmerkung: X=Vorhandensein eines Weiterbildungselements, ind. Std.=Anzahl der Stunden in individueller Unterstützungsformate; Gruppe Std.=Anzahl der Stunden in Gruppenformaten

### **Beschreibung der verwendeten Weiterbildungsdidaktik**

Das Vorhandensein unterschiedlicher Didaktik und Weiterbildungsmethoden wurde codiert (Tabelle 7). Dabei ist auffällig, dass viele der Methoden, die in der Literatur als bedeutsam für den Weiterbildungserfolg beschrieben werden, auch in den untersuchten Maßnahmen angewendet wurden. So wurde Verhaltensmodellierung und –demonstration bei 12 Weiterbildungen angeboten. Rollenspiel wurde lediglich bei einer Studie erwähnt. Selbstreflexion und Selbstevaluation wurde in 9 Maßnahmen durchgeführt. Auch Videos kamen häufig zum Einsatz (n=10). Dabei wurden u.a. externe Videos mit Best Practice Beispielen zum Kennenlernen unterschiedlicher Interaktionsstrategien oder auch eigene Videos für Reflexion oder Feedback verwendet. In 13 Weiterbildungen wurde gezielt Feedback in schriftlicher oder mündlicher Form angeboten.

Tabelle 7 Überblick zur spezifischen Weiterbildungsdidaktik

Studie	Demonstration	Rollenspiel	Selbstreflexion/ -evaluation	Feedback	Video
Barnett et al., 2008					
Breffni, 2011	x		x		
Domitrovich et al., 2009	x		x	x	x
Early et al., 2014; 2017 (1)					x
Early et al., 2014; 2017 (2)	x		x	x	x
Hamre et al., 2012; 2010	x		x	x	x
Hemmeter et al., 2016			x	x	x
Morris et al., 2010	x				
Morris et al., 2014 (1)	x			x	
Morris et al., 2014 (2)				x	
Morris et al., 2014 (3)				x	
Pianta et al., 2008	x			x	x
Pronchenko-Jain, 2012	x			x	x
Raver et al., 2008				x	
Wasik & Hindman, 2011	x		x	x	x
Whittaker et al., 2015	x		x		x
Yoshikawa et al., 2015	x		x	x	
Zan & Donegan-Ritter, 2014	x	x	x	x	x

Anmerkung: x=Vorhandensein eines Weiterbildungselements

### 4.3 Ergebnisse der Metaanalyse

#### Ergebnisse zu Weiterbildungseffekten und Erwartungsspielräume

Insgesamt gingen 115 Effektstärken aus CLASS-Ratings auf Dimensions- oder Item-Ebene in die Berechnung der Metaanalyse ein (Abbildung 3). Die 115 Effektstärken basieren auf standardisierten Mittelwertdifferenzen aus dem Kontrast zwischen Interventions- und Kontrollgruppe. Die Weiterbildungswirkung variiert von  $g=-0,38$  bis 3,22 auf unterschiedliche Aspekte der Interaktionsqualität. Etwa 10,4% ( $n=12$ ) der Effektstärken liegen im negativen Bereich und 30 % ( $n=31$ ) deuten auf marginale bzw. statistisch nicht bedeutsame Effekte ( $g<.0.2$ ) hin.

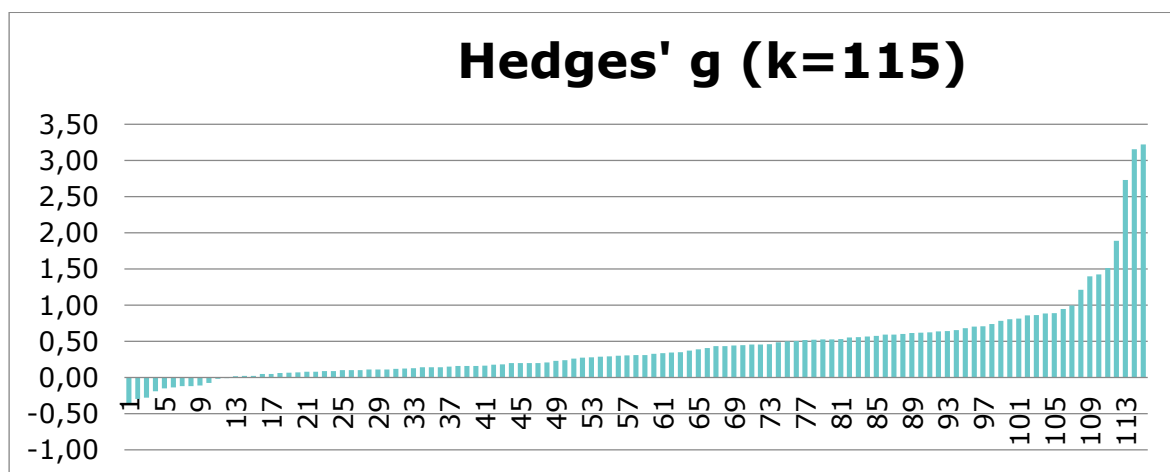


Abbildung 3 Übersicht der einzelnen Effektstärken

Bei einem Gesamteffekt von  $g'=0,42$  ( $SE=0,06$ ) pendelte sich die zu erwartende Durchschnittswirkung von Weiterbildungen auf Interaktionsqualität ein (Tabelle 8). Dieser Wert ist nach Cohen (1988) im Bereich einer kleinen-bis-mittelgroßen Effektstärke zu verorten<sup>2</sup>. Der Erwartungsspielraum der einzelnen Maßnahmen liegt dabei im Bereich von  $g'=-0,08$  bis  $g'=3,02$ . Dies bedeutet, dass Weiterbildungen sowohl keine Wirkung erzielen können als auch das Potential haben große Effektstärken zu erreichen. Eine Heterogenität zwischen den Weiterbildungen ist statistisch vorhanden und als substantiell einzuschätzen ( $Q=212,35$ ;  $df=17$ ;  $p<.000$ ;  $T^2=.055$ ;  $SE=.031$ ). Insgesamt findet sich eine hohe systematische Varianz von 92% zwischen den Interventionen ( $I^2=91,99$ ). Bei einer solch hohen Streuung schlägt Borenstein et al. (2009) vor, den Fokus der Metaanalyse auf die Erklärung dieser zu legen anstelle auf die aggregierte Effektstärke.

*Tabelle 8 Weiterbildungseffekte auf Interaktionsqualität*

	Hedges's g	SE	CI Low	CI up	Z	p	Relatives Gewicht in %
Barnett et al., 2008	0,40	0,16	0,08	0,72	2,47	0,01	4,97
Breffni, 2011	3,02	0,40	2,23	3,80	7,53	0,00	1,87
Domitrovich et al., 2009	0,38	0,10	0,18	0,59	3,65	0,00	6,14
Early, Pan & Maxwell, 2014 (1)	0,32	0,08	0,16	0,47	4,04	0,00	6,62
Early, Pan & Maxwell, 2014 (2)	0,15	0,08	-0,01	0,30	1,83	0,07	6,60
Hamre et al., 2012; 2010	0,32	0,03	0,26	0,38	11,06	0,00	7,25
Hemmeter et al., 2016	0,14	0,18	-0,21	0,50	0,81	0,42	4,64
Morris et al., 2010	0,62	0,09	0,45	0,79	7,04	0,00	6,45
Morris, Mattera et al., 2014 (1)	0,15	0,04	0,07	0,23	3,61	0,00	7,14
Morris, Mattera et al., 2014 (2)	0,12	0,04	0,04	0,20	2,87	0,00	7,14
Morris, Mattera et al., 2014 (3)	0,21	0,04	0,13	0,29	5,14	0,00	7,14
Pianta et al., 2008	-0,08	0,07	-0,23	0,06	-1,14	0,26	6,71
Pronchenko-Jain, 2012	0,81	0,58	-0,32	1,95	1,40	0,16	1,03
Raver et al., 2008	0,64	0,11	0,43	0,84	6,06	0,00	6,13
Wasik & Hindmann, 2011	1,40	0,41	0,59	2,20	3,40	0,00	1,81
Whittaker et al., 2015	0,34	0,09	0,16	0,52	3,73	0,00	6,40
Yoshikawa et al., 2015	0,56	0,13	0,30	0,82	4,18	0,00	5,55
Zan & Donegan-Ritter, 2014	1,00	0,09	0,83	1,18	11,23	0,00	6,43
<b>Gesamteffekt</b>	<b>0,42</b>	<b>0,06</b>	<b>0,29</b>	<b>0,54</b>	<b>6,58</b>	<b>0,00</b>	

### **Ergebnisse zur Robustheit und zum Risiko einer Publikationsverzerrung**

Da Metaanalysen das Ziel haben, eine statistisch akkurate Synthese von Studien zu berechnen, erfolgen immer Analysen zur Robustheit und zum Risiko einer systematischen Publikationsverzerrung. Die One-Treatment-Remove-Analyse zur Einschätzung der Robustheit zeigt, dass der aggregierte Gesamteffekt zwischen  $g'=0,36$  und  $0,45$  schwankt, wenn jeweils eine Weiterbildungsmaßnahme systematisch ausgeschlossen wird. Der höchste aggregierte Weiterbildungseffekt wurde erzielt, wenn

<sup>2</sup>  $g \geq 0,2$  kleiner Effekt,  $g \geq 0,5$  mittelgroßer Effekt,  $g \geq 0,8$  großer Effekt



die Intervention von Pianta et al. (2008) ausgeschlossen wurde und die niedrigsten Wert beim Ausschluss der Maßnahme von Breffni (2011).

Eine Publikationsverzerrung tritt auf, wenn der aggregierte Effekt auf einer verzerrten Anzahl an publizierten vs. nicht-publizierten Untersuchungen basieren würde. In der Forschung werden hauptsächlich drei Probleme thematisiert, die mit einer Publikationsverzerrung einhergehen können (Borenstein et al., 2009; Rosenthal, 1979). Dabei wird angenommen, dass Untersuchungen mit nicht signifikanten Ergebnissen häufiger nicht publiziert werden. Weiter wird auf das Risiko verwiesen, dass dadurch möglicherweise publizierte Untersuchungen häufiger in Metaanalysen aufgenommen werden. Darüber hinaus gibt es Hinweise dafür, dass die Tendenzen für Effektstärken sich zwischen kleineren Pilotstudien und groß angelegten Untersuchungen unterscheiden. Dabei ist die Verzerrung von Effekten bei kleineren Untersuchungen mit niedrigerer Studienqualität größer (Sterne, Egger, & Moher, 2008). Auch die Präzision der Effektberechnung erhöht sich mit steigender Stichprobenzahl (Harbord, Harris, & Sterne, 2009). Um den Zusammenhang zwischen der Stichprobengröße und dem Standardfehler abzuschätzen, werden Funnel Plots analysiert. Eine Publikationsverzerrung wird hauptsächlich bei einer Asymmetrie angenommen, wenn sich am Boden der Abbildung viele Studien bündeln (Borenstein et al., 2009; Harbord, Harris, & Sterne, 2009).

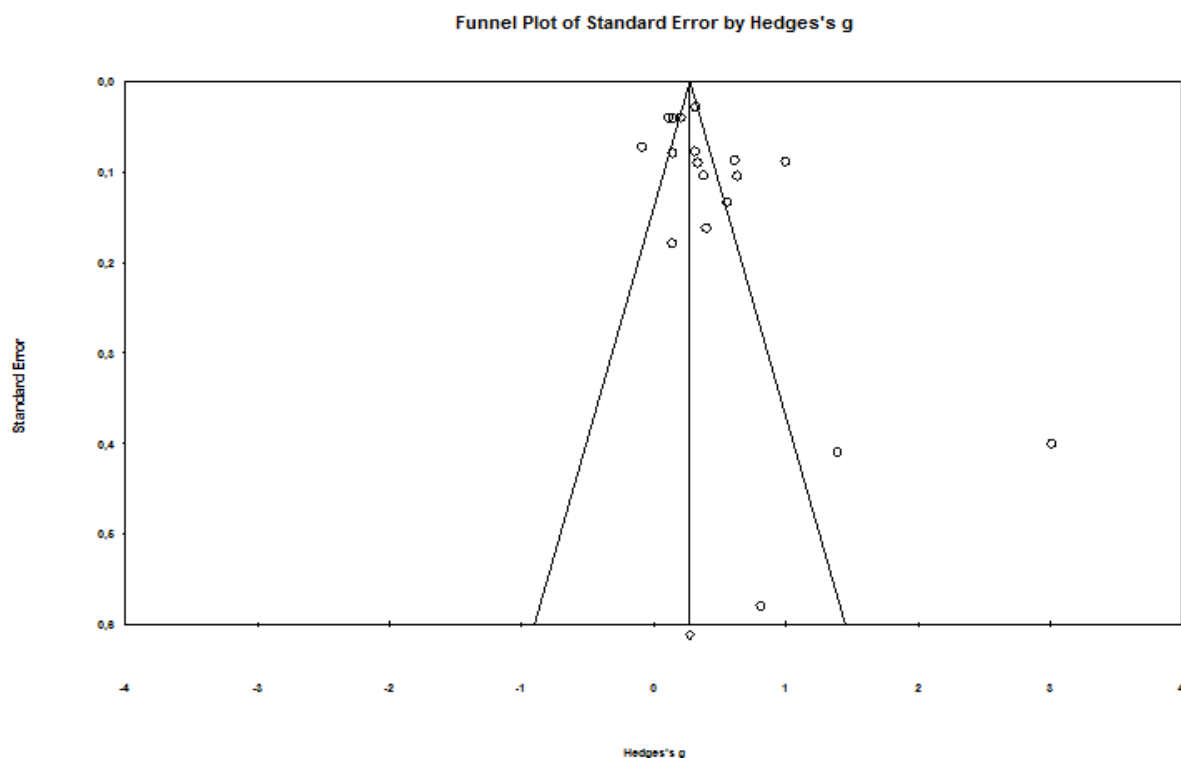


Abbildung 4 Funnel Plot zur Publikationsverzerrung

Aus dem Funnel Plot (Abbildung 4) ist ersichtlich, dass sich die meisten Weiterbildungen mit kleinem Standardfehler im Bereich der kleinen bis mittleren Effektstärken bündeln, aber auch einzelne Ausreißer vorhanden sind. Die Egger Regression (Intercept=2,95; SE=1,31;  $p=.003$ ) macht deutlich, dass die Weiterbildungswirkung asymmetrisch verteilt ist. Ein kleiner Überhang von Studien mit

großem Standardfehler und großen Effektstärken ist ersichtlich. Die Orwin's fail-safe N Analyse zeigt jedoch, dass 1586 Studien mit Nulleffekten benötigt würden, um den aggregierten Gesamteffekt zu nivellieren. In einer Subgruppenanalyse<sup>3</sup> fand sich kein signifikanter Unterschied zwischen Befunden zu Weiterbildungen aus publizierten Artikeln  $g'=0,48$  und Ergebnissen aus grauer Literatur  $g'=0,29$  ( $Q=1,74$ ,  $df=1$ ,  $p=.187$ ). Die Effektstärke variiert auch nicht bedeutsam zwischen Studien, die vor oder nach 2010 veröffentlicht wurden ( $Q=0,746$ ,  $df=1$ ,  $p=.388$ ). Alternative Hypothesen zur Erklärung der Heterogenität der Studien müssen herangezogen werden.

#### 4.4 Ergebnisse der Subgruppen- und Moderatorenanalysen

##### Die Qualität der Untersuchungen als Effektmoderatoren

Da alle Untersuchungen über eine Zufallsauswahl verfügten, konnten lediglich Analysen zwischen unterschiedlichen Zuweisungsebenen durchgeführt werden. Es ergaben sich keine bedeutsamen Unterschiede ( $Q=2,01$ ,  $df=1$ ,  $p=.156$ ) zwischen der zufälligen Zuordnung von einzelnen Fachkräften ( $g'=0,35$ ) und von Gruppen bzw. Einrichtungen ( $g'=0,54$ ). Bei den meisten Evaluationsstudien wurde eine business as usual-Bedingung für die Kontrollgruppe gewählt und bei fünf Studien erhielt die Vergleichsgruppe ein alternatives Weiterbildungstreatment. Wie erwartet ergaben sich Tendenzen, dass Weiterbildungen mit unbehandelter Kontrollgruppe besser abschnitten. Dieser Unterschied war allerdings statistisch nicht bedeutsam ( $g'=0,24$  vs.  $0,48$ ;  $Q=2,70$ ;  $df=1$ ;  $p=.101$ ). Auch die Dreiteilung der Vergleichsbedingungen (BAU, BAU+C und AT) ergab keine signifikanten Unterschiede ( $Q=2,93$ ,  $df=2$ ,  $p=.231$ ). Da die Bewertung mit der CLASS als Live-Beobachtung und über Videos erfolgen kann, wurde dies als Effektmoderator geprüft. Es ergab sich eine nicht signifikante Tendenz ( $Q=2,95$ ,  $df=1$ ,  $p=.086$ ) zugunsten von Videoanalysen ( $g'=0,59$ ) im Vergleich zu Live Beobachtungen ( $g'=0,35$ ). In den meisten Fällen waren die Erheber\_innen der CLASS-Ratings verblindet in Bezug auf die Versuchsbedingungen. Es fand sich kein statistischer Unterschied zwischen den Studien mit Verblindung und den drei Studien ohne ( $g'=0,47$  vs.  $0,28$ ;  $Q=1,02$ ;  $df=1$ ;  $p=.312$ ). Die Information zur externen Evaluation sollte zur Einschätzung des Risikos einer Allegiance-Verzerrung (Vermischung von Entwicklung, Durchführung und Evaluation der Weiterbildung) bzw. eines sogenannten „Conflict of Interest“ beitragen. Signifikant höhere Effektstärken erzielten Untersuchungen ( $Q=13,36$ ,  $df=2$ ,  $p=.001$ ), bei denen keine Informationen dazu verfügbar waren ( $g'=1,20$ ) im Vergleich zu externen Evaluationen ( $g'=0,31$ ) und Studien, bei denen die Autoren gleichzeitig auch in der Entwicklung oder Durchführung der Weiterbildung involviert waren ( $g'=0,41$ ). Eine Gegenüberstellung der letzten beiden Konditionen wurde nicht signifikant ( $Q=0,56$ ;  $df=1$ ,  $p=.456$ ). Die Meta-Regression ergab keinen Hinweis auf einen linearen Zusammenhang zwischen Weiterbildungswirkung und Anzahl der Interventionsteilnehmer (Koeffizient=-.002;  $Q=3,10$ ;  $df=1$ ,  $p=.079$ ). Informationen zur Stichprobenszusammensetzung bezüglich der Ausgangsqualifikation verzerren möglicherweise die Effekthöhe ( $Q=8,83$ ,  $df=2$ ,  $p=.012$ ). So divergierten Studien, bei denen keine Information zur

<sup>3</sup> Da identisch berichtete Ergebnisse für Early, Pan und Maxwell (2014), Early, Maxwell, Skinner, Kraus, Hume und Pan (2014) und Early, Maxwell, Ponder und Pan (2017) vorliegen, wurde die Studie als publiziert bewertet.

Qualifikation vorlag, am deutlichsten von allen anderen ( $g'=0,87$ ). Untersuchungen von Fachkräften mit hohem Prozentsatz (<90%) akademisch ausgebildeter Fachkräfte ( $g'=0,25$ ) und Stichproben mit wenigen Fachkräften mit akademischem Abschluss ( $g'=0,43$ ) erzielten eine ähnlich hohe Weiterbildungswirkung. Zwischen den beiden Konditionen waren die Effektstärken vergleichbar ( $Q=1,93$ ;  $df=1$ ;  $p=.165$ ).

### **Weiterbildungsdesign als Effektmoderatoren**

Von 11 Maßnahmen waren Informationen erhältlich, dass die Weiterbildung mit hoher Umsetzungsqualität (fidelity) durchgeführt wurde. Für 7 Interventionen war keine Information vorhanden. Überraschenderweise erzielten die Weiterbildungen mit guter Umsetzungsqualität ( $g'=0,29$ ) in der Subgruppenanalyse ( $Q=6,20$ ;  $df=1$ ;  $p=.013$ ) niedrigere Werte als Maßnahmen ohne Information zur Umsetzung ( $g'=0,59$ ).

Die Analysen zur Dauer und Intensität waren wenig hilfreich, um die Varianz der Weiterbildungseffekte zu erklären. Es fand sich kein linearer Zusammenhang zwischen Effekthöhe und Dauer in Monaten (Koeffizient=-.036;  $p=0.142$ ) sowie Anzahl der individuellen Weiterbildungsstunden (Koeffizient=-.003;  $p=0.173$ ) und Anzahl der Weiterbildungsstunden im Gruppenformat (Koeffizient=-.014;  $p=0.078$ ). Bei einer Neuberechnung der durchschnittlichen Intensität der Weiterbildungsstunden pro Monat im Gruppenformat ergab sich ein linearer Zusammenhang (Koeffizient=.097;  $p=0.003$ ). Für die durchschnittliche Intensität der individuellen Unterstützung im Monat blieb dieser jedoch aus (Koeffizient=-.026;  $p=0.114$ ).

Die Bedeutung einzelner Weiterbildungselemente sowie die Kombination aus mehreren Formaten wurde durch eine Reihe von Subgruppenanalysen geprüft. Der Einsatz von Workshops ( $Q=4,49$ ;  $df=1$ ,  $p=.484$ ) ergab keinen nennhaften Vorteil. Aber die Durchführung von Seminaren mit mehreren Kurseinheiten ( $Q=3,01$ ;  $df=1$ ;  $p=.078$ ) ergab tendenziell höhere Effektstärken für Maßnahmen mit Kursformat ( $g'=0,48$ ) als Weiterbildungen ohne dieses Merkmal ( $g'=0,23$ ). Weiterbildungen mit Online-Elementen erzielten identische Weiterbildungseffekte wie Maßnahmen ohne solche Blended Learning Angebote<sup>4</sup> (web-basierte Plattform für Selbststudium, Austausch, Videoanalysen und Online-Support mit dem My Teaching Partner Concept) ( $Q=2,18$ ;  $df=1$ ;  $p=.140$ ). Überraschenderweise ließ sich im Vergleich zu anderen Metaanalysen (Egert, 2015, Markussen-Brown et al., 2017) hier keine statistische Überlegenheit von Weiterbildungen, bei denen individuelle Unterstützungsformate angeboten werden, finden ( $Q=2,68$ ;  $df=1$ ;  $p=.102$ ). Ähnliche Weiterbildungseffekte erzielte die Kombination aus Kursen und begleitender individueller Unterstützung ( $g'=0,47$ ) im Vergleich zu anderen Kombinationen ( $g'=0,43$ ;  $Q=0,071$ ;  $df=1$ ;  $p=.791$ ).

Die detailliertere Analyse zum Einsatz von spezifischen Weiterbildungsmethoden gibt weitere Anhaltspunkte für höhere Weiterbildungseffekte zur Steigerung der Interaktionsqualität. So waren Maßnahmen effektiver ( $Q=4,31$ ;  $df=1$ ;  $p=.037$ ), wenn Selbstreflektion und Selbstevaluation als Methode eingesetzt wurde ( $g'=0,56$ ) im Unterschied zu Interventionen ohne ( $g'=0,29$ ). Die Studie von Zan und Donegan-Ritter (2014) enthielt als einzige Maßnahme Rollenspiele und erzielte statistisch höhere Werte als die Weiterbildungen ohne ( $Q=8,57$ ;  $df=1$ ;  $p=.003$ ). Aufgrund des Einzelbefundes

<sup>4</sup> Pianta et al. 2008 wurde von der Analyse ausgeschlossen, da die Kontrollgruppe auch Zugang zum webbasierten My Teaching Partner Qualifizierungsportal hatte und somit zwei Interventionen mit Blended Learning verglichen werden würden.

muss dieses Ergebnis mit Vorsicht interpretiert werden. Es ergaben sich keine Überlegenheit für Maßnahmen mit Feedback ( $g'=0,35$  vs.  $0,60$ ;  $Q=3,07$ ;  $df=1$ ;  $p=.080$ ) oder mit Verhaltensdemonstration und -modellierung ( $g'=0,50$  vs.  $0,30$ ;  $Q=1,98$ ;  $df=1$ ;  $p=.159$ ). Auch der pauschale Einsatz von Videos in Weiterbildungen zu unterschiedlichen Zwecken führte nicht zwangsläufig zu höheren Effekten ( $Q=0,70$ ;  $df=1$ ;  $p=.404$ ).

### **Kontextbedingungen als Effektmoderatoren**

Insgesamt war es schwierig Informationen zu Kontextbedingungen der Kindertageseinrichtungen zu erhalten, in denen die an der Studie beteiligten Fachkräfte arbeiten. Bezüglich der Zusammensetzung der Kinder in der Gruppe, der Standards zum Qualifikationsniveau der pädagogischen Fachkräfte, der administrativen Organisation, dem Service für Familien und den finanziellen Beiträgen unterschieden sich Head Start Einrichtungen<sup>5</sup> von regulären Kindertagesstätten. Es wurde deshalb geprüft, ob die Wirkung von Weiterbildung bei regulären Kindertagesstätten im Gegensatz zu Head Start Einrichtungen anders ausfällt. Bei der Subgruppenanalyse fand sich jedoch kein signifikanter Unterschied für die Arbeitskontextbedingungen ( $g'=0,45$  vs.  $0,41$ ;  $Q=0,082$ ;  $df=1$ ;  $p=.775$ ).

## **5 Diskussion**

Aus der Metaanalyse geht hervor, dass die Quantität und Qualität der vorhandenen Primärstudien zur Evaluation von Weiterbildungseffekten auf die Steigerung der Interaktionsqualität in Kindertageseinrichtungen zufriedenstellend ist und einige belastbare Aussagen zulässt. So fand in allen Studien eine Randomisierung statt, um eine potentielle Motivationsverzerrung auszuschließen. Auch eine systematische Verzerrung aufgrund der Stichprobengröße bildete sich nicht ab. Weiter ergaben sich keine entscheidenden Unterschiede zwischen Stichproben mit unterschiedlichen Anteilen an akademisch ausgebildetem Personal oder mit unterschiedlichen Einrichtungsträgern (z.B. Head Start). Ähnliche Befunde zum Träger fanden sich auch bei Werner et al. (2016). Dies erscheint insbesondere für die Einschätzung der Generalisierbarkeit der Weiterbildungswirkung für unterschiedliche Zusammensetzungen der Weiterbildungsteilnehmer, Kindergruppen und Kontextbedingungen von Bedeutung. Derzeit scheint es, dass die Effektivität der evaluierten Maßnahmen nicht von Personalstruktur und Trägerkonstellation beeinflusst werden.

Aus der Übersicht der Maßnahmen wird deutlich, dass nicht nur die Studienqualität auf einem hohen Niveau ist, sondern auch die didaktischen Überlegungen zu den Weiterbildungen ziemlich ausgereift erscheinen. Insgesamt werden viele der didaktischen Elemente zu aktiven Lernmöglichkeiten in den vorhanden Weiterbildungen umgesetzt, die in der Forschungsliteratur als bedeutende Wirkmechanismen bekannt sind (Buysse, Winton, & Rous, 2009; Elmore, 2002; Ingvarson et al., 2005; Yoon et al., 2007; Zaslow et al., 2010). Dazu zählen u.a. die individuelle Unterstützung, mündliches und schriftliches Feedback, Verhaltensdemonstration und -modellierung, und Selbstreflexion. Insgesamt scheint es, dass die untersuchten Weiterbildungen bisherige Empfehlungen von Forschungsbefunden aus der Lern- und Trainingstransferforschung (Baldwin & Ford, 1988; Baldwin,

---

<sup>5</sup> Head Start Einrichtungen sind Programme der kompensatorischen Erziehung für sozial-schwache und benachteiligte Kinder und Familien in den USA, die staatlich finanziert sind (<https://www.acf.hhs.gov/ohs>)

Ford, & Blume, 2009; Blume, Ford, Baldwin, & Huang, 2010) sowie Metaanalysen aus der Frühpädagogik (Egert, 2015; Fukkink & Lont, 2007; Markussen-Brown et al., 2017; Werner et al., 2016) prototypartig umgesetzt haben. Es ist deshalb auch nicht verwunderlich, dass die meisten Weiterbildungen im Durchschnitt kleine bis mittelgroße Effekte erzielen. Dies macht deutlich, dass das Feld der Frühpädagogik die Chance hatte, die Qualität der Fachkraft-Kind-Interaktionen in Kindertageseinrichtungen systematisch zu steigern. Aus der Metaanalyse von Egert, Fukkink und Eckhardt (2018) wurde deutlich, dass die zeitgleiche Wirkung auf die Qualitätssteigerung in der Kita-Gruppe 53% der Varianz der Weiterbildungseffekte auf die kindliche Entwicklungsmaße erklärte. Die aktuellen positiven metaanalytischen Ergebnisse geben somit Anlass zu optimistischen Einschätzungen, dass die kurzzeitigen Effekte auf Fachkräftebene auch indirekt den Kindern in der Gruppe zu Gute kommen.

Ein Anliegen der Metaanalyse war es, die Wirksamkeit unterschiedlicher Weiterbildungsformate und didaktischer Elemente zu untersuchen und Implikationen für die Praxis zu generieren. In Bezug auf strukturell-organisatorische Merkmale konnte keine Überlegenheit eines Weiterbildungsformats gefunden werden. So erzielten beispielsweise die Kombinationsweiterbildungen vergleichbar hohe Effekte mit und ohne individuelles Unterstützungsformat. Eine generelle Überlegenheit der Weiterbildungen mit individueller Unterstützung z.B. durch Coaching zeigte sich bei unseren Analysen nicht. Dies widerspricht den positiven Befunden von anderen Metaanalysen (Egert, 2015; Markussen-Brown et al., 2017; Werner et al., 2016). Dies ist u.a. darin begründet, dass lediglich bei drei Weiterbildungen keine individuelle Unterstützungsform zum Einsatz kam. Somit wurde Coaching nicht mit wenig intensiven Weiterbildungen verglichen, sondern als Referenzgruppe diente in den meisten Fällen intensive Kombinationsweiterbildungen aus Kurs, Workshop und/oder Blended Learning Angebot. Eine Gegenüberstellung zu traditionellen, rein rezeptiven Weiterbildungstypen, konnte innerhalb der Untersuchung nicht vorgenommen werden.

Auch die Wirkung von neuen Medien, u.a. einer Online-Bibliothek mit Best-Practice-Videos und der Möglichkeit zum Austausch und Selbststudium, wurden bereits mehrfach untersucht. Die vorliegenden Befunde zur Nutzung von Blended Learning Angeboten sind allerdings nur bedingt generalisierbar, da sie sich lediglich auf die My Teaching Partner Plattform und das dazugehörige Distance Coaching beziehen. In den vier Studien mit Online-Elementen wurde somit immer das gleiche Blended Learning Angebot von My Teaching Partner verwendet, so dass keine Aussage über die Wirkung anderer Online-Maßnahmen möglich ist. Jedoch zeigen die bisherigen Befunde auch das positive Potential von Blended Learning-Angeboten auf, auf die andere Maßnahmen aufbauen können.

Die meisten Weiterbildungen (n=10) waren für einen längeren Zeitraum von mindestens 8 bis 9 Monaten veranschlagt (was in etwa einem Kindergartenjahr in den USA entspricht) oder hatten mindestens 20 Weiterbildungsstunden. Derzeit sind somit keine Aussagen über Kurzzeit-Maßnahmen oder Vergleiche zwischen kurzen und längeren Maßnahmen möglich. Die Dauer und Intensität stand nicht per se im linearen Zusammenhang mit der Weiterbildungswirkung, wie auch schon andere Metaanalysen belegten (Egert, Fukkink & Eckhardt, 2018; Fukkink & Lont, 2007; Werner et al., 2016). Eine pauschale Aussage über eine Mindestdauer und Stundenzahl erscheint somit nicht ratsam. Jedoch fand sich ein linearer Zusammenhang der monatlichen Weiterbildungsdosis von

Gruppenangeboten (wie Workshops, Kursen) und der Effektstärke zugunsten von intensiveren Weiterbildungen. Die Intensität und Regelmäßigkeit scheint ein Schlüsselmechanismus zu sein, wobei die Dauer einer Weiterbildung an den Inhalt angepasst sein sollte.

## Limitationen

Ziel von Metaanalysen ist es, den Forschungsstand zu einem Thema zu systematisieren und Hypothesen zu prüfen. Für die Durchführung von Subgruppen- und Moderatorenanalysen bedarf es der Codierung studien- und interventionsbezogener Informationen. Somit ist jede Metaanalyse nicht nur abhängig von der Quantität sondern auch der Qualität der Berichterstattung. Aus der bisherigen Forschungsliteratur zu Wirkungsmechanismen von Weiterbildungen lassen sich eine Vielzahl an Effektmoderatoren ableiten (Egert & Schauland, in Review). Ebenso sind viele Merkmale der Studienqualität als Einflussfaktoren bekannt (Wilson & Lipsey, 2001).

Leider war es aus Gründen der vagen bzw. nicht sehr detaillierten Berichterstattung nicht möglich, einige der Hypothesen zu testen. Möglicherweise bedingt durch die limitierte Zeichenzahl in wissenschaftlichen Zeitschriften wurden die verwendeten Weiterbildungen nur undifferenziert beschrieben. Detaillierte Informationen, welche didaktischen Prinzipien mit welchem Umfang stattgefunden haben (z.B. Art der individuellen Unterstützungsform, Austausch- und Diskussionsmöglichkeiten, aktive Übungen oder Planungseinheiten), waren oftmals nicht vorhanden. Auch die Prüfung oder Dokumentation der Qualität der Umsetzung der Weiterbildungen fehlte bei einigen Untersuchungen. Diese sehr vagen Weiterbildungsbeschreibungen erschwerten den Codierprozess, um reliable Codes zu erstellen. Um möglichst zuverlässige Codes über Weiterbildungsmerkmale zu erhalten, wurde ein einfaches Codiersystem mit überwiegend dichotomen Items erstellt, bei dem lediglich das Vorhandensein einer bekannten Weiterbildungsmethode verneint oder bejaht werden konnte. Anhand der bestehenden Informationslage war es trotz Vereinfachung der Codes teilweise nicht möglich eine zufriedenstellende Gutachterübereinstimmung zu erzielen. Von fehlenden oder undifferenzierten Informationen waren auch Moderatoren zur Studienqualität betroffen. So war es u.a. nicht möglich, aus den Texten zuverlässig herauszulesen, ob die Beobachterübereinstimmung zufriedenstellend war.

Ein weiterer Aspekt, der die Prüfung unterschiedlicher Hypothesen bezüglich potentieller Moderatoren erschwert, ist das Auftreten (a) einer sehr homogenen Datenlage oder (b) einer zahlenmäßig geringen Ausprägung bestimmter Variablen. Die Aussagekraft und Generalisierbarkeit der Ergebnisse ist somit eingeschränkt, wenn ein statistischer Vergleich beispielsweise auf einem Befund einer oder weniger Weiterbildungen fußt (z.B. Rollenspiel). Zudem ist es deutlich schwieriger, bedeutsame Unterschiede zwischen Maßnahmen zu finden, die sich sehr ähnlich sind. Dies bezieht sich überwiegend auf individualisierte und dem Kontext angepasste Weiterbildungselemente. In den meisten Weiterbildungen wurden individuelle Unterstützung durch Coaching oder Mentoring (n=15), Feedback (n=13) oder Verhaltensmodellierung und –demonstration (n=12) angeboten, so dass in der Metaanalyse keine einzige Studie einging, die traditionelle Weiterbildungen mit rein rezeptiven Vermittlungsformen verwendete. Detaillierte Informationen zur Umsetzungsqualität (fidelity) der

individualisierten und kontextangepassten Unterstützungselemente war nicht vorhanden und konnte nicht weiterführend analysiert werden.

Eine weitere Einschränkung der Metaanalyse ist die regionale Begrenztheit der Befunde auf Nord- und Südamerika. Keiner der gefundenen Weiterbildungsstudien stammt aus Deutschland und die Übertragbarkeit der Ergebnisse muss diskutiert werden. Als generelle Problematik bei Wirkungsanalysen im Bildungskontext ist die Heterogenität der Kontextbedingungen und die eingeschränkte Kontrollierbarkeit der Lehr-Lernsituation bekannt (Pant, 2014). Für eine Generalisierbarkeit werden oftmals Bedenken wegen unterschiedlicher Organisations- und Personalstruktur sowie einer anderen Zuordnung des frühkindlichen Bildungswesens entweder zum Schul- oder Sozialwesen geäußert (Kuger, Sechtig & Anders, 2012). So waren die Anzahl der pädagogischen Fachkräfte mit akademischer Ausbildung in den Stichproben der meisten Studien deutlich höher als die derzeitigen 5,3% in Deutschland (Autorengruppe Fachkräftebarometer, 2017). Möglicherweise haben verschiedene Gruppenkompositionen (u.a. Altersmischung, sozialer und kultureller Hintergrund) einen Einfluss auf die Gestaltung des Alltags und Fachkraft-Kind-Interaktionen. Übertragen auf die Situation in Deutschland, finden sich allerdings auch innerhalb der Bundesrepublik länderspezifische Bedingungen in Bezug auf Gruppengröße, Personen-Kind-Schlüssel und Ausbildungsniveau der Fachkräfte (Bock-Famulla & Lange, 2013; Statistisches Bundesamt, 2018), die ebenso nur einschränkende Interpretationen zulassen.

Von praktischer Relevanz sind die internationalen Befunde insofern, da ein Konsens über gelingende und entwicklungsfördernde Fachkraft-Kind-Interaktionen sowie ein gemeinsames Qualitätsverständnis besteht. So wird das Classroom Assessment Scoring System (CLASS) national und international zur Messung der Interaktionsqualität in Kindertageseinrichtungen eingesetzt (Kammermeyer, Roux & Stuck, 2011; Suchodoletz et al., 2014; Wildgruber, Wirts & Wertfein, 2014). Ebenso gibt es aktuelle Bestrebungen, die Weiterbildungswirkung im Rahmen der BiSS E1 und BiSS E2 Projekte mit der CLASS in Deutschland zu untersuchen. Darüber hinaus ermöglicht die CLASS eine Operationalisierung des Weiterbildungsziels, die Interaktionsqualität in Kindertageseinrichtungen zu steigern und lässt ländervergleichende Untersuchungen zu.



## Literatur

\* In die Metaanalyse eingeschlossene Referenzen sind mit einem Asterisk gekennzeichnet

- Autorengruppe Fachkräftebarometer (2017). *Fachkräftebarometer Frühe Bildung 2017*. Weiterbildungsinitiative Frühpädagogische Fachkräfte. München.
- Baldwin, T. T., Ford, J. K., & Blume, B. D. (2009). Transfer of training 1988-2008: An updated review and new agenda for future research. In G. P. Hodgkinson & J. K. Ford (Eds.), *International review of industrial and organizational psychology* (pp. 41-70). Chichester, UK: Wiley.
- Baldwin, T. T., & Ford, K. J. (1988). Transfer of training: A review and directions for future research. *Personnel Psychology*, 41(1), 63-105.
- \*Barnett, S. W., Jung, K., Yarosz, D. J., Thomas, J., Hornbeck, A., Stechuk, R., et al. (2008). Educational effects of the tools of the mind curriculum: A randomized trial. *Early Childhood Research Quarterly*, 23(3), 299-313.
- Baumeister, K., & Grieser, A. (2011). *Berufsbegleitende Fort- und Weiterbildung frühpädagogischer Fachkräfte – Analyse der Programmangebote*. Weiterbildungsinitiative Frühpädagogische Fachkräfte. WiFF Studien, Band 10. Verfügbar unter: [http://www.weiterbildungsinitiative.de/uploads/media/Baumeister\\_Grieser\\_pdf.pdf](http://www.weiterbildungsinitiative.de/uploads/media/Baumeister_Grieser_pdf.pdf)
- Beher, K., & Walter, M. (2012). *Qualifikationen und Weiterbildung frühpädagogischer Fachkräfte. Bundesweite Befragung von Einrichtungsleitungen und Fachkräften in Kindertageseinrichtungen: Zehn Fragen - Zehn Antworten*. Weiterbildungsinitiative Frühpädagogische Fachkräfte. WiFF Studien, Band 15. Verfügbar unter: [http://www.weiterbildungsinitiative.de/uploads/media/Studie\\_BeherWalter.pdf](http://www.weiterbildungsinitiative.de/uploads/media/Studie_BeherWalter.pdf)
- Blume, B. D., Ford, K., Baldwin, T. T., & Huang, J. L. (2010). Transfer of training: A meta-analytic review. *Journal of Management*, 36(4), 1065-1105.
- Bock-Famulla, K. & Lange, J. (2013). *Länderreport Frühkindliche Bildungssysteme 2013. Transparenz schaffen - Governance stärken*. Gütersloh: Bertelsmann.
- Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2014). *Comprehensive Meta Analysis Software Version 3.0*. Biostat, Englewood: NJ.
- Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. Chichester, UK: Wiley.
- \*Breffni, L. (2011). Impact of curriculum training on state-funded prekindergarten teachers' knowledge, beliefs, and practices. *Journal of Early Childhood Teacher Education*, 32(2), 176-193.
- Buysse, V., Winton, P. J., & Rous, B. (2009). Reaching consensus on a definition of professional development for the early childhood field. *Topics in Early Childhood Special Education* 28(4), 235-243.
- Campbell Collaboration (2014). Campbell systematic reviews: Policies and guidelines. *Campbell Systematic reviews (Supplement 1)*. doi, 10. Verfügbar unter: [https://campbellcollaboration.org/media/k2/attachments/Campbell\\_Policies\\_and\\_Guidelines.pdf](https://campbellcollaboration.org/media/k2/attachments/Campbell_Policies_and_Guidelines.pdf)
- Cohen, J. (1988). *Statistical power analysis for the behavioral science (2nd Edition)*. Hillsdale, NJ: Lawrence Erlbaum.

- \*Domitrovich, C. E., Gest, S. D., Gill, S., Bierman, K. L., Welsh, J. A., & Jones, D. (2009). Fostering high-quality teaching with an enriched curriculum and professional development support: The Head Start REDI program. *American Educational Research Journal*, 46(2), 567-597.
- Early, D. M., Bryant, D. M., Pianta, R. C., Clifford, R. M., Burchinal, M. R., Ritchie, S., et al. (2006). Are teachers' education, major, and credentials related to classroom quality and children's academic gains in pre-kindergarten? *Early Childhood Research Quarterly*, 21(2), 174-195.
- Early, D. M., Maxwell, K. L., Burchinal, M., Alva, S., Bender, R. H., Bryant, D., et al. (2007). Teachers' education, classroom quality, and young children's academic skills: Results from seven studies of preschool programs. *Child Development*, 78(2), 558-580.
- \*Early, D. M., Maxwell, K. L., Ponder, B.D., & Pan, Y. (2017). Improving teacher-child interactions: A randomized control trial of Making the Most of Classroom Interactions and My Teaching Partner professional development models. *Early Childhood Research Quarterly*, 38, 57-70.
- \*Early, D. M., Maxwell, K. L., Skinner, D., Kraus, S., Hume, K., & Pan, Y. (2014). *Georgia's Pre-K Professional Development Evaluation: Final report*. Chapel Hill, NC: University of North Carolina at Chapel Hill.
- \*Early, D. M., Pan, Y., & Maxwell, K. L. (2014). *Georgia's Pre-K professional development evaluation. Technical appendix*. Chapel Hill, NC: University of North Carolina at Chapel Hill.
- Eckhardt, A., Tietze, W., Bensel, J., Haug-Schnabel, G., Keller, H., Leyendecker, B., Kalicki, B., & Becker-Stoll, F. (2013). Zusammenfassung: Anlage und Ergebnisse der Untersuchung, Empfehlungen für Politik und Praxis, wissenschaftlicher Ausblick. In W. Tietze, F. Becker-Stoll, J. Bensel, A. G., Eckhardt, G. Haug-Schnabel, B. Kalicki, H. Keller, & B. Leyendecker (Hrs.) Nationale Untersuchung zur Bildung, Betreuung und Erziehung in der frühen Kindheit NUBBEK (S. 139-157). Weimar: Verlag das Netz
- Egert, F. (2015). *Meta-analysis on the impact of in-service professional development programs for preschool teachers on quality ratings and child outcomes*. Doctoral dissertation. University of Bamberg Press. Verfügbar unter: <https://opus4.kobv.de/opus4-bamberg/frontdoor/index/index/docId/45682>
- Egert, F., Eckhardt, A., & Fukkink, R. G. (2017). Zentrale Wirkmechanismen von effektiven Weiterbildungen zur Steigerung der Prozessqualität in Kindertageseinrichtungen. Ein narrativer Review. *Frühe Bildung*, 6(2), 58–66.
- Egert, F., Fukkink, R., & Eckhardt, A. G. (2018). Impact of in-service professional development programs for preschool teachers on quality ratings and child outcomes. *Review of Educational Research*, 88(3), 401-433.
- Egert, F., & Schauland, N. (in Review). Wirksame Weiterbildungen für pädagogische Fachkräfte. Ein schwieriges Unterfangen? *Pädagogische Rundschau*,
- Elmore, R. F. (2002). *Bridging the gap between standards and achievement. The imperative for professional development in education*. Washington, D.C.: Albert Shaker Institute.
- Fukkink, R. G., & Lont, A. (2007). Does training matter? A meta-analysis and review of caregiver training studies. *Early Childhood Research Quarterly*, 22(3), 294-311.

- \*Hamre, B. K., Pianta, R. C., Burchinal, M., Field, S., LoCasale-Crouch, J., & Downer, J. T. (2012). A course on effective teacher-child interactions: Effects on teacher beliefs, knowledge, and observed practice. *American Educational Research Journal*, 49(1), 88-123.
- \*Hamre, B. K., Pianta, R. C., Burchinal, M., J. & Downer, J. T. (2010). A course on supporting early language and literacy development through effective teacher-child interactions: Effects on teacher beliefs, knowledge, and practice. *Paper presented at the SREE Conference in Washington D.C.*
- Harbord, R. M., Harris, R. J., & Sterne, J. A. C. (2009). Updated test for small-study effects in meta-analyses. *The Stata Journal*, 9(2), 197-210.
- \*Hemmeter, M. L., Snyder, P.A., Fox, L., & Algina, J. (2016). Evaluating the implementation of the Pyramid Model for promoting social-emotional competence in early childhood classrooms. *Topics in Early Childhood Special Education*, 36(3), 133-146.
- Higgins, J., & Altman, D. (2008). Assessing risk of bias in included studies. *Cochrane Handbook for Systematic Reviews of Interventions: Cochrane Book Series*, 187-241.
- Higgins, J., & Green, S. (2008). *Cochrane handbook for systematic reviews of interventions* (Vol. 5): Wiley Online Library.
- \*Hindman, A. H., & Wasik, B. A. (2012). Unpacking an effective language and literacy coaching intervention in Head Start. *The Elementary School Journal*, 113(1), 131-154.
- Ingvarson, L., Meiers, M., & Beavis, A. (2005). Factors affecting the impact of professional development programs on teachers' knowledge, practice, student outcomes and efficacy. *Education Policy Analysis Archives*, 13(1), 1-28.
- Justice, L., Mashburn, A. J., Hamre, B. K., & Pianta, R. C. (2008). Quality of language and literacy instruction in preschool classrooms serving at-risk pupils. *Early Childhood Research Quarterly*, 23, 51-68.
- Kammermeyer, G., Roux, S. & Stuck, A. (2011). "Was wirkt wie?"- *Evaluation von Sprachfördermaßnahmen in Rheinland-Pfalz. Zweiter Zwischenbericht (Juli 2011)*. Landau: Universität Landau.
- Kuger, S., Sechtig, J., & Anders, Y. (2012). Kompensatorische (Sprach-)Förderung: Was lässt sich aus US-amerikanischen Projekten lernen? *Frühe Bildung*, 1(4), 181-193.
- Markussen-Brown, J., Juhl, C. B., Piasta, S. B., Bleses, D., Hojen, 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.
- Mashburn, A. J., Pianta, R., Hamre, B. K., Downer, J. T., Barbarin, O., Bryant, D., et al. (2008). Measures of classroom quality in pre-kindergarten and children's development of academic, language and social skills. *Child Development*, 79(3), 732-749.
- \*Morris, P., Raver, C. C., Millenky, M., Jones, S., Llyod, C. M. (2010). Making preschool more productive. How classroom management training can help teachers. Retrieved, January 15th, 2011, from <http://files.eric.ed.gov/fulltext/ED514651.pdf>
- \*Morris, P., Mattera, S.K., Castells, N., Bangser, M., Bierman, K., & Raver, C. (2014). *Impact Findings from the Head Start CARES Demonstration: National Evaluation of Three Approaches to*

- Improving Preschoolers' Social and Emotional Competence*. OPRE Report 2014-44. Washington, DC: Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.
- Pant, H. A. (2014). Aufbereitung von Evidenz für bildungspolitische und pädagogische Entscheidungen: Metaanalysen in der Bildungsforschung. *Zeitschrift für Erziehungswissenschaft, Sonderheft 17*, 79–99.
- Pianta, R. C., La Paro, K. M. & Hamre, B. K. (2008). *Classroom Assessment Scoring System. Manual Pre-K* (7. Aufl.). CLASS Pre-K. Baltimore: Paul Brookes.
- \*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(4), 431-451.
- Pronchenko-Jain, Y. (2012). Child teacher relationship training (CRTR) with children exhibiting disruptive behavior: Effects on teacher's ability to provide emotional and relational support to students and on student-teacher relationship stress. Doctoral dissertation. University of North Texas, Denton TX.
- \*Raver, C. C., Jones, S. M., Li-Grining, C. P., Metzger, M., Champion, K. M., & Sardin, L. (2008). Improving preschool classroom processes: Preliminary findings from a randomized trial implemented in Head Start settings. *Early Childhood Research Quarterly*, 23(1), 10-26.
- Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin*, 86(3), 638-642.
- Statistisches Bundesamt (2018). *Der Personalschlüssel in Kindertageseinrichtungen. Methodische Grundlagen und aktuelle Ergebnisse*. Wiesbaden. Statistisches Bundesamt. Verfügbar unter: [https://www.destatis.de/DE/Publikationen/Thematisch/Soziales/KinderJugendhilfe/KindertageseinrichtungenPersonalschluesel5225409179004.pdf?\\_\\_blob=publicationFile](https://www.destatis.de/DE/Publikationen/Thematisch/Soziales/KinderJugendhilfe/KindertageseinrichtungenPersonalschluesel5225409179004.pdf?__blob=publicationFile)
- Sterne, J. A. C., Egger, M., & Moher, D. (2008). Addressing reporting bias. In J. P. T. Higgins & S. Green (Eds.), *Cochrane handbook for systematic reviews of interventions* (pp. 297-333). Chichester, UK: Wiley-Blackwell.
- 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(4), 509-519. doi: 10.1016/j.ecresq.2014.05.010.
- Tietze, W., Lee, H.-J., Bensel, J., Haug-Schabel, G., Aselmeier, M., & Egert, F. (2013). Pädagogische Qualität in Kindertageseinrichtungen und Tagespflegestellen. In W. Tietze, F. Becker-Stoll, J. Bensel, A. G. Eckhardt, G. Haug-Schnabel, B. Kalicki, H. Keller & B. Leyendecker (Hrsg.), *NUBBEK. Nationale Untersuchung zu Bildung, Betreuung und Erziehung in der frühen Kindheit* (S. 69-87). Weimar, Berlin: Verlag das netz.
- \*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(2), 455-469.

- Werner, C. D., Linting, M., Vermeer, H. J., & van IJzendoorn, M. H. (2016). Do intervention programs in child care promote the quality of caregiver-child interactions? A meta-analysis of randomized controlled trials. *Prevention Science, 17*, 259–273.
- \*Whittaker, J.V., Kinzie, M.B., Williford, A. & DeCoster, J. (2015). Effects of MyTeachingPartner–Math/Science on teacher–child interactions in prekindergarten classrooms. *Early Education and Development, 27*(1), 110-127.
- Wildgruber, A., Wirts, C. & Wertfein, M. (2014). Interaktionsqualität in Kindertageseinrichtungen – Forschung mit dem Classroom Assessment Scoring System (CLASS Pre-K). In A. Prengel & M. Winklhofer (Hrsg.). *Kinderrechte in pädagogischen Beziehungen - Studien und Forschungsmethoden*. Opladen: Barbara Budrich.
- Wilson, D. B., & Lipsey, M. W. (2001). The role of method in treatment effectiveness research: Evidence from meta-analysis. *Psychological Methods, 6*(4), 413-429.
- Wirtz, M., & Casper, F. (2002). *Beurteilerübereinstimmung und Beurteilerreliabilität: Methoden zur Bestimmung und Verbesserung der Zuverlässigkeit von Einschätzungen mittels Kategoriensystemen und Ratingskalen*. Göttingen: Hogrefe.
- 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). Verfügbar unter: [https://ies.ed.gov/ncee/edlabs/regions/southwest/pdf/REL\\_2007033.pdf](https://ies.ed.gov/ncee/edlabs/regions/southwest/pdf/REL_2007033.pdf)
- \*Yoshikawa, H., Leyva, D., Snow, C.E., Treviño, E., Barata, M.C., Weiland, C., Gomez, C.J., Moreno, L., Rolla, A., D'Sa, N., & Arbour, M.C. (2015). Experimental impacts of a teacher professional development program in Chile on preschool classroom quality and child outcomes. *Developmental Psychology, 51*(3), 309-22.
- \*Zan, B., & Donegan-Ritter, M. (2014). Reflecting, coaching and mentoring to enhance teacher-child interactions in Head Start classrooms. *Early Childhood Education Journal, 42*, 93-104.
- Zaslow, M., Tout, K., Halle, T., Whittaker, J. V., & Lavelle, B. (2010a). Towards the identification of features of effective professional development for early childhood educators. Literature review. <https://www2.ed.gov/rschstat/eval/professional-development/literature-review.pdf>