



## UvA-DARE (Digital Academic Repository)

### Impact of In-Service Professional Development Programs for Early Childhood Teachers on Quality Ratings and Child Outcomes: A Meta-Analysis

Egert, F.; Fukkink, R.G.; Eckhardt, A.G.

**DOI**

[10.3102/0034654317751918](https://doi.org/10.3102/0034654317751918)

**Publication date**

2018

**Document Version**

Final published version

**Published in**

Review of Educational Research

**License**

Article 25fa Dutch Copyright Act

[Link to publication](#)

**Citation for published version (APA):**

Egert, F., Fukkink, R. G., & Eckhardt, A. G. (2018). Impact of In-Service Professional Development Programs for Early Childhood Teachers on Quality Ratings and Child Outcomes: A Meta-Analysis. *Review of Educational Research*, 88(3), 401-433. <https://doi.org/10.3102/0034654317751918>

**General rights**

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

**Disclaimer/Complaints regulations**

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

*UvA-DARE is a service provided by the library of the University of Amsterdam (<https://dare.uva.nl>)*

# Impact of In-Service Professional Development Programs for Early Childhood Teachers on Quality Ratings and Child Outcomes: A Meta-Analysis

Franziska Egert 

*State Institute of Early Childhood Research (IFP Bayern)*

Ruben G. Fukkink

*University of Amsterdam/Amsterdam University of Applied Sciences*

Andrea G. Eckhardt

*University of Applied Sciences Zittau/Görlitz*

*High pedagogical quality in early childhood education and care (ECEC) is related to developmental outcomes in young children. This review summarizes findings from (quasi)-experimental studies that evaluated in-service training effects for ECEC professionals on external quality ratings and child development. The aggregation of findings at teacher level (including 36 studies with 2,891 teachers) revealed a medium in-service training effect on process quality (effect size [ES] = 0.68, SE = 0.07,  $p < .001$ ). Furthermore, a subset of nine studies (including 486 teachers and 4,504 children) that provided data on both quality ratings and child development were analyzed, and they showed a small effect at child level (ES = 0.14; SE = 0.02,  $p < .001$ ) and a medium effect at the corresponding classroom level (ES = 0.45, SE = 0.11,  $p < .001$ ). Variance in effect sizes at child level was significantly related to in-service effects on quality ratings (53% explained variance). The results show that quality improvement is a key mechanism to accelerate the development of young children.*

**KEYWORDS:** professional development, in-service training, early childhood education and care, meta-analysis

There is increased public interest in the professionalization of the early childhood workforce given the augmented policy emphasis on the early years as a foundation for later school success (Buysse, Winton, & Rous, 2009; Sheridan, Edwards, Marvin, & Knoche, 2009). Significant public investments in pre- and in-service professional development (PD) for early childhood educators are being made all over the world to improve child care quality and foster the development of young children in early childhood education and care (ECEC; Martinez-Beck & Zaslow, 2006; Oberhuemer, 2013; OECD, 2012; Whitebook & Ryan, 2011). State and federal governments employ high-quality early education to address achievement gaps of vulnerable children (Burchinal, Hyson, & Zaslow, 2008). Underlying these investments, there is consensus that the quality of ECEC classrooms contributes substantially to children's learning and development (Eurydice, 2014). Longitudinal and experimental studies showed the link between high-quality center-based care and child development, well-being, and school adjustment of young children in preschool years and later school success (National Institute of Child Health and Human Development Early Child Care Research Network, 2000, 2003; National Institute of Child Health and Human Development & Duncan, 2003; Peisner-Feinberg & Burchinal, 1997; Sammons, 2010; Schweinhart, Barnes, & Weikart, 1993; Schweinhart et al., 2005; Vandell & Wolfe, 2000). However, most children around the globe attend child care centers of average or mediocre quality (Peisner-Feinberg et al., 1999; Sylva, 2010; Tietze et al., 2013; Vermeer, van IJzendoorn, Cárcamo, & Harrison, 2016). Preservice preparation and ongoing in-service PD for early childhood educators are seen as essential for the provision of high-quality service for children and their families (National Association for the Education of Young Children, 2011). However, the correlations between teacher formal qualifications and quality in ECEC are weak (Burchinal et al., 2008; Kelley & Camilli, 2007; Manning, Garvis, Fleming, & Wong, 2017). In particular, high-quality child care experience, defined as teacher-child interactions and effective implementation of instruction and a curriculum that stimulates development, is not reliably produced by teacher preparation (Pianta, Barnett, Burchinal, & Thornburg, 2009). Recently, PD is seen as an important intervention to improve low child care quality. Many in-service trainings are on the market for the ECEC workforce, but only a few programs are evaluated and relatively little is known about their effectiveness. Furthermore, little is known about PD mechanisms that foster success for teachers in ECEC. In particular, questions related to effective components, the instructional design, and delivery of PD for ECEC teachers remain unanswered. This meta-analysis aims to fill these gaps in summarizing the results of experimental research on PD effects in ECEC on quality ratings and child outcomes. Additionally, we explore whether characteristics of the training design, instructional content, and participants moderate experimental effects.

### *Professional Development*

In ECEC, PD covers the entire spectrum of education and training opportunities for ECEC teachers, ranging from a workshop to a university degree (Whitebook, Gombay, Bellm, Sakai, & Kipnis, 2009). It encompasses different facilitated learning opportunities to support the acquisition of professional knowledge, skills, and disposition, aimed at the improvement of teaching and, related to

this, to beneficial child outcomes. Whereas preservice training covers formal education required to become a certified teacher, PD is defined as in-service training opportunities for teachers who work in center-based child care. These in-service programs do not contribute to the attainment of a formal credential or college/university degree. In the United States and Canada, qualifications for lead teachers differ by state or territory and range from no or little preservice training to university degree. Furthermore, some states and territories have strict requirements on the number of in-service training hours (e.g., 120 hours over 5 years in the Western Territories), while others do not require ongoing training (Friendly, Grady, MacDonald, & Forer, 2015; Whitebook et al., 2009). The only exception is the national Head Start program in the United States, serving at-risk children, which has high qualification and training standards (Whitebook et al., 2009). Heterogeneity in the requirements for teacher preparation and ongoing training can also be found in several other countries in Europe (Eurydice, 2014).

In ECEC, PD activities are widely diverse and the theoretical discussion of what constitutes high-quality PD needs to be grounded with empirical research (Schachter, 2015). To reach a consensus for planning and evaluating PD in ECEC, Buysse et al. (2009) surveyed teachers, administrators, and researchers and developed a conceptual framework. Snyder et al. (2012) validated the framework in their review and used it to characterize key components of PD activities in ECEC. According to them, a shared vision is needed to interpret the relations between teacher training, improved child care practice, and desired child outcomes. Therefore, the logic of our coding procedure and the interpretation of our results also underlies this framework.

In the framework of Buysse et al. (2009), three intersecting core components are distinguished, labeled as the “who,” the “what,” and the “how” of in-service programs. The “who” component emphasizes heterogeneity of both the learner and provider of PD, including their organizational context (see also Blank & de las Alas, 2009; Elmore, 2002; Garet, Porter, Desimone, Birman, & Yoon, 2001; Ingvarson, Meiers, & Beavis, 2005; Yoon, Duncan, Lee, Scarloss, & Shapley, 2007; Zaslow, Tout, Halle, Whittaker, & Lavelle, 2010a). Training sessions are offered by different providers with various qualifications and experiences (Snyder et al., 2012) and also early childhood practitioners are widely diverse with respect to their qualification and experience. This variety characterizes the ECEC field (Buysse et al., 2009). Furthermore, contextual variables such as the workplace environment and organizational support are considered effect modifiers in this framework (Blume, Ford, Baldwin, & Huang, 2010; Harvard Family Research Project, 2006; Klein & Gomby, 2008).

The “what” component from the framework comprises the content, defined as specific knowledge, skills, and dispositions targeted by PD programs. The implementation of a specific curriculum has been a frequently studied topic in PD (Klein & Gomby, 2008). In the last decade, scale-based PD approaches have become very common in which content of training is aligned to a quality rating scale and in which in-service providers offered individual scale-based feedback to teachers (see McNerney, Nielsen, & Clay, 2006). It is presumed that these content-specific programs are more effective than PD focusing on general instruction (Buysse et al., 2009).

The “how” component encompasses the organization and facilitation of learning experience and describes training intensity and delivery. In practice, PD varies widely with respect to format (traditional vs. modern, individual vs. standardized, direct vs. indirect feedback) and delivery, ranging from the absence of teachers from classrooms to modern integrated approaches that provide intensive training with onsite support (Snyder et al., 2012). It is assumed that PD is more effective if it is intensive and sustained over time, including guidance and feedback aligned with instruction goals and curricula (Buysse et al., 2009). The results from the meta-analyses of Markussen-Brown et al. (2017) and Werner, Linting, Vermeer, and van IJzendoorn (2016) with regard to the “how” component of PD are divergent (e.g., for intensity, duration, training formats), although meta-analytic results suggest that individualized training and coaching intensity are related to positive outcomes.

The current study seeks to gather evidence on the predictive value of these three theoretical PD components as potential effect modifiers (a) by including unpublished studies to avoid publication bias, (b) by exclusively focusing on center-based teachers for context-specific interpretation of results, and (c) by including experimental studies older than 15 years.

### *In-Service Training Effects*

Research reinforces the importance of ongoing PD to support ECEC teachers. Several studies show the capacity of in-service programs to improve observed quality in the areas of age-appropriate activities (Burchinal et al., 2008; Cassidy, Buell, Pugh-Hoesse, & Russell, 1995; Howe, Jacobs, Vukelich, & Recchia, 2011), instructional support, classroom management, and caregiver responsiveness (Breffni, 2011; Wasik & Hindman, 2011), as well as language and literacy-specific classroom practices (Dickinson & Caswell, 2007; Neuman & Cunningham, 2009). Some reviews have summarized findings of in-service training or coaching effects on child development or quality ratings in a narrative way (Isner et al., 2011; Tout, Isner, & Zaslow, 2011; Zaslow, Tout, Halle, Whittaker, & Lavelle, 2010a, 2010b). Based on a limited number of studies published between 1980 and 2005, the meta-analysis by Fukkink and Lont (2007) showed statistically positive training effects at caregiver level (effect size  $[ES] = 0.45$ ). For a small subset of studies, a medium nonsignificant effect of training programs was found at child level ( $ES = 0.55$ ). Restricting the meta-analysis to randomized controlled trials and published articles and dissertations, Werner et al. (2016) showed that trainings are generally effective in improving child care quality, caregiver interaction skills, and children’s development. Furthermore, Markussen-Brown et al. (2017), including studies from both center-based and family-based settings, also indicated positive effects of language and literacy-related in-service training on teacher knowledge, as well as structural and process quality in ECEC. In sum, all reviews and meta-analyses indicate aggregated positive training effects, but showed also that PD impact ranged from negative to positive outcomes.

Important questions remain concerning which PD approach is most effective for whom, for which outcome, and under what condition. It is, therefore, important to gain deeper insight into mechanisms that are related to the diverse

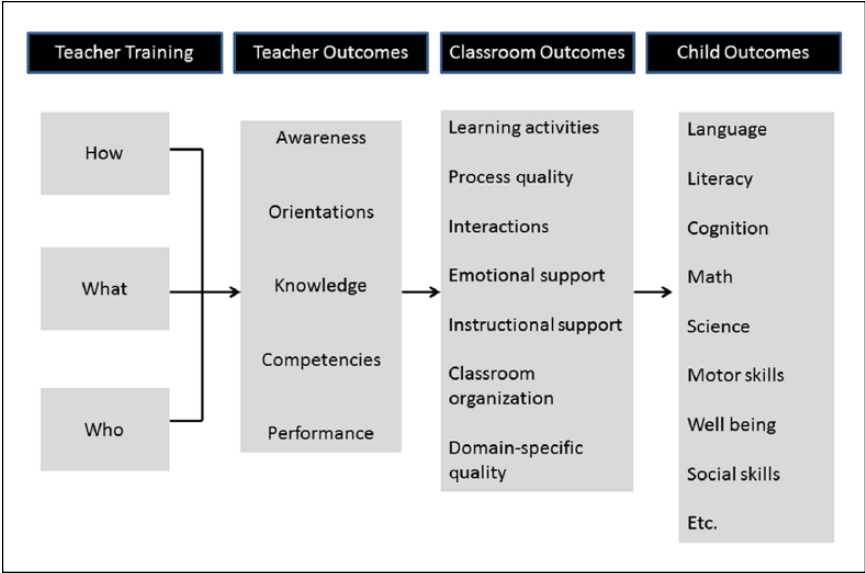


FIGURE 1. *Model of change.*

outcomes of PD that have been reported in the ECEC literature. Moreover, a deeper understanding between the relationship between quality improvements through PD training and child development is needed (Markussen-Brown et al., 2017). Theoretical models of change imply a multistep path from caregiver training to enhanced child outcomes (see Figure 1). It is assumed that teacher training and its components (who, what, and how) change latent (e.g., awareness, knowledge, and orientations) and observable teacher outcomes (e.g., teaching practice). The improved teacher outcomes and enhanced classroom quality subsequently improve student learning (see Fukkink & Lont, 2007; Hamre et al., 2012; Harvard Family Research Project, 2006; Klein & Gomby, 2008; Yoon et al., 2007). Only a small number of intervention studies provide data on the different kinds of outcomes (Fukkink & Lont, 2007; Markussen-Brown et al., 2017; Zaslow et al., 2010a), and based on a limited number of studies, the relation between teacher improvements in process quality and children's outcomes has not been convincingly demonstrated (Fukkink & Lont, 2007; Markussen-Brown et al., 2017). The present study aims to fill this gap by aggregating effects of PD both at teacher and child levels, including child measures from different developmental domains, because research has shown that child care quality predicts the development of young children in different domains (Burchinal, 2010). In sum, a meta-analytic review of experimental findings is needed to draw reliable conclusions on the benefits of in-service trainings, to explore mechanisms that influence experimental results, and to guide future investments in ECEC.

### *Aims of the Study*

First, the primary research objective of this study was to evaluate the impact of in-service programs for ECEC teachers on standardized quality ratings. Only standardized quality instruments were included for several reasons. These instruments (a) ensure the quality of ratings through intensive rater training and reliability check procedures before data collection, (b) have stable test-retest reliabilities to measure changes, and (c) predict child development.

The second research objective was to explore program characteristics that moderate the effects on quality ratings. To this end, we investigated effect modifiers based on methodological features of the studies and the most promising in-service approaches to improve child care quality. Third, this review investigated the link between in-service programs to child outcomes. Of particular interest was the amount of variance in developmental effects that could be explained by simultaneous changes in classroom quality through in-service training.

In accordance with other meta-analyses in educational research (Camilli, Vargas, Ryan, & Barnett, 2010; Fukkink & Lont, 2007; Piasta & Wagner, 2010), and also taking into account the limited number of randomized studies, intervention studies with and without control conditions (e.g., randomized, quasi-experimental, and nonexperimental studies) were included in the meta-analysis.

### **Method**

#### *Systematic Literature Search*

Relevant studies were retrieved through a systematic literature search procedure and the timeframe was set at the starting year of most databases in 1970, ending with 2011. The systematic literature search was double coded by two independent reviewers. First, an electronic search in English-language databases was conducted. Second, individual studies were sought in bibliographies, review articles, and renowned journals to complement our systematic electronic search; the selection of journals was based on the number of hits for a number of journals in our electronic search. Additionally, conference and convention programs were systematically screened. The electronic search was applied in ERIC, PsycINFO, ProQuest Dissertations and Theses, and ProQuest Educational Journals. The search included keywords related to in-service professional development (in-service, “professional development,” coaching, consulting, “technical assistance,” mentoring, “teacher training,” or “teacher education”), outcome measure (quality, performance, “teacher behavior,” “child outcomes,” “program improvement,” enrichment, or environment) and target (impact or effect\*), and type of education (“early childhood education” or kindergarten or preschool, or “child care”).

A manual search was executed to complement the systematic electronic search. We also searched prior published systematic reviews and meta-analyses for individual studies (i.e., Fukkink & Lont, 2007; Klein & Gomby, 2008; Zaslow et al., 2010b). Furthermore, the reference lists of studies that were identified as relevant for meta-analysis were also reviewed to identify additional studies. According to Rosenthal (1994), conferences play an important role in the dissemination of recently completed research and should be integrated in systematic search procedures. The following conference proceedings and venues were thus screened for

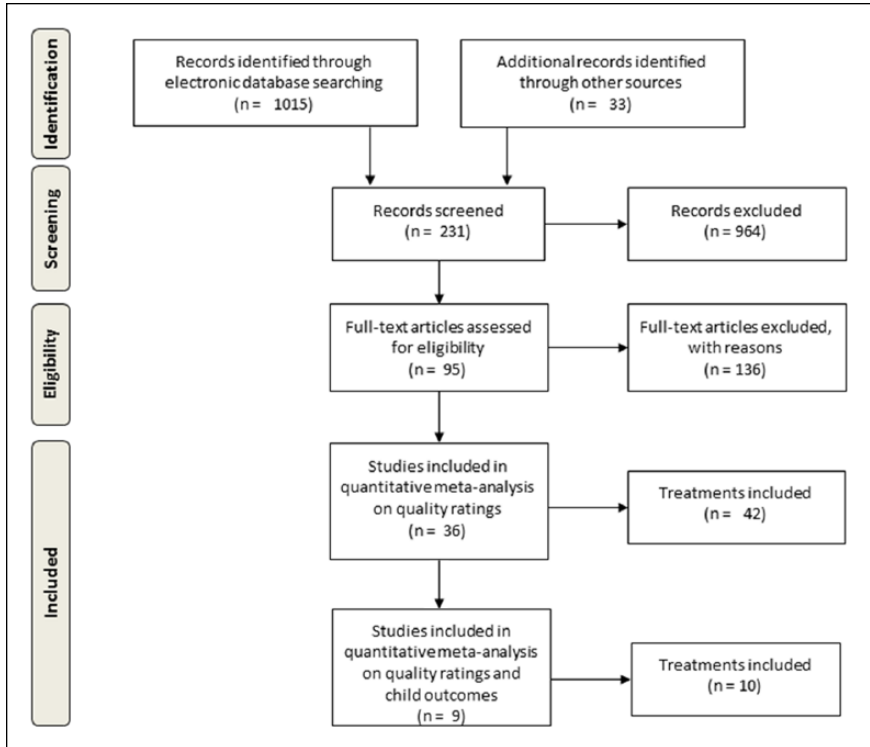


FIGURE 2. *Flow diagram.*

additional research: the biennial meeting of the European Association for Learning and Instruction, biennial meeting of the European Association for Learning and Instruction Special Interest Group 5, and conferences of the Society for Research on Educational Effectiveness; these conferences emerged from our systematic electronic search and could possibly yield additional reports. The goal of the conference proceedings search was to find relevant additional research projects; conference presentations were not included and only reports were used as primary sources for our review. Furthermore, we did a web search through Google, using the keywords from the electronic search and other combinations of keywords (i.e., “impact of in-service professional programs for preschool teachers”).

The search in the electronic databases resulted in an initial set of 1,015 hits (see Figure 2). ERIC revealed 703 hits ( $n = 124$  coded as relevant for full-text coding after title and abstract screening), PsycINFO 77 hits ( $n = 25$  relevant), ProQUEST D&T 84 hits ( $n = 20$  relevant), and ProQUEST Educational Journals 151 hits ( $n = 32$  relevant). Studies were included in the review if they met the following criteria: (a) the investigation of in-service programs was designed to improve the quality of child care or child development; (b) the sample included preschool, Pre-K, or



kindergarten teachers, as well as educators working in center-based care; (c) studies used a quantitative research design (control group comparison or one group pre–post measure); and (d) studies were published in English, as Egger, Zellweger, and Antes (1996) indicated that publications relevant to meta-analysis are uncommon in non–English-language journals. After the title and abstract screening by two independent coders, 235 references were identified as relevant and ordered or downloaded, but only 231 were available. Subsequently, a full-text review procedure with a short coding form was used to evaluate the 231 studies. All texts were coded by two independent reviewers. Further inclusion criteria were the following:

1. A focus on in-service PD for early childhood practitioners working in center-based child care
2. Quantitative outcome measures related to child care quality and/or child development
3. The sample contained children age 0 to 7 years in center-based care
4. Experimental or quasi-experimental designs (including one group pre–posttest designs)
5. Studies must report sufficient statistic information to compute effect sizes
6. Studies must be published in English for transparency and the possibility of replicating our search results, but studies were not limited to English-speaking countries

From the 231 studies, 85 were identified as relevant and as also providing sufficient statistical information. A further 44 papers were categorized as relevant but did not provide sufficient statistical data. The main authors of these 44 papers were contacted via email three times, 21 responded and 10 provided the missing statistical information. In summary, the multistep literature search resulted in a total of 95 studies that included data on in-service effects on quality ratings and/or child outcomes. Of these studies, 36, covering 42 different in-service treatments, were found that evaluated the impact of quality ratings. Finally, 9 studies covering 10 treatments provided both data on quality ratings and child outcomes.

We excluded qualitative studies, correlational studies, and studies that did not provide sufficient information on statistical data to estimate effect sizes (e.g., Armstrong, Cusumano, Todd, & Cohen, 2008; Fiene, 2002; Haskell, 1994; Honig & Hirallal, 1998, Miller & Bogatova, 2009). In particular, information on standard deviation and number of participants per condition were mostly missing. Because of the fact that self-evaluations of teachers and external ratings on quality are not reliable (Sheridan, 2000), we excluded studies with self-evaluation (e.g., Thornton, Crim, & Hawkins, 2009). In ECEC, classrooms are usually constituted of more than one teacher, and teacher behavior/performance (including teacher–child interactions and relationship, instructional practices) can vary significantly between teachers within one classroom. Measuring the behavior/performance of one teacher would not be sufficient to capture the mean quality of learning experience of children in a classroom. Therefore, we also excluded studies that did not measure child care quality externally with certified raters at classroom level (e.g.,

Arnett, 1989; Cain, Rudd, & Saxon, 2007; Fantuzzo et al., 1996; Girolametto, Weitzman, & Greenberg, 2003).

### *Coding of Studies*

Based on the guidelines from the Campbell Collaboration (n.d.), all studies were double coded by two independent coders, and disagreements between the coders were resolved through discussion and consensus. Interrater reliability was estimated with *Kappa* ( $\kappa$ ) for nominal variables and with intraclass correlations (ICCs) for ordinal and interval scaled variables.

A multistep full-text coding procedure was used to systematize information. First, studies were coded with a short screening form to evaluate the quality of studies and the sufficient provision of statistical data. The short screening form included coding on study design (e.g., control-group design, randomization, sample size), instructional features of PD, sufficient statistical information, and relevance to meta-analysis. Interrater reliability was excellent for study relevance ( $\kappa = 1$ ).

Second, relevant studies were coded with an extended coding schema, which gathered additional information on study methodology, in-service program, and potential effect modifiers (see Tables 1 and 3 for an overview of coded variables).

### *Coding of Study and Methodology*

We coded year of study, country, and type of publication (published in a peer-reviewed journal or not) for each in-service program. A number of methodological variables are considered meaningful effect modifiers in the psychological, behavioral, and educational treatment literature (see Cooper, Hedges, & Valentine, 2009; Lipsey & Wilson, 1993). At study and treatment levels, we coded assignment (random or not) and type of experimental design, distinguishing between randomized controlled trial, clustered randomized trial, quasi-experimental design with matching, and nonexperimental studies with pre–post test of intervention group. We also coded sample size for experimental and control group. At effect size level, we coded whether there were differences at pretest (experimental or control group had highest score) or the scale used in general. We also made a dummy variable for recognized scale (1 = *Classroom Assessment Scoring System [CLASS]*; *Early Language and Literacy Classroom Observation [ELLCO]*; *Environmental Rating Scales [ERS]*, including *ECERS* or *ITERS*; and *Individualized Classroom Assessment Scoring System [inCLASS]*). Reliability of the instruments was also coded and dummy variables were made that distinguished between interobserver agreement (1 = *Kappa* or ICC  $\geq .80$ ; 0 = *not*) and internal consistency (1 = *Cronbach's*  $\alpha \geq .70$ ; 0 = *not*).

### *Coding of Professional Development Components*

Various characteristics of PD may be relevant to our understanding of the implementation and effects of ECEC training programs (see Schachter, 2015; Sheridan et al., 2009; Snyder et al., 2011). The codes for PD delivery (how) and content (what) as well as participants (who) were based on the framework from Buyse et al. (2009).

To systemize and compare the “how” component, delivery formats of the training programs were coded into the following categories: workshop, coursework,

**TABLE 1**  
*Overview of studies included in meta-analysis on quality ratings*

Author(s)	Year	Country	Pubtype	Design	Sample	BA	AG	WS	CS	ONL	COP	ON	Month	Topic	Scale	Hours
Algozzine et al.	2011	USA	Pub	CRT	72		PK		x			x	12.00	CUR	ELLCO	360
Barnett et al.	2008	USA	Pub	RCT	18	x	PK	x		x			9.00	CUR	ERS, CLASS, SELA	59
Bloom and Sheerer	1992	USA	Pub	QNE	44	x	PK		x				16.00	LEAD	ECCOS	308
Boller et al.	2010	USA	Unpub	RCT	129		M			x			7.00	QI	ERS	56
Breffini	2011	USA	Pub	RCT	20		PK	x					2.00	CUR	CLASS	32
Brinks	2007	USA	Unpub	OGPP	31		PK	x	x				9.00	L&L	ELLCO	—
Bryant et al.	2009	USA	Unpub	RCT	108		PK			x			9.00	QI	ERS	30
Burchinal, Cryer, Clifford, and Howes	2002	USA	Pub	QNE	553		M	x					0.10	GEN	ERS	—
Buyse, Castro, and Peisner-Feinberg	2010	USA	Pub	CRT	55		PK	x		x			2.00	L&L	ELLCO	40
Campbell and Milbourne	2005	USA	Pub	QNE	180		IT		x				3.50	QI	ERS	18
Cassidy et al.	1995	USA	Pub	QNE	41		M	x					7.50	GEN	ERS	—
Clancy-Menchetti	2006	USA	Unpub	RCT	77		PK	x		x			24.00	L&L	ELLCO	276
Cunningham* (a)	2007	USA	Unpub	RCT	304		PK		x				3.00	L&L	ELLCO	30
Cunningham* (b)	2007	USA	Unpub	RCT	304		PK	x		x			9.00	L&L	ELLCO	94
Dickinson and Caswell	2007	USA	Pub	QNE	70		PK		x				5.00	L&L	ELLCO, AP	45
Domitrovich et al.	2009	USA	Pub	CRT	88		PK	x		x			9.00	CUR	CLASS	68
Edgar <sup>b</sup>	2008	USA	Unpub	OGPP	94		PK			x			2.25	L&L	ELLCO	18
Englund	2010	USA	Unpub	OGPP	6		PK			x			0.13	L&L	CLASS	4
Grace et al.	2008	USA	Pub	CRT	40		PK		x				36.00	L&L	ELLCO	—
Howe et al. (a)	2011	CAN	Pub	QNE	94		PK			x			3.75	CUR	ERS	45
Howe et al. (b)	2011	CAN	Pub	QNE	94		PK	x					3.75	CUR	ERS	8

(continued)

**TABLE 1 (continued)**

Author(s)	Year	Country	Pubtype	Design	Sample	BA	AG	WS	CS	ONL	COP	ON	Month	Topic	Scale	Hours
Japel (a)	2009	CAN	Unpub	CRT	30	PK		x					12.00	QI	ERS	—
Japel (b)	2009	CAN	Unpub	CRT	30	PK		x					0.10	QI	ERS	—
Lonigan, Farver, Phillips, and Clancy-Menchetti (a)	2011	USA	Pub	CRT	51	PK		x					9.00	CUR	ELCO, EC-COM	68
Lonigan et al. (b)	2011	USA	Pub	CRT	51	PK		x					9.00	CUR	ELCO, EC-COM	32
McNerney et al.	2006	USA	Pub	OGPP	23	PK		x					6.00	L&L	ELCO	—
Morris, Raver, Millenky, Jones, and Lloyd	2010	USA	Unpub	CRT	51	PK	x	x					9.00	CUR	(im)CLASS	162
Neuman and Cummingham (a)	2009	USA	Pub	RCT	304	PK			x				3.75	L&L	ELCO	45
Neuman and Cummingham (b)	2009	USA	Pub	RCT	304	PK			x				8.00	L&L	ELCO	85
Neuman and Wright (a)	2010	USA	Unpub	RCT	148	PK		x					2.50	L&L	ELCO	30
Neuman and Wright (b)	2010	USA	Unpub	RCT	148	PK				x			2.50	L&L	ELCO	30
Onchwari and Keengwe	2010	USA	Pub	QNE	44	PK				x			5.00	L&L	ELCO	40
Palsha and Wesley	1998	USA	Pub	OGPP	73	M		x					7.68	QI	ERS	30
Powell, Diamond, Burchinal, and Koehler (a)	2010	USA	Pub	RCT	88	PK			x				9.00	L&L	ELCO	—
Raver et al.	2008	USA	Pub	RCT	94	PK			x				7.00	SOC	ERS, CLASS	132
Roehrig, Dubosarsky, Mason, Carlson, and Murphy	2011	USA	Pub	OGPP	37	PK		x	x				24.00	SCI	CLASS	—

(continued)

TABLE 1 (continued)

Author(s)	Year	Country	Pubtype	Design	Sample	BA	AG	WS	CS	ONL	COP	ON	Month	Topic	Scale	Hours
Sheridan, Clarke, Knoche, and Edwards	2006	USA	Pub	QNE	61	PK		x					12.00	QI	ERS	—
Sibley and Sewell	2011	USA	Pub	QNE	72	PK	x	x					36.00	CUR	ELCO	—
Sipp	2010	USA	Unpub	OGPP	24	PK			x				1.50	QI	CLASS	12
Sprague, Glantz, Koppelli, Raya-Carlton, and Schilder	2009	USA	Unpub	RCT	20	PK			x				5.00	CUR	ELCO, ELA, LOC	105
Uttley and Horm	2008	USA	Pub	OGPP	21	M	x	x					8.00	QI	ERS	—
Wasik and Hindman	2011	USA	Pub	CRT	30	PK	x	x					9.00	L&L	ELCO, CLASS	50

*Note.* Pubtype = publication type; Pub = published in journals or as book chapter; Unpub = unpublished paper, e.g., technical paper or report, dissertation; design = study design; BA = bachelor's degree or higher; AG = age group; PK = preschool and kindergarten classrooms; IT = infant and toddler classrooms; M = mixed classrooms; WS = workshop; CS = course and in-service with ongoing sessions; ONL = online training or blended learning; COP = community of learning or community of practice; ON = onsite support like coaching or consulting; FOC = focus of training; CUR = curriculum implementation; L&L = language and literacy; SOC = social and behavioral development; QI = quality improvement initiative; SCI = science; GEN = general focus of early childhood education (e.g., developmental appropriate practice); RCT = randomized control trial assigning teacher to groups; CRT = clustered randomized trial assigning classrooms or center to groups; QWM = quasi-experimental design with matching procedure/equivalent groups; QNE = quasi-experimental design without randomization or matching/non-equivalent groups design; OGPP = one group pre-post measure; AP = Assessment Profile Learning Environment; CLASS = Classroom Assessment Scoring System; ECCOM = Early Childhood Classroom Observation Measure; ECCOS = Early Childhood Classroom Observation Scale; ELARS = Early Literacy Activity Rating Scale; ELCO = Early Language and Literacy Classroom Observation; ERS = Environmental Rating Scales (e.g., ECERS or ITERS); (in)CLASS = Individualized Classroom Assessment Scoring System; LOC = Literacy Observation Checklist; SELA = Support for Early Literacy Assessment.

a. Cunningham (2007) and Neuman and Cunningham (2009) use data from the Project Great Start Professional Development Initiative, but report different results

b. Edgar (2008) compared three different coaching approaches, but statistical data were only available for master coaching.

online and distance training, community of practice, and onsite support. Based on this coding, we determined whether programs included multiple or single formats. Additionally, we differentiated onsite support in coaching, mentoring, technical assistance, or consulting. The intensity of the program was also determined, including both dosage (number of hours) and the duration of the program (in months).

Coding the “what” component, training focus could be curriculum implementation, language and literacy, social and behavioral development, quality improvement initiative, science, or a general focus on ECEC (e.g., developmental appropriate practice). For moderator analysis, we dummy-coded whether the training focused on curriculum implementation or not. We also dummy-coded whether the training was scale-based (e.g., when training content was based on the quality rating scale that is used to measure in-service training effects), and whether participants received feedback or quality profiles based on standardized quality ratings.

Finally, we coded the “who” component concerning classroom compositions with two dummies: whether the children in the classrooms were below the age of 3 years and whether they were considered an at-risk group. If possible, we coded whether all teachers in the sample had a university degree or not.

### *Analysis*

Statistical information was transformed into the effect size Hedges’  $g$  using Comprehensive Meta-Analysis Software V2 (Borenstein et al., 2005). An aggregated effect size was calculated for each in-service treatment evaluated. Furthermore, all treatment effects were integrated into a summary effect size with a random effects model, nesting the effect sizes under treatments. With CMA, a weighted summary effect  $g'$  was calculated and weights were given based on the variance of effect sizes within a treatment. In addition to the aggregation of research results, meta-regression was used to examine theoretical moderators and methodological potential effect modifiers with MLwiN (Rasbash, Charlton, Browne, Healy, & Cameron, 2005). In Model 1, we checked for any statistically significant relation between effect sizes and methodological variables or PD features. By means of hierarchical regression analysis, in Model 2 we first included significant methodological characteristics and tested whether training-related characteristics explained additional variance in study results. The models were determined using the restricted maximum likelihood method. A chi-square test was used to test the significance of residual variance (Bryk & Raudenbush, 2002; Hox, 2010).

## **Results**

### *Description of Studies*

Approximately one third of the studies were unpublished (e.g., research reports, conference papers, or dissertations) and two thirds were published studies ( $n = 25$ ). Almost all studies were conducted in the United States, except two from Canada (Howe et al., 2011; Japel, 2009). The majority of studies ( $n = 19$ ) had experimental designs: randomized controlled trial with randomization at individual level ( $n = 11$ )

and cluster randomized trials ( $n = 8$ ) with randomized assignment of classrooms or centers. A further nine studies applied a quasi-experimental design with nonrandomized assignment of participants, while eight investigations used a nonexperimental design evaluating the progress of a treatment group without control condition.

Overall, data from 2,891 teachers were included in the analysis. Sample size varied from 6 to 553 educators at the beginning of the studies. In three samples, all teachers had a university degree such as Bachelor or Master (Barnett et al., 2008; Bloom & Sheerer, 1992; Morris et al., 2010). The other study samples consisted of teachers with mixed qualification level. Thirty-six training programs were evaluated in preschool and kindergarten classrooms and six training programs included data from infant/toddler classrooms. Twenty-three of all in-service approaches applied multiple delivery formats. One third of in-service programs used a single delivery strategy, either a workshop ( $n = 4$ ), a course ( $n = 4$ ), or onsite support ( $n = 5$ ). The duration of the PD approaches ranged from very short programs of half-day workshops or 4 days of coaching (e.g., Burchinal et al., 2002; Englund, 2010; Japel, 2009) to 3-year programs. Most PD programs focused on language and literacy (14 studies with 17 different treatments), while 8 studies covering 9 treatments focused on quality improvement in ECEC classrooms. In 9 studies covering 11 treatments, curriculum implementation was the major focus. The other studies emphasized general topics of ECEC ( $n = 2$ ), social-emotional development of young children, or science. Most of the studies ( $n = 35$ ) used internationally recognized quality rating scales.

In total, 17 studies measured the impact of PD programs at classroom level with ELLCO, 12 with ERS, and 9 with CLASS. Almost all programs with a focus on language and literacy themes measured the impact with ELLCO (16 of 17) and nearly all quality improvement approaches applied ERS (8 of 9). Overall, training dosage ranged from 4 hours to 308 hours, 8 studies, reporting 9 treatments, evaluated scale-based training sessions. For 13 studies reporting the results of 16 treatments, a curriculum or a set of specific manual-based activities were included in the PD. However, only 9 studies covering 11 treatments reported that PD focused on curriculum implementation as a major theme. Descriptive data for the 42 different in-service treatments, reported in 36 studies, are listed in Table 1.

### *Meta-Analysis of Quality Ratings*

Overall, 289 effect sizes could be extracted from contrasts of experimental and control groups at posttest, or change of scores from pre- and posttest on external quality ratings. Effect sizes ranged from strong negative impact ( $g = -0.82$ ) to very large positive effects ( $g = 6.62$ ; see Figure 3). Approximately 25% of effect sizes ( $k = 73$ ) showed no or a negative impact ( $g < 0.20$ ), 68 were small effects ( $g \geq 0.20$ ), 62 medium effect sizes ( $g \geq 0.50$ ), and approximately 30% ( $k = 86$ ) revealed large effects ( $g \geq 0.80$ ) of in-service programs on quality ratings, according to Cohen's (1988) rules of thumb.

A heterogeneity analysis, conducted with CMA (Borenstein, Hedges, Higgins, & Rothstein, 2005), indicated significant between-study variance ( $Q = 746.11$ ;  $df = 41$ ;  $p < .001$ ;  $T^2 = .189$ ;  $SE = 0.072$ ), with a high percentage (95%) of systematic variance between treatments ( $I^2 = 94.505$ ). These heterogeneity statistics motivated the decision to use a random effects model and to conduct a meta-regression.

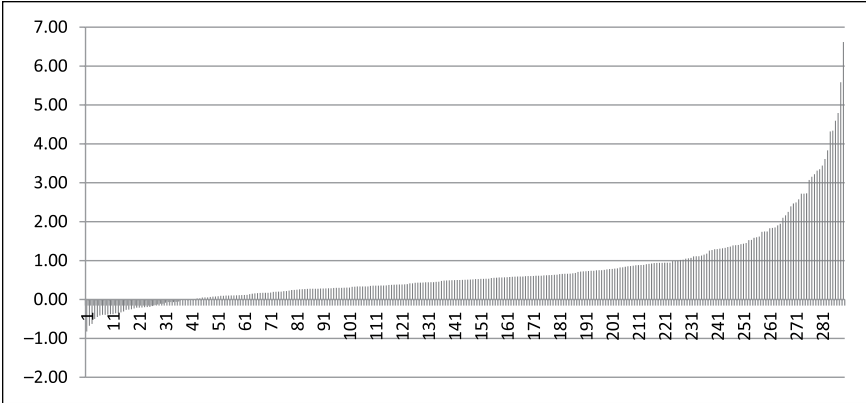


FIGURE 3. *Distribution of effect sizes for educational quality ratings.*

The meta-analysis showed a significant medium weighted summary effect size ( $g' = 0.68$ ;  $SE = 0.071$ ;  $k = 289$ ; 95% confidence interval [CI] = 0.55–0.82;  $Z = 9.70$ ;  $p < .001$ ; relative weight per treatment 0.55% to 2.79%) for in-service programs on quality ratings. There was no significant difference ( $Q = 0.001$ ;  $df = 1$ ,  $p = .973$ ) between the summary effect of experimental studies ( $g' = 0.69$ ;  $SE = 0.079$ ;  $k = 207$ ; 95% CI = 0.54–0.84;  $Z = 8.77$ ;  $p < .001$ ) and the summary effect of non-experimental studies without control group ( $g' = 0.68$ ;  $SE = 0.072$ ;  $k = 82$ ; 95% CI = 0.32–1.04;  $Z = 9.52$ ;  $p < .001$ ), and we decided to combine both study types in the data set. Orwin’s Fail Safe  $N$  analysis indicated that at least 50 studies with a mean effect size of zero are needed to decrease the summary effect size to a negligible impact of  $g < 0.20$ . A test of sensitivity suggested a robust aggregated effect that slightly varied between  $g = 0.62$  and  $g = 0.71$  if one treatment was removed. The funnel plot (see Figure 4) and the Egger regression (intercept = 3.91;  $SE = 1.08$ ;  $p < .001$ ) suggested an asymmetry, with small studies with high effects missing from the data.

*Moderator Analysis on Quality Ratings*

Conducting an unweighted multilevel analysis with MLwiN, the summary effect showed a similar result ( $g = 0.73$ ;  $SE = 0.122$ ;  $k = 289$ ; RIGLS = 566.6;  $p < .001$ ) and can be used as a reference for subgroup effect sizes in Table 2. First, methodological variables were examined through multilevel moderator analysis. Table 2 displays direct effect size estimates of subgroup analysis for dichotomous moderators. For continuous interval scaled moderators, we report the regression coefficients that indicate a change in effect size with a unit increase in the exploratory moderator variable.

In Model 1 (see Table 2), all moderator variables were analyzed individually. This analysis showed that effect sizes in unpublished studies were considerably lower than effect sizes in published studies. Considering methodological issues, the use of different quality rating scales moderated the outcomes. The impact was



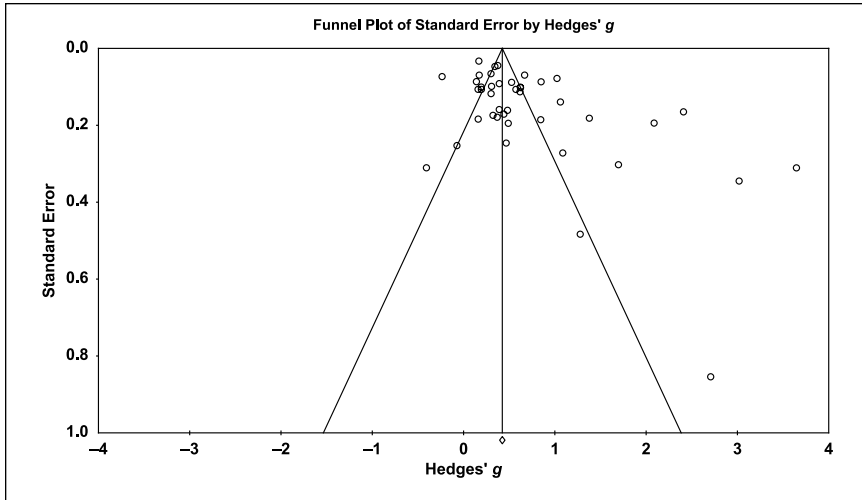


FIGURE 4. *Funnel plot of in-service effects on quality ratings.*

significantly higher when recognized scales ( $g = 0.76$ ), such as CLASS, ERS, or ELLCO, were applied, compared with newly developed rating instruments ( $g = 0.41$ ). Comparing the recognized scales, outcomes using CLASS had higher values ( $g = 1.04$ ), while outcomes measured with ERS or ELLCO were in line with the summary effect. All reported outcomes were coded for instrument reliability and differences in mean value at pretest on effect size level. Meta-regression showed that outcomes where the experimental group had slightly higher values than the control group at the beginning resulted in higher effect sizes at posttest ( $g = 0.99$ ). Furthermore, effect sizes were significantly lower for outcomes ( $g = 0.53$ ) where teachers in the control group had lower values in quality ratings at pretest. The effect size of outcomes with a highly reliable quality rating scale, defined as Cronbach's  $\alpha > .70$ , was significantly higher in comparison with outcomes where lower reliability scores or no reliability values were displayed. Interrater reliability was also coded as a potential effect modifier and effect sizes were lower for less reliable observational measures. Effect sizes were not related to treatment comparison type or the procedures used to assign teachers to the treatment or control group. Finally, the effect sizes from studies published before and after the millennium were similar.

*Background Characteristics Related to “Who”*

Focusing on trainee characteristics, in-service effectiveness did not differ between samples consisting exclusively of teachers with a university degree and those with mixed qualification level. Background variables of child care centers and classrooms were also investigated as moderators of training effectiveness, but were found to be nonsignificant. Neither classroom conditions with children at risk nor conditions with children under the age of 3 years were significantly related to the effect size of interventions.

**TABLE 2***Meta-analysis on quality ratings*

Study	<i>g</i>	<i>SE</i>	Lower CI	Upper CI	Relative Weight in %
Algozzine et al. (2011)	0.59	0.11	0.37	0.80	2.63
Barnett et al. (2008)	0.63	0.12	0.40	0.86	2.61
Bloom and Sheerer (1992)	1.07	0.14	0.80	1.35	2.52
Boller et al. (2010)	3.66	0.31	3.05	4.28	1.81
Breffini (2011)	3.03	0.35	2.35	3.71	1.67
Brinks (2007)	2.10	0.20	1.71	2.49	2.30
Bryant et al. (2009)	0.21	0.11	-0.01	0.42	2.63
Burchinal et al. (2002)	0.36	0.05	0.26	0.45	2.77
Buysse et al. (2010)	0.17	0.11	-0.04	0.39	2.63
Campbell and Milbourne (2005)	0.17	0.19	-0.19	0.54	2.35
Cassidy et al. (1995)	0.50	0.20	0.11	0.89	2.30
Clancy-Menchetti (2006)	-0.39	0.31	-1.01	0.22	1.81
Cunningham (a) (2007)	0.38	0.18	0.02	0.74	2.37
Cunningham (b) (2007)	0.86	0.19	0.49	1.23	2.34
Dickinson and Caswell (2007)	1.04	0.08	0.88	1.20	2.71
Domitrovich et al. (2009)	0.40	0.10	0.22	0.59	2.67
Edgar (2008)	0.39	0.05	0.29	0.48	2.77
Englund (2010)	2.72	0.86	1.04	4.40	0.55
Grace et al. (2008)	2.42	0.17	2.09	2.75	2.42
Howe et al. (a) (2011)	1.10	0.28	0.56	1.64	1.97
Howe et al. (b) (2011)	-0.06	0.26	-0.56	0.44	2.05
Japel (a) (2009)	0.49	0.16	0.17	0.82	2.44
Japel (b) (2009)	0.40	0.16	0.09	0.72	2.45
Lonigan et al. (a) (2011)	0.32	0.10	0.12	0.52	2.65
Lonigan et al. (b) (2011)	0.21	0.10	0.00	0.41	2.65
McNerney et al. (2006)	0.18	0.07	0.04	0.33	2.73
Morris et al. (2010)	0.18	0.04	0.11	0.25	2.79
Neuman and Cunningham (a) (2009)	0.31	0.07	0.18	0.45	2.74
Neuman and Cunningham (b) (2009)	0.68	0.07	0.54	0.82	2.73
Neuman and Wright (a) (2010)	0.15	0.09	-0.02	0.33	2.69
Neuman and Wright (b) (2010)	0.54	0.09	0.36	0.72	2.68
Onchwari and Keengwe (2010)	0.45	0.17	0.11	0.80	2.40
Palsha and Wesley (1998)	0.86	0.09	0.69	1.04	2.69
Powell et al. (2010)	1.39	0.19	1.03	1.75	2.36
Raver et al. (2008)	0.64	0.11	0.43	0.84	2.64
Roehrig et al. (2011)	0.64	0.10	0.43	0.84	2.65
Sheridan et al. (2006)	1.29	0.49	0.34	2.24	1.20
Sibley and Sewell (2011)	0.31	0.12	0.08	0.55	2.60
Sipp (2010)	-0.22	0.08	-0.37	-0.07	2.72
Sprague et al. (2009)	0.34	0.18	-0.01	0.68	2.39
Uttley and Horm (2008)	0.48	0.25	-0.01	0.97	2.08
Wasik and Hindman (2011)	1.71	0.31	1.11	2.31	1.84
Weighted summary effect <i>g'</i>	0.68	0.07	0.55	0.82	

Our analysis did not provide evidence that large-scale PD is less effective than small-scale programs, using the number of providers (regression coefficient =  $-0.004$ ;  $SE = 0.007$ ;  $k = 120$ ) and the number of training participants (regression coefficient =  $-0.001$ ;  $SE = 0.001$ ;  $k = 289$ ) as large-scale indicators. The training effect did not decrease systematically with larger numbers of participants or trainers; however, information on the number of in-service providers was only available for some studies.

#### *Intervention Characteristics Related to “How”*

Investigating in-service delivery format as a modifier, in-service programs that solely used coaching as a training strategy ( $g = 1.98$ ) were nearly three times more effective than other programs ( $g = 0.67$ ). Findings did not show superior results for interventions with onsite support only or onsite support in combination with other delivery formats (e.g., workshops or courses). In addition, the combination of multiple delivery formats in a treatment was not found to be more successful than intervention formats using a single training strategy.

Our analysis did not support the suggestion that a longer duration (regression coefficient =  $0.010$ ;  $SE = 0.015$ ) and higher in-service dosage in hours (regression coefficient =  $-0.001$ ;  $SE = 0.002$ ) lead to higher in-service effects in general. An explorative analysis that split duration and dosage showed that in-service programs with a medium training dosage of 45 to 60 hours ( $g = 1.93$ ;  $n = 6$ ) were significantly more effective than other programs that were either shorter or longer in duration.

#### *Intervention Characteristics Related to “What”*

The content of scale-based training is teaching and intervening practices that are assumed to predict beneficial child outcomes and are related to a specific quality rating scale. It is hypothesized that scale-based PD approaches guide in-service providers in relation to work with teachers and provide feedback to teachers on learning materials and instructional sequencing and delivery (see McNerney et al., 2006). However, our moderator analysis did not support the assumption that scale-based training is more effective than training which is not based on quality rating or performance scales (regression coefficient =  $0.435$ ;  $SE = 0.295$ ;  $k = 279$ ). Furthermore, Klein and Gomby (2008) suggested that the curriculum might contribute to the positive effects of in-service programs. Our analysis did not provide significant evidence for the superiority of trainings with a focus on implementation of a curriculum with prescribed training content, material, and directions for actions. Related to this, a focus on the implementation of a curriculum (regression coefficient =  $-0.169$ ;  $SE = 0.278$ ), or curriculum and manual-based strategies as part of the training, were also not significant predictors (regression coefficient =  $-0.172$ ;  $SE = 0.260$ ).

In Model 2 (see Table 3), we tested the significance of theoretical effect modifiers, adjusting for methodological characteristics. Statistically significant methodological moderators, such as recognized scale (1 = *recognized scales* [CLASS; ERS; ELLCO]; 0 = *not*), the reliability of the single outcome reported (1 = Cronbach's  $\alpha > .70$ ), and mean differences at pretest (1 = *when control group scored higher*), were included as covariates. All covariates remained statistically significant when testing the exploratory variables below, while correlations between covariates of

**TABLE 3***Moderator analysis of in-service effects on quality ratings*

Characteristic	<i>k</i>	<i>g</i>	Model 1 ( <i>k</i> = 289)		Model 2 ( <i>k</i> = 289)	
			Residual variance	$\Delta X^2$	Significance of moderator	Residual variance
<b>Publication type</b>						
Published	195	0.75	.597	780*		
Unpublished	94	0.69*	.597	780*		
Published after 1999	261	0.75	.598	1,447*		
<b>Methodology</b>						
Recognized scale						
Yes	260	0.76*	.579	1,303*		
No	29	0.41*	.579	1,303*		
CLASS	66	1.04*	.580	248*		
ELLCO	127	0.77	.590	740*		
ERS	67	0.85	.594	279*		
Difference at pretest						
Experimental group higher score	101	0.85*	.565	447*		
Control group higher score	25	0.53*	.567	156*		
Reliability						
Cronbach's $\alpha \geq .70$	92	0.99*	.589	470*		
<.70 or not reported	197	0.64*	.589	470*		
Interrater reliability						
<i>Kappa</i> or <i>ICC</i> $\geq .80$	157	0.81	.693	903*		
<i>Kappa</i> or <i>ICC</i> < .80 or not reported	132	0.68*	.603	903*		
Treatment comparison type						
Contrast EG vs. CG	207	0.73	.601	1,075*		
Change from pre to post	82	0.75	.601	485*		
Random assignment	181	0.74	.599	948*		
<i>The "who" component</i>						
Teacher with university degree						
Classroom composition/setting						
At-risk population (children)	204	0.65	.596	1,021*		
Children under age 3	57	0.71	.600	319*		
<i>The "how" component</i>						
In-service intensity						
Dosage						
Less than 30 hours	53	0.44	.562	410*		
30 to <45 hours	43	0.85	.597	210*		
45 to <60 hours	28	1.93*	.484	139*	Sig.	.489
60 to <100 hours	30	0.45	.591	150*		
100 to <120 hours	6	0.34	.596	5		

*(continued)*

**TABLE 3 (continued)**

Characteristic	k	g	Model 1 (k = 289)		Model 2 (k = 289)	
			Residual variance	$\Delta X^2$	Significance of moderator	Residual variance
120 to <170 hours	27	0.41	.594	212*		
>170 hours	16	0.46	.596	34		
<i>Duration</i>						
Less than 3 months	50	0.69	.598	499*		
3 to <6 months	61	0.44	.580	241*		
6 to <9 months	129	0.87	.584	662*		
9 to <12 months	19	0.74	.599	24		
12 to <24 months	19	0.47	.596	26		
24 to 36 months	11	1.35	.574	108*		
<i>In-service delivery</i>						
Onsite is part of program	226	0.76	.597	1,281*		
Coaching is part of program	79	0.11	.541	465*		
Consultation is part of program	55	0.37	.577	313*		
Mentoring is part of program	45	0.52	.595	172*		
Only coaching			.520	123*	Sig.	.524
Yes	14	1.98*				
No	275	0.67*				
Onsite support only			.568	190*		
Yes	41	1.20				
No	248	0.64				
Multiple delivery formats	186	0.65				
<i>The "what" component</i>						
Scale-based training	94	0.62	.593	420*		
<i>Curriculum implementation</i>						
Mayor focus of training	102	0.61	.594	384*		
Content part of the training	118	0.72	.596	434*		

Note. Nonsignificant results in Model 2 were blank.

\* $p < .05$ . Sig = significant result at  $p < .05$  level.

Model 2 were modest ( $r < .22$ ). After controlling for methodological covariates, delivery format and in-service program intensity remained significant moderators in Model 1 and Model 2. More specifically, findings on PD programs with an intensity of 45 to 60 hours were still related to favorable effect sizes. Furthermore, ECEC programs that solely provided coaching were again associated with superior quality improvement. After adjusting for methodological characteristics, the moderating effect of publication type was not confirmed in Model 2.

*Meta-Analysis on In-Service Effects on Both Quality and Child Development*

Nine studies covering 10 treatments evaluated the impact of in-service programs on both quality ratings and child outcomes. Overall, data from 486 teachers

and 4,504 children were included in this meta-analytic review. All studies were conducted in the United States and had been recently published (2006–2011). The in-service programs consisted of intensive PD approaches with a duration of 4 to 24 months and 30 to 216 training hours. Most of the treatments ( $n = 7$ ) used a combination of delivery methods (e.g., workshop, course, and onsite).

For the aggregation of in-service effects on quality ratings for the studies with teacher and child outcomes, 82 effect sizes were available at classroom level. The aggregation of effect sizes per treatment showed that treatment effects, ranging from  $g = -0.39$  to 1.71, showed serious variation; only the study by Clancy-Menchetti (2006) reported a negative effect of in-service training on quality ratings. The meta-analysis of teacher effects for the subset of studies revealed a significant weighted summary effect of  $g' = 0.45$  ( $SE = 0.11$ ;  $k = 82$ ; 95% CI = 0.24–0.67;  $Z = 4.21$ ;  $p < .001$ ; relative weight per treatment 6.13% to 12.51%).

The studies yielded 68 effect sizes based on posttest contrast immediately after the in-service training at child level, including 53 effects based on language and literacy scores; 8 on social-behavioral ratings; 6 on assessment of cognition, knowledge, and school readiness; and one effect size on early mathematics testing. Effect sizes ranged from  $g = -0.24$  to 0.44. The aggregated effect sizes at study level showed a moderate range from  $g = 0.07$  to 0.22 (see Table 4). The weighted summary effect of  $g' = 0.14$  ( $SE = 0.019$ ; 95% CI = 0.10–0.17; relative weight per treatment 5.38% to 15.77%) was in the lower range, but was statistically significant ( $Z = 7.30$ ;  $p < .001$ ). The linear regression analyzed the empirical relationship between observed quality improvements and observed effects on child development after training. The aggregated effects per treatment on quality ratings were related to all effect sizes on child outcomes. Quality gains during training predicted the in-service effects on child outcomes ( $\beta = 0.73$ ;  $B = 0.06$ ;  $p = .017$ ). Of the variance in effect sizes on child development, 53% was “elucidated” by in-service effects on quality ( $R^2 = .53$ ). The model thus explained a substantial proportion of student achievement, although the summary effect on child outcomes was modest.

## **Discussion**

The first meta-analysis in our study showed that in-service PD improves the quality of ECEC. The second meta-analysis based on a subset of studies indicated that in-service PD supports child development when substantial improvements in ECEC quality appeared during the program. These studies allow an investigation of the assumed two-step causal link between quality improvements of ECEC and developmental outcomes for young children that is of interest to policy makers and stakeholders around the globe (see Fukkink & Lont, 2007; Klein & Gomby, 2008; Werner et al., 2016; Yoon et al., 2007). The findings of our meta-analytic review show that PD of ECEC staff enhances pedagogical quality, which is a key mechanism to accelerate developmental outcomes in young children.

The magnitude of a medium effect size for in-service programs at classroom level, as found in this meta-analysis, is in line with other meta-analyses and systematic reviews that synthesized findings on the impact of PD at teacher level (see Fukkink & Lont, 2007; Isner et al., 2011; Markussen-Brown et al., 2017; Werner et al., 2016; Yoon et al., 2007; Zaslow et al., 2010a, 2010b). According to general guidelines for the interpretation of effect sizes in the educational domain (see

**TABLE 4**  
*In-service effect on child development*

Study	Child outcomes					Classroom quality				
	<i>g</i>	<i>SE</i>	Lower CI	Upper CI	Relative weight in %	<i>g</i>	<i>SE</i>	Lower CI	Upper CI	Relative weight in %
Barnett et al. (2008)	0.12	0.07	0.00	0.25	6.05	0.63	0.12	0.40	0.86	11.07
Bryant et al. (2009)	0.10	0.07	-0.04	0.24	5.38	0.21	0.11	-0.01	0.42	11.22
Buyse et al. (2010)	0.12	0.04	0.05	0.19	12.46	0.17	0.19	-0.01	0.42	11.22
Clancy-Menchetti (2006)	0.07	0.03	0.01	0.12	15.77	-0.39	0.31	-1.01	0.22	6.13
Lonigan et al. (a) (2011)	0.15	0.03	0.09	0.22	13.30	0.32	0.10	0.12	0.52	11.40
Lonigan et al. (b) (2011)	0.13	0.04	0.06	0.20	12.73	0.21	0.10	0.00	0.41	11.36
Morris et al. (2010)	0.07	0.06	-0.04	0.18	7.32	0.18	0.04	0.11	0.25	12.51
Onchwari and Keengwe (2010)	0.22	0.06	0.11	0.34	7.35	0.45	0.17	0.11	0.80	9.54
Powell et al. (a) (2010)	0.21	0.03	0.14	0.27	13.85	1.39	0.19	1.03	1.75	9.25
Wasik and Hindman (2011)	0.19	0.07	0.06	0.32	5.73	1.71	0.31	1.11	2.31	6.29
Weighted summary effect <i>g'</i>	0.14	0.02	0.10	0.17		0.45	0.11	0.23	0.66	

*Note.* *g* = Hedges' *g*/effect size; *SE* = standard error; CI = confidence interval.

Lipsey & Wilson, 1993; What Works Clearinghouse, 2013), in-service programs have important effects on classroom quality ratings. In addition, the small effects on developmental outcomes in the cognitive and socioemotional domain in early childhood may result in meaningful financial returns for society in the long-term, as the cost–benefit analysis by Heckman (2008) demonstrated for early intervention programs for at-risk populations. The in-service effect on quality ratings based on the small subset of studies that provided data at child and classroom levels was smaller in comparison to the first meta-analysis ( $g' = 0.45$  for the subset of studies, compared with 0.68 for all included studies). It seems that the teacher interventions included in the second meta-analysis were less effective in improving classroom practice and this might explain why the impact of PD on child development was relatively small in comparison with other meta-analyses (see Fukkink & Lont, 2007; Markussen-Brown et al., 2017; Werner et al., 2016). Nevertheless, the discrepancy between teacher and child effects suggests that major improvements are needed at teacher and classroom levels before enhanced child development in ECEC may be expected.

*Implications for Practice and Future Research Based on Findings*

*Effect Moderators Based on Training Design (“How”)*

The dosage was related to the effects of the in-service programs, although not in a straightforward manner. In contrast to Markussen-Brown et al. (2017), our moderator analysis suggested that the dosage is associated with PD outcomes, not in a linear but a curvilinear relationship. More specifically, programs with a duration of 45 to 60 hours appeared most effective in changing classroom practice, with both shorter and longer programs showing less positive results. Werner et al. (2016) set the bar at 10 hours, but found a similar curvilinear relationship for training with more or less intensity. Short-term programs might be sufficient when a set of specific skills is the target of a program, and long-term and intensive training may only be needed when the focus of PD is broad and comprehensive (Hamre & Hatfield, 2012). Process quality measures focus on several facets of child care (e.g., learning materials, daily structure, learning activities, interactions; see Burchinal, 2010) and offer various goals for PD. This might explain why greater intensity efforts are related to significant changes in classroom practice. In addition, Markussen-Brown et al. (2017) found that course intensity was not related to process quality, but coaching intensity did. Although the number of studies was relatively small, interventions that only consisted of coaching appeared to be relatively effective; this positive effect remained after correcting for methodological moderators. Coaching seems, therefore, an effective element of in-service programs (Isner et al., 2011; Markussen-Brown et al., 2017; Werner et al., 2016). For the planning of effective PD, we suggest therefore a training intensity of PD programs that is related to the scope of the program and the inclusion of a substantial number of hours for individual coaching.

*Effect Moderators Based on Content (“What”)*

Klein and Gomby (2008) illustrated that curriculum implementation is a frequent target in PD. In our analysis, we could not find differences in effects of in-service programs when the content of training concerned the



implementation of a specific curriculum with explicit classroom practices and instructions. The findings on the beneficial effect of curriculum implementation through in-service training are still not conclusive and further research is needed. Scale-based training seems an interesting intervention and offers the possibility of providing individual and concrete feedback to teachers on proven learning materials and instructional practices (see McNerney et al., 2006). From a methodological point of view, scale-based PD with its close alignment between training and outcome measures may be considered as “training to the test” and one may consider the magnitude of the experimental effects to be biased. Despite this, regular and scale-based training showed similar findings. Thus, a positive effect or a positive bias effect of scale-based training was not found in our review. This result suggests that it is not sufficient for quality improvement to show teachers a quality rating scale and encourage them to practice what is defined as high-quality child care without further guidance and coaching.

#### *Effect Moderators Based on Participants and Context (“Who”)*

Several authors have emphasized that teacher characteristics and the workplace context affect the results of PD (Buysse et al., 2009; Klein & Gomby, 2008). In our review, we found similar effect sizes for teachers with or without a university degree, and the positive effects of PD may benefit teachers with different backgrounds. The training impact at teacher level was also equal for interventions for teachers in infant/toddler, preschools, or kindergarten classrooms. In line with Werner et al. (2016), teachers working in centers serving at-risk children or working in regular child care centers showed equal progress in our review. This may indicate that applied in-service training is adaptive to the learning needs of the target participants and their professional context, which is an important characteristic of such training (Buysse et al., 2009). However, several factors that promote training transfer effects (see Blume et al., 2010), including trainee characteristics (e.g., motivation, self-efficacy, or knowledge) need further study. Furthermore, the majority of research papers did not provide sufficient information on the providers of PD. It was not possible, therefore, to analyze the role of the trainer or coach in the PD processes. In line with Sheridan et al. (2009), we suggest that more research is needed to investigate the influence of the characteristics of the trainer (e.g., experience, qualification, profession) and the relationship between trainer and trainee.

#### *Methodological Effect Moderators*

Our meta-analysis also showed some relationships between methodological study characteristics and experimental outcomes. The outcomes were moderated by pretest differences between the experimental and control groups, the reliability of outcome measures, and unpublished studies, with all relationships in the expected direction. This finding stresses the importance of experimental studies with rigorous evaluation designs, reliable measures, and statistical adjustment for initial differences between conditions (Slavin, 2008). It should be noted, though, that also after adjustment for statistically significant methodological study characteristics, PD effects on classroom quality and student development remained substantial.

There were also significant differences in training effects between studies using different quality rating instruments. Studies that used standardized international instruments (e.g., CLASS, ERS, ELLCO) resulted, on average, in higher quality improvements, which might be explained in part by their favorable psychometric properties. Moreover, training interventions that were evaluated with the CLASS measure reported higher effects. This is not surprising, because the CLASS measure focuses mainly on teacher behavior (e.g., teacher sensitivity and responsiveness, quality of instructions, and classroom management), whereas ELLCO and ERS include structural quality indicators related to the physical classroom (Burchinal, 2010; Markussen-Brown et al., 2017). The main goal of the PD programs is changing teaching behavior, and a transfer effect to the physical classroom environment (e.g., the number of books or literacy-related material) is therefore not likely to occur if the intervention program does not include monetary support.

### *Limitations of This Review*

Meta-analyses are dependent on the number and quality of primary studies available. In relation to the limitations of this study, the meta-analysis was based on studies that evaluated short-term effects. Follow-up data allowing analysis of long-term effects or the maintenance of quality improvements and student achievement are generally lacking.

Second, our search and including criteria (e.g., recognized quality rating scales as outcome measures) resulted in the inclusion of many studies from North America. As a consequence, our findings might be restricted to the predominant U.S. context, and experimental studies in other countries are thus needed to extend the current evidence base. Also, the generalization of findings to other Western countries must be discussed with regard to contextual factors, including regulations and requirements for teacher qualification in different countries (Eurydice, 2014; Friendly et al., 2015; Whitebook et al., 2009). It should be noted, however, that the evaluated in-service training formats and the quality rating scales that were included in our review are available and are also used in European countries (Oberhuemer, 2012).

Third, as mentioned above, there was an insufficient number of experimental studies providing data at teacher, classroom, and child levels (see also Markussen-Brown et al., 2017). The small number of studies that measure both classroom quality and child outcomes was also a concern for other researchers (Fukkink & Lont, 2007; Markussen-Brown et al., 2017), and further experimental studies are urgently required. Some of these studies reported effects of in-service PD on child development without a clear mechanism of change in classroom interactions and practices. Seen from this perspective, research into PD is still at an early stage, as Zaslow et al. (2010b) also concluded, and future studies of PD at both classroom and student level are needed. This line of study should reveal which intervention characteristics promote changes in teachers' knowledge, attitudes, awareness, and skills, which subsequently result in enhanced student–teacher interactions and classroom quality.

### *Conclusion*


This review provides meta-analytic evidence that the ECEC field has a unique chance to invest in the development of practitioners through in-service training. Our review confirms the assumed link between PD of teachers and the levels of

student achievement via increased classroom ECEC quality, demonstrating a strong relationship between improvement of pedagogical quality and the development of young children. Based on our analysis driven by the conceptual PD framework from Buysse et al. (2009), we can make some suggestion for the design of PD with regard to training format and duration. Training intensity should be related to the scope of the program and we suggest that a duration of more than 45 hours should be scheduled to produce significant improvement of global quality. Furthermore, the format of PD delivery seems to matter: PD with coaching seems effective in enhancing classroom quality.

### Note

This research was part of a dissertation project at the Otto-Friedrich-University of Bamberg and supervised by Prof. Dr. Hans-Guenther Roßbach.

### ORCID iD

F. Egert  <http://orcid.org/0000-0003-0972-602X>

### References

- \**Studies included in the meta-analysis with quality measures are marked with an asterisk.*  
\*\**Studies included in the meta-analysis with both quality and child outcomes measures are marked with a double asterisk.*
- \*Algozzine, B., Babb, J., Algozzine, K., Mraz, M., Kissel, B., Spano, S., & Foxworth, K. (2011). Classroom effects of an early childhood educator professional development partnership. *NHSA Dialog: A Research-to-Practice Journal for the Early Childhood Field*, 14, 246–262.
- Armstrong, K., Cusumano, D. L., Todd, M., & Cohen, R. (2008). Literacy training for early childhood providers: Changes in knowledge, beliefs, and instructional practices. *Journal of Early Childhood Teacher Education*, 29, 297–308.
- Arnett, J. (1989). Caregiver in day-care centers: Does training matter? *Journal of Applied Developmental Psychology*, 10, 541–552.
- \*\*Barnett, S. W., Jung, K., Yarosz, D. J., Thomas, J., Hornbeck, A., Stechuk, R., & Burns, S. (2008). Educational effects of the tools of the mind curriculum: A randomized trial. *Early Childhood Research Quarterly*, 23, 299–313.
- Blank, R. K., & de las Alas, N. (2009). *Effects of teacher professional development on gains in student achievement. How meta analysis provides scientific evidence useful to education leaders*. Washington, DC: Council of Chief State School Officers.
- \*Bloom, P. J., & Sheerer, M (1992). The effects of leadership training on child care program quality. *Early Childhood Research Quarterly*, 7, 579–594.
- Blume, B. D., Ford, J. K., Baldwin, T. T., & Huang, J. L. (2010). Transfer of training. A meta-analytic review. *Journal of Management*, 36, 1065–1105.
- \*Boller, K., Del Grosso, P., Blair, R., Jolly, Y., Fortson, K., Paulsell, D., . . . Kovac, M. (2010). *The Seeds to Success Modified Field Test: Findings from the impact and implementation studies*. Princeton, NJ: Mathematica Policy Research.
- Borenstein, M., Hedges, L., Higgins, J., & Rothstein, H. (2005). *Comprehensive Meta-analysis (CMA) Version 2*. Englewood, NJ: Biostat.
- \*Breffini, L. (2011). Impact of curriculum training on state-funded prekindergarten teachers' knowledge, beliefs, and practices. *Journal of Early Childhood Teacher Education*, 32, 176–193.

- \*Brinks, R. (2007). *Intensive professional development in early literacy instruction for preschool teachers* (Unpublished dissertation). Western Michigan University, Kalamazoo.
- \*\*Bryant, D., Wesley, P., Burchinal, M., Sidersis, J., Taylor, K., Fenson, C., & Iruka, I. U. (2009). *The QUINCE-PFI Study. An evaluation of a promising model for child care provider training*. Chapel Hill: University of North Carolina at Chapel Hill, FPG Child Development Institute.
- Bryk, A. S., & Raudenbush, S. W. (2002). *Hierarchical linear models: Applications and data analysis methods*. Thousand Oaks, CA: Sage.
- Burchinal, M. (2010). *Differentiating among measures of quality: Key characteristics and their coverage in existing measures* (Research-to-Policy, Research-to-Practice Brief OPRE 2011-10b). Retrieved from [https://www.acf.hhs.gov/sites/default/files/opre/differ\\_measures.pdf](https://www.acf.hhs.gov/sites/default/files/opre/differ_measures.pdf)
- Burchinal, M., Hyson, M., & Zaslow, M. (2008). Competencies and credentials for early childhood educators: What do we know and what do we need to know? *NHSA Dialog Briefs*, *11*(1), 1–8.
- \*Burchinal, M. R., Cryer, D., Clifford, R. M., & Howes, C. (2002). Caregiver training and classroom quality in child care centers. *Applied Developmental Science*, *6*(1), 2–11.
- \*\*Buisse, V., Castro, D. C., & Peisner-Feinberg, E. (2010). Effects of a professional development program on classroom practices and outcomes for Latino dual language learners. *Early Childhood Research Quarterly*, *25*, 194–206.
- Buisse, 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*, 235–243.
- Cain, D. W., Rudd, L. C., & Saxon, T. F. (2007). Effects of professional training on joint attention engagement in low-quality child care centers. *Early Child Development and Care*, *177*, 159–185.
- Camilli, G., Vargas, S., Ryan, S., & Barnett, S. W. (2010). Meta-analysis of the effects of early education interventions on cognitive and social development. *Teachers College Record*, *112*, 579–620.
- Campbell Collaboration. (n.d.). *Campbell Collaboration systematic reviews: Policies and guidelines*. Retrieved from <https://www.campbellcollaboration.org/library/campbell-collaboration-systematic-reviews-policies-and-guidelines.html>
- \*Campbell, P. H., & Milbourne, S. A. (2005). Improving the quality of infant? Toddler care through professional development. *Topics in Early Childhood Special Education*, *25*(1), 3–14.
- \*Cassidy, D. J., Buell, M. J., Pugh-Hoesse, S., & Russell, S. (1995). The effect of education on child care teachers' beliefs and classroom quality: Year one evaluation of the TEACH early childhood associate degree scholarship program. *Early Childhood Research Quarterly*, *10*, 171–183.
- \*\*Clancy-Menchetti, J. L. (2006). *Early literacy professional development: Exploring the effects of mentoring for preschool teachers* (Unpublished dissertation). The Florida State University, Tallahassee.
- Cohen, J. (1988). *Statistical power analysis for the behavioral science* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.
- Cooper, H., Hedges, L. V., & Valentine, J. C. (2009). *The handbook of research synthesis and meta-analysis* (2nd ed.). New York, NY: Russell Sage Foundation.
- \*Cunningham, L. C. (2007). *Toward better early childhood education: Links between professional development and language and literacy practices* (Unpublished dissertation). The University of Michigan, Ann Arbor.

- \*Dickinson, D. K., & Caswell, L. (2007). Building support for language and early literacy in preschool classrooms through in-service professional development: Effects of the Literacy Environment Enrichment Program (LEEP). *Early Childhood Research Quarterly*, 22, 243–260.
- \*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, 567–597.
- \*Edgar, T. (2008). *The impact of a research-based professional development project on the language and literacy environment of pre-kindergarten teachers' classrooms* (Unpublished dissertation). University of Houston, TX.
- Egger, M., Zellweger, T., & Antes, G. (1996). Randomised trials in German language journals. *Lancet*, 347, 1047–1048.
- Elmore, R. F. (2002). *Bridging the gap between standards and achievement. The imperative for professional development in education*. Washington, DC: Albert Shaker Institute.
- \*Englund, L. W. (2010). *Evaluating and improving the quality of teacher's language modeling in early childhood classrooms*. Las Vegas: University of Nevada.
- Eurydice. (2014). *Key data on early childhood education and care 2014*. Retrieved from [http://eacea.ec.europa.eu/education/eurydice/documents/key\\_data\\_series/166EN.pdf](http://eacea.ec.europa.eu/education/eurydice/documents/key_data_series/166EN.pdf)
- Fantuzzo, J., Childs, S., Stevenson, H., Coolahan, K. C., Ginsburg, M., Gay, K., . . . Watson, C. (1996). The Head Start Teaching Center: An evaluation of an experiential, collaborative training model for Head Start teachers and parent volunteers. *Early Childhood Research Quarterly*, 11, 79–99.
- Fiene, R. (2002). Improving child care quality through an infant caregiver mentoring project. *Child & Youth Care Forum*, 31(2), 79–87.
- Friendly, M., Grady, B., MacDonald, L., & Forer, B. (2015). *Early childhood education in Canada 2014*. Retrieved from <http://childcarecanada.org/sites/default/files/ECEC-2014-full-document-revised-10-03-16.pdf>
- 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.
- Garet, M. S., Porter, A. C., Desimone, L., Birman, B. F., & Yoon, K. S. (2001). What makes professional development effective? Results from a national sample of teachers. *American Educational Research Journal*, 38, 915–945.
- Girolametto, L., Weitzman, E., & Greenberg, J. (2003). Training day care staff to facilitate children's language. *American Journal of Speech-Language Pathology*, 12, 299–311.
- \*Grace, C., Bordelon, D., Cooper, P., Kazelskis, R., Reeves, C., & Thames, D. G. (2008). Impact of professional development on the literacy environments of preschool classrooms. *Journal of Research in Childhood Education*, 23(1), 52–75.
- Hamre, B. K., & Hatfield, B. E. (2012). Moving evidenced-based professional development into the field: Recommendations for policy and research. In C. Howes, B. Hamre, & R. Pianta (Eds.), *Effective early childhood professional development. Improving teacher practice and child outcomes* (pp. 213–228). Baltimore, MD: Brookes.
- 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.

- Harvard Family Research Project. (2006). Pathways from workforce development to child outcomes. *Evaluation Exchange*, 11(4), 2–4.
- Haskell, M. (1994). *Using a training as a mean to improve the level of quality in child care facilities* (Practicum paper). Nova Southeastern University, Davie, FL.
- Heckman, J. (2008). Schools, skills, and synapses. *Economic Inquiry*, 40, 289–324.
- Honig, A. S., & Hirallal, A. (1998). Which counts more for excellence in childcare staff—Years in service, education level or ECE coursework? *Early Child Development and Care*, 145, 31–46.
- \*Howe, N., Jacobs, E., Vukelich, G., & Recchia, H. (2011). In-service professional development and constructivist curriculum: Effects on quality of child care, teacher beliefs, and interactions. *Alberta Journal of Educational Research*, 57, 353–378.
- Hox, J. (2010). *Multilevel analysis: Techniques and applications*. Mahwah, NJ: Erlbaum.
- 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.
- Isner, T., Tout, K., Zaslow, M., Soli, M., Quinn, K., Rothenberg, L., . . . Burkhauser, M. (2011). *Coaching in early care and education programs and Quality Rating and Improvement Systems (QRIS): Identifying promising features*. Washington, DC: Child Trends.
- \*Japel, C. (2009). *How can we enhance child care quality? Paper presented at the ECERS International Network Meeting, Porto, Portugal*.
- Kelley, P., & Camilli, G. (2007). *The impact of teacher education on outcomes in center-based early childhood education programs: A meta-analysis* (NIEER Working Paper). Retrieved from <http://nieer.org/wp-content/uploads/2016/08/TeacherEd.pdf>
- Klein, L., & Gomby, D. S. (2008). *A synthesis of federally funded studies on school readiness: What are we learning about professional development? Prepared for "A Working Meeting on Recent School Readiness Research: Guiding the Synthesis of Early Childhood Research."* Washington, DC: U.S. Department of Health and Human Services, Office of the Assistant Secretary for Planning and Evaluation and the Administration for Children and Families, Office of Planning Research and Evaluation.
- Lipsey, M. W., & Wilson, D. B. (1993). The efficacy of psychological, educational, and behavioral treatment: Confirmation from meta-analysis. *American Psychologist*, 48, 1181–1209.
- \*\*Lonigan, C. J., Farver, J. M., Phillips, B. M., & Clancy-Menchetti, J. (2011). Promoting the development of preschool children's emergent literacy skills: A randomized evaluation of a literacy-focused curriculum and two professional development models. *Reading and Writing: An Interdisciplinary Journal*, 24, 305–337.
- Manning, M., Garvis, S., Fleming, C., & Wong, G. T. W. (2017). *The relationship between teacher qualification and the quality of the early childhood education and care environment* (A Campbell Systematic Review 2017:1). Retrieved from [https://www.campbellcollaboration.org/media/k2/attachments/0224\\_ECG-Manning-Teacher\\_qualifications.pdf](https://www.campbellcollaboration.org/media/k2/attachments/0224_ECG-Manning-Teacher_qualifications.pdf)
- 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.
- Martinez-Beck, I., & Zaslow, M. (2006). Introduction: The context for critical issues in early childhood professional development. In M. Zaslow & I. Martinez-Beck

- (Eds.), *Critical issues in early childhood professional development* (pp. 1–16). Baltimore, MD: Brookes.
- \*McNerney, S., Nielsen, D. C., & Clay, P. (2006). Supporting literacy in preschool: Using a teacher-observation tool to guide professional development. *Journal of Early Childhood Teacher Education*, 27(1), 19–34.
- Miller, J. A., & Bogatova, T. (2009). Quality improvements in the early and education workforce: Outcomes and impact of the T.E.A.C.H. Early Childhood® Project. *Evaluation and Program Planning*, 32, 257–277.
- \*\*Morris, P., Raver, C. C., Millenky, M., Jones, S., & Lloyd, C. M. (2010). *Making preschool more productive. How classroom management training can help teachers*. Retrieved from <http://files.eric.ed.gov/fulltext/ED514651.pdf>
- National Association for the Education of Young Children, & National Association of Child Care Resources & Referral Agencies. (2011). *Early childhood education professional development: Training and technical assistance glossary*. Retrieved from [http://www.naeyc.org/GlossaryTraining\\_TA.pdf](http://www.naeyc.org/GlossaryTraining_TA.pdf)
- National Institute of Child Health and Human Development Early Child Care Research Network. (2000). The relation of child care to cognitive and language development. *Child Development*, 74, 960–980.
- National Institute of Child Health and Human Development Early Child Care Research Network. (2003). Does quality of child care affect child outcomes at age 4 1/2? *Developmental Psychology*, 39, 451–469.
- National Institute of Child Health and Human Development, Early Child Care Research Network, & Duncan, G. (2003). Modeling the impacts of child care quality on children’s preschool cognitive development. *Child Development*, 74, 1454–1475.
- \*Neuman, S. B., & Cunningham, L. (2009). The impact of professional development and coaching on early language and literacy instructional practices. *American Educational Research Journal*, 46, 532–566.
- \*Neuman, S. B., & Wright, T. S. (2010). *Promoting language and literacy development for early childhood educators: A mixed-methods study of coursework and coaching*. Retrieved from <http://www-personal.umich.edu/~sbneuman/pdf/coachingInEarly-Childhood.pdf>
- Oberhuemer, P. (2012). *Fort- und Weiterbildung frühpädagogischer Fachkräfte im europäischen Vergleich* [Continuing education of early childhood education professionals in European comparison]. Retrieved from [http://www.weiterbildungsinitiative.de/uploads/media/Studie\\_Oberhuemer.pdf](http://www.weiterbildungsinitiative.de/uploads/media/Studie_Oberhuemer.pdf)
- Oberhuemer, P. (2013). Continuing professional development and the early years workforce. *Early Years: An International Research Journal*, 33, 103–105.
- OECD. (2012). *Starting Strong III: A quality toolbox for early childhood education and care*. Paris, France: Author.
- \*\*Onchwari, G., & Keengwe, J. (2010). Teacher mentoring and early literacy learning: A case study of a mentor-coach initiative. *Early Childhood Education Journal*, 37, 311–317.
- \*Palsha, S. A., & Wesley, P. W. (1998). Improving quality in early childhood environments through on-site consultation. *Topics in Early Childhood Special Education*, 18, 243–253.
- Peisner-Feinberg, E., & Burchinal, M. (1997). Relations between preschool children’s child care experiences and concurrent development: The Cost, Quality and Outcomes Study. *Merrill-Palmer Quarterly*, 43, 451–477.
- Peisner-Feinberg, E., Burchinal, M., Clifford, R., Yazejian, N., Culkin, M., Zelazo, J., . . . Rustici, J. (1999). *The children of the Cost, Quality, and Outcomes Study go to*

- school. Chapel Hill: University of North Carolina at Chapel Hill, FPG Child Development Institute.
- Pianta, R. C., Barnett, W. S., Burchinal, M. P., & Thornburg, K. R. (2009). The effects of preschool education: What we know, how public policy is or is not aligned with the evidence base, and what we don't know. *Psychological Science in the Public Interest*, 10(2), 49–88.
- Piasta, S. B., & Wagner, R. K. (2010). Developing early literacy skills: A meta-analysis of alphabet learning and instruction. *Reading Research Quarterly*, 45(1), 8–38.
- \*\*Powell, D. R., Diamond, K. E., Burchinal, M. R., & Koehler, M. J. (2010). Effects of an early literacy professional development intervention on Head Start teachers and children. *Journal of Educational Psychology*, 102, 299–312.
- Rasbash, J., Charlton, C., Browne, W. J., Healy, M., & Cameron, B. (2005). *MLwiN Version 2.02*. Bristol, England: Centre for Multilevel Modelling, University of Bristol.
- \*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.
- \*Roehrig, G. H., Dubosarsky, M., Mason, A., Carlson, S., & Murphy, B. (2011). We look more, listen more, notice more: Impact of sustained professional development on Head Start teachers' inquiry-based and culturally-relevant science teaching practices. *Journal of Science Education and Technology*, 20, 566–578.
- Rosenthal, M. L. C. (1994). The fugitive literature. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 85–94). New York, NY: Russell Sage.
- Sammons, P. (2010). Does pre-school make a difference? Identifying the impact of pre-school on children's cognitive and social behavioral development at different ages. In K. Sylva, E. Melhuish, P. Sammons, I. Siraj-Blatchford, & B. Taggart (Eds.), *Early childhood matters: Evidence from the Effective Pre-school and Primary Education Project* (pp. 92–113). Oxford, England: Routledge.
- Schachter, R. E. (2015). An analytic study of the professional development research in early childhood education. *Early Education and Development*, 26, 1057–1085.
- Schweinhart, L. J., Barnes, H. V., & Weikart, D. P. (1993). *Significant benefits. The High/Scope Perry Preschool Study through age 27*. Ypsilanti, MI: High Scope Press.
- Schweinhart, L. J., Montie, J., Xiang, Z., Barnett, W. S., Belfield, C. R., & Nores, M. (2005). *Lifetime effects. The High Scope Perry Preschool Study through age 40*. Ypsilanti, MI: High Scope Press.
- Sheridan, S. (2000). A comparison of external and self-evaluations of quality in early childhood education. *Early Child Development and Care*, 164, 63–78.
- \*Sheridan, S. M., Clarke, B. L., Knoche, L. L., & Edwards, C. P. (2006). The effects of conjoint behavioral consultation in early childhood settings. *Early Education and Development*, 17, 593–617.
- Sheridan, S. M., Edwards, C. P., Marvin, C. A., & Knoche, L. L. (2009). Professional development in early childhood programs: Process issues and research needs. *Early Education and Development*, 20, 377–401.
- \*Sibley, A., & Sewell, K. (2011). Can multidimensional professional development improve language and literacy instruction for young children? *NHSA Dialog: A Research-to-Practice Journal for the Early Childhood Field*, 14, 263–274.
- \*Sipp, R. B. (2010). *The efficacy of Emotional and Instructional Support (EIS) training and consultation on Head Start teacher-child interactions* (Unpublished dissertation). University of Nevada, Las Vegas.



- Slavin, R. E. (2008). Perspectives on evidence-based research in education. What works? Issues in synthesizing educational program evaluations. *Educational Researcher*, 37(1), 5–14.
- Snyder, P., Hemmeter, M. L., Meeker, K. A., Kinder, K., Pasia, C., & McLaughlin, T. (2012). Characterizing key features of the early childhood professional development literature. *Infants & Young Children*, 25, 188–212.
- \*Sprague, K., Glantz, F., Kokopelli, F. G., Raya-Carlton, P., & Schilder, D. (2009). *What is gained from a randomized controlled trial and what is lost? Planning, policy, and fear: Effects and lessons learned from a cluster-randomized trial assessing a community-advocacy program's professional development initiative for early childhood educators*. Paper presented at the Society for Research on Educational Effectiveness Conference, Washington, DC.
- Sylva, K. (2010). Quality in early childhood settings. In K. Sylva, E. Melhuish, P. Sammons, I. Siraj-Blatchford, & B. Taggart (Eds.), *Early childhood matters: Evidence from the Effective Pre-school and Primary Education Project* (pp. 70–91). Oxford, England: Routledge.
- Thornton, J. S., Crim, C. L., & Hawkins, J. (2009). The impact of an ongoing professional development program on prekindergarten teachers' mathematic practices. *Journal of Early Childhood Teacher Education*, 30, 150–161.
- Tietze, W., Lee, H.-J., Bensel, J., Haug-Schabel, G., Aselmeier, M., & Egert, F. (2013). Pädagogische Qualität in Kindertageseinrichtungen und Tagespflegestellen [Educational quality in day care centers and day care centers]. In W. Tietze, F. Becker-Stoll, J. Bensel, A. G. Eckhardt, G. Haug-Schnabel, B. Kalicki, . . . B. Leyendecker (Eds.), *NUBBEK. Nationale Untersuchung zu Bildung, Betreuung und Erziehung in der frühen Kindheit* [NUBBEK. National Investigation into Education, Care and Education in Early Childhood] (pp. 69–87). Weimar, Berlin: Verlag das Netz.
- Tout, K., Isner, T., & Zaslow, M. (2011). *Coaching for quality improvement: Lessons learned from quality rating and improvement systems*. Retrieved from <http://www.childtrends.org/wp-content/uploads/2013/05/2011-35CoachingQualityImprovement1.pdf>
- \*Uttley, C. M., & Horm, D. M. (2008). Mentoring in early childhood professional development: Evaluation of the Rhode Island Child Development Specialist Apprenticeship Program. *Journal of Early Childhood Teacher Education*, 29, 237–252.
- Vandell, D. L., & Wolfe, B. (2000). *Child care quality: Does it matter and does it need to be improved?* Retrieved from <https://aspe.hhs.gov/report/child-care-quality-does-it-matter-and-does-it-need-be-improved-full-report>
- Vermeer, H. J., van IJzendoorn, M., Cárcamo, R. A., & Harrison, L. (2016). Quality of child care using the Environmental Rating Scales: A meta-analysis of international studies. *International Journal of Early Childhood*, 48, 33–60.
- \*\*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.
- What Works Clearinghouse. (2013). *Procedures and standards Handbook* (Version 3.0). Retrieved from [https://ies.ed.gov/ncee/wwc/Docs/referenceresources/wwc\\_procedures\\_v3\\_0\\_standards\\_handbook.pdf](https://ies.ed.gov/ncee/wwc/Docs/referenceresources/wwc_procedures_v3_0_standards_handbook.pdf)

- Whitebook, M., Gomby, D., Bellm, D., Sakai, L., & Kipnis, F. (2009). *Preparing teachers of young children: The current state of knowledge, and a blueprint for the future. Part I: Teacher preparation and professional development in grades K-12 and in early care and education: Differences and similarities, and implications for research* (Policy report). Retrieved from [http://www.irle.berkeley.edu/cscce/wp-content/uploads/2009/01/teacher\\_prep\\_1.pdf](http://www.irle.berkeley.edu/cscce/wp-content/uploads/2009/01/teacher_prep_1.pdf)
- Whitebook, M., & Ryan, S. (2011). *Degrees in context: Asking the right questions about preparing skilled and effective teachers of young children* (Preschool Policy Brief 11). New Brunswick, NJ: National Institute for Early Education Research. Retrieved from <http://nieer.org/resources/policybriefs/23.pdf>
- 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). Retrieved from [http://ies.ed.gov/ncee/edlabs/regions/southwest/pdf/rel\\_2007033.pdf](http://ies.ed.gov/ncee/edlabs/regions/southwest/pdf/rel_2007033.pdf)
- 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*. Retrieved from <http://www2.ed.gov/rschstat/eval/professional-development/literature-review.pdf>
- Zaslow, M., Tout, K., Halle, T., Whittaker, J. V., & Lavelle, B. (2010b). Emerging research on early childhood professional development. In S. B. Neuman & M. L. Kamil (Eds.), *Preparing teachers for the early childhood classroom: Proven models and key principles* (pp. 19–47). Baltimore, MD: Brookes.

### **Authors**

- FRANZISKA EGERT is a senior scientist at the State Institute of Early Childhood Research (IFP Bayern), Winzererstr. 9, 80797 München, Germany; email: [franziska.egert@ifp.bayern.de](mailto:franziska.egert@ifp.bayern.de). Her research focusses on the effectiveness of professional development, early education programs, and language interventions for mono- and multilingual children in early childhood.
- RUBEN G. FUKKINK is professor at the University of Amsterdam and the Amsterdam University of Applied Sciences, Nieuwe Achtergracht 127, 1018 WS Amsterdam; email: [r.g.fukkink@uva.nl](mailto:r.g.fukkink@uva.nl). His research interests include the pedagogical quality of child care and professional development of staff in early childhood education and care.
- ANDREA G. ECKHARDT is a professor at the Hochschule Zittau/Görlitz, University of Applied Sciences in Görlitz, Furtstr. 2, 02826 Görlitz, Germany; email: [a.eckhardt@hszg.de](mailto:a.eckhardt@hszg.de). Her research interests are quality, professional development, and educational programs in early childhood education and care.