# The impact of in-service professional development on the quality of teacher-child interactions in early education and care

*A meta-analysis*

Egert, F.; Dederer, V.; Fukkink, R.G.

**DOI**
10.1016/j.edurev.2019.100309

**Publication date**
2020

**Document Version**
Final published version

**Published in**
Educational Research Review

Educational Research Review 29 (2020) 100309

Contents lists available at ScienceDirect

Educational Research Review

journal homepage: www.elsevier.com/locate/edurev

# The impact of in-service professional development on the quality of teacher-child interactions in early education and care: A meta-analysis

Franziska Egert[a,*], Verena Dederer[b], Ruben G. Fukkink[c]

[a] Catholic University of Applied Sciences Munich (KSH München), Campus Benediktbeuern, Don-Bosco-Str. 1, 83671, Benediktbeuern, Germany
[b] State Institute of Early Childhood Research (ifp Bayern), Winzererstr. 9, 80797, München, Germany
[c] University of Amsterdam & Amsterdam University of Applied Sciences, Nieuwe Achtergracht 127, 1018 WS, Amsterdam, the Netherlands

## 1. Introduction

A growing body of evidence reveals the importance of high-quality teacher-child interactions in early childhood education and care (ECEC) for learning and the behavioral engagement of young children before they enter school (Mashburn et al., 2008; Ponitz, Rimm-Kaufman, Grimm, & Curby, 2009). In general, teachers' performance and effectiveness is seen a function of their behavior in classrooms and their interactions with students. In line with that, the "Teaching Through Interactions" framework (Hamre et al., 2013) assumes that the quality of interactions between children and adults is the primary mechanism to promote developmental gains. This paper therefore focuses on efforts to improve the quality of teacher-child-interactions in ECEC classrooms. According to the framework, interactions between teacher and child are distinguished in three domains, namely emotional support (ES), classroom organization (CO), and instructional support (IS). The observational measure Classroom Assessment Scoring System (CLASS Pre-K) was developed to assess the quality of teacher-child interactions in ECEC for these three domains based on this framework (Hamre et al., 2013; Pianta, La Paro, & Hamre, 2008a). Emotional support (ES) involves teacher sensitivity and their responsiveness to children's signals, as well as regard for student perspectives, interests, and motivation (Pianta et al., 2008a). Classroom organization (CO) comprises a variety of teacher-child interactions that help students to organize their behavior and attention related to classroom routines and academic goals (Hamre et al., 2013). A teacher who sets clear rules, organizes stable classroom routines, uses proactive approaches to discipline, and provides interesting and diversified activities contributes significantly to children's motivation and engagement in learning tasks (Pianta et al., 2008a). Instructionally supportive teachers (IS) offer a wide range of learning activities to enhance knowledge of concepts and language, relate new information to students' background knowledge and real-world experiences, and provide learning feedback that is immediate, corrective, and specific (Hamre et al., 2013; Pianta et al., 2008a). High levels of ES, CO or IS indicate that a teacher provides high-quality teacher-child interactions consistently throughout the day and during different activities (Hamre et al., 2013; Pianta et al., 2008a).

### 1.1. Teacher-child interactions and child development

Empirical research has found consistent associations between the quality of teacher-child interactions measured with the CLASS and children's outcomes. The magnitude of this association is small (see Keys et al., 2013; Perlman et al., 2016). Recent research indicates that a certain quality threshold must be reached to affect students' school readiness skills (Hatfield, Burchinal, Pianta, & Sideris, 2016; Wieland, Ulvestad, Sachs, & Yoshikawa, 2013). Further, different quality domains of teacher-child interactions seem to be predictive for specific competencies of children in preschool and kindergarten (Perlman et al., 2016). In particular, ES can foster

* Corresponding author.
*E-mail addresses:* franziska.egert@ksh-m.de (F. Egert), verena@dederer.com (V. Dederer), r.g.fukkink@uva.nl (R.G. Fukkink).

https://doi.org/10.1016/j.edurev.2019.100309
Received 30 March 2019; Received in revised form 10 November 2019; Accepted 17 December 2019
Available online 20 December 2019
1747-938X/ 

students' social and emotional functioning, facilitates positive student-teacher interactions and peer interactions in classrooms (Hamre et al., 2013) and can reduce problem behavior (Curby, Brock, & Hamre, 2013; Mashburn et al., 2008). There is some evidence that CO is related to children's learning motivation and self-regulation (Pakarinen et al., 2010; Rimm-Kaufman, Curby, Grimm, Nathanson, & Brock, 2009). Children in kindergarten classrooms with a higher quality of classroom management generally have superior behavioral and cognitive self-control, show more on-task behavior, and are more engaged in learning activities than children in classrooms with lower levels of classroom management (Rimm-Kaufman et al., 2009). Finally, research indicates that the quality of IS correlates with a wide range of cognitive, language, and social-emotional skills (Mashburn et al., 2008). In particular, children in classrooms with higher levels of instructive interaction are assumed to have advanced competencies in receptive and expressive vocabulary, oral and written language skills, rhyming, letter naming, inhibitory control, problem solving, mathematics, and empathy, and lower levels of disruptive behavior (Curby et al., 2009; Hatfield et al., 2016; Keys et al., 2013; Siekkinen et al., 2013). Thus, high levels of IS seems to be a key aspect of high-quality interactions and effective teaching, which contributes to children's academic learning and school readiness.

### *1.2. Professional development to improve the quality of teacher-child interactions*

In policy and practice, there is a growing interest in effective professional development interventions for ECEC teachers to improve classroom quality (Buysse, Winton, & Rous, 2009; Sheridan, Edwards, Marvin, & Knoche, 2009; Zaslow, Tout, Halle, Whittaker, & Lavelle, 2010). Especially as significant differences in the quality of teacher-child interactions between the three quality domains occurs in ECEC. Studies around the world (in, e.g., Australia, China, Chile, Finland, Germany, the Netherlands, Spain, and the US) have demonstrated that the quality of ES and CO is, on average, in the medium-to-high range, whereas the quality of IS is generally low (e.g., Hu, Dieker, Yang, & Yang, 2016; Leyva et al., 2015; Salminen et al., 2012; Sandstrom, 2012; Slot, Boom, Verhagen, & Leseman, 2017; Tayler, Ishimine, Cloney, Cleveland, & Thorpe, 2013; Von Suchodoletz, Fäsche, Gunzenhauser, & Hamre, 2014). These findings suggest that ECEC teachers are, generally speaking, sensitive to children's needs and ensure clear classroom routines, but instructional rich interactions that effectively stimulate young children's language learning and higher order thinking are observed less often.

Professional development (hereafter: PD) for ECEC teachers focuses on a specific practice in the classroom, acknowledges the everyday pedagogical work, and often contains multiple components to provide sustained learning opportunities, including individual guidance and feedback (Buysse et al., 2009). In general, PD faces the task to support teachers to learn and to retain pedagogical skills, to generalize the new skills to the job context and to maintain them (Baldwin & Ford, 1988). In our case, PD stimulates teachers to develop the skills to provide emotional, organizational, and instructional supportive interactions consistently throughout the day in different activities and routines in ECEC classrooms.

Recent meta-analyses have demonstrated that ECEC classroom practice in general can be improved through in-service interventions (Egert, Fukkink, & Eckhardt, 2018; Fukkink & Lont, 2007; Markussen-Brown et al., 2017; Werner, Linting, Vermeer, & van IJzendoorn, 2016). Pianta, Hamre, and Downer (2011) argue that improvement in ECEC programs may be the result of the alignment of professional development goals, the measurement of quality, and goals for child development. Indeed, there is meta-analytic evidence of the assumed two-stage link from PD to child development, in which PD has a positive effect on improved classroom practice which subsequently affects child outcomes; improvement of classroom quality through PD explained half of the variance in child outcomes in the study of Egert et al. (2018). These findings thus support the theoretical model of change, starting with the investment in PD to improve the quality of pedagogical processes in ECEC and, consequently, to support children's development.

Meta-analytic evidence suggests that experimental outcomes are more positive if PD includes an individual component such as coaching (Egert et al., 2018; Markussen-Brown et al., 2017; Werner et al., 2016). There is only tentative evidence that a combination of PD formats (that might reflect various learning opportunities) is more effective than the use of a single PD type. The inclusion of several didactical elements that can be used across PD formats (e.g. feedback, self-reflection, and modeling) is also assumed to affect professional growth. In particular, the possibility for teachers to reflect on their teaching, defined as “a self-critical, investigative process wherein teachers consider the effect of their pedagogical decisions on their situated practice with the aim of improving those practices” (Tripp & Rich, 2012, p. 678), is seen as crucial to their professional development (Schön, 1983). Further, current meta-analytic evidence does not show a linear relation between PD effects and training duration. No moderation effects have been found either related to the focus of training, the use of videos in PD or the implementation of a specific curriculum in the classroom (Egert et al., 2018; Werner et al., 2016).

At methodological level, several moderators are related to experimental effects (Egert et al., 2018; Werner et al., 2016). In particular, the use of the CLASS was associated with larger experimental gains ($d$ = 0.19–0.27) in comparison to other standardized scales (Egert et al., 2018). This difference in PD effects might occur because of the way classroom quality is measured. Whereas some measures combine teacher indicators with other features like the availability of learning materials and the physical environment of the classroom, CLASS ratings are exclusively based on markers of teacher behavior, and do not include structural elements or other indicators that are related to equipment, financing and administration of the center. Relatedly, there is evidence that the constructive alignment of PD and outcome measures matters in the effective educational design of PD for ECEC teachers (Fukkink & Lont, 2007). In ECEC, PD typically focuses on classroom practices and should therefore be matched with instructional goals, learning standards and material teachers use in both the in-service training, practice and the evaluation (Buysse et al., 2009).

### 1.3. Focus of this study

The meta-analysis is part of the national BiSS initiative (Bildung durch Sprache und Schrift/Education through Language and Writing) in Germany. The federal initiative aims at the development and evaluation of effective PD programs to improve teacher-child interactions and to foster child development in the areas of language, literacy, reading and writing.

Our study synthesizes experimental studies that evaluated PD effects in preschool and kindergarten by means of the widely used CLASS Pre-K measure. It is grounded in theory through the "Teaching Through Interactions" framework (Hamre et al., 2013) to improve effective teaching that is linked to student learning, and has the unique potential to evaluate PD goals that are aligned with the design of recent, evidence-based PD interventions. In addition, our study aims to provide insight into the sensitivity of the measure to capture PD effects on teacher behavior. Further, the large differences observed in various descriptive studies between the relatively high levels of ES and CO and the relatively low level of IS prompt the question whether PD is (equally) effective across the three domains in improving ECEC teachers' interactions with children. This study also allowed us to chart the effects of PD systematically for ES, CO, and IS separately.

Our study investigates the following questions:

1 How effective is PD in improving the quality of teacher-child interactions at a global level (CLASS composite score) and in specific CLASS domains (ES, CO, IS)?
2 Is the PD effect, at global and domain level, moderated by methodological variable (e.g., sample size, randomization procedure, control group condition, blind data collection, and rating procedure) or context (e.g., Head Start, settings with high qualified teachers)?
3 Are specific PD formats, didactic elements, or PD intensity associated with greater experimental effects at global and domain level?

## 2. Method

We included all CLASS domains to estimate the overall effect on PD of the system-level quality of teacher-child interactions. We then conducted three separate meta-analyses to determine outcomes for ES, CO, and IS and to explore possible moderators.

### 2.1. Search strategy

First, an electronic search (ending with 2017) was conducted in the English-language databases ERIC, Dissertation Express, PsycINFO, ProQuest Dissertations and Theses, and SocINDEX. Following recently published review studies, the search included various keywords related to in-service professional development, (e.g., 'in-service training for teachers', 'professional development', 'coaching'), outcome measure (e.g., 'interaction quality', 'teacher child interaction', 'interaction skills'), target (e.g., 'impact', 'effect', 'experiment', 'influence'), and type of education (e.g., 'early childhood education', 'kindergarten', 'preschool') that were combined with Boolean operators (for detailed description see technical report (Egert & Dederer, 2018). Second, a manual search was conducted to complement the systematic electronic search. Here, we looked for individual studies in previously published systematic reviews and meta-analyses (i.e., Egert et al., 2018; Fukkink & Lont, 2007; Klein & Gomby, 2008; Markussen-Brown et al., 2017; Werner et al., 2016; Zaslow et al., 2010), and used web engines (e.g., Google) with the above-mentioned keywords.

The search in the electronic databases resulted in an initial set of 511 hits without duplicates (see Fig. 1). PsycINFO yielded 117 hits ($n$ = 35 coded as relevant for full-text coding), ERIC another 153 hits ($n$ = 33 relevant, excluding duplicates), ProQUEST D&T 143 ($n$ = 12 relevant), DISSEXPRESS 80 ($n$ = 7 relevant), and SocINDEX 18 hits ($n$ = 0 relevant). Another 32 papers were identified in the manual search.

### 2.2. Inclusion criteria

Studies were included in the review if they met the following criteria. First, the professional development programs were in-service trainings designed to improve the interactional quality of child care or teacher-child interactions. Pre-service training to prepare prospective teachers was excluded. Second, interaction quality and teacher-child interactions were measured using CLASS, including the versions for infants, toddlers, pre-k, and K-3. In addition, the sample had to include preschool, pre-k, or kindergarten teachers, as well as educators working in center-based care. Unlike other meta-analyses (e.g. Fukkink & Lont, 2007; Werner et al., 2016), we focused exclusively on ECEC settings (infant-toddler, preschool, kindergarten, or pre-kindergarten classrooms). Family child care providers were excluded. Fourth, experimental designs had to include at least one intervention condition and a control or comparison group. Given the limited number of randomized studies at teacher level, we also included experimental or quasi-experimental studies with or without randomization at this level; previous meta-analytic reports have found no significant differences in effects between randomized and non-randomized designs (Egert et al., 2018; Markussen-Brown et al., 2017). Furthermore, although studies were not limited to English-speaking countries they had be published in English for the sake of transparency and to facilitate the possibility of replicating our search results. Finally, studies had to report sufficient statistical information in order to compute effect sizes.

After title and abstract screening by two independent coders, 93 references were identified as potentially relevant. Subsequently, a full-text review procedure with a short coding form was used to evaluate the quality of the studies. Several papers were excluded

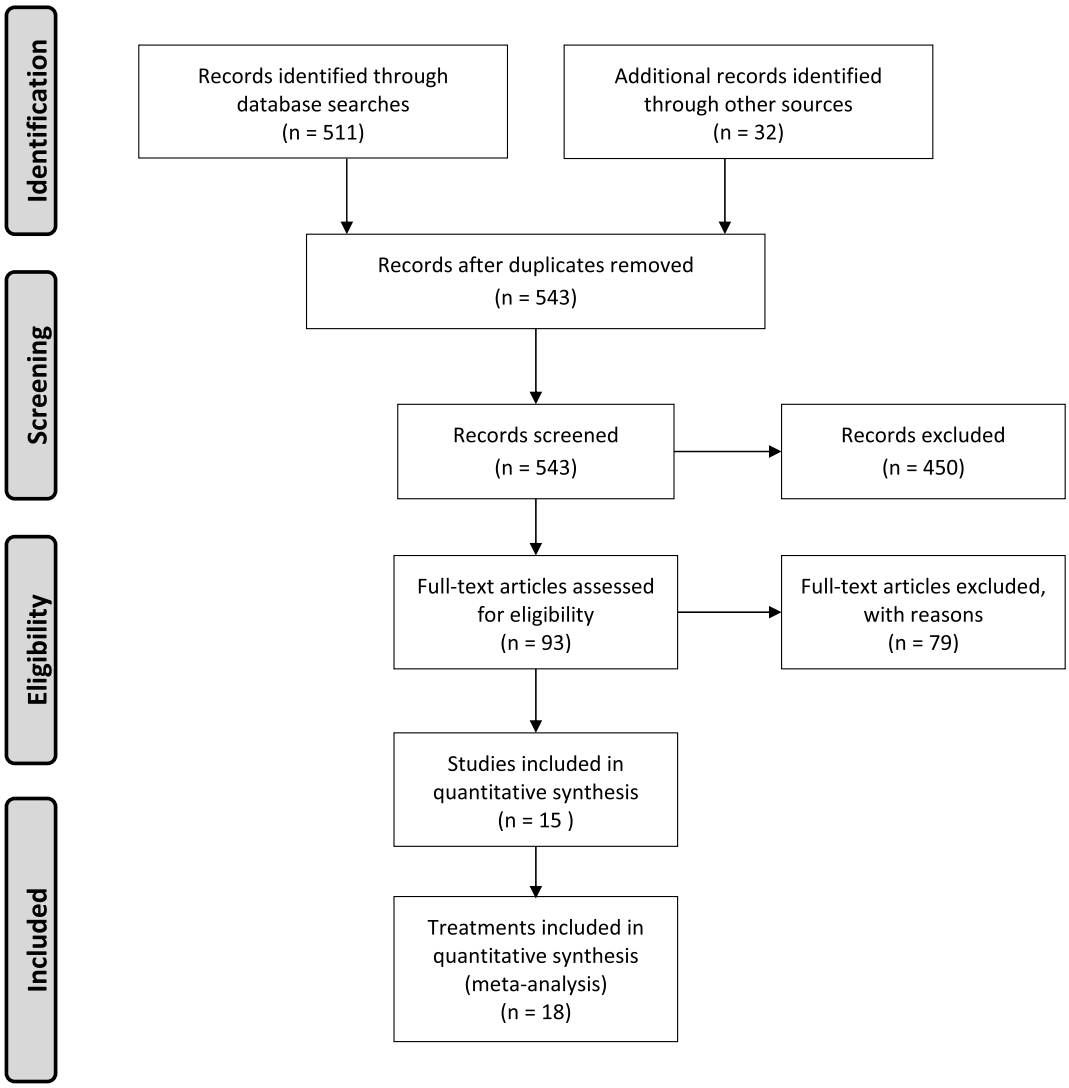


**Fig. 1.** PRISMA flow diagram reporting the systematic literature search (Moher et al., 2009).

after full-text analysis, for the following reasons: (1) the absence of an control/comparison group (e.g., Englund, 2010); (2) the use of an instrument other than the CLASS system to measure the quality of teacher-child-interactions in classrooms (e.g., Girolametto, Weitzman, & Greenberg, 2003; Sterling Honig & Martin, 2009); or (3) the absence of relevant statistical data (e.g., Downer et al., 2013). Of the 93 original studies, 15 studies which reported effects of 18 in-service treatments were included in the final sample (see Fig. 1).

### *2.3. Coding of studies*

Following the guidelines from the Campbell Collaboration (2019), all studies were double-coded by two independent coders. For the interpretation of the level of inference and objectivity of our codes, inter-rater reliability of initial scores for the independent coding procedure was estimated with kappa (κ) for nominal variables and with intra-class correlations (ICC) for interval-scaled variables. Coder disagreements were resolved through discussion and reaching consensus; as a consequence, the final codes reached a Kappa and ICC value of 1.

A two-step coding procedure was used to systematize information from the studies. First, full texts were coded using a short screening form to evaluate the quality of the studies and their provision of sufficient statistical data. The short screening form included coding for study design (e.g., controlled group design, randomization, sample size, equality of groups, drop-out, and fidelity), structural features of PD (e.g., duration in months, training hours, and format), sufficient statistical information, and relevance for meta-analysis. Inter-rater reliability was excellent for study relevance (kappa = 1). Second, the full texts of relevant studies were coded using an extended coding schema, which gathered additional information on the instructional features and

**Table 1**
Study description (at intervention level): Study and methodological characteristics.

| Study | Pub. type | Group design evidence standards | Design | Control | Exp (N) | Con (N) | Ratings | Blind | Alignment | Academic degree (%) | Head Start |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Barnett et al. (2008) | Pub | w.r. | CRT | BAU + PD | 14 | 22 | Video | Yes | No | – | No |
| Breffni (2011) | Pub | w.r. | RCT | BAU | 11 | 8 | Obs. | Yes | No | – | No |
| Domitrovich et al. (2009) | Pub | fully | CRT | BAU + PD | 44 | 44 | Obs. | Yes | No | 32 | Yes |
| Early et al., 2014, 2017 (1) | Pub | fully | RCT | BAU | 175 | 160 | Obs. | Yes | Yes | 91 | No |
| Early et al., 2014, 2017 (2) | Pub | fully | RCT | BAU | 151 | 160 | Obs. | Yes | Yes | 91 | No |
| Hamre et al., 2012; 2010 | Pub | w.r. | RCT | BAU | 223 | 217 | Obs. | – | Yes | 61 | No |
| Hemmeter et al. (2016) | Pub | fully | RCT | BAU | 20 | 20 | Obs. | – | No | 100 | No |
| Morris, Raver, Millenky, Jones, and Llyod (2010) | Gray | w.r. | CRT | BAU | 26 | 25 | Obs. | Yes | Yes | 98 | No |
| Morris et al., 2014 (1) | Gray | w.r. | CRT | BAU | 77 | 77 | Obs. | Yes | Yes | 62 | Yes |
| Morris et al., 2014 (2) | Gray | w.r. | CRT | BAU | 77 | 77 | Obs. | Yes | Yes | 62 | Yes |
| Morris et al., 2014 (3) | Gray | w.r. | CRT | BAU | 76 | 77 | Obs. | Yes | Yes | 62 | Yes |
| Pianta et al. (2008b) | Pub | w.r. | RCT | AT | 82 | 91 | Obs. | Yes | Yes | 100 | No |
| Pronchenko-Jain, 2012 | Gray | fully | CRT | AT | 12 | 12 | Obs. | Yes | No | 50 | Yes |
| Raver et al. (2008) | Pub | fully | CRT | BAU | ≈44 | ≈44 | Obs. | Yes | Yes | 25 | Yes |
| Wasik & Hindman, 2011; Hindman & Wasik, 2012 | Pub | w.r. | CRT | BAU | 19 | 11 | Obs. | Yes | No | 63 | Yes |
| Whittaker, Kinzie, Williford, and DeCoster (2015) | Pub | w.r. | CRT | BAU | 14 | 11 | Video | – | Yes | 100 | No |
| Yoshikawa et al. (2015) | Pub | w.r. | CRT | AT | 120 | 93 | Video | Yes | Yes | – | No |
| Zan and Donegan-Ritter (2014) | Pub | fully | CRT | BAU | 38 | 22 | Video | Yes | Yes | 58 | Yes |
| Summary | 73% Pub | 39% fully<br>61% w.r. | 33% RCT<br>67% CRT | 17% AT 83% BAU | 11–223 | 8–217 | 22% Videos | 83% | 67% | 32–100% | 44% |

*Note*. Design: Pub = published in peer review journals; Gray = gray literature; fully = study meet evidence standard criteria; w.r. = study meet evidence standards with reservations; CRT = cluster-randomized trial; RCT = randomized controlled trial; BAU = business as usual; BAU + PD = business as usual with regular PD; AT = alternative treatment, Obs. = live observation.

**Table 2**
Study description (at intervention level): Characteristics of professional development.

| Study | Workshop | Course | Online | Individual component | Individual hours | Group hours | Duration (month) | Modeling | Video | Self-reflection | Feedback |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Barnett et al. (2008) | X | X | – | X | 18 | 41 | 9 | – | – | – | – |
| Breffni (2011) | – | X | – | – | 0 | 32 | 2 | X | – | X | – |
| Domitrovich et al. (2009) | X | – | – | X | 144 | 32 | 9 | X | X | X | X |
| Early et al., 2014; 2017 (1) | – | X | X | – | 0 | 25 | 5 | – | X | – | – |
| Early et al., 2014; 2017 (2) | – | – | X | X | – | – | 8 | X | X | X | X |
| Hamre et al., 2012; 2010 | – | X | – | X | 0 | 42 | 18 | X | X | X | X |
| Hemmeter et al. (2016) | X | – | – | X | 35.2 | 19.5 | 4 | – | X | X | X |
| Morris et al. (2010) | X | X | – | X | 162 | 30 | 10 | X | – | – | – |
| Morris et al., 2014 (1) | – | X | – | X | 51.1 | 48 | 8 | X | – | – | X |
| Morris et al., 2014 (2) | – | X | – | X | 52.6 | 32 | 6 | – | – | – | X |
| Morris et al., 2014 (3) | – | X | – | X | 50.8 | 40 | 8 | – | – | – | X |
| Pianta et al. (2008) | – | – | X | X | – | – | – | X | X | – | X |
| Pronchoenko-Jain (2012) | X | X | – | X | 18.5 | 33 | 6 | X | X | – | X |
| Raver et al. (2008) | – | X | – | X | 82 | 30 | 9 | – | – | – | X |
| Wasik & Hindman, 2011; Hindman & Wasik, 2012 | X | – | – | X | – | 43 | 9 | X | X | X | X |
| Whittaker et al. (2015) | – | X | X | – | 0 | 25.5 | – | X | X | X | – |
| Yoshikawa et al. (2015) | – | X | – | X | – | 48 | 6 | X | – | X | X |
| Zan and Donegan-Ritter (2014) | – | X | – | X | 12.3 | 12 | 8 | X | X | X | X |
| Summary | 33% | 72% | 22% | 83% | 0–162 | 12–48 | 2–18 | 67% | 56% | 50% | 72% |

*Note.* X = component is included in the intervention.

didactics of the evaluated in-service program and other potential effect modifiers (see Tables 1 and 2 for an overview).

At study level, we coded publication year (ICC = 1), the country where the in-service program was undertaken (κ = 1), and type of publication (1 = published in a peer-reviewed journal, 0 = not; κ = 0.98). We further included methodological variables that have been found to serve as moderators in the psychological, behavioral, and educational treatment literature (see Cooper, Hedges, & Valentine, 2009). We also coded the assignment procedure (RCT, randomized controlled trial with assignment at individual level = 1, or CRT, clustered randomized trial with assignment at classroom or center level = 2; κ = 0.77). We also coded the attrition related to intervention (ICC = 0.99) and baseline equivalence of the intervention and control group based on reported information on teacher and classroom characteristics (κ = .77). We subsequently determined for each study an evidence standard index for group designs based on reported information on randomization, attrition, and baseline equivalence (see What Works Clearinghouse [WWC], 2014).

The activities of the control group were coded (κ = 0.79) as business-as-usual without PD (BAU = 1), business-as-usual with regular PD with similar training hours (BAU + PD = 2), or an alternative PD program (AT = 3). We also determined the sample sizes for the experimental and control groups (ICC = 0.94 and 0.93, respectively). In addition, we coded whether the quality ratings were made through observations in classrooms or through videos (κ = 0.88), and whether the study reported that data collectors were blind to condition or not (κ = 0.77). Further, the constructive alignment of PD content and the outcome measures was coded (κ = 1). The percentage of teachers with a Bachelor's or Master's degree in the study sample was coded (ICC = 1). Two contextual variables were coded, too. We coded the percentage of teachers with bachelor or higher in sample (κ = 1) and made dummy code for studies with teachers with predominant academic qualification (≥90%). In respect of classroom compositions, we dummy coded whether or not the PD was implemented and evaluated at Head Start centers (κ = 1).

With regard to the professional development components, we coded various characteristics that involve the implementation and effects of PD programs in ECEC (see Schachter, 2015; Sheridan, Edwards, Marcin, & Knoche, 2009; Snyder et al., 2012). We distinguished between four types of delivery format: workshop (a half- or full-day format), coursework (multiple periodical sessions), online or blended learning, and individual support (e.g., coaching, mentoring) (κ = .76, .87, 0.82, and 0.90, respectively), and determined how many of these components the PD programs included. Moreover, based on the training description we coded whether or not participants received feedback (κ = 0.73), defined as any constructive comment (written or oral) of a trainer related to teachers' performance as part of the PD process (Shute, 2008). The feedback may be related to the teacher, the task, or the process of learning; examples are direct instructions, appraisals, and encouragement [examples: "coaches returned for a second session in which they provided feedback to the teacher and aide", Yoshikawa et al., 2015, p. 313; "the coach providing supportive and constructive performance feedback to the teacher about observed practices", Hemmeter, Snyder, Fox, & Algina, 2016, p. 6]. We also coded whether teachers received demonstrations or modeling of good practice (κ = 0.73), defined as a pedagogical practice where an instructor demonstrates high-quality performance in teacher-student interaction for the teachers during the PD program, possibly complemented with explanations or other instructions [examples: "the mentor modeled the strategy", Domitrovich et al., 2009, p. 571; "coach serving as an expert who explained and demonstrated best practices", Hindman & Wasik, 2012, p. 135], and whether trainees had the opportunity for self-reflection and self-assessment (κ = 0.75) [examples: "The debriefing involved reflection by the teacher and the coach", Hemmeter et al., 2016, p. 6; "Monthly video-based teacher self-reflection was a main component of the professional development", Zan & Donegan-Ritter, 2014, p. 95]. Additionally, we analyzed the general usage of videos in PD (κ = 0.92) [example: "watching videos highlighting effective language instruction", Hamre et al., 2012, p. 101]. We also determined the number of training hours (ICC = 1) and duration of the program (in months, ICC = 0.95). Also, the intensity of the individual component (hours, ICC = 0.98) and the hours spent in a group training setting (ICC = 0.83) were coded and intensity scores were estimated for average individual training hours per month and average training hours in group settings per month. Additionally, we dummy-coded whether the training was scale-based (e.g., when training content was aligned with the rating scale used to measure PD effects) (κ = 1) [example: "The first three sessions provided teachers with information …. and introduction to the three broad domains of the CLASS", Hamre et al., 2012, p. 101; "written teacher reflection guides developed by project staff to focus on specific CLASS dimension", Zan & Donegan-Ritter, 2014, p. 96].

### 2.4. Analysis

Statistical information was transformed into the Hedges' g effect size using Comprehensive Meta-Analysis Software V3 (Borenstein, Hedges, Higgins, & Rothstein, 2014). Effect size estimates were predominantly based on raw descriptive data (pretest and posttest means, standard deviations, standard error, and sample size).

A weighted summary effect g was calculated. A random effects model was applied, which accounted for sampling errors between and within studies. When double and redundant information for the outcome measure was available (i.e., both items and domain score), we used information from single items excluding the domain score.

The homogeneity of effect sizes was tested using the Q statistic. The $I^2$ statistic represents the proportion of observed variance that reflects real differences in effect sizes, indicates the extent of overlap of confidence intervals and quantifies the inconsistencies (Deeks, Higgins, & Altman, 2008). In addition, between-study variance in effect sizes was estimated using Kendall's $\tau^2$ (Borenstein, 2009). A sensitivity analysis was conducted to estimate the robustness of the average effect size. Specifically, a leave-one-out analysis was used to estimate the stability of the average effect size, where one treatment was iteratively removed to identify outlier studies. Furthermore, classic fail-safe N test was carried out to calculate how many studies with null results are needed to reduce the average effect to a non-significant value. Finally, we investigated potential moderators by analyzing subgroups for categorical variables, applying a random model, and with meta-regressions for continuous variables.

**Table 3**
Effects of professional development on CLASS categories: Effect sizes (and standard errors).

| | PC | NC | TS | RSP | ES | PD | BM | ILF | CO | CD | QF | LM | IS | Overall effect |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Barnett et al. (2008) | 0.65 (0.49) | 0.27 (0.48) | 0.95 (0.50) | −0.20 (0.51) | | 1.00 (0.51)* | 0.52 (0.49) | −0.38 (0.48) | | 0.35 (0.48) | 0.57 (0.49) | | | 0.49 (0.49) |
| Breffni (2011) | | | | | 3.22 (0.72)* | | | | 3.15 (0.71)* | | | | 2.73 (0.66)* | 3.04 (0.70)* |
| Domitrovich et al. (2009) | 0.61 (0.32) | 0.16 (0.31) | 0.57 (0.32) | | 0.48 (0.31) | 0.49 (0.31) | 0.07 (0.31) | 0.30 (0.31) | | 0.37 (0.31) | 0.44 (0.31) | | 0.53 (0.32) | 0.38 (0.31) |
| Early et al., 2014; 2017 (1) | | | | | 0.41 (0.14)* | | | | 0.24 (0.13) | | | | 0.30 (0.14) | 0.32 (0.14)* |
| Early et al., 2014; 2017 (2) | | | | | 0.21 (0.14) | | | | 0.11 (0.14) | | | | 0.12 (0.14) | 0.15 (0.14) |
| Hamre et al., 2012; 2010 | 0.31 (0.10)* | 0.12 (0.10) | 0.29 (0.10)* | 0.46 (0.10)* | 0.40 (0.10)* | 0.23 (0.10)* | 0.16 (0.10) | 0.35 (0.10)* | 0.29 (0.10)* | 0.64 (0.10)* | 0.52 (0.10)* | 0.56 (0.10)* | 0.65 (0.10)* | 0.36 (0.10)* |
| Hemmeter et al. (2016) | | | | | 0.18 (0.31) | | | 0.60 (0.30)* | 0.34 (0.31) | | | | −0.08 (0.31) | 0.15 (0.14) |
| Morris et al. (2010) | 0.59 (0.39) | 0.89 (0.20)* | 0.45 (0.30) | 0.28 (0.30) | | 0.62 (0.30)* | 0.71 (0.39) | 0.60 (0.30)* | | | 0.43 (0.39) | 0.53 (0.49) | | 0.57 (0.35) |
| Morris et al., 2014 (1) | 0.20 (0.18) | 0.26 (0.14) | 0.14 (0.18) | −0.02 (0.15) | 0.17 (0.14) | 0.08 (0.18) | 0.39 (0.19)* | 0.13 (0.15) | 0.22 (0.15) | 0.14 (0.11) | 0.15 (0.14) | −0.15 (0.17) | 0.04 (0.12) | 0.13 (0.16) |
| Morris et al., 2014 (2) | −0.01 (0.18) | 0.18 (0.13) | 0.02 (0.18) | 0.11 (0.15) | 0.13 (0.14) | −0.11 (0.18) | 0.10 (0.19) | 0.08 (0.15) | 0.02 (0.15) | 0.33 (0.11)* | 0.29 (0.14)* | 0.16 (0.16) | 0.27 (0.12)* | 0.11 (0.16) |
| Morris et al., 2014 (3) | 0.05 (0.18) | 0.12 (0.13) | 0.07 (0.18) | 0.06 (0.15) | 0.09 (0.14) | 0.10 (0.18) | 0.09 (0.19) | 0.20 (0.15) | 0.14 (0.15) | 0.10 (0.11) | 0.20 (0.14) | 0.02 (0.17) | 0.12 (0.12) | 0.10 (0.16) |
| Pianta et al. (2008) | 0.11 (0.21) | | 0.05 (0.21) | | | −0.12 (0.18) | −0.28 (0.21) | −0.30 (0.20) | | 0.09 (0.18) | | −0.14 (0.18) | | −0.08 (0.20) |
| Pronchoenko-Jain (2012) | | | | | 0.81 (0.58) | | | | | | | | | 0.82 (0.58) |
| Raver et al. (2008) | 0.88 (0.21)* | 0.63 (0.21)* | 0.53 (0.21)* | | | | 0.52 (0.21)* | | | | | | | 0.64 (0.21)* |
| Wasik & Hindman, 2011; Hindman & Wasik, 2012; | | | | | | | | | | | | | 1.40 (0.41)* | 1.40 (0.41) |
| Whittaker et al. (2015) | | | | | 0.14 (0.13) | | | | 0.31 (0.22) | | | | 0.62 (0.15)* | 0.36 (0.17) |
| Yoshikawa et al. (2015) | | | | | 0.81 (0.24)* | | | | 0.43 (0.23) | | | | 0.45 (0.23) | 0.57 (0.23)* |
| Zan and Donegan-Ritter (2014) | 0.70 (0.27)* | 0.78 (0.27)* | 0.50 (0.27) | 0.86 (0.28)* | | 0.68 (0.27)* | 0.86 (0.28)* | 1.42 (0.29)* | | 1.21 (0.08)* | 1.51 (0.30)* | 1.89 (0.32)* | | 1.04 (0.28)* |

*Note.*$p < .05$; PC = Positive Climate; NC = Negative Climate; TS = Teacher Sensitivity; RSP = Regard for Student Perspective; ES = Emotional Support; BM= Behavior Management; PD = Productivity; ILF = Instructional Learning Formats; CO = Classroom Organization; CD = Concept Development; QF = Quality of Feedback; LM = Language Modeling; IS = Instructional Support; LIT = Literacy; A positive effect size in the column for NC indicates a reduction of negative climate in experimental group.

## 3. Results

The methodological characteristics of the sample are summarized in Table 1. All studies were conducted in the US, with the exception of Yoshikawa et al. (2015) which was conducted in Chile. Six studies with seven treatments met evidence standards for group design 'fully' and without reservation, whereas eight studies with 11 treatments met evidence standard with reservation (see Table 1). These studies had at least a random assignment and reported baseline equivalence for groups in analytic sample. Unfortunately, drop out was not reported in most studies. Moderator analyses for the composite score and the three domains sets revealed no significant difference in effect size between studies that met the evidence standard fully or with reservation. All experiments were controlled studies with random assignment at either individual teacher or location level. The training focus of the most evaluated programs (n = 12) was aligned with the CLASS measure. The percentage of teachers with a bachelor degree or higher varied substantially between studies. Six experiments were conducted in Head Start centers. Table 2 presents an overview of the PD characteristics of the included interventions. This overview shows that course formats with an individual component and feedback on interaction behavior are predominant. However, the interventions also show unique differences in their specific curricular design and the intensity of the program. Most of the interventions provided individual support (83%) and combined it with either a workshop or a course. Teachers received training in group settings (e.g., course or workshop) from 12 to 48 h. Individual support varied greatly from none to 162 h. Most of the PD interventions (n = 8) took place over the course of a school year (approx. 8–9 months).

Overall, data was available to estimate 120 effect sizes related to a specific CLASS item or domain (see Table 3), each based on a comparison between an experimental group and a control group. The effect size estimates showed marked differences, ranging from g = −0.38 to 3.22.

### 3.1. Overall effects on the quality of teacher-child interactions

For the aggregation of the PD effect on the CLASS, 107 effect sizes (Hedges' g) were analyzed at domain or item level. For comparability reasons, we included only the original items and domains of the commercial version of the CLASS Pre-K (Pianta et al., 2008a). The aggregated effect size is g = 0.39 (SE = 0.08; 95% CI = 0.23–0.55), corresponding with a robust outcome with a small-to-medium effect size. An iterative analysis leaving out one study did not alter outcomes significantly; adjusted outcomes varied between 0.33 (leaving out Breffni, 2011) and 0.42 (leaving out Pianta, Mashburn, Downer, Hamre, & Justice, 2008b). A funnel plot of effect sizes with sample size (see Fig. 2) and an Egger regression (intercept = 2.00; SE = 0.73, p = .015) indicated that effect sizes were distributed asymmetrically with a slight overrepresentation of studies with relatively large effect sizes and relatively small samples (Harbord, Harris, & Sterne, 2009; Sterne, Egger, & Moher, 2008). In particular, the study by Breffni (2011) is an outlier in several ways. The pilot study had the smallest sample, reported the highest effect sizes, and was the only PD that consisted of a single component.

A file-drawer problem (Rosenthal, 1979) does not seems likely, as fail-safe N analysis indicated that a very large number of studies with null effects, 277, is required to reduce the overall effect to zero. Related to this, our meta-analysis included five treatments from three unpublished studies. We found no statistically significant difference between treatment outcomes from the 'gray' literature ($g = 0.22$; n = 5) and peer-reviewed journals ($g = 0.46$; $n = 13$), $Q = 1.76$, $df = 1$, $p = .184$.

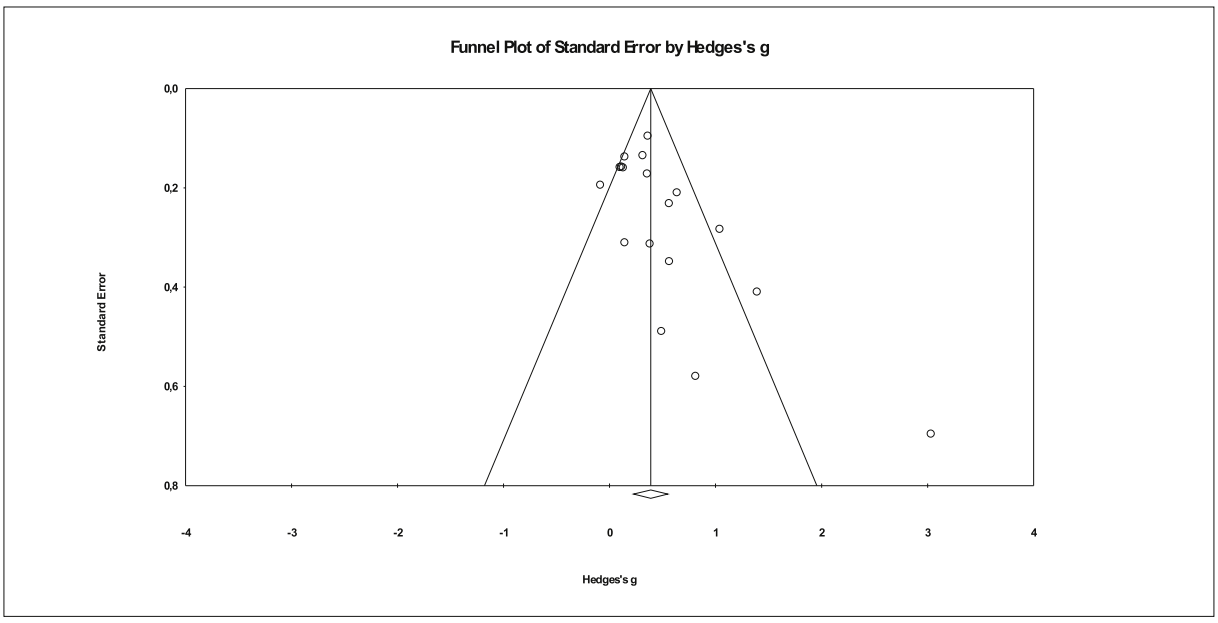


**Fig. 2.** Funnel plot for effect sizes for all CLASS items and dimensions.

#### 3.1.1. Moderator analysis

A heterogeneity analysis showed significant variation between study outcomes ($Q = 44.81$, $df = 17$, $p < .000$; $T^2 = 0.063$; $SE = 0.04$) and a high percentage of systematic between-study variance ($I^2 = 62.06$). We therefore conducted a moderator analysis to explore possible relations between methodological and program-related study characteristics and experimental outcomes (Borenstein, 2009).

None of the following methodological characteristics of the studies were statistically significant moderators of their outcomes: random assignment at individual versus other level ($Q = 0.73$, $df = 1$, $p = .393$); untreated control groups with business as usual or a control group with an alternative treatment ($Q = 1.09$, $df = 1$, $p = .297$); blind observers or not ($Q = 0.23$, $df = 1$, $p = .629$); or comparison of CLASS ratings through live or video observation ($Q = 1.63$, $df = 1$, $p = .201$). Moreover, neither a high percentage (> 90%) of teachers in the sample with a Bachelor's or higher degree ($Q = 1.41$, $df = 1$, $p = .236$) nor evaluation of the treatment at a Head Start facility moderated the PD effects by comparison with other samples (Q = 0.10, df = 1, p = .748). Studies from the CLASS developers (e.g., Hamre et al., 2012; Pianta et al., 2008b) where equal effective than studies from other authors ($Q = 1.22$, $df = 1$, $p = .269$). Unexpectedly, PD that was aligned with outcome measures showed smaller effect sizes than studies without alignment of PD and measures ($Q = 4.13$, $df = 1$, $p = .042$). However, after deleting the outlier study from Breffni (2011) from the analysis, the contrast was no longer significant ($Q = 1.60$, $df = 1$, $p = .206$). Overall, our moderator analyses did not suggest that the meta-analytic outcomes may be systematically biased by methodological artefacts or contextual variables.

Subsequently, we conducted meta-regressions to test possible associations between experimental outcomes and PD-related characteristics. Program duration (*coefficient* = −0.021; $p = .507$), number of PD hours for individual instruction (*coefficient* = −0.002; $p = .506$), and hours of group instruction (*coefficient* = −0.017; $p = .108$) were unrelated to study outcome. There was a trend towards a positive linear relationship with the intensity of the program in group training settings, defined as the number of PD hours per month in workshops or courses (i.e., number of hours of PD, divided by program duration in months; *coefficient* = .087, $p = .091$).

We also tested the possible effect of an individual component as part of the PD program, based on Egert et al. (2018) and Werner et al. (2016). Inclusion of an individual component in the PD program proved not to be related to study outcomes, $Q = .65$; $df = 1$, $p = .420$. Also, the combination of a course with an individual component was not superior to other formats, $Q = 0.82$, $df = 1$, $p = .366$. Nor were the other coded program characteristics significant moderators (i.e., inclusion of a workshop, course with regular instruction unit, blended-learning format with online component). However, the number of program components does seem to matter, $Q = 5.41$, $df = 1$, $p = .020$. In particular, programs with three components (namely a workshop, coursework, and individual support) had an average effect of $g = 0.90$ ($SE = 0.26$), compared with $g = 0.29$ ($SE = 0.06$) for programs with two PD components. Further, interventions with combinations with two or three PD components differed significantly from the single-component intervention evaluated by Breffni (2011), $Q = 19.88$; $df = 2$, $p < .000$.

Other didactic elements were studied in an exploratory fashion. There were no significant differences related to feedback, ($Q = 0.93$; $df = 1$; $p = .335$), modeling and demonstration of target behavior ($Q = 1.08$; $df = 1$, $p = .300$), the use of video in the PD program ($Q = 0.02$ $df = 1$; $p = .891$). There was trend that PD is more effective when the possibility for self-reflection or self-assessment was offered ($Q = 2.76$; $df = 1$; $p = .097$).

### 3.2. Effects on emotional support

In total, 17 treatments with 44 effect sizes, ranging from −0.12 to 3.22, were included in the meta-analysis on emotional support. In line with the first meta-analysis at CLASS level, the aggregated effect size is $g = 0.35$ ($SE = 0.08$; 95% $CI = 0.20$–$0.50$), corresponding to a small-to-medium effect size. Iteratively excluding one treatment from the analysis, this aggregated outcome seemed robust, varying between 0.30 and 0.38. The funnel plot and the Egger regression (intercept = 2.00; $SE = 0.71$, $p = .013$) indicated some asymmetry (see Figure A.1). A fail-safe *N* analysis indicated that 234 studies with null effects are required to reduce the overall effect to zero, and no statistically significant difference was found between study outcomes from the 'gray' literature ($g = 0.15$; $n = 5$) and peer-reviewed journals ($g = 0.34$, $n = 12$; $Q = 0.05$, $df = 1$, $p = .218$. There was a high percentage of systematic between-study variance ($I^2 = 57.28$) and moderate heterogeneity, $Q = 37.45$, $df = 16$, $p = .002$; $T^2 = 0.047$; $SE = 0.033$. We explored the same moderators with regards to methodology, context factors, and PD characteristics. Only the intensity of training, defined as hours spent in group training sessions per month, was negatively related to study outcomes for ES, coefficient = −.081; $p = .028$. Leaving out the study by Breffni (2011) as a statistical outlier, no statistically significant moderators on ES were found.

### 3.3. Effects on classroom organization

Sixteen different treatments with 33 effect sizes on CO, which ranged from $g = -0.38$ to 3.15, resulted in an aggregated effect of $g = 0.30$ ($SE = 0.08$; 95% $CI = 0.14$–$0.46$). The impact was relatively small, but proved robust and varied only from $g = 0.24$ to 0.33 when one PD intervention was iteratively excluded from analysis. Again, some significant asymmetry (see Figure A.2) was indicated, with an overrepresentation of small studies with larger effect sizes and standard errors (intercept = 2.03; $SE = 0.82$, $p = .027$). In line with previous fail-safe *N* and publication-bias analyses, several studies ($N = 139$) with null effects were needed to neglect the impact, and aggregated effect sizes did not differ substantially with regard to publication type (Q = 0.61; $df = 1$, $p = .437$; peer-reviewed $g = 0.34$; $n = 12$ vs. gray literature $g = 0.20$; $n = 4$). Although there was moderate heterogeneity ($Q = 36.87$, $df = 15$, $p = .001$; $T^2 = 0.05$; $SE = 0.037$; $I^2 = 59.31$), none of the moderators significantly explained the variance.

### 3.4. Effects on instructional support

The quality of instructive teacher-child interactions that foster cognitive and language development in young children received great attention in PD, as an important aspect of ECEC. Sixteen different PD interventions were evaluated, with 30 effect sizes with a range of $g = -0.15$ to 2.73. The overall effect was close to medium, with $g = 0.43$ ($SE = 0.10$; 95% *CI*: 0.23–0.63). This aggregated impact was robust, varying between 0.36 and 0.47 when iteratively omitting a single treatment from the analysis. The funnel plot revealed some studies with a large effect size and standard error at the bottom right of the diagram (see Figure A.3), but, according to the Egger regression, the asymmetry was not significant (intercept = 1.66; $SE = 1.09$, $p = .151$). The classic fail-safe $N$ analysis showed that 280 studies with null effects are needed to neglect the overall impact. Moreover, the publication-type analysis showed no evidence of a file-drawer problem ($Q = 2.57$, $df = 1$, $p = .109$; peer-reviewed $g = 0.52$; $n = 12$; gray literature $g = 0.18$; $n = 4$).

We found substantial variance between and within the studies, $Q = 61.22$, $df = 15$, $p < .000$; $T^2 = 0.102$; $SE = 0.06$; $I^2 = 75.50$). Again, the contextual variables, study methodology, PD components, and didactical elements were analyzed but – with two exceptions – did not contribute systematically to the variation in effect sizes. First, we found a tendency for the PD effects to larger when evaluated through video ratings rather than live observations, $Q = 3.74$, $df = 1$, $p = .053$; $g = 0.76$ *vs.* $g = 0.32$. This finding became even more robust when Breffni was excluded, $Q = 6.08$, $df = 1$, $p = .014$. Second, PD interventions that facilitate self-reflection were significantly associated with larger effect sizes (with self-reflection: $g = 0.63$; $SE = 0.13$; without: $g = 0.18$; $SE = 0.13$; $Q = 6.04$, $df = 1$, $p = .014$).

## 4. Discussion

In contrast to previous reviews that aggregated PD effects from heterogeneous primary studies (published until, 2014) with a variety of quality ratings and a mixture of ECEC settings, our meta-analysis is grounded in a well-defined quality construct of teacher-child interaction. Our meta-analytic results from recent studies show an average small to medium effect of professional development in early childhood education and care on the different quality domains distinguished by the CLASS measure. The magnitude of the in-service effect is in line with other reviews (Fukkink & Lont, 2007; Werner et al., 2016). Our review further indicates that extensive PD programs enhance important teacher skills in ECEC classrooms across a wide range of areas, including instructional support, which appears to be a weaker part of teacher's performance (Perlman et al., 2016). The validity of our meta-analytic findings is strengthened by the methodological quality of the integrated experiments, which included a relatively homogeneous set of high-quality studies with consistent construct validity of the evaluated quality of the teacher-child-interactions. Further, significant effects were found across heterogeneous teacher qualifications, service agencies and administrations (e.g., Head Start), and different PD goals.

The PD programs included in our review, aimed to improve the quality of teacher-child interactions in ECEC and shared various program characteristics, like the combination of a course format with an individual component and feedback. Seen from this perspective, the programs from recently published studies are relatively homogeneous. This may indicate that the field is moving towards conceptual alignment of PD design and PD input that is based on empirical findings on classroom behavior and instructional practice to foster child development, as proposed by Pianta et al. (2011). Additionally, PD interventions for ECEC may have also become more homogenous, because of the application of adult learning principles in in-service training, as suggested by Buysse et al. (2009). Further, in two-thirds of the included studies there was a clear constructive alignment between the PD program and the evaluation measures. From a measurement perspective, our meta-analysis demonstrates the sensitivity of the CLASS measure to chart individual teachers' progress as a result of in-service training; this was found at both composite level and for all three domains (i.e., ES, CO, IS). The raise of teachers' CLASS scores after the intervention is found for PD with different formats and varying intensity and in study designs with rating teachers' performance based on either video or observations in classrooms.

### 4.1. Implications for future research and practice

With this review, policy makers are able to make evidence-based decisions about the implementation of PD interventions that is grounded in empirical knowledge on teacher-child-interactions and provides good learning opportunities with individual children or in groups in ECEC practice. Our review provides insight into some of the characteristics of effective PD for ECEC teachers for policy makers and program developers. Often multiple instructional formats were combined when PD focused on the implementation of a new curriculum and transfer of comprehensive knowledge and skills to classroom practice. Further, extensive PD includes self-reflection, feedback or modeling and seems equally effective for different ECEC settings and teachers with varying baseline qualifications. We replicated the finding from Markussen-Brown et al. (2017) that extensive programs with more components seem more effective than isolated, single component programs. The combination of workshop, course, and on-site support may support professional growth effectively by offering a variety of individual learning opportunities. For example, teachers received a general orientation and became attuned to the PD content in an initial workshop first, which was followed by sustained learning opportunities on task over a period of time in a course, and, finally, they received guidance and feedback to fine-tune their interaction skills in practice through coaching. It seems unlikely that a single component can offer these learning experiences for ECEC teachers in a systematic sequence (see Buysse et al., 2009). However, more experimental research is needed to unpack the 'black box' how various formats and elements of PD, individually or in combination, promote professional growth of teachers (Sheridan et al., 2009).

Another exploratory outcome from our moderator analysis was that self-reflection, as part of a broader PD program, seems to contribute to high-quality instructive interactions between ECEC teacher and children. Schön (1983) already mentioned that professional growth begins with critical reflection 'in action' or 'on action'. Self-reflection can therefore be seen as an important tool to

improve teachers' learning and the application of specific teaching and instructional strategies. In several cases, reflection involved a guided self-assessment of teacher performance related to concrete CLASS domains and/or items. Such CLASS-supported reflection may help teachers to focus on the critical aspects of the target behavior practiced within the training program. Further experimental studies should replicate and extend these findings.

Our review shows that PD is generally effective in improving teachers' IS. However, instruction, guided practice and feedback within the IS domain may still require additional effort and innovative tools to support teachers' performance, taking into account the relatively lower (initial) levels. Our review suggests, specifically for IS, that studies with video ratings are more effective than live observations. Instructional supportive interactions often involve fine-grained, linguistically advanced behavior (e.g., why and how questions, back-and-forth exchanges, integrating of previous knowledge, clarification, etc.). Videos may be more suited than live observations for an in-depth analysis of these verbal and cognitive stimulating interactions as part of IS.

With our meta-analysis we could provide first insights to some process variables and functional mechanism (e.g. individual support, self-reflection) and setting (e.g., Head Start) that are related to the effectiveness of PD. However, we are still at an early stage to encode central PD mechanism (Sheridan et al., 2009) related to (1) personal and relational variables of trainer and trainee (e.g., personal characteristics, relation and best fit), and (2) contextual and systematic variables to promote (sustained) change (e.g., agency culture, work environment, knowledge management).

Future studies, in and outside the US, should explore new ways and seek further evidence to effectively support ECEC teachers in offering high-quality learning experiences to their students.

### 4.2. Limitations of this review

First, most studies stem from the United States and only one non-US study was included in our study even though the CLASS is used in over 50 countries (see https://teachstone.com/international/). There are several plausible explanations for the uneven distribution. The CLASS was first invented in the US and is mandatory in Quality Rating and Improvement System in at least 17 states (Hsu, 2017).

Second, the findings of our review are restricted to CLASS studies and so, strictly speaking, cannot be generalized to other outcome measures. The results may, however, be generalized to Quality Rating and Improvement Systems in which the CLASS is included (see Early et al., 2017).

Third, the substantial intercorrelations of CLASS domains (see Pianta et al., 2008a) made it difficult to detect possible differential effects of PD.

Fourth, our moderator findings on PD characteristics were found within multiple, exploratory analyses. Chance capitalization can, therefore, not be excluded. Moderator analysis in a meta-analysis clearly has its limits to reveal effective components in the optimal design of interventions. The analysis is, by definition, correlational. The moderator findings from our review need therefore to be studied in future experimental studies to substantiate the positive effect of significant moderators. This line of future study into effective interventions may profit from different experimental designs that are suited to test theories of change, like the use of multiple baseline designs (see Kennedy, 2005), component study designs (see Bell, Marcus, & Goodlad, 2013) or microtrials (see Howe, Beach, & Brody, 2010).

Fifth, the statistical power to detect significant differences between studies was limited due to the relatively small number of analyzed studies. In addition, the sample of studies was relatively homogenous with respect to a number of PD characteristics, in particular the number of group training hours, the provision of individual support, and feedback. The similarities in didactic approach in our sample of experimental studies may provide tools for future PD interventions, but they do not allow a strong test of possible moderators at program level.

### 4.3. Conclusion

Teachers face a challenging task in taking care of young children's need for security, structure, and learning throughout the daily program. Enhancing the level of these complex teacher-child interaction skills in this dynamic ECEC context is an equally serious challenge. Our review demonstrates that trainers and ECEC teachers can work together effectively in extensive PD formats to enhance key interaction skills in the domains of emotional support, classroom organization and instructional support.

## Funding

The project was supported by a grant from the Federal Ministry for Social Affairs, Senior Citizens, Women and Youth (BMFSFJ) in Germany (BIS.00.00005.15) as part of the national BiSS initiative (Bildung durch Sprache und Schrift/Education through Language and Writing). The BiSS E2 Evaluation project was undertaken at the ifp Bayern.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.edurev.2019.100309.

## References


Studies included in the meta-analysis are marked with an asterisk.

Baldwin, T. T., & Ford, K. J. (1988). Transfer of training: A review and directions for future research. *Personnel Psychology, 41*(1), 63–105.

(*)Barnett, S. W., Jung, K., Yarosz, D. J., Thomas, J., Hornbeck, A., Stechuk, R., et al. (2008). Educational effects of the tools of the mind curriculum: A randomized trial. *Early Childhood Research Quarterly, 23*(3), 299–313. https://doi.org/10.1016/j.ecresq.2008.03.001.

Bell, E. C., Marcus, D. K., & Goodlad, J. K. (2013). Are the parts as good as the whole? A meta-analysis of component treatment studies. *Journal of Consulting and Clinical Psychology, 81*(4), 722–736.

Borenstein, M. (2009). Effect size for continuous data. In H. Cooper, L. V. Hedges, & J. C. Valentine (Eds.). *The handbook of research synthesis and meta-analysis* (pp. 221–235). (2nd ed.). New York: Russell Sage Foundation.

Borenstein, M., Hedges, L., Higgins, J., & Rothstein, H. (2014). *Comprehensive meta-analysis (CMA) version 3.* Englewood, NJ: Biostat.

(*)Breffni, L. (2011). Impact of curriculum training on state-funded prekindergarten teachers' knowledge, beliefs, and practices. *Journal of Early Childhood Teacher Education, 32*(2), 176–193. https://doi.org/10.1080/10901027.2011.572226.

Buysse, V., Winton, P. J., & Rous, B. (2009). Reaching consensus on a definition of professional development for the early childhood field. *Topics in Early Childhood Special Education, 28*(4), 235–243. https://doi.org/10.1177/0271121408328173.

Campbell Collaboration (n.d.)Campbell Collaboration systematic reviews: Policies and guidelines. Retrieved, July 15th 2019 http://www.campbellcollaboration.org/artman2/uploads/1/C2_Protocols_guidelines_v1.pdf.

Cooper, H., Hedges, L. V., & Valentine, J. C. (2009). *The handbook of research synthesis and meta-analysis* (2nd ed.). New York: Russell Sage Foundation.

Curby, T. W., Brock, L. L., & Hamre, B. K. (2013). Teachers' emotional support consistency predicts children's achievement gains and social skills. *Early Education & Development, 24*(3), 292–309. https://doi.org/10.1080/10409289.2012.665760.

Curby, T. W., LoCasale-Crouch, J., Konold, T. R., Pianta, R. C., Howes, C., Burchinal, M., et al. (2009). The relations of observed pre-K classroom quality profiles to children's achievement and social competence. *Early Education & Development, 20*, 346–372. https://doi.org/10.1080/10409280802581284.

Deeks, J. J. H., Higgins, J. P. T., & Altman, D. G. (2008). Analysing data and undertaking meta-analyses. In J. P. T. Higgins, & S. Green (Eds.). *Cochrane handbook for systematic reviews of interventions* (pp. 243–296). Chichester, UK: Wiley-Blackwell.

(*)Domitrovich, C. E., Gest, S. D., Gill, S., Bierman, K. L., Welsh, J. A., & Jones, D. (2009). Fostering high-quality teaching with an enriched curriculum and professional development support: The Head Start REDI program. *American Educational Research Journal, 46*(2), 567–597. https://doi.org/10.3102/0002831208328089.

Downer, J., Brown, J., Herrera, M. J., Stuhlman, M., Bourassa, K., Gologor, B., et al. (2013). Coaching Quality and Teachers' Implementation of the 4Rs social-emotional and literacy curriculum; Testing the link between two levels of intervention fidelity. Society for Research on Educational Effectiveness. Retrieved, December 15th, 2018, from https://eric.ed.gov/?id=ED563032.

(*)Early, D. M., Maxwell, K. L., Ponder, B. D., & Pan, Y. (2017). Improving teacher-child interactions: A randomized control trial of making the most of classroom interactions and my teaching partner professional development models. *Early Childhood Research Quarterly, 38*, 57–70. https://doi.org/10.1016/j.ecresq.2016.08.005.

(*)Early, D. M., Maxwell, K. L., Skinner, D., Kraus, S., Hume, K., & Pan, Y. (2014). *Georgia's pre-K professional development evaluation: Final report*Chapel Hill, NC: University of North Carolina at Chapel Hill.

Egert, F., & Dederer, V. (2018). *Metaanalyse zur Wirkung von Weiterbildungen für pädagogische Fachkräfte zur Steigerung der Interaktionsqualität*. München: IFP Bayern. https://www.ifp.bayern.de/imperia/md/content/stmas/ifp/metaanalyse_interaktionsqualitat_egert__dederer_2018_final.pdf.

Egert, F., Fukkink, R. G., & Eckhardt, A. G. (2018). Impact of in-service professional development programs for early childhood teachers aon quality ratings and child outcomes: A meta-analysis. *Review of Educational Research, 88*(3), 401–433. https://doi.org/10.3102/0034654317751918.

Englund, L. W. (2010). *Evaluating and improving the quality of teacher's language modeling in early childhood classrooms.* Las Vegas, NV: University of Nevada.

Fukkink, R. G., & Lont, A. (2007). Does training matter? A meta-analysis and review of caregiver training studies. *Early Childhood Research Quarterly, 22*, 294–311. https://doi.org/10.1016/j.ecresq.2007.04.005.

Girolametto, L., Weitzman, E., & Greenberg, J. (2003). Training day care staff to facilitate children's language. *American Journal of Speech-Language Pathology, 12*(3), 299–311. https://doi.org/10.1044/1058-0360(2003/076.

(*)Hamre, B. K., Pianta, R. C., Burchinal, M. J., & Downer, J. T. (2010). A course on supporting early language and literacy development through effective teacher-child interactions: Effects on teacher beliefs, knowledge, and practice. *Paper presented at the SREE conference in Washington D.C.*

(*)Hamre, B. K., Pianta, R. C., Burchinal, M., Field, S., Locasale-Crouch, J. L., Downer, J. T., ... Scott-Little, C. (2012). A course on effective teacher-child interactions: Effects on teacher beliefs, knowledge, and observed practice. *American Educational Research Journal, 49*(1), 88–123. https://doi.org/10.3102/0002831211434596.

Hamre, B. K., Pianta, R. C., Downer, J. T., DeCoster, J., Mashburn, A. J., Jones, S. M., et al. (2013). Teaching through Interactions. Testing a developmental framework of teacher effectiveness in over 4,000 classrooms. *The Elementary School Journal, 113*(4), 461–487. https://doi.org/10.1086/669616.

Harbord, R. M., Harris, R. J., & Sterne, J. A. C. (2009). Updated test for small-study effects in meta-analyses. *STATA Journal, 9*(2), 197–210. https://doi.org/10.1177/1536867X0900900202.

Hatfield, B. E., Burchinal, M. R., Pianta, R. C., & Sideris, J. (2016). Threshold in the association between quality of teacher-child interactions and preschool children's school readiness. *Early Childhood Research Quarterly, 36*, 561. https://doi.org/10.1016/j.ecresq.2015.09.005 7´571.

(*)Hemmeter, M. L., Snyder, P. A., Fox, L., & Algina, J. (2016). Evaluating the implementation of the Pyramid Model for promoting social-emotional competence in early childhood classrooms. *Topics in Early Childhood Special Education, 36*(3), 133–146. https://doi.org/10.1177/0271121416653386.

(*)Hindman, A. H., & Wasik, B. A. (2012). Unpacking an effective language and literacy coaching intervention in Head Start. *The Elementary School Journal, 113*(1), 131–154. https://doi.org/10.1086/666389.

Howe, G. W., Beach, S. R. H., & Brody, G. H. (2010). Microtrial methods for translating gene-environment dynamics into preventive interventions. *Prevention Science, 11*(4), 343–354.

Hsu, N. (2017). Highlights and updates from QRIS. Retrieved, July 15th, 2019, from https://info.teachstone.com/blog/highlights-and-updates-from-qris.

Hu, B. Y., Dieker, L., Yang, Y., & Yang, N. (2016). The quality of classroom experience in Chinese kindergarten classrooms across settings and learning activities: Implications for teacher preparation. *Teaching and Teacher Education, 57*, 39–50. https://doi.org/10.1016/j.tate.2016.03.001.

Kennedy, C. H. (2005). *Single-case designs for educational research.* Boston, MA: Pearson Education.

Keys, T. D., Farkas, G., Burchinal, M. R., Duncan, G. J., Vandell, D. L., Weilin, L., et al. (2013). Preschool center quality and school readiness: Quality effects and variation by demographic and child characteristic. *Child Development, 84*(4), 1171–1190. https://doi.org/10.1111/cdev.12048.

Klein, L., & Gomby, D. S. (2008). *A synthesis of federally funded studies on school readiness: What are we learning about professional development?* Washington, D.C.: U.S. Department of Health and Human Services (ASPE) and OPRE).

Leyva, D., Weiland, C., Barata, M., Yoshikawa, H., Snow, C., Treviño, E., et al. (2015). Teacher-child interactions in Chile and their associations with pre-kindergarten outcomes. *Child Development, 86*, 781–799. https://doi.org/10.1111/cdev.12342.

Markussen-Brown, J., Juhl, C. B., Piasta, S. B., Bleses, D., Højen, A., & Justice, L. M. (2017). The effects of language- and literacy-focused professional development on early educators and children: A best-evidence meta-analysis. *Early Childhood Research Quarterly, 38*, 97–115. https://doi.org/10.1016/j.ecresq.2016.07.002.

Mashburn, A. J., Pianta, R. C., Hamre, B. K., Downer, J. T., Barbarin, O., Bryant, D., ... Howes, C. (2008). Measures of classroom quality in prekindergarten and children's development of academic, language, and social skills. *Child Development, 79*(3), 732–749. https://doi.org/10.1111/j.1467-8624.2008.01154.x.

Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. the PRISMA Group. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Med, 6*(7), e1000097. https://doi.org/10.1371/journal.pmed100009.

(*)Morris, P., Mattera, S. K., Castells, N., Bangser, M., Bierman, K., & Raver, C. (2014). *Impact findings from the Head Start CARES demonstration: National evaluation of three approaches to improving preschoolers' social and emotional competence. OPRE report 2014-44*Washington, DC: Office of Planning, Research and Evaluation.

(*)Morris, P., Raver, C. C., Millenky, M., Jones, S., & Llyod, C. M. (2010). Making preschool more productive: How classroom management training can help teachers.

Retrieved, June 15th, 2019, from https://files.eric.ed.gov/fulltext/ED514648.pdf.
Pakarinen, E., Kiuru, N., Lerkkanen, M.-K., Poikkesu, A.-M., Siekkienen, M., & Nurmi, J.-E. (2010). Classroom organization and teacher stress predicts learning motivation in kindergarten children. *European Journal of Psychology of Education, 25*(3), 281–300. https://doi.org/10.1007/s10212-010-0025-6.
Perlman, M., Falenchuk, O., Fletcher, B., McMullen, E., Beyene, J., & Shah, P. S. (2016). A systematic review and meta-analysis of a measure of staff/child interaction quality (the classroom assessment scoring system) in early childhood education and care settings and child outcomes. *PLoS One, 11*(12), e0167660. https://doi.org/10.1371/journal.pone.0167660.
Pianta, R. C., Hamre, B. K., & Downer, J. (2011). Aligning measures of quality with professional development goals and goals for children's development. In M. Zaslow, I. Martinez-Beck, K. Tout, & T. Halle (Eds.). *Quality measurement in early childhood settings* (pp. 297–315). Baltimore, MD: Paul H. Brookes Publishing Co.
Pianta, R. C., La Paro, K. M., & Hamre, B. K. (2008a). *Classroom assessment scoring system (CLASS)*. Baltimore, MA: Brookes.
(*)Pianta, R. C., Mashburn, A. J., Downer, J. T., Hamre, B. K., & Justice, L. (2008b). Effects of web-mediated professional development resources on teacher-child interactions in pre-kindergarten classrooms. *Early Childhood Research Quarterly, 23*, 431–451. https://doi.org/10.1016/j.ecresq.2008.02.001.
Ponitz, C. C., Rimm-Kaufman, S. E., Grimm, K. J., & Curby, T. W. (2009). Kindergarten classroom quality, behavioral engagement, and reading achievement. *School Psychology Review, 38*(1), 102–120.
(*). Pronchenko-Jain, Y. (2012). *Child teacher relationship training (CRTR) with children exhibiting disruptive behavior: Effects on teacher's ability to provide emotional and relational support to students and on student-teacher relationship stress*Doctoral dissertation. Denton, TX: University of North Texas.
(*)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. https://doi.org/10.1016/j.ecresq.2007.09.001.
Rimm-Kaufman, S. E., Curby, T. W., Grimm, K., Nathanson, L., & Brock, L. L. (2009). The contribution of children's self-regulation and classroom quality to children's adaptive behaviors in the kindergarten classroom. *Developmental Psychology, 45*, 958–972. https://doi.org/10.1037/a0015861.
Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin, 86*(3), 638–642. https://doi.org/10.1037/0033-2909.86.3.638.
Salminen, J., Lerkkanen, M., Poikkeus, A., Pakarinen, E., Siekkinen, M., Hannikainen, M., ... Rasku-Puttonen, H. (2012). Observed classroom quality profiles of kindergarten classrooms in Finland. *Early Education & Development, 23*(5), 654–677. https://doi.org/10.1080/10409289.2011.574267.
Sandstrom, H. (2012). The characteristics and quality of pre-school education in Spain. *International Journal of Early Years Education, 20*(2), 130–158. https://doi.org/10.1080/09669760.2012.715012.
Schachter, R. E. (2015). An analytic study of the professional development research in early childhood education. *Early Education & Development, 26*(8), 1057–1085. https://doi.org/10.1080/10409289.2015.1009335.
Schön, D. (1983). *The reflective practitioner*. New York: Basic Books.
Sheridan, S. M., Edwards, C. P., Marcin, C. A., & Knoche, L. L. (2009). Professional development in early childhood programs: Process issues and research needs. *Early Education & Development, 20*(3), 377–401. https://doi.org/10.1080/10409280802582795.
Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research, 78*(1), 153–189. https://doi.org/10.3102/0034654307313795.
Siekkinen, M., Pakarinen, E., Lerkkanen, M.-K., Poikkeus, A.-M., Salminen, J., Poskiparta, E., et al. (2013). Social competence among 6-year-old children and classroom instructional support and teacher stress. *Early Education & Development, 24*(6), 877–897. https://doi.org/10.1080/10409289.2013.745183.
Slot, P. L., Boom, J., Verhagen, J., & Leseman, P. P. M. (2017). Measurement properties of the CLASS toddler in ECEC in The Netherlands. *Journal of Applied Developmental Psychology, 48*, 79–91. https://doi.org/10.1016/j.appdev.2016.11.008.
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*(3), 188–212. https://doi.org/10.1097/IYC.0b013e31825a1ebf.
Sterling Honig, A., & Martin, P. M. (2009). Does brief in-service training help teachers increase turn-talking talk and socratic questions with low-income preschoolers? *NHSD Dialog, 12*(1), https://doi.org/10.1080/15240750802590776 33–33.
Sterne, J. A. C., Egger, M., & Moher, D. (2008). Addressing reporting bias. In J. P. T. Higgins, & S. Green (Eds.). *Cochrane handbook for systematic reviews of interventions* (pp. 297–333). Chichester, UK: Wiley-Blackwell.
Tayler, C., Ishimine, K., Cloney, D., Cleveland, G., & Thorpe, K. (2013). The quality of early childhood education and care services in Australia. *Australasian Journal of Early Childhood, 38*(2), 13–21.
Tripp, T., & Rich, P. (2012). Using video to analyze one's own teaching. *British Journal of Educational Technology, 43*(4), 678–704. https://doi.org/10.1111/j.1467-8535.2011.01234.x.
Von Suchodoletz, A., Fäsche, A., Gunzenhauser, C., & Hamre, B. K. (2014). A typical morning in preschool: Observations of teacher-child interactions in German preschools. *Early Childhood Research Quarterly, 29*(4), 509–519. https://doi.org/10.1016/j.ecresq.2014.05.010.
(*)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. https://doi.org/10.1037/a0023067.
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. https://doi.org/10.1007/s11121-015-0602-7.
What Works Clearinghouse [WWC] (2014). Procedures and standards handbook (Version 3.0). Retrieved November 5th, 2019, from https://ies.ed.gov/ncee/wwc/Docs/referenceresources/wwc_procedures_v3_0_standards_handbook.pdf.
(*)Whittaker, J. V., Kinzie, M. B., Williford, A., & DeCoster, J. (2015). Effects of MyTeachingPartner–Math/Science on teacher–child interactions in prekindergarten classrooms. *Early Education & Development, 27*(1), 110–127. https://doi.org/10.1080/10409289.2015.1047711.
Wieland, C., Ulvestad, K., Sachs, J., & Yoshikawa, H. (2013). Associations between classroom quality and children‘s vocabulary and executive function skills in urban public prekindergarten progam. *Early Childhood Research Quarterly, 28*, 199–209. https://doi.org/10.1016/j.ecresq.2012.12.002.
(*)Yoshikawa, H., Leyva, D., Snow, C. E., Trevino, E., Barata, M. C., Weiland, C., et al. (2015). * Experimental impacts of a teacher professional development program in Chile on preschool quality and child outcomes. *Developmental Psychology, 51*(3), 309–322. https://doi.org/10.1037/a0038785.
(*)Zan, B., & Donegan-Ritter, M. (2014). Reflecting, coaching and mentoring to enhance teacher-child interactions in Head Start classrooms. *Early Childhood Education Journal, 42*, 93–104. https://doi.org/10.1007/s10643-013-0592-7.
Zaslow, M., Tout, K., Halle, T., Whittaker, J. V., & Lavelle, B. (2010). Towards the identification of features of effective professional development for early childhood educators; Literature review. Retrieved June 5th, 2019, from https://files.eric.ed.gov/fulltext/ED527140.pdf.