Professional Development and its Impact on Children in Early Childhood Education and Care: A Meta-Analysis Based on European Studies

Peter Jensen and Astrid Würtz Rasmussen

Peter Jensen: Department of Economics and Business Economics, Aarhus University, Fuglesangs Allé 4, DK-8210 Aarhus V, Denmark. Telephone: (+45) 87165053. E-mail: pje@econ.au.dk

Astrid Würtz Rasmussen (**Corresponding author**): Department of Economics and Business Economics, Aarhus University, Fuglesangs Allé 4, DK-8210 Aarhus V, Denmark. Telephone: (+45) 87164983. E-mail:awr@econ.au.dk

NOTE: This is the final version before proofs (from Feb. 2018). Accepted for publication at *Scandinavian Journal of Educational Research*.

Article DOI (when available): https://doi.org/10.1080/00313831.2018.1466359

Abstract: This study reviews the European evidence on the impact of professional development (PD) of pre-school educators on child outcomes. A meta-analysis investigates how PD of pre-school educators in formal pre-school centers in Europe affects child outcomes. The European studies are quite recent and limited in numbers, and our results show a significant positive effect of PD on child outcomes with an overall effect size of 0.35 (with a 95 % confidence interval from 0.20 to 0.51). The magnitude is slightly smaller than corresponding results based on US studies, but indicates a general positive effect of PD on child outcomes.

Keywords: Professional development, pre-schooling, Europe, meta-analysis

Introduction

In this study, we review the currently existing evidence on the impact of professional development (PD) in early childhood education and care (ECEC) on child outcomes in Europe. We conduct a meta-analysis of the results of European studies focusing on investigating the impact of PD on child outcomes.

In general, PD covers pre-service training at all levels as well as ongoing professional development such as in-service training. The aim of PD is to increase the knowledge and skills of ECEC practitioners and professionals (see also Jensen et al., 2015). PD can take different forms, e.g. formal and informal training, workshops, consultation and coaching, reflective supervision, or collaborative group work between educators. Technological innovations improve accessibility of and content of PD interventions, which e.g. enables popular online based coaching and mentoring (Powell, Diamond, & Cockburn, 2013).[1] The focus of this current study is solely on PD in the form of in-service training, i.e. professional development of educated (and/or working) ECEC practitioners, professionals and teachers. We are able to distinguish and measure ongoing upgrading of skills of these individuals, but do not directly measure impacts from basic education of ECEC professionals.

Rapidly growing evidence shows that all children, disadvantaged children in particular, benefit from early childhood education and care (Melhuish et al., 2015). Positive effects of ECEC are observed for cognitive skills as well as for socio-emotional skills of the children. Much of this evidence relates to the US (see e.g. Duncan and Magnuson, 2013). There is also evidence that ECEC in Europe has positive effects (Hagen, 2017; Ulferts, Anders, Leseman, & Melhuish, 2016), but generally there is

[1] An overview of coaching-based PD interventions is provided by Powell and Diamond (2010).

substantially less research on European ECEC than for the US. Burger (2010) e.g. finds considerable fewer European studies than US studies for his review on effects of various pre-school programs. In the US, one of the main reasons for providing pre-school is to reduce the achievement gap related to income disparity, and to improve children's school readiness as measured by various child outcome measures. To improve child outcomes, their educators need to demonstrate improved knowledge and skills of effective teaching practices. In the European setting, there is (at least in some of the included countries) somewhat less focus on preparing children for school and more focus on developing the children's social skills. Despite European countries generally having a smaller gap in e.g. income between groups (compared to the US), there are still achievement gaps between children of different income groups. Politicians are interested in using ECEC as one way of reducing such gaps (for a more elaborate discussion, see OECD, 2006).

Professional development is a promising way of improving the teaching quality of ECEC and thereby improving child outcomes (Schachter, 2015). Combined with growing interest in evaluating the impact of PD interventions, it has led to the emergence of a number of (mainly US based) reviews of the key features and effects of PD in ECEC (e.g. Powell and Diamond, 2010; Schachter, 2015; Snyder et al., 2012; Zaslow, Tout, Halle, Whittaker, & Lavelle, 2010). Politicians increasingly express interest in and earmark resources for PD interventions in ECEC settings, often focusing at disadvantaged children (Farran and Hofer, 2013). These resources sometimes contain targeted money for evaluation purposes (Hamre, Downer, Jamil, & Pianta, 2012). Our careful study of evidence based on European experiences of PD creates a link between US and European evidence and is able to shed light on the potential importance of the underlying institutional settings for effects of PD on children's outcomes.

PD taking the shape of interventions in center-based ECEC provides a clear way of evaluating the effect of a given new initiative (on the targeted population). Duncan and Magnuson (2013) point out that center-based care is generally more beneficial than home-based care. Datta Gupta and Simonsen (2010) find that formal center-based care outperforms family day care for the general population, however with some variation across different groups of children. Thus, a next step is to investigate possibilities for improving the quality of existing programs of formal center-based ECEC. PD interventions has most systematically been introduced in the US, e.g. in relation to the Head Start program or specific state funded programs, but has also started to appear more systematically in Europe, for example in the Netherlands (e.g. Henrichs and Leseman, 2014). Also in Denmark, Germany, France, and Finland several programs are currently carried out or recently completed (Ecalle et al., 2015; Gasteiger, 2014; Jensen, Holm, & Bremberg, 2013; Jensen, Jensen, & Rasmussen, 2017; Pakarinen et al., 2011). As the US has been first mover in evaluations of PD in pre-schools, most of the existing evaluations of the effect of PD are related to US experiences. Comparisons of effects between studies can be a challenge due to the different types of PD interventions and differences in the implementation and exact dosage of PD within the same type of PD. Even within center-based ECEC, some pre-schools are mainly focusing on childcare, whereas others focus more on actual learning. This complicates the definition of a common quality measure suitable for both care- and learning-types of programs (Farran and Hofer, 2013). It also makes it more likely to observe outliers in the empirical analyses.

Our aim of this study is to collect the currently existing evidence on in-service PD in center-based ECEC specifically focusing on the European research. We are not aware of previous studies collecting *general* evidence on European PD experiences.

However, we are aware of several studies investigating *specific* PD experiences in European countries. We conduct a meta-analysis in which we initially include the nine European studies on PD fulfilling all of our inclusion criteria. These studies represent five different countries: Denmark, France, Germany, The Netherlands, and Wales, and the studies are comparable on several parameters: they focus on a European setting, investigate in-service training, and restrict focus to ECEC in a specific age range and for children in formal ECEC centers. The studies have somewhat different focus points (details are given below), so we test and confirm the robustness of our results to various specifications.

We generally focus on ECEC for children in the ages 0 to 6 years which is the relevant age-range for pre-schooling in Europe (however, with country-specific differences). We focus solely on PD of the pre-school educators within formal ECEC centers (henceforth denoted *pre-schools*).[2] Finally, we investigate the effect only on the children, as positive effects on children have to be the ultimate goal of engaging in PD. Positive effects of PD on e.g. pre-school educators' satisfaction with work could also lead to better child outcomes, but we are not able to address such indirect effects in the current study.[3]

Our contribution is to quantitatively analyze the effect of in-service professional development on child outcomes in Europe and relate the findings to the existing findings from (mainly) the US, thus providing evidence on the importance of the underlying institutional settings. We systematically search for and collect a range of relevant studies of European experiences and conduct a meta-analysis of the results of

[2] We thus do not compare pre-schools with e.g. home care.

[3] See also the later discussion of Inclusion Criterion 2.

these studies. There are already recent comprehensive systematic reviews that cast a wider net, as e.g. Zaslow et al. (2010), Snyder et al. (2012), and Melhuish et al. (2015), but especially for the European literature, there is room for a more detailed focus on the measured child outcomes. A meta-analysis is ideal for such an analysis.

Specifically, based on quantitative evidence this study aims to answer the following two research questions in the European context: (1) For children age 0-6 years in formal childcare, can child outcomes be improved by professional development of pre-school educators? And (2) how large is the effect of in-service training of pre-school educators on child outcomes, and does the effect differ across child outcome domains and/or depend on intensity of the professional development?

Method

Search Strategies

To acquire knowledge on the effect of professional development on child outcomes in a European context, it is important to investigate whether the effects found for US-based interventions are comparable to effects found in Europe where the institutional settings typically differ from those in the US. Thus, we designed our search strategy to be able to collect studies comparable to the existing studies on PD in the US.

Through extensive piloting, three categories of search terms were created and combined in each database using the boolean operators OR within each category, and the operators AND between the three categories. The search categories were: *childcare, professional development,* and *outcome and effect.* The full list of search terms is

included in Appendix A. We searched several relevant databases[4] over the period August 14-24, 2015. The database searches returned a total of 8,507 studies, ranging from 8 to 2,370 hits in each database. The full list of hits in each of the databases is available in Table A2 in Appendix A. We searched for studies from January 1990 to August 2015 (the end-date being the before-mentioned end of the search period). The start date was determined in order to collect all potentially relevant studies for pre-schooling today, and we expected that the results of studies older than 25 years would not be relevant due to potential restructuring of pre-schooling during the latest decades across Europe. However, the start date is clearly debatable, and the studies that turned out to meet our search and inclusion criteria were all from 2009 and later. Thus, with the benefit of hindsight, to reduce the initial number of studies collected, we could have used a much later start date.

The screening and selection of studies was conducted by the authors of this article, who are both experienced researchers specialized in the field of child development and education. Studies were screened using the inclusion criteria defined and described below and in many cases an excluded study would fail on several of the criteria. Most often a study would be excluded due to geographic restrictions, no child outcomes measured, or the investigated children were too old. Home pages and CVs of authors of included studies were investigated to ensure that all relevant studies were found. This resulted in finding updated or published versions of some of the initially included studies. The reference lists of included studies were also checked for additional studies that would meet the inclusion criteria, but no relevant studies were added during

[4] These databases were chosen based on recommendations by university librarians. The databases cover a wide range of studies in education research, psychology, economics, etc.

this process. Finally, our inclusion criteria along with the list of included studies were distributed to a forum of European experts in ECEC and PD from more than 10 countries. No further studies were included during this final process. In the end, a total of 9 studies were included for the meta-analysis.

Inclusion Criteria

All studies generated by the search process were manually screened and selected for the meta-analysis using the seven inclusion criteria described below.

(1) European studies. The most important criterion is that the study is from a European setting, i.e. the intervention is performed in one or more European countries,[5] which ensures fairly similar institutional settings. Then we can be relatively confident that the estimated effects will be applicable to a general European context. Most studies retrieved using the categories of search terms for the database search are related to randomized controlled trials (RCT) on PD in the US (and will thus be excluded from the empirical analysis). Studies combining results of PD in a European and non-European context are excluded if we cannot directly identify the effect of PD in the European case (as is the case for the study by Montie, Xiang, & Schweinhart, 2006).

(2) Studies examining the effect of professional development. The applied definition of PD covers all studies investigating an in-service PD intervention with the aim of improving children's outcomes. Studies aimed

[5] We have decided not to include Russia and Turkey in our definition of Europe as both countries are large and span over two continents; Europe and Asia.

to improve educators' skills or working conditions will only be included if they also measure and examine child outcomes.[6]

(3) *Studies examining the effect on child outcomes*. Child outcomes cover literacy, math skills, or behavior. The latter can be captured by measures such as the Strengths and Difficulties Questionnaire (SDQ), which is targeted at measuring socio-emotional outcomes for children (Goodman, 1997). We include studies investigating both child and teacher outcomes, but exclude studies that *only* measure teacher outcomes.

(4) *Studies focusing on regular care*. Participants are pre-school educators providing care to 0-6 year-old children in pre-schools, kindergartens, or ECEC centers. Studies evaluating PD in both pre-school and primary school, and measuring *the combined effect* of these, are excluded.

(5) *Studies not focusing on specific risk groups of children*. The children included in the current study are not *solely* targeted as special-needs children or disadvantaged children, e.g. children with ADHD, immigrant children, or children with physical disabilities etc. Among the included children *some* of them could have special needs or be defined as disadvantaged based on their observable characteristics. This criterion ensures that the results of the meta-analysis are generally applicable and generalizable beyond a subgroup of disadvantaged children.

---

[6] Many studies on PD in early childhood only investigate teacher outcomes. In Schachter (2015), for example, only 49 % of the included studies contain measures on children's outcome. 40 % of the studies use measures of children's learning while only 11 % use measures of children's behavior.

(6) *Studies reporting effect sizes or the necessary information to calculate effect sizes*. The included studies need to report the data necessary for calculating effect sizes (i.e. the standardized mean difference). If these numbers are not included or can be calculated from the included information in the study, we have to disregard the study.

(7) *Language, time period, and peer-review.* To cover also the 'grey' literature, studies included in this review do not have to be published in a scientific, peer-reviewed journal. However, in practice most of them are because the included studies need to have information that can be coded into effect sizes (inclusion criterion 6). Such information was not available for e.g. the summary articles and various reports initially detected in the search process. Cross-sectional and longitudinal studies from 1990 or later[7] are considered for the analysis as long as the studies are published in English. The language requirement ensures that our study can be replicated and further ensures that we are not asymmetrically including or excluding studies based on our own language skills.

Study Coding

Each included study is coded by the authors for a number of variables including main outcomes of children and covariates/moderators. Since several studies are inaccurately described, we have had to exercise a fair amount of judgement in the coding of the covariates. An example of such inaccurate information is the lack of details about the amount and duration of PD in some of the studies.

[7] See earlier discussion of start- and end-dates under 'search strategies'.

Methods for the meta-analysis

We use meta-analysis to perform a statistical synthesis of the results from the individual studies included in the review. From each study, we extract or calculate an intervention effect in the form of an effect size. Since the studies use different instruments to measure the children's outcome, we use the standardized mean difference as an effect size that is comparable across studies. The standardized mean difference is estimated using the mean scores and standard deviations in the treatment and control groups (and is referred to as Cohen's d – see Borenstein, Hedges, Higgins, & Rothstein, 2009).[8]

Differential experimental designs cause the included studies to vary in the number of treatment groups being tested and the number of outcomes being measured. Such variation can result in statistical dependence between effect sizes. In the main meta-analysis, we therefore aggregate the effect sizes within each study, typically by averaging, before we combine the results over studies. We investigate the sensitivity of our results to this aggregation.

Since the included studies are diverse in some dimensions and we pool all the studies in one meta-analysis, it is highly implausible to assume that the true effect size is identical across studies. Hence, we perform a random-effects meta-analysis to address the variation across studies. All the analyses are conducted using STATA 14 using the add-on command metaan written by Kontopantelis and Reeves (2010).[9]

---

[8] For studies with unequally sized treatment and control groups, the sample sizes are used to calculate the within-groups standard deviation, pooled across groups.

[9] This add-on command on meta-analysis provides similar results and outputs as e.g. the meta-analysis program developed by Borenstein. It provides forest plots, tables with effect sizes and the overall effect, tests of common hypotheses, etc.

Extracting and calculating effect sizes

Few of the studies report an effect size that directly can be included in the meta-analysis. Thus, in most cases we had to extract the necessary information from the studies and calculate the effect size ourselves. In most cases, the studies report t-tests or F-tests for significance of the intervention effects (together with the number of observations in the experimental group and the control group) and based on this information we can calculate the effect sizes using the prescriptions given by Thalmeier and Cook (2002).[10] For outcomes measured on a decreasing scale, we transformed the effect sizes to an increasing scale by changing the sign. This ensures that all effect sizes are comparable. Due to the relatively low number of studies passing the main inclusion criteria, we are not able to conduct separate analyses on each of the outcome domains (literacy, mathematics, and behavior), despite this would have been our preferred choice. Instead, we conduct a suggestive moderator analysis using only two domains, behavior and cognition (the latter being a combination of literacy and mathematics outcomes).

Several studies reported more than one outcome (e.g. in two different domains), or reported both a total score and several subscores. In the former case, we combined the different outcomes by averaging the outcomes, whereas, in the latter case, we use the total outcome measure. Combining all outcomes in a study into an aggregated effect size might reduce the estimated impact on the child outcomes directly impacted by the PD. For example, a behavioral skills PD is designed to directly impact behavior and to

[10] For some studies this strategy was not possible due to lack of information. We e.g. had to exclude the study by Ahsam, Shepherd and Warren-Adamson (2006) because it was impossible to calculate an effect size based on their results.

indirectly impact cognitive outcomes. Hence, aggregating outcomes could obscure larger impacts on the directly impacted outcomes. We perform sensitivity analyses to assess the magnitude of this potential bias (see later e.g. on 'cherry picking').

Results

Description of studies

After the selection process, we ended up with only nine studies included for the meta-analysis. The characteristics of these nine studies are described systematically in Table 1.

[Insert Table 1 about here]

We are particularly interested in investigating effects of PD across child outcome domains and for different intensities of PD. Four studies are related to the cognitive domain with two studies focusing on language and literacy, one study focusing on mathematics, and one study investigates integration of science and language learning (Henrichs and Leseman, 2014). The other five studies are related to the socio-emotional domain focusing on various aspects of the children's behavior. The included studies range from a very low-intensity intervention of three hours of group training (Henrichs and Leseman, 2014) to a much more comprehensive training program of 17 full days running over a two-year period (Jensen et al., 2017). The

majority of the PD interventions involve several training sessions (typically between 4 and 6) organized over a longer period of time.[11]

The majority of the included studies are published during the period 2013-2015. This clearly indicates a research area in its infancy in the European context. All studies rely on an experimental or quasi-experimental design, where children (and educators) are allocated to a control group and an experimental group, respectively. The sample sizes range from using a very small sample of only 38 children (Gasteiger, 2014) to using more than 3,000 children (Ecalle et al., 2015) and the included studies evaluate PD interventions for children being at least 3 years of age. The studies are concentrated on only five European countries. Denmark, Germany and the Netherlands each contribute more than one study whereas France and Wales contribute one study each. There are no studies from Southern or Eastern Europe. Furthermore, no studies on the UK are included (except the one study on Wales).[12]

In the following we describe the studies more detailed. Cviko, McKenney and Voogt (2014) study a technology-rich PD implemented over 8 weeks by a RCT. The PD centers on teacher roles and qualifies as a modern type of PD where the educators play an active role in designing and implementing new technology in ECEC. 196 (162) children participate in the experimental (control) group of this study, and the results

[11] To investigate the importance of a) the domain and b) the intensity of training, we conduct a suggestive moderator analysis on each of these parameters, and we discuss the results in the final Discussion and Conclusion section.

[12] Many UK studies do not rely on interventions with pre- and post-intervention tests as is e.g. the case for the British Effective Pre-School and Primary Education (EPPE) project study (see Sylva et al., 2010).

show generally positive effects of the different teacher roles on children's literacy. The PD intervention is successful for all three sub-groups of the interventions.

Ecalle et al. (2015) evaluates the effect of evidence-based literacy practices by a RCT and finds global effects on different components of literacy skills in the experimental group of children. This study on French kindergartens finds a stronger impact on children in the lower part of the distribution. The study is large in terms of participating children (1932 in the experimental group and 1357 in the control group), but it is a low-intensive PD in the sense that the training session for teachers only lasts one day. The teachers are trained in the use of new pedagogical practices inspired and recommended by evidence-based research.

Gasteiger (2014) evaluates PD targeted at mathematics for children 3 to 6 years old. The PD is implemented over two years and focuses on learning mathematics in natural learning situations, e.g. through every-day activities or just simply while playing. The PD intervention also offers room for reflection. The study includes less than 20 children in each of the experimental and control groups. No significant general effect on the children's mathematical competencies was found, but there were significant improvements for experimental group children in one of the sub-domains.

Henrichs and Leseman (2014) evaluates a very low-intensive PD, which during one training session makes teachers focus on academic language use. The intention is to integrate science and language learning. The study shows modest but promising results for 5-year old children's understanding and use of academic/scientific words and reasoning. 29 children participate in the experimental group and 26 children are allocated to a control group.

Hutchings, Martin-Forbes, Daley and Williams (2013) study a PD intervention implemented over 5 months with the purpose of improving, among other things, child

behavior. 3-7 year old children are investigated in the study (53 in the experimental group and 54 in the control group).[13] Results show a larger effect on children with behavioral problems and PD thus implies a reduction in children's negative behavior.

The two studies from Denmark (Jensen et al., 2013 and Jensen et al., 2017) are quite similar in their PD intervention. The intervention includes a training programme where the pre-school staff learns how to critically reflect on current practices and change these if necessary. The reflective tools are implemented over a period of about 1½ -2 years. The new insights should lead to changes in daily activities and result in a positive impact on SDQ-scores of the included children. In Jensen et al. (2013) a RCT is implemented in 59 pre-schools affecting 524 children in the experimental group and 521 children in the control group. The children are 3 to 6 years old. Jensen et al. (2017) follows the outline of Jensen et al. (2013), but puts more emphasis on evidence-based knowledge in the pre-school staff training program. Furthermore, the children (396 in the experimental group and 290 in the control group) are followed for a longer period of time and in other areas of Denmark. Both studies document positive effects on children's behavior due to professional development of the pre-school educators.

The study by Perels, Merget-Kullmann, Wende, Schmitz, and Buchbinder (2009) shows clear positive effects when evaluating a 5-week PD intervention. The aim of the PD is to positively impact children's self-regulation, and the sample consists mainly of middle-class children which are 5-6 years old. The RCT includes less than 50

[13] Reception classes (ages 4 to 5) as well as multiyear classes (ages 3 to 7) are included, but we cannot separate the 7-year olds from the rest of the children. Since they are not dominating the groups (the mean age of children in both groups is less than 5 years), we decided to include this study.

children in each of the experimental and control groups. In this PD the educators get theoretical knowledge and combine it with practical exercises both at the training sessions and in their kindergarten group. The children in the experimental group become better at understanding and describing different phases of activities, i.e. they improve their self-regulation.

Finally, Rönnau-Böse and Fröhlich-Gildhoff (2009) investigate a RCT PD intervention lasting two years. There are around 100 children in each of the experimental and control groups. 3-6 year old children's self-esteem, behavioral stability, and cognitive development are positively affected because of PD. Thus, this PD intervention accomplished its purpose of systematically involving pre-schools and its professionals in a preventive program aimed at promoting the development of children and the promotion of resilience.

The studies differ in a number of aspects, but more importantly, they have in common that they evaluate the effect of a PD intervention on (young) children's outcomes within the institutional settings provided by formal pre-schooling and ECEC centers in Europe. Hence, we apply a meta-analysis to gain further insight into the overall effect of PD on child outcomes in Europe.

Results of the meta-analysis

We first include all 9 studies in a meta-analysis using a random-effects model. From the forest plot and the heterogeneity statistics (available from the authors upon request), it is obvious that these studies are too heterogeneous to be described by the same model. Cochrane's Q takes the value 83.68 (p-value of 0.000) and thus clearly rejects homogeneity. More specifically, the study by Perels et al. (2009) appears to be an outlier with a very large effect size. The study itself mentions that the measurement

procedure of self-regulatory skills of the children has some flaws, which could perhaps explain the result, despite the procedure is validated using another instrument.

In our main specification, we therefore exclude Perels et al. (2009) from the meta-analysis. There is still a considerable amount of heterogeneity in the effect sizes of the remaining studies (forest plot shown in Figure 1 and results in Table 2), but that is no surprise since we have pooled all effect sizes together regardless of the measured domain. We keep the remaining 8 studies as our preferred final sample and later conduct a sensitivity analysis of this decision. The meta-analysis aggregates the effects of these 8 studies into an overall effect size by weighting each study mainly according to its sample size (see Borenstein et al., 2009). The final result is that the 8 studies contribute to an overall effect size of 0.35 (with a 95 % confidence interval from 0.20 to 0.51). This overall estimated effect size of 0.35 corresponds to PD improving the children's outcome such that instead of 50% of the children being above the previous median value, after the intervention 64% will be above the previous median value.

We are restricted in the possibilities to conduct subgroup analysis of moderator effects due to the few included studies, but we did run meta-regression analyses with the moderators PD intensity and outcome domain. Obviously, the low number of studies hindered us in obtaining any significant results, so we are limited to discussing suggestive evidence based on these estimates (see the discussion and conclusion).

[Insert Table 2 about here]

[Insert Figure 1 about here]

Sensitivity Analyses

From the forest plot in Figure 1, it is obvious that the study by Cviko et al. (2014) may be a possible outlier. To deal with the heterogeneity in the effect sizes of the included studies, we therefore perform a sensitivity analysis of excluding this study from the meta-analysis. The new forest plot and heterogeneity statistics reveal that the remaining 7 studies can be described by the same model (see Table B1 and Figure B1 in Appendix B) with Cochran's Q taking the value 7.00 (p-value 0.321) and thus not indicating heterogeneity. We find an overall effect size of 0.20 (with a 95 % confidence interval from 0.13 to 0.27). This is lower than in the main meta-analysis, reflecting the higher effect size in Cviko et al. (2014) than in the remaining 7 studies. Cviko et al. (2014) reports on 4 separate experiments based on 4 different samples. We averaged the effect sizes over the 4 different samples but instead could have treated them as 4 separate studies. The result from a meta-analysis where Cviko et al. (2014) supplies 4 effect sizes instead of 1 is that the overall effect size increases from 0.35 to 0.40 (with a 95 % confidence interval from 0.26 to 0.54).[14]

In the study by Henrichs and Leseman (2014) there is also some ambiguity in how to select the main effect size for the meta-analysis. Two different tasks were used as basis for the experiment, but in the main meta-analysis, we have only used the effect size from the first task. If we instead average the effect sizes from both tasks, the overall effect size decreases slightly from 0.35 to 0.34 (with a 95 % confidence interval from 0.19 to 0.50). Thus, the results are not very sensitive to the input from this small study.

An alternative to using the average effect size from studies where more than one outcome was measured is to select the largest effect size from each study, i.e. 'cherry

[14] Three of the four effect sizes are above 0.8 and thus quite large.

picking' the results. The overall effect size from the cherry-picking exercise increases to 0.39 (with a 95 % confidence interval from 0.26 to 0.53) and this overall effect size can be thought of as reflecting the maximum potential impact of PD on child outcomes (results available upon request). The result further indicates that the potential bias arising from aggregating directly and indirectly impacted outcomes is rather limited.

Finally, we include all possible effect sizes (from all outcomes) in the meta-analysis (results available upon request). When including all 36 effect sizes from the 8 studies, the result from the random-effects model is an overall estimated effect size of 0.25 (with a 95 % confidence interval from 0.19 to 0.31). The overall effect size decreases compared to the main meta-analysis, because the studies with many outcome measures are those studies with low effect sizes and large sample sizes. Moreover, the standard error of the overall effect size is underestimated, resulting in a narrow confidence interval. Both of these changes are spurious and are consequences of ignoring the dependence between effect sizes from the same study resulting from evaluating on multiple outcome measures for the same sample of children (Van den Noortgate, López-López, Marin-Martinez, and Sánchez-Meca, 2015). A more appropriate way to perform the analysis with all possible effect sizes would be by multivariate or multilevel meta-analysis, requiring knowledge about the correlation between outcome measures (or that the same number of outcome measures is observed for each study). Neither of these requirements is fulfilled, and we therefore could not perform this kind of analysis.

Based on the sensitivity analyses showing quite small differences in overall effect sizes, the main meta-analysis of the previous section remains our preferred specification.

Discussion and Conclusion

This study of the impact of PD in ECEC shows that there exists very little quantitative research on the relation between PD and child outcomes in Europe. A broad search strategy covering more than 8,500 studies, results in less than 10 studies fulfilling the inclusion criteria. The available research on European experiences is quite recent and it is thus likely that the use of RCTs in research on PD is on the rise in Europe.

The included studies cover a number of outcome domains, and if we categorize them broadly into studies in either the cognitive domain or the socio-emotional (/behavioral) domain, a suggestive moderator analysis indicates that the focus of the PD influences the impact of the PD intervention. When focusing PD interventions on the cognitive domain, effects on the children's outcomes are higher (measured in comparable effect sizes) than for PD interventions focused on the socio-emotional domain (although the difference is not significant, due to the low number of studies).

The characteristics of the included studies (reported in Table 1) show that PD has positive effects on child outcomes both for very comprehensive PD interventions and for interventions with a much smaller dosage of PD. The very low-intensity intervention investigated by Henrichs and Leseman (2014) for example shows that only three hours of group training of the educators result in an effect size of 0.64 on the children’s academic language development. Zaslow et al. (2010) also noted in their review of PD interventions that not only comprehensive programs, but even small dosages of PD have been associated with positive child outcomes. When categorizing the studies into two groups according to the intensity of the PD, we find that the intensity of the PD is not strongly linked to the children's outcomes. In fact, the weak relationship is negative (although not significant), such that shorter interventions with a

lower dosage of PD produce the best results in terms of effect sizes. Hence, the focus of the PD is potentially more important than the dosage or intensity.

Another important issue in relation to the impact of PD on child outcomes is implementation. Most of the included studies are silent on implementation issues, especially with regard to fidelity of implementation, but some of the studies attempt to measure e.g. teachers' behavior or attitude, although it is not linked directly to child outcomes. An exception to this is the study by Cviko et al. (2014) which pays particular attention to the implementation, but concludes that "a link between pupil learning outcomes and implementation findings cannot be made easily". This resembles the findings of Zaslow et al. (2010) regarding the lack of analyses of the link between child outcomes and fidelity of implementation. The Danish study (Jensen et al., 2017) investigates the importance of staff turnover (or staff stability) for the effects of PD and finds that a high turnover rate among the staff is a barrier for successful implementation. Thus, the implementation of PD interventions is likely important for later child outcomes, but most of the current studies do not provide enough information on the implementation phase to provide clear suggestions for policy makers.

The studies using relatively large samples (Ecalle et al., 2015; Jensen et al., 2013; Jensen et al., 2017) have smaller effect sizes than the other studies. This feature is often seen in meta-analyses and discussed in detail by Slavin and Smith (2009). Also Fukkink and Lont (2007) observed that large-scale training programs appeared less effective. Our study is thus in line with this previous research. In the main analysis we removed the study with the highest estimated effect size from the sample, i.e. the overall effect size could therefore be viewed as a conservative estimate of the possible effect. Hence, our results strongly point towards a general positive effect of PD on child outcomes in Europe. Even if we also disregard the study by Cviko et al. (2014), we end

up with an effect size of 0.20 (with a 95 % confidence interval from 0.13 to 0.27). The overall estimated effect size of 0.35 corresponds to PD improving the children's outcome such that after the intervention 64 % (or almost two thirds) of the children will be above the previous median value.

Estimated effect sizes based (primarily) on US studies are similar or slightly higher than our estimated overall effect size of 0.35; Fukkink and Lont (2007) report an overall positive effect size of 0.55, although not statistically significant at a 5 % level. Conclusions of Zaslow et al. (2010) point in the direction that PD interventions have positive impacts on child language and literacy, mathematics, and social behavior outcomes. Many US-based studies of specific PD interventions are evaluating randomized controlled trials in relation to Head Start, e.g. the Exceptional Coaching for Early Language and Literacy (ExCell) model described by Wasik and Hindman (2010). These studies may offer great inspiration for interventions that could be adopted to a European formal pre-school or ECEC setting as their results are encouraging for the potential effects of PD in Europe.

Thus, to conclude, our meta-analysis of the few available European studies has promisingly shown that PD has an overall positive effect on children's outcomes in Europe. We find a somewhat smaller overall effect size than what is found in the US-based studies, but our conservative yet positive results could be an indication that the institutional settings are not crucial for adopting effective professional development. Due to the relatively few studies included in the meta-analysis, we cannot provide definitive policy recommendations regarding the most efficient type of PD or the necessary amount of PD needed. Our suggestive evidence implies that for relatively young children, focusing on cognitive outcomes is a promising route. Also, relatively short PD interventions do seem to have an effect so it is not necessarily important to

provide very expensive PD interventions. We cannot say anything about the combination of different types of domains or a combination of short and long interventions, i.e. there is definitely more to the story than our results reveal. Furthermore, despite the fact that the review by Duncan and Magnuson (2013) did not find any variation in effect sizes according to starting age, we will not generalize our results beyond a starting age of 3 in ECEC. New insights into human capital investments generally indicate that "earlier is better" when it comes to investments in children's cognitive and non-cognitive outcomes (Cunha, Heckman, Lochner, and Masterov, 2006), and our estimated positive effect of PD should therefore probably be interpreted as a lower bound of the true effect for children under the age of 3.

More research is clearly needed in order to shed further light on the effect of PD on child outcomes in European pre-schools. We recommend developing a common, flexible design framework for analyses of the impact of PD that can be implemented in different contexts and countries but still provide comparable analyses and effect sizes. A unified framework with comparable studies decreases the importance of culturally-sensitive questions and provides power and degrees of freedom to carry out a complete moderator analysis. This enables focus on identifying the most important and efficient aspects of PD for child outcomes, e.g. the outcome domain and dose of treatment (amount and type of PD and the time period over which the PD is implemented), and provides a promising route for further research.

Acknowledgements: This study is conducted as part of the *Curriculum Quality Analysis and Impact Review of European ECEC (CARE)* project. We thank the European Commission for financial support within the Seventh Framework Programme (Grant Agreement 613318). We would like to thank Lars Lund-

Thomsen and Anne Catharine Andersen for assisting with the systematic search for studies, and Yvonne Anders, Bente Jensen, Paul Leseman, Ted Melhuish, Hannah Ulferts and the rest of the CARE-group for providing very helpful comments. Finally, we would like to thank four anonymous referees for their helpful suggestions and comments. The usual disclaimer applies.

Declaration of interest: None of the authors has any financial interests or benefits arisen from the direct application of the research described in this paper.

References

Ahsam, S., Shepherd, J., & Warren-Adamson, C. (2006). Working with pre-school practitioners to improve interactions. *Child Language Teaching and Therapy, 22*(2), 197-217.

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to Meta-Analysis*. Wiley.

Burger, K. (2010). How does early childhood care and education affect cognitive development? An international review of the effects of early interventions for children from different social backgrounds. *Early Childhood Research Quarterly* 25, 140-165.

Cunha, F., Heckman, J.J., Lochner, L., and Masterov, D.V. (2006). Interpreting the Evidence on Life Cycle Skill Formation. In E. Hanushek & F. Welsh (Eds.), *Handbook of the Economics of Education* (Vol. 1, ch. 12). Elsevier.

Cviko, A., McKenney, S., & Voogt, J. (2014). Teacher roles in designing technology-rich learning activities for early literacy: A cross-case analysis. *Computers & Education, 72*, 68-79.

Datta Gupta, N. and Simonsen, M. (2010), Non-cognitive child outcomes and universal high quality child care, *Journal of Public Economics 94*, 30-43.

Duncan, G. J., & Magnuson, K. (2013). Investing in Preschool Programs. *Journal of Economic Perspectives, 27*(2), 109-132.

Ecalle, J., Labat, H., Le Cam, M., Rocher, T., Cros, L., & Magnan, A. (2015). Evidence-based practices to stimulate emergent literacy skills in kindergarten in France: A large-scale study. *Teaching & Teacher Education, 50*, 102-113.

Farran, D.C., & Hofer, K.G. (2013). Evaluating the quality of early childhood education programs. In Saracho, O. N., & Spodek, B. (Eds), *Handbook of Research on the Education of Young Children* (3rd ed.). Routledge.

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.

Gasteiger, H. (2014). Professionalization of Early Childhood Educators with a Focus on Natural Learning Situations and Individual Development of Mathematical Competencies: Results from an Evaluation Study. In C. Benz, B. Brandt, U. Kortenkamp, G. Krummheuer, S. Ladel & R. Vogel (Eds.), *A Mathematics Education Perspective on Early Mathematics Learning Between the Poles of Instruction and Construction*. Heidelberg: Springer.

Goodman, Robert (1997). The strengths and difficulties questionnaire: a research note. *Journal of Child Psychology and Psychiatry, 38*, 581–586.

Hagen, Å.M. (2017). Improving the Odds: Identifying Language Activities that Support the Language Development of Preschoolers with Poorer Vocabulary Skills. *Scandinavian Journal of Educational Research*. Advance online publication. doi: 10.1080/00313831.2016.1258727

Hamre, B.K, , Downer, J.T., Jamil, F. M. & Pianta, R.C. (2012). Enhancing Teachers' Intentional Use of Effective Interactions with Children: Designing and Testing Professional Development Interventions. In Pianta, R. C. (Ed), *Handbook of Early Childhood Education*. Guilford Press.

Henrichs, L. F., & Leseman, P. P. M. (2014). Early Science Instruction and Academic Language Development Can Go Hand in Hand. The Promising Effects of a Low-Intensity Teacher-Focused Intervention. *International Journal of Science Education, 36*(17), 2978-2995.

Hutchings, J., Martin-Forbes, P., Daley, D., & Williams, M. E. (2013). A randomized controlled trial of the impact of a teacher classroom management program on the classroom behavior of children with and without behavior problems. *Journal of School Psychology, 51*(5), 571-585.

Jensen, B., Holm, A., & Bremberg, S. (2013). Effectiveness of a Danish early year preschool program: A randomized trial. *International Journal of Educational Research, 62*, 115-128.

Jensen, B., Iannone, R.L., Mantovani, S., Bove, C., Struczyk, M.K., & Wyslowska, O. (2015). *Comparative review of professional development approaches.* CARE project; Curriculum Quality Analysis and Impact Review of European Early Childhood Education and Care (ECEC). Retrieved from **http://ecec-care.org/resources/publications/**

Jensen, B., Jensen, P., & Rasmussen, A. W. (2017). Does professional development of preschool teachers improve childrens socio-emotional outcomes? *Labour Economics, 45*, 26-39.

Kontopantelis, E., & Reeves, D. (2010). metaan: Random-effects meta-analysis. *The Stata Journal, 10*(3), 395-407.

Melhuish, E., Ereky-Stevens, K., Petrogiannis, K., Ariescu, A., Penderi, E., Rentzou, K., Tawell, A., Broekhuizen, M., Slot, P., & Leseman, P. (2015). *A review of research on the effects of Early Childhood Education and Care (ECEC) upon child development*. CARE project; Curriculum Quality Analysis and Impact Review of European Early Childhood Education and Care (ECEC). Retrieved from **http://ecec-care.org/resources/publications/**

Montie, J. E., Xiang, Z., & Schweinhart, L. J. (2006). Preschool Experience in 10 Countries: Cognitive and Language Performance at Age 7. *Early Childhood Research Quarterly, 21*(3), 313-331.

OECD (2006). *Starting Strong II: Early Childhood Education and Care*. Paris: OECD Publishing.

Pakarinen, E., Kiuru, N., Lerkkanen, M.-K., Poikkeus, A.-M., Ahonen, T., & Nurmi, J.-E. (2011). Instructional support predicts children's task avoidance in kindergarten. *Early Childhood Research Quarterly, 26*(3), 376-386.

Perels, F., Merget-Kullmann, M., Wende, M., Schmitz, B., & Buchbinder, C. (2009). Improving self-regulated learning of preschool children: Evaluation of training for kindergarten teachers. *British Journal of Educational Psychology, 79*(2), 311-327.

Powell, D. R., & Diamond, K. E. (2010). Improving the Outcomes of Coaching-Based Professional Development Interventions. In Neuman, S. B., & Dickinson, D. K. (Eds), *Handbook of Early Literacy Research* (Vol. 3). Guilford Press.

Powell, D. R., Diamond, K. E., & Cockburn, M. K. (2013). Promising Approaches to Professional Development for Early Childhood Educators. In Saracho, O. N., & Spodek, B. (Eds), *Handbook of Research on the Education of Young Children* (3rd ed.). Routledge.

Rönnau-Böse, M., & Fröhlich-Gildhoff, K. (2009). The promotion of resilience: A person-centered perspective of prevention in early childhood institutions. *Person-Centered and Experiential Psychotherapies, 8*(4), 299-318.

Schachter, R. E. (2015). An Analytic Study of the Professional Development Research in Early Childhood Education. *Early Education and Development*, *26*(8), 1057-1085.

Slavin, R., & Smith, D. (2009). The Relationship Between Sample Sizes and Effect Sizes in Sytematic Reviews in Education. *Educational Evaluation and Policy Analysis*, *31*(4), 500-506.

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.

Sylva, K. , Chan, L. L. S., Melhuish, E., Sammons, P., Siraj-Blatchford, I. & Taggart, B. (2010). Emergent Literacy Environments: Home and Preschool Influences on Children's Literacy Development. In Neuman, S. B., & Dickinson, D. K. (Eds), *Handbook of Early Literacy Research* (Vol. 3). Guilford Press.

Thalmeier, W., & Cook, S. (2002). *How to calculate effect sizes from published research: A simplified methodology*. Work-Learning Research Publication.

Ulferts, H., Anders, Y., Leseman, P., & Melhuish, E. (2016). *Effects of ECEC on academic outcomes in literacy and mathematics: Meta-analysis of European longitudinal studies*). CARE project; Curriculum Quality Analysis and Impact Review of European Early Childhood Education and Care (ECEC). Retrieved from **http://ecec-care.org/resources/publications/**

Van den Noortgate, W., López-López, J. A., Marin-Martinez, F., & Sánchez-Meca, J. (2015). Meta-analysis of multiple outcomes: a multilevel approach. *Behavior Research Methods 47*, 1274-1294.

Wasik, B. A., & Hindman, A. H. (2010). Identifying Critical Components of an Effective Preschool Language and Literacy Coaching Intervention. In Neuman, S. B., & Dickinson, D. K. (Eds), *Handbook of Early Literacy Research* (Vol. 3). Guilford Press.

Zaslow, M., Tout, K., Halle, T., Whittaker, J. V., & Lavelle, B. (2010). *Toward the identification of features of effective professional development for early childhood educators. Literature review.* Washington, D.C: U.S. Department of Education.

Appendix A

[Insert Table A1 about here]

[Insert Table A2 about here]

Appendix B

[Insert Table B1 about here]

[Insert Figure B1 about here]

Tables and Figures

Tables appear in the following order (originally in a separate file, now below figures):

Table 2

Table A1

Table A2

Table B1

Table 1

Figures appear in the following order (originally in a separate file)

Figure 1

Figure B1

Figure 1. Forest plot for final meta-analysis

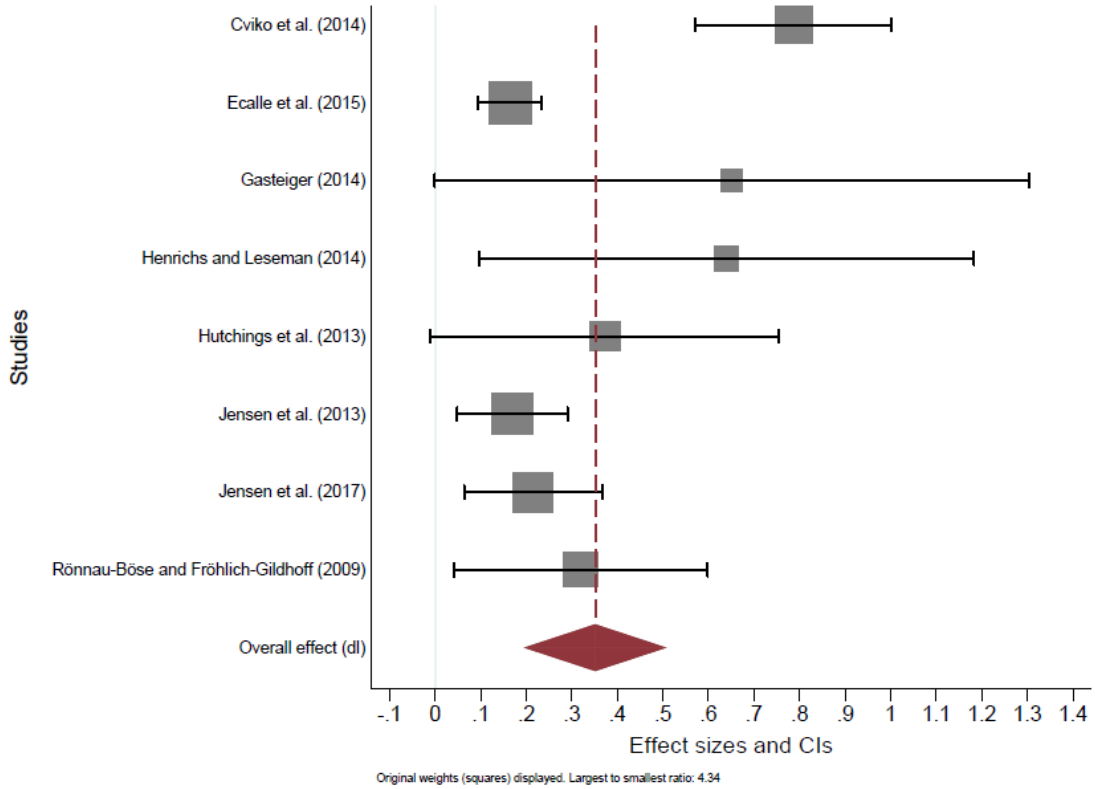

Figure B1. Forest plot for meta-analysis with 7 studies

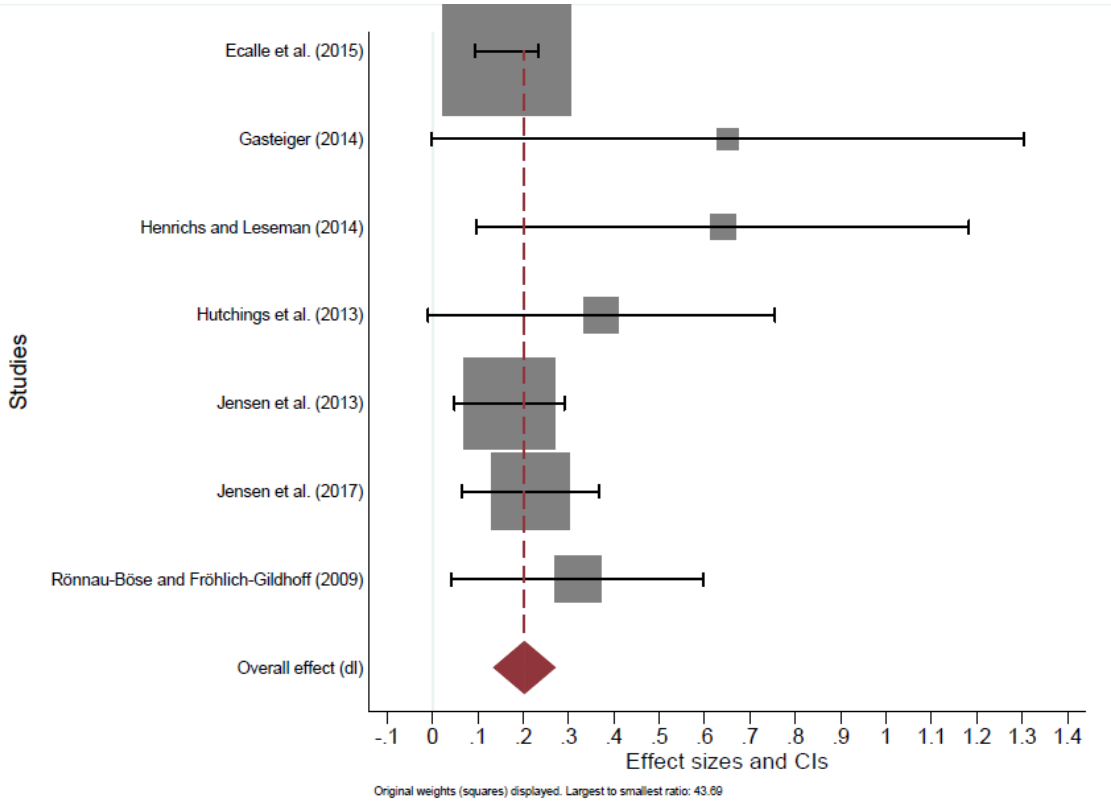

Table 2

Table A1

Table A2

Table B1

Table 1

Table 2. Results from final meta-analysis

Study	Effect	[95% Conf. Int.]		% Weight
Cviko et al. (2014)	0.785	0.571	1.003	14.47
Ecalle et al. (2015)	0.165	0.095	0.235	19.26
Gasteiger (2014)	0.652	-0.001	1.305	4.44
Henrichs and Leseman (2014)	0.640	0.097	1.183	5.84
Hutchings et al. (2013)	0.373	-0.009	0.755	9.08
Jensen et al. (2013)	0.170	0.049	0.291	17.86
Jensen et al. (2017)	0.216	0.064	0.368	16.83
Rönnau-Böse and Fröhlich-Gildhoff (2009)	0.320	0.042	0.598	12.22
Overall effect (dl)	0.352	0.196	0.507	100.00
Test of homogeneity: χ^2 (7) = 34.58, p-value=0.000				

Table A1. Search terms

Childcare
"Child* care" or childcare or "Early childhood" or preschool* or "pre-school*" or "pre school" or "human capital development" or kindergarten or "pre k" or "pre-kindergarten" or "primary education*" or "early education*"
Professional development
"Centre based" or "Classroom practice" or "education of teach*" or "Embedded instruction" or "in service training" or "inservice training" or "Professional development" or "Teach* education" or "Teach* improvement*" or "Teach* training" or "training of teach*" or "Teach* skill*"
Outcome and effect
"Academic achievement*" or "cognitive competence*" or "cognitive skill*" or "cognitive development" or "cognitive outcome*" or "Conceptual knowledge" or "Educational attainment" or "knowledge gain*" or "Outcome measur*" or "Outcome* of education*" or "Science education intervention*" or "Scientific knowledge" or SDQ or "Socio-emotional" or "Socioemotional" or "Socio emotional" or literacy or numeracy or vocabulary or "language development*" or "language skill*" or math* or "self-control" or "pro-social"

Table A2. Databases and no. of hits

Database	No. of hits
ERIC	2370
PsycINFO	742
Econlit	8
ABI/inform	243
ProQuest Education Journals	609
LLBA	239
Scopus	704
Web of Science	375
Business Source Complete	38
Academic search elite	947
British education index	163
Education research complete	1654
Teacher reference center	415
Total	8507

Table B1. Results from meta-analysis with 7 studies

Study	Effect	[95% Conf. Int.]		% Weight
Ecalle et al. (2015)	0.165	0.095	0.235	47.82
Gasteiger (2014)	0.652	-0.001	1.305	1.09
Henrichs and Leseman (2014)	0.640	0.097	1.183	1.57
Hutchings et al. (2013)	0.373	-0.009	0.755	3.12
Jensen et al. (2013)	0.170	0.049	0.291	23.87
Jensen et al. (2017)	0.216	0.064	0.368	16.80
Rönnau-Böse and Fröhlich-Gildhoff (2009)	0.320	0.042	0.598	5.72
Overall effect (dl)	0.203	0.134	0.272	100.00
Test of homogeneity: χ^2 (6) = 7.00, p-value=0.321				

Table 1. Characteristics of included studies

Study	Country	Domain	Age of children	Focus	Type of PD	Amount of PD
Cviko et al. (2014)	The Netherlands	Literacy	4-6 years	Technology-rich early literacy learning activities	Teacher roles in designing technology-rich activities	8 weeks of implementation
Ecalle et al. (2015)	France	Literacy /reading	5-6 years	Evidence-based literacy practices	Educator training	One day
Gasteiger (2014)	Germany	Mathematics	3-6 years	Early mathematics instruction	Educator training	4 modules, total of 24 hours, over two years
Henrichs and Leseman (2014)	The Netherlands	Language /science	5 years	Early science instruction	Group training of educators	Three hours
Hutchings et al. (2013)	Wales	Behavior	3-7 years	Behavior	Incredible Years – Teacher Classroom Management	5 times a full day, over 5 months
Jensen et al. (2013)	Denmark	Behavior	3-6 years	Socio-emotional competences	Workshops and reflection groups	41 hours, over 2 years (imprecise reported)
Jensen et al. (2017)	Denmark	Behavior	3-6 years	Socio-emotional competences	Educator training, workshops	17 full days, over 2 years
Perels et al. (2009)	Germany	Self-regulation	5-6 years	Self-regulated learning	Training of kindergarten educators	5 times 120 minutes, over 5 weeks
Rönnau-Böse and Fröhlich-Gildhoff (2009)	Germany	Resilience	3-6 years	Building resilience within children	Training and supervision of educators	6 training sessions, over 2 years (imprecise reported)

Table 1, continued

Study	Outcome measure	Instrument	Sample size, Exp./con. group	Design	Statistical method	Effect sizes
Cviko et al. (2014)	Early literacy	Own instrument	196/162	Quasi-experimental	Analysis of covariance	Calculated from F-test, avr. calculated (from 4 groups)
Ecalle et al. (2015)	Literacy skills	Own instrument	1932/1357	Experimental	Analysis of variance, matching	Average calculated (from 6 measures)
Gasteiger (2014)	Mathematics competence	Own instrument	19/19	Quasi-experimental	Analysis of variance	Calculated from basic information
Henrichs and Leseman (2014)	Language developm.: lexical diversity	Own instrument, group level	29/26	Experimental	Analysis of variance	Calculated from F-test, avr. calculated (from 3 measures)
Hutchings et al. (2013)	Behavior	TPOT	53/54	Experimental	HLM analysis	Average calculated (from 6 measures)
Jensen et al. (2013)	Behavior	SDQ (educator rated)	524/521	Experimental	Value-added analysis	Average calculated (from 4 measures)
Jensen et al. (2017)	Behavior	SDQ (educator rated)	396/290	Experimental	Value-added analysis	Extracted from paper
Perels et al. (2009)	Self-regulation	Own instrument	48/49	Experimental	Analysis of covariance	Calculated from F-test
Rönnau-Böse and Fröhlich-Gildhoff (2009)	Self-esteem a.o.	Standardized tests (SKF)	99/102	Experimental	Multivariate variance analysis	Calculated from F-test