Curriculum and Quality Analysis and Impact Review of European Early Childhood Education and Care

613318

CARE

Curriculum Quality Analysis and Impact Review of European ECEC

Instrument: Collaborative project
Call Identifier: FP7-SSH-2013-2
Early childhood education and care: promoting quality for individual, social and economic benefits

D3.2: Professional Development and its Impact on Children in Early Childhood Education and Care: A Meta-Analysis Based on European Studies

Final report: March 31, 2016

Short summary: This study reviews the European evidence on the impact of professional development (PD) of pre-school educators on child outcomes. A meta-analysis is used to investigate how PD of pre-school educators affects child outcomes. The studies included in the meta-analysis all concern PD in formal pre-school centers in Europe, and focus is on effects directly on child outcomes. Thus, potential effects of PD on the pre-school educators' skills, beliefs, attitudes and work conditions are not investigated. The geographical restriction excludes the wide range of studies from the US, but these studies still form the base for comparison of the estimated effect sizes. The range of relevant European studies is quite small, but despite the small number of studies included, our results show a significantly positive effect of PD on child outcomes. We find an overall effect size of 0.35 (with a 95 % confidence interval from 0.20 to 0.51).

DUE DATE OF DELIVERABLE: DEC. 31, 2015
ACTUAL SUBMISSION DATE: DEC. 30, 2015

Start date of project: 01-01-2014 Duration: 36 Months
CARE contractor:
Utrecht University

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

Organisation: Aarhus University
Authors (main authors in bold): **Peter Jensen** and **Astrid Würtz Rasmussen**
Email: pje@econ.au.dk and awr@econ.au.dk
Number of PM: 8
Dissemination Level: PU

This study was conducted with a grant from the Seventh Framework Programme of the European Union, Grant Agreement 613318.

Project co-funded by the European Commission within the Seventh Framework Programme (2014-2017)		
Dissemination Level		
PU	Public	x
PP	Restricted to other programme participants (including the Commission Services)	
RE	Restricted to a group specified by the consortium (including the Commission Services)	
CO	Confidential, only for members of the consortium (including the Commission Services)	

Executive Summary

This report is part of the project *Curriculum Quality Analysis and Impact Review of European Early Childhood Education and Care (CARE)*, funded by the European Union within the Seventh Framework Programme, and the current study is part of WP3 *Professional Development: Impact and Innovation*.

This report reviews the currently existing evidence on the impact of professional development (PD) in early childhood education and care (ECEC) on child outcomes in Europe. An essential part of this study is a meta-analysis of the results of existing European studies focusing on investigating the impact of PD of pre-school educators on child outcomes, i.e. we conduct a quantitative analysis of the effect of professional development on child outcomes in Europe. We systematically search for and collect a range of relevant studies of European experiences and conduct a meta-analysis of the results of these studies.

The studies included in the meta-analysis all concern PD in formal pre-school centers in Europe. Focus is on effects directly on child outcomes and we thus ignore potential effects of PD on the pre-school educators' own outcomes. We have chosen to focus primarily on the effects on child outcomes, as positive effects for children have to be the ultimate goal of engaging in PD.

Professional development is increasingly being addressed as a potential way of improving the teaching quality of ECEC and thereby improving child outcomes (Schachter, 2015). This has led to a growing interest in evaluating the impact of PD interventions and to the emergence of a number of reviews of the effects of PD in ECEC (e.g. Schachter, 2015; Powell and Diamond, 2010; Zaslow et al., 2010). These reviews are mainly based on US research.

The aim of this study is therefore to review the currently existing evidence on PD specifically focusing on the European research. This is of particular relevance in the context of the project CARE. Furthermore, we are not aware of previous studies collecting *general* evidence on European PD experiences, despite having knowledge of several studies investigating specific PD experiences in European countries. In summary, this review and meta-analysis aims to answer the following 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?
2. What is the effect of in-service training of pre-school educators on child outcomes?

The meta-analysis in the current study initially includes nine European studies that pass some rather strict inclusion criteria. All included studies investigate PD in formal pre-schools (i.e. they only focus on changes within pre-schools) and are based on data from Denmark, France, Germany, The Netherlands, and Wales. The studies have different focus points despite all being concerned with child outcomes: Literacy, reading or language (Ecalle et al., 2015; Cviko et al., 2014, and Henrichs and Leseman, 2014), mathematics (Gasteiger, 2014), or behavior and self-regulation (Hutchings et al., 2013; Jensen et al., 2013; Jensen et al., 2015; Perels et al., 2009, and Rönnau-Böse et al., 2009). Thus, there is a lot of variation in the outcomes investigated. Despite the small number of studies included and the wide variety of outcomes, our results show a significantly positive effect of PD on child outcomes. We find an overall effect size of 0.35 (with a 95 % confidence interval from 0.20 to 0.51).

Previous reviews primarily based on US studies (e.g. Fukkink and Lont, 2007; Zaslow et al., 2010) have similarly concluded that PD has a positive impact on child outcomes. Fukkink and Lont (2007) perform a meta-analysis and report an overall effect size of 0.55, although this is not statistically significant (at 5% level). This is slightly higher than our estimated overall effect size of 0.35 from European studies, but illustrates that the European studies do not necessarily deviate much from the US studies. Moreover, a recent meta-analysis of the effects of PD on child outcomes in language and emergent literacy (Markussen-Brown et al., 2015, also relying on studies from North America) finds statistically positive effects for different outcomes, with overall effect sizes in the range of 0.18-0.46.

To conclude, more research in a European context is clearly needed in order to shed more light on the effect of PD in pre-schools on child outcomes. The results from the US are interesting and encouraging but not necessarily directly comparable to effects of PD in Europe. It is therefore very important that we also have research-based evidence from Europe. The results from the current meta-analysis point in the direction that professional development has a clear positive impact on child outcomes. The exact mechanism(s) need to be further explored, but it can definitely be concluded that a greater focus on PD will be beneficial for children across Europe. With this evidence, and hopefully evidence from more studies in the future, we can compare to the lessons learned from the US context in more detail.

Table of Contents

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

Peter Jensen, Aarhus University
Astrid Würtz Rasmussen, Aarhus University

Abstract
This study reviews the European evidence on the impact of professional development (PD) of pre-school educators on child outcomes. A meta-analysis is used to investigate how PD of pre-school educators affects child outcomes. The studies included in the meta-analysis all concern PD in formal pre-school centers in Europe, and the focus is on effects directly on child outcomes. Thus, potential effects of PD on the pre-school educators' skills, beliefs, attitudes and work conditions are not investigated. The geographical restriction excludes the wide range of studies from the US, but these studies still form the base for comparison of the estimated effect sizes. The range of relevant European studies is quite small, but despite the small number of studies included, our results show a significantly positive effect of PD on child outcomes. We find an overall effect size of 0.35 (with a 95 % confidence interval from 0.20 to 0.51).

Keywords: Professional development, pre-schooling, Europe, meta-analysis
JEL classification: I21, J13, J24

I N T R O D U C T I O N

This report is part of the project *Curriculum Quality Analysis and Impact Review of European Early Childhood Education and Care (CARE)*, funded by the European Union within the Seventh Framework Programme. The task reported on here is part of WP3, entitled *Professional Development: Impact and Innovation*. More specifically, the aim of this report is to deliver a *"Report and scientific paper on the effects of professional development, based on quantitative analyses: Report and scientific paper on the impact of professional development among ECEC practitioners upon children's short-, medium-, and long-term developmental outcomes based on existing data (longitudinal studies, programmes compared to findings from literature review) and other studies on professional development systems and strategies."* (according to Annex I – "Description of Work" in the CARE grant agreement). Thus, in the following 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. As an essential part of this task, we conduct a meta-analysis of the results of existing European studies focusing on investigating the impact of PD on child outcomes.

In general, PD covers pre-service training at all levels, and, more importantly for this current study, 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., 2015b). PD can take different forms, e.g. training, workshops or collaborative group work between educators. Technological innovations can be used to improve accessibility of and content of PD interventions, and finally, coaching or mentoring has also become a popular, however rather expensive, type of PD (Powell et al., 2013).[1] Thus, PD applies to a widespread range of activities such as formal coursework, consultation, coaching, and reflective supervision.

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, or Elango et al., 2015). More recently there is also evidence that ECEC in Europe has positive effects (see Anders et al., 2015), but generally there is 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.

Center-based childcare is an important type of ECEC, and it is thus interesting to focus on the quality of center-based childcare and how to further improve the quality in order to enhance child outcomes. As pointed out by Duncan and Magnuson (2013), center-based care is generally more beneficial than home-based care. Datta Gupta and Simonsen (2010) find that formal center-based care also outperforms family day care for the general

[1] A recent overview of coaching-based PD interventions is provided by Powell and Diamond (2010), whereas Wasik and Hindman (2010) describe one specific and promising coaching model of PD for pre-school educators, the Exceptional Coaching for Early Language and Literacy (ExCell) model.

population, however with some variation across different groups of children. The next step now is thus investigating possibilities for improving the quality of existing programs of formal center-based ECEC.

Professional development is increasingly being addressed as a potential way of improving the teaching quality of ECEC and thereby improving child outcomes (Schachter, 2015). This has naturally led to a growing interest in evaluating the impact of PD interventions and to the emergence of a number of reviews of the effects of PD in ECEC (e.g. Schachter, 2015; Powell and Diamond, 2010; Zaslow et al., 2010), again mainly based on the US research. Another explanation for the growing literature evaluating PD interventions in relation to ECEC is the interest expressed from politicians. They to a larger degree earmark resources for ECEC settings, often focusing at disadvantaged children (Farran and Hofer, 2013), and these resources sometimes contain targeted money for evaluation purposes (Hamre et al., 2012).

PD taking the shape of interventions in center-based ECEC provides a clear and relatively simple way of evaluating the effect of a given new initiative (on the targeted population). PD interventions has most systematically been introduced in the US, e.g. in relation to the Head Start program, but has also recently 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 (Jensen et al., 2013; Jensen et al., 2015; Gasteiger, 2014; Ecalle et al., 2015; and Pakarinen et al., 2011).[2] 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.

Thus, experiences from the US generally inspire new PD interventions around the world. Due to the many different types of interventions and the different ways of implementing a specific type of intervention, comparisons of effects between studies have proven to be a challenge. Not only do the PD interventions differ in type, they also differ in duration, and the exact dosage of PD can therefore vary within the same type of PD. Furthermore, a specific type of PD can be implemented in rather different types of ECEC. Even within center-based ECEC, some are mainly focusing on child care (with learning being a sort of by-product), whereas others focus more on actual learning. This makes it extremely difficult to define a general or common quality measure suitable for both care- and learning-types of programs (Farran and Hofer, 2013).

Our aim of this study is to review the currently existing evidence on PD specifically focusing on the European research. This is of particular relevance in the context of the project CARE. 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. Some of these studies focus on effects of PD on educator outcomes, some on child outcomes, and some on both. In the current study, we primarily focus on the effects on children, as positive effects on children have to be the ultimate goal of engaging in PD. We conduct a meta-analysis in which we initially include nine European studies on PD based on data from Denmark, France, Germany, The Netherlands, and Wales. The studies have different focus points despite all being concerned with child outcomes: Literacy, reading or language (Ecalle et al., 2015; Cviko et al., 2014 and Henrichs and Leseman, 2014), matematics (Gasteiger, 2014), or behavior and self-regulation (Hutchings et al., 2013; Jensen et al., 2013; Jensen et al., 2015; Perels et al., 2009 and Rönnau-Böse et al., 2009). Thus, after restricting the studies of

[2] The UK has implemented the British Effective Pre-School and Primary Education (EPPE) project study (for further descriptions see Sylva et al., 2010). EPPE is not based on interventions but instead systematically measures quality of pre-schools and relates these measures to child outcomes.

interest to be concerned with PD effects on child outcomes, there is a lot of variation in the actual outcomes investigated. Some studies even include several outcomes. We therefore investigate the sensitivity of our results to aggregation or averaging of outcomes, and we also test different specifications of the meta-analysis.

CONTRIBUTION AND RESEARCH QUESTIONS

In the following we focus on ECEC for children in the ages 0 to 6 years, but only if the children are not already enrolled in school. In some countries, such as e.g. UK, the compulsory age of starting school is lower than 6 years (and some children are enrolled in school as early as from age 4). Thus, children in these countries already enrolled in school are disregarded unless it is clear that the focus is more on care than school activities. These country-specific differences inevitably imply a varying age-range of the children of interest, and we therefore specify the age range of the children in each of the included studies (see Table 1 in the Results section). In this study we focus solely on PD within pre-schools and thus do not compare pre-schools with e.g. home care. To simplify, we will in the following refer to child care in formal ECEC centers as *pre-schooling* in short. We focus on PD of the pre-school educators in these centers, and then investigate the effect on the children. Most often the PD is introduced as an intervention in pre-schools.

The contribution of the current study is to quantitatively analyze the effect of professional development on child outcomes in Europe.[3] Based on a brief review of key international literature on the impact of PD on child outcomes, we systematically search for and collect a range of relevant studies of European experiences and conduct a meta-analysis of the results of these studies.

Specifically, this review aims to answer the following 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?
2. What is the effect of in-service training of pre-school educators on child outcomes?

M E T H O D

SEARCH STRATEGIES

This study is part of the *Curriculum Quality Analysis and Impact Review of European ECEC (CARE)* project. The purpose of this current study is to acquire knowledge on the effect of professional development on child outcomes in a European context. A substantial amount of studies of PD already exists in the US, but there is a clear lack of evidence on European PD initiatives. It is important to investigate whether the effects found for US-based interventions, e.g. in relation to the Head Start Program, 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

[3] We do not include Russia and Turkey in the European countries covered in this review.

collect studies for the current review of PD in Europe, using content that is hopefully 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. The search strings contained the necessary terms for conducting the current review of PD.

The following databases were searched over the period August 14-24, 2015: ERIC, PsycINFO, EconLit, ABI/Inform, ProQuest Education Journals, Linguistics and Language Behavior Abstracts, Business Source Complete, Academic Search Elite, British Education Index, Education Research Complete, Teacher Reference Center, Web of Science, and Scopus. The database searches returned a total of 8,507 studies, ranging from 8 to 2,370 hits in each database. The number of hits in each of the databases can be seen in Table A2 in Appendix A.

The screening and selection of studies was conducted by the authors of this report, who both are experienced researchers specialized in the field of child development and education. Studies passing an initial screening were further screened using the inclusion criteria defined and described below.

In order to make sure that all relevant studies were found, home pages and CVs of authors of included studies were further investigated. The reference lists of included studies were also checked for additional studies that would meet the inclusion criteria. Finally, in order to further extend the list of included studies, our inclusion criteria along with the list of included studies were distributed to a forum of European experts in ECEC and PD. The forum contained experts 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

As described above, the goal of this report is to analyze the effect of professional development in ECEC on child outcomes in Europe. The studies generated by the search process were screened and selected for the meta-analysis using the inclusion criteria that are described in this section.

***European studies*.** The study has to be done in a European setting, i.e. the intervention should have been performed in one or more European countries. Most studies retrieved using the categories of search terms for the database search are related to PD in the US, as there are few relevant European studies. In the US, (long-term) strategic funds for investigating PD in pre-schooling (often Head Start centers) has resulted in a large range of randomized controlled trials (RCT) in US pre-schools. These RCTs have led to numerous studies of effects of PD e.g. in the shape of coaching of pre-school educators, special intervention programs for children and pre-school educators, or programs involving also the parents. As we specifically want to focus on European experiences with PD, we are excluding these US studies from the empirical analysis (even though they may contain interesting results on the impact of PD, but as mentioned above there are numerous reviews of the literature based on US studies).

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. This is for example the case for the study by Montie, Xiang and Schweinhart (2006) where results from either 7 or 10 different countries have been combined into one measure.

Studies examining the effect of professional development. The applied definition of PD is quite broad and covers all studies investigating a PD intervention with the aim of improving children's outcomes (broadly defined). In principle, we would also include studies investigating PD interventions even if they did not have as a specific target to improve children's outcomes (but e.g. aimed to improve educators' skills or working conditions), but such studies will only be included if they also measure and examine child outcomes (cf. the next inclusion criterion). Originally, we also intended to include studies investigating the effect of pre-service training of educators on child outcomes, using e.g. pre-school educators' years of formal training as a measure of structural quality (cf. Bauchmüller et al., 2014, and studies of EPPE in the UK). However, this turned out to be impossible, since we did not find any studies with interventions on pre-service training and that would leave us with only observational studies from which we could not extract effect sizes that would fit our methodological approach.

***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). Schachter (2015) in her recent review of the literature on professional development in early childhood education notes that less than half (49 %) of her included studies contain measures on children's outcome. She also reports that 40 % of the studies use measures of children's learning while only 11 % use measures of children's behavior. Since we focus on the effect on children's outcomes, we exclude all studies that only measure educators' outcomes.

***Studies focusing on regular care*.** In the included studies, participants are pre-school educators providing care to 0-6 year-old children in pre-schools, kindergartens, or ECEC centers. PD of nannies or childminders is not considered as the focus is on child care in formal child care centers.

We want to identify the pure effect of PD in *pre*-schooling and thus do not consider children already enrolled in school. This implies that studies evaluating PD in both pre-school and primary school, and measuring *the combined effect* of these, are excluded. If an effect of only PD in pre-school can be clearly identified, the study will be included.

***Studies not focusing on specific risk groups of children*.** The children included in the current study are not targeted as special-needs children or disadvantaged children, e.g. children with ADHD, children of immigrant parents, children with physical disabilities etc. However, many studies are targeted at very specific groups, and thus the results from these studies are not necessarily generalizable to a broader population of children. We cannot include such studies in the meta-analysis if we want to be sure that the results of the meta-analysis are generally applicable.[4]

***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 by the authors from the included information in the study, we have to disregard the study. Having such a clear measure of child outcomes turns out to be a strict requirement.

[4] Furthermore, many of these studies on specific groups of disadvantaged children are carried out in the US.

Language, time period, and peer-review. Studies included in this review do not have to be published in a scientific, peer-reviewed journal, but in practice most of them are. Both cross-sectional and longitudinal studies from 1990 or later are considered for the analysis as long as the studies focus at PD of pre-school educators in a European setting. Moreover, we restrict the literature search to studies published in English. This latter requirement is included to ensure that our study can be replicated and to ensure that we are not asymmetrically including or excluding studies based on our own language skills. The main concern with this criterion could be whether we are ignoring reports e.g. from ministries which are often only published in the country's own language.

Ideally, we would have liked to state precisely how many studies were excluded due to each of the inclusion criteria, but this is unfortunately not possible. One obvious reason is that some of the initially collected studies from the database search fail several of the inclusion criteria. Another reason is the hands-on procedure used for screening the studies. All studies were manually screened, and in some cases studies could be excluded just by looking at the title. In other cases, studies were excluded based on the abstract of the study (if an abstract existed). Finally, many studies could not be excluded until the full text (i.e. book, thesis, report, journal article, newspaper article, etc.) was retrieved and investigated closer. Therefore, it was not possible to program a screening device for sorting and counting studies of each type failing the inclusion criteria described above.

To sum up, the major restriction from the inclusion/exclusion criteria turns out to be the focus on European studies. Furthermore, within the studies of PD in Europe we want to clearly identify the effect of PD in pre-schooling only – not in combination with PD in schools – and it is required that the effect of PD is measured on the children (not the pre-school educator). This implies that there are several of the initially collected studies that cannot be used in the current meta-analysis.

STUDY CODING

After the final selection of studies to be included in the review, each study has been coded by the authors for a number of variables including main outcomes and covariates/moderators. We concentrate on effects on child outcomes, as described in the Introduction. Several studies report educator outcomes, but we do not include these in the analysis. The study coding results in a set of descriptive tables that will be reported in the following section.

Also note that (as discussed above) several studies are poorly described, making it difficult as well as time-consuming to code their characteristics. We have had to exercise a fair amount of judgement to be able to code all the necessary covariates for the included studies.

In order to code a study for the meta-analysis it is required that an effect size is directly specified in the study in question, or, alternatively, that the necessary elements needed to calculate the effect size are available from the study. This latter issue is described in further details below.

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 will use the standardized mean difference as an effect size that is comparable across studies. The standardized mean difference is in principle

estimated using the mean scores, standard deviations, and number of participants in the treatment and control groups (and is referred to as Cohen's d – see Borenstein et al., 2009).

Many of the included studies use experimental designs that cause them to vary in the number of treatment groups being tested and the number of different outcomes being measured. Such variation in experimental design will (at least, in some cases) result in statistical dependence between effect sizes, and this problem needs to be dealt with in the meta-analysis. In the main meta-analysis we 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 we expect the included studies to be quite diverse and we pool all the studies in one meta-analysis, it would be 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).

EXTRACTING AND CALCULATING EFFECT SIZES

Extracting data from the studies (in particular, the effect sizes to be used in the meta-analysis) was a difficult and challenging task.

Very few of the studies report an effect size that directly can be included in the meta-analysis. In the remaining cases, we had to extract the necessary information from the studies and calculate the effect size ourselves. Even this was not always possible (and as mentioned below, we had to exclude one study because it turned out to be impossible to calculate an effect size). 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). Unfortunately, this introduces an additional source of imprecision in the calculated effect sizes as values of the t-tests and F-tests are often not given very precisely.

Several studies measured and reported more than one outcome. In general, this happens as a result of two different situations, either because it was possible to measure more than one outcome (e.g. in two different domains) or because both a total score and several subscores were reported. In the former case, we combined the different outcomes into a single one by averaging the outcomes of the relevant study, whereas, in the latter case, we decided to use the total outcome measure. We perform sensitivity analyses where we investigate various other choices. 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 in a meaningful way.

In some cases we found more than one study reporting results from the same project or experiment. In those cases, we selected one of these studies, typically the one with the most complete set of results (as e.g. in the case of Cviko et al. (2015) who reported a subset of the results in Cviko et al. (2014)) or the final publication from the project (as e.g. in the case of Gasteiger (2012) which was an earlier version of Gasteiger (2014)).

For those outcomes that were 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.

One study (Ahsam et al., 2006) had to be excluded because it did not give enough information to enable us to calculate an effect size, even though the study satisfied all other inclusion criteria. This study reported a significant effect of a training program on children's interaction and communication skills. It was, however, a very small-scale study only including 9 children, and furthermore, it only used a pre-test post-test design (i.e. with no control group). Hence, we do not consider the exclusion of this study to be an important omission.

R E S U L T S

DESCRIPTION OF STUDIES

After the selection process described in the previous section, we ended up with only nine studies included for the meta-analysis. These nine studies are first described briefly below by the country of the PD intervention, and thereafter their various characteristics are described more systematically in Tables 1-4.

BRIEF OVERVIEW OF INCLUDED STUDIES

The nine included studies are briefly described here by the country of the PD intervention. The studies are concentrated on few countries, with Denmark, Germany and the Netherlands each contributing more than one study and France and Wales each contributing one study. There is an apparent lack of studies from Southern Europe and Eastern Europe, probably reflecting that research on ECEC is less common in these regions. Furthermore, no studies on the UK are included (except the one study on Wales). We noticed during the search and selection process that many studies on the UK had to be excluded due to the setting being formal schooling rather than pre-school or ECEC, and in other cases the studies do not rely on interventions with pre- and post-intervention tests as is e.g. the case for the EPPE project study.

Two studies from Denmark are included in the analysis, and they are quite similar in their type of PD intervention and general implementation using a RCT. The PD 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. Thus, the aim is to improve the reflective skills of the pre-school staff and these new insights should lead to changes in daily activities which positively affect SDQ-scores of the included children. The first study, Jensen et al. (2013), evaluates a RCT implemented in 59 pre-schools affecting a total of 524 children in the experimental group and 521 children in the control group. The children are in the age range of 3 to 6 years old and the study finds a positive effect on several aspects of children's behavior. The second Danish study, Jensen et al. (2015), follows the outline of the Jensen et al. (2013) study, but implements some changes to the pre-school staff training program (with more emphasis on evidence-based knowledge) and furthermore follows the children for a somewhat longer period of time and in other municipalities in Denmark. Estimation methods are also different across the two studies. In the second study, the sample size used in the final estimations is 396 in the experimental group and 290 in the control group. The results of the second study also show positive effects on child behavior due to professional development of the pre-school teachers, but do no reveal stronger effects for disadvantaged children than for other children.

Only one study from France fulfilled the inclusion criteria (Ecalle et al., 2015). This study evaluates the effect of evidence-based literacy practices in a RCT and finds global effects on different components of literacy skills in the experimental group of children. Contrary to the Danish studies, the study on French kindergartens shows stronger impact on children in the lower part of the distribution. The study is quite large, having 1932 children in the experimental group and 1357 in the control group, but the training session for teachers only lasts one day, i.e. it is a low-intensive PD in that sense. During this day the teachers are trained in the use of new pedagogical practices inspired and recommended by evidence-based research. These new practices are then implemented such that each child in the experimental group received approximately 9 hours of oral comprehension training along with 18 hours of alphabetic code training and 18 hours of phonological training.

Germany is the country supplying the most studies to our analysis (Gasteiger, 2014; Perels et al., 2009; and Rönnau-Böse et al., 2009). The studies are concerned with different outcome domains ranging from mathematics to self-regulation and resilience, and the included children are 3 to 6 years old. The studies by Gasteiger (2014) and Perels et al. (2009) both include less than 50 children in each of the experimental and control groups, whereas the study by Rönnau-Böse et al. (2009) is somewhat larger with around 100 children in each of the groups. The PD evaluated in Gasteiger 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. This intervention did not have a significant general effect on the children's mathematical competencies, but there were significant improvements for experimental group children in one of the domains. The author interprets this as a positive and promising result as this type of PD typically takes a lot of time for the teachers to be familiar with and to implement in their daily activities. The study by Perels et al., on the other hand, 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 of children is mainly middle-class children. In this PD the teachers get theoretical knowledge and combine it with practical exercises both at the training sessions and when they are working in their kindergarten group. The result is that the children in the experimental group become better at understanding and describing different phases of activities, i.e. they improve their self-regulation. The final included study from Germany, Rönnau-Böse et al. (2009), also investigates a RCT of PD. This PD intervention lasted for two years and showed a positive effect on the 3-6 year old children's self-esteem, behavioral stability, and cognitive development. Thus, this PD intervention accomplished its purpose of systematically involving pre-schools and its professionals in a general preventive program aimed at promoting the development of children and the promotion of resilience.

The Netherlands contribute with two studies (Cviko et al., 2014 and Henrichs and Leseman, 2014). Both studies are concerned with implementing more technology/science-rich learning activities for improving early language and literacy skills, and they both contain RCTs. The Henrichs and Leseman study evaluates a very low-intensive PD, which during one training session makes teachers focus on their (and others) academic language use. Despite being a low intensive PD, 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. Cviko et al. (2014) studies a PD implemented over 8 weeks, i.e. a more intensive PD than that evaluated in Henrichs and Leseman. This PD centers on teacher roles. When engaging in technology-rich activities teachers can either be 'executor-only', 're-designer', or 'co-designer' of the activities. 196 (162) children participate in the experimental (control) group of this study, and, compared to results of the control group, the results show general positive effects on children's literacy of each of the teacher roles. Thus, the PD intervention has proven successful in all three sub-groups of the interventions.

The UK contributes with a study from Wales (Hutchings et al., 2013). In this study a PD intervention is implemented over 5 months in order to improve, among other things, child behavior. 3-7 year old children are used in the study with 53 allocated to the experimental group and 54 to the control group.[5] Some of the children have behavioral problems, and these children are also allocated to experimental and control groups to investigate if they react similarly to the teachers' PD as children who do not exhibit behavioral problems. It turns out that the PD has a larger effect on children with behavioral problems and thus reduces their negative behavior.

SYSTEMATIC OVERVIEW OF INCLUDED STUDIES

In the systematic overview of the nine included studies, the characteristics of the studies are described in Tables 1-4. These tables are organized according to the type of characteristics, distinguishing between basic, educational, outcome and methodological characteristics.

Table 1. Included studies, basic characteristics

Study	Country	Domain	Age of children
Cviko et al (2014)	The Netherlands	Literacy	4-6 years
Ecalle et al (2015)	France	Literacy/reading	5-6 years
Gasteiger (2014)	Germany	Mathematics	3-6 years
Henrichs and Leseman (2014)	The Netherlands	Language/science	5 years
Hutchings et al (2013)	Wales	Behavior	3-7 years
Jensen et al (2013)	Denmark	Behavior	3-6 years
Jensen et al (2015)	Denmark	Behavior	3-6 years
Perels et al (2009)	Germany	Self-regulation	5-6 years
Rönnau-Böse et al (2009)	Germany	Resilience	3-6 years

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

Table 2. Included studies, educational characteristics

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

Table 3. Included studies, outcome characteristics

Study	Outcome measure	Instrument	Sample size, Exp./con. group
Cviko et al (2014)	Early literacy	Own instrument	196/162
Ecalle et al (2015)	Literacy skills	Own instrument	1932/1357
Gasteiger (2014)	Mathematics competence	Own instrument	19/19
Henrichs and Leseman (2014)	Language development: lexical diversity	Own instrument, group level	29/26
Hutchings et al (2013)	Behavior	TPOT	53/54
Jensen et al (2013)	Behavior	SDQ (educator rated)	524/521
Jensen et al (2015)	Behavior	SDQ (educator rated)	396/290
Perels et al (2009)	Self-regulation	Own instrument	48/49
Rönnau-Böse et al (2009)	Self-esteem a.o.	Standardized tests (SKF)	99/102

Table 4. Included studies, methodological characteristics

Study	Design	Statistical method	Effect sizes
Cviko et al (2014)	Quasi-experimental	Analysis of covariance	Calculated from F-test, average calculated (from 4 groups)
Ecalle et al (2015)	Experimental	Analysis of variance, matching	Average calculated (from 6 measures)
Gasteiger (2014)	Quasi-experimental	Analysis of variance	Calculated from basic information
Henrichs and Leseman (2014)	Experimental	Analysis of variance	Calculated from F-test, average calculated (from 3 measures)
Hutchings et al (2013)	Experimental	HLM analysis	Average calculated (from 6 measures)
Jensen et al (2013)	Experimental	Value-added analysis	Average calculated (from 4 measures)
Jensen et al (2015)	Experimental	Value-added analysis	Extracted from paper
Perels et al (2009)	Experimental	Analysis of covariance	Calculated from F-test
Rönnau-Böse et al (2009)	Experimental	Multivariate variance analysis	Calculated from F-test

It is seen that all included studies are very recent studies, which have been published within the last 6 years and with the majority being published during the period 2013-2015. This clearly shows that this research area is in its infancy in Europe, but that research activities have started to being initiated. Possibly, as a result of this infancy, the studies are characterized by a lack of consistency in reporting their results (despite the fact that most of the European research is heavily inspired by the US research). Also, as mentioned earlier, only five countries contribute with studies fulfilling the inclusion criteria.

In terms of focus or targeted domain, the studies show a lot of diversity: about half are related to the cognitive domain (with three studies focusing on language and literacy and one study focusing on mathematics[6]) and the other half are related to the socio-emotional domain focusing on various aspects of the children's behavior.

All included studies evaluate PD interventions in settings for children above 3 years of age. In terms of this aspect the studies do not seem to differ very much. However, this clearly illustrates that very little research exist on the youngest children below 3 years of age. Naturally, this is related to the problems of measuring child outcomes for the very young children, but it may also be problematic since some of the new insights into human capital investments in young children indicate that "earlier is better" (Cunha et al., 2006). However, the review by Duncan and Magnuson (2013) did not find any variation in effect sizes according to starting age.

Looking at the educational characteristics reported in Table 2, it is seen that the studies have examined very different PD interventions. These 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., 2015). The majority of the PD interventions involve several training sessions (typically between 4 and 6) organized over a longer period of time. It is remarkable that the French study (Ecalle et al., 2015) entailing a large-scale intervention including more than 3,000 children only supplied one day of training to the educators (see also Table 3).

Although not reported in the tables, it should also be noted that several of the studies also measure and analyze teacher outcomes (e.g. Cviko et al., 2014; Gasteiger, 2014; Perels et al., 2009; Hutchings et al, 2013; Henrichs and Leseman, 2014). In this review and meta-analysis, we do not consider teacher outcomes, since our main focus is on child outcomes, which are the ultimate goal of any ECEC intervention.

Looking at the methodological characteristics reported in Table 4, it is seen that all the studies rely on an experimental or quasi-experimental design, where children (and educators) are allocated to a control group and an experimental group, respectively. However, the sample sizes (see Table 3) are very different, as the studies range from using a very small sample of only 38 children (Gasteiger, 2014) to using more than 3,000 children (Ecalle et al., 2015).

[6] The study by Henrichs and Leseman (2014) has a combined focus, since it investigates whether science and language learning can be integrated.

RESULTS OF THE META-ANALYSIS

As a first analysis, we have included all 9 studies in a single meta-analysis, using a random-effects model. The results from this meta-analysis are shown in Table 5 and the corresponding forest plot is shown in Figure 1. From this forest plot and the heterogeneity statistics, 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.

Looking more closely at this study, it is apparent that the very large effect size may be a result of the way the self-regulatory skills of the children are measured. The self-regulatory behavior for each child was coded as 1, 2 or 3, based on a videotaped structured interview with the child. The study itself mentions that this measurement procedure has some flaws, although it is validated using another instrument to assess self-regulation.

As a next step, we therefore excluded the study from the meta-analysis and reran it with the remaining 8 studies. Again, we see that there is a considerable amount of heterogeneity in the effect sizes of the studies (forest plot shown in Figure 2 and results in Table 6), but that is no surprise since the studies are quite different and we have pooled all effect sizes together regardless of the domains that have been measured. We decided to keep the remaining 8 studies as our final sample, but we will perform a sensitivity analysis of this decision in the next section. Hence, the final result of the meta-analysis 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).

We are severely restricted in the possibilities to conduct meta-regression analysis or subgroup analysis of moderator effects due to the few included studies. For example, our initial plan was including measures of PD intensity, type of PD, and the outcome domain in the meta-analysis. These would all be interesting aspects to investigate further. However, due to the few included studies, we instead have to limit the discussion to comments on various aspects of the studies.

When the analysis of effect sizes is compared to the characteristics of the studies reported in Table 2 it is seen 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 most obvious example of the latter is the very low-intensity intervention investigated by Henrichs and Leseman (2014), where only three hours of group training of the educators result in a sizeable effect size of 0.64 on the children' academic language development. Zaslow et al. (2010) also noted in their review of PD interventions that even small dosages of PD have been associated with positive child outcomes (at the same time as much more comprehensive programs have the same type of effects) and they suggest that this ambiguous result on the appropriate dosage of PD is related to the goals or focus of the PD.

It is also interesting to note, that we observe some differences across sample size. Some of the studies are performed on small samples, while three studies are performed on relatively large samples (Ecalle et al. (2015), Jensen et al. (2013), Jensen et al. (2015)). It is obvious from the forest plot that these three studies have smaller effect sizes than the other studies. This is a feature also often seen in other meta-analyses (and discussed in detail by Slavin and Smith, 2009). Fukkink and Lont (2007) observed that large-scale training programs appeared less effective. Our study is thus in line with this previous research.

Table 5. Results from initial meta-analysis

Study	Effect	[95% Conf.	Interval]	% Weight
Cviko et al (2014)	0.787	0.571	1.003	12.89
Ecalle et al (2015)	0.165	0.095	0.235	14.53
Gasteiger (2014)	0.652	-0.001	1.305	6.37
Henrichs and Leseman (2014)	0.640	0.097	1.183	7.73
Hutchings et al (2013)	0.373	-0.009	0.755	10.16
Jensen et al (2013)	0.170	0.049	0.291	14.10
Jensen et al (2015)	0.216	0.064	0.368	13.77
Perels et al (2009)	1.970	1.485	2.455	8.54
Rönnau-Böse et al (2009)	0.320	0.042	0.598	11.90
Overall effect (dl)	0.514	0.296	0.733	100.00

Heterogeneity Measures

Cochrane Q = 83.68 df=8 p-value=0.000

	[95% Conf.	Interval]
I^2 (%) = 90.44	84.12	94.25
H^2=10.46	6.30	17.38
tau^2 est(dl) = 0.084		

Figure 1. Forest plot for initial meta-analysis

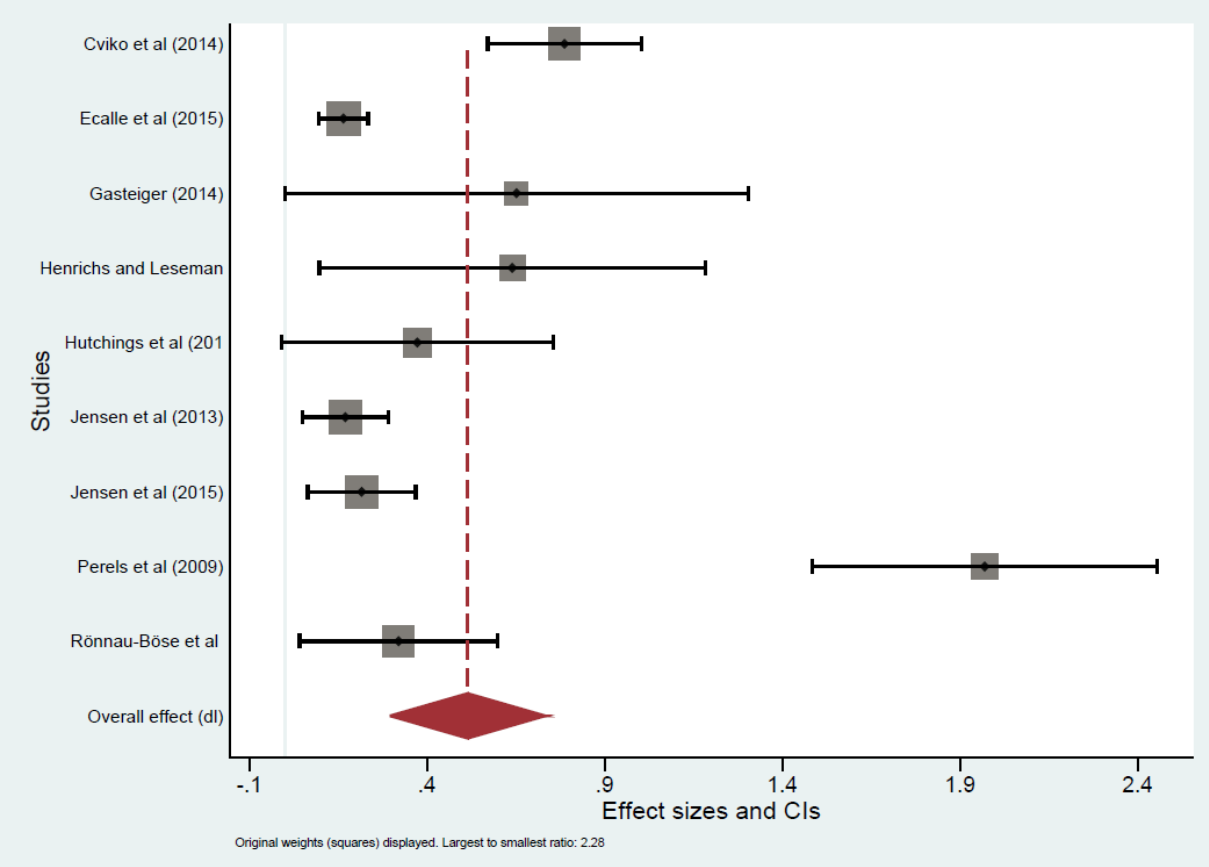


Table 6. Results from final meta-analysis

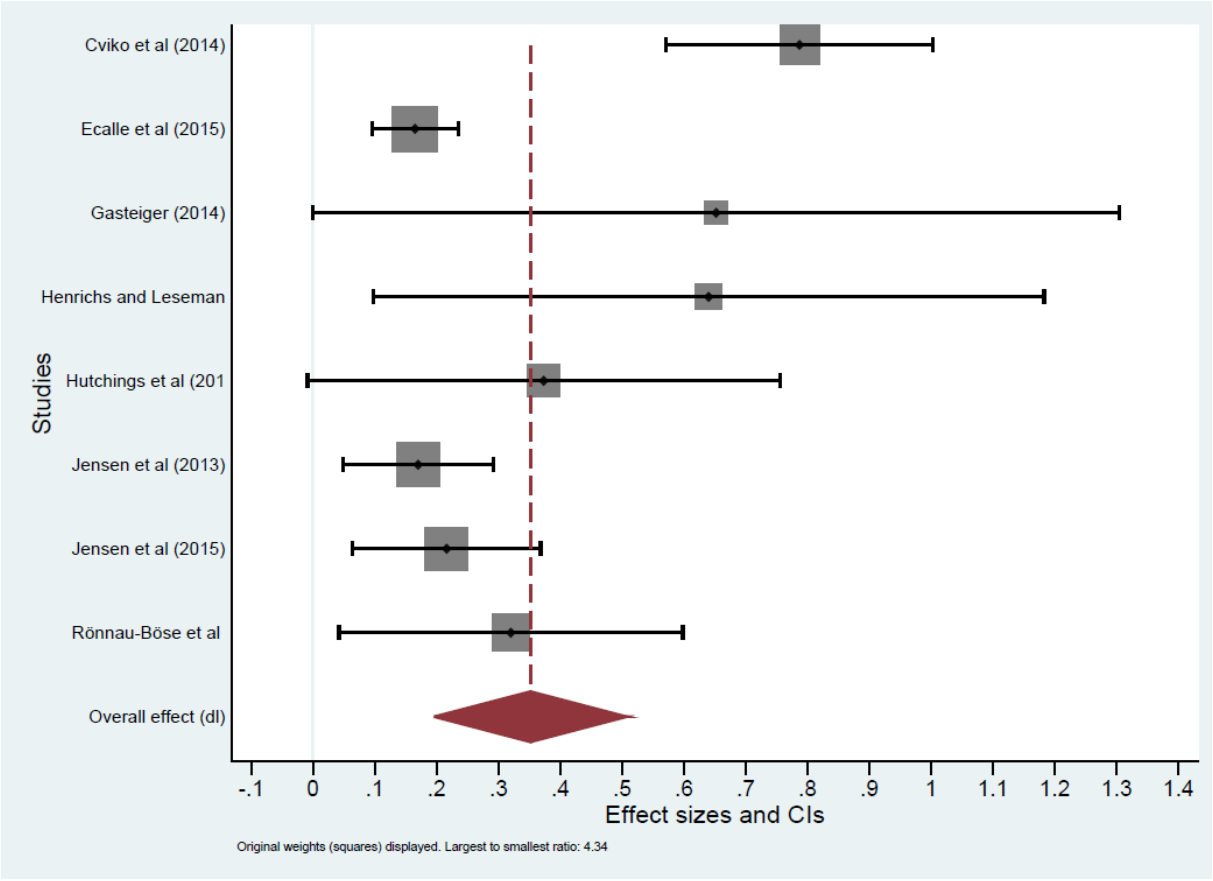
Study	Effect	[95% Conf.	Interval]	% Weight
Cviko et al (2014)	0.785	0.571	1.003	14.47
Ecalle et al (2015)	0.165	0.095	0.235	19.26
Gasteiger (2014)	0.652	-0.001	1.305	4.44
Henrichs and Leseman (2014)	0.640	0.097	1.183	5.84
Hutchings et al (2013)	0.373	-0.009	0.755	9.08
Jensen et al (2013)	0.170	0.049	0.291	17.86
Jensen et al (2015)	0.216	0.064	0.368	16.83
Rönnau-Böse et al (2009)	0.320	0.042	0.598	12.22
Overall effect (dl)	0.352	0.196	0.507	100.00

Heterogeneity Measures

Cochrane Q = 34.58 df=7 p-value=0.000

	[95% Conf.	Interval]
I^2 (%) = 79.76	60.65	89.58
H^2=4.94	2.54	9.60
tau^2 est(dl) = 0.032		

Figure 2. Forest plot for final meta-analysis



As mentioned in the method section, we have extracted effect sizes from the studies by using the average effect size for each study if more than one outcome was measured. Clearly, we could alternatively have selected the most important effect size from each study, but it might be rather arbitrary which effect size should be considered the most important (apparently, most authors of the studies seem to believe that the most important effect size coincides with the most significant effect size). Unless this would lead us to select larger effect sizes, in which case the overall effect size would be larger, it is *a priori* unclear if and how this would bias the overall effect size. In the following section, we perform a number of sensitivity analyses, including a 'cherry picking' version.

It should also be noted that we have been very conservative in the way we have treated heterogeneity in the meta-analysis. We have actually removed the study with the highest estimated effect size from the sample. This of course means that the overall effect size will be smaller and it could therefore in some sense be viewed as a conservative estimate of the possible effect.

SENSITIVITY ANALYSES

In the previous section, we ended up with a final sample consisting of 8 studies. However, we were still left with a considerable amount of heterogeneity in the effect sizes of the included studies. From the forest plot (shown in Figure 2), it is obvious that the study by Cviko et al. (2014) may be a possible outlier.

In the current section, we therefore perform a sensitivity analysis of excluding this study from the meta-analysis. Both the forest plot and the heterogeneity statistics now reveal that these 7 studies can be described by the same model (see Table 7 and Figure 3). In this specification, Cochran's Q is 7.00 with a p-value of 0.321 and thus not indicating heterogeneity.

The result of the meta-analysis is that the 7 studies contribute to an overall effect size of 0.20 (with a 95 % confidence interval from 0.13 to 0.27). This is lower than in the final meta-analysis of the previous section, but that simply reflects that the study by Cviko et al. (2014) has a much higher effect size than the remaining 7 studies.

As an additional argument for excluding the study by Cviko et al. (2014), it should be noted that the type of PD in this study might not be considered a real PD intervention, but rather an investigation into different educator roles in implementing new technology in early childhood education. Hence, it could be argued that this study is on the borderline of inclusion in the analysis, but we think that it qualifies as a modern type of PD where the educators play an active role in designing and implementing new technology in ECEC.

In fact, the study by Cviko et al. (2014) can be considered as reporting on 4 separate experiments based on 4 different samples. In the final meta-analysis reported in the previous section, we have pooled these 4 separate experiments into one effect size (by averaging the effect sizes over the 4 different samples). However, it is instead possible to treat them as 4 separate studies in the meta-analysis. Each of these will then obviously have an effect size with a higher variance due to the smaller number of observations in each of the 4 separate samples, but the random-effects method assigns more weight in total to the 4 separate effect sizes than to the common pooled effect size. Hence, the result from a meta-analysis where the study by Cviko et al. (2014) supplies 4 effect sizes is that the overall effect size increases from 0.35 to 0.40 (with a 95 % confidence interval from 0.26 to 0.54).

Table 7. Results from meta-analysis with only 7 studies

Study	Effect	[95% Conf.	Interval]	% 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 (2015)	0.216	0.064	0.368	16.80
Rönnau-Böse et al (2009)	0.320	0.042	0.598	5.72
Overall effect (dl)	0.203	0.134	0.272	100.00

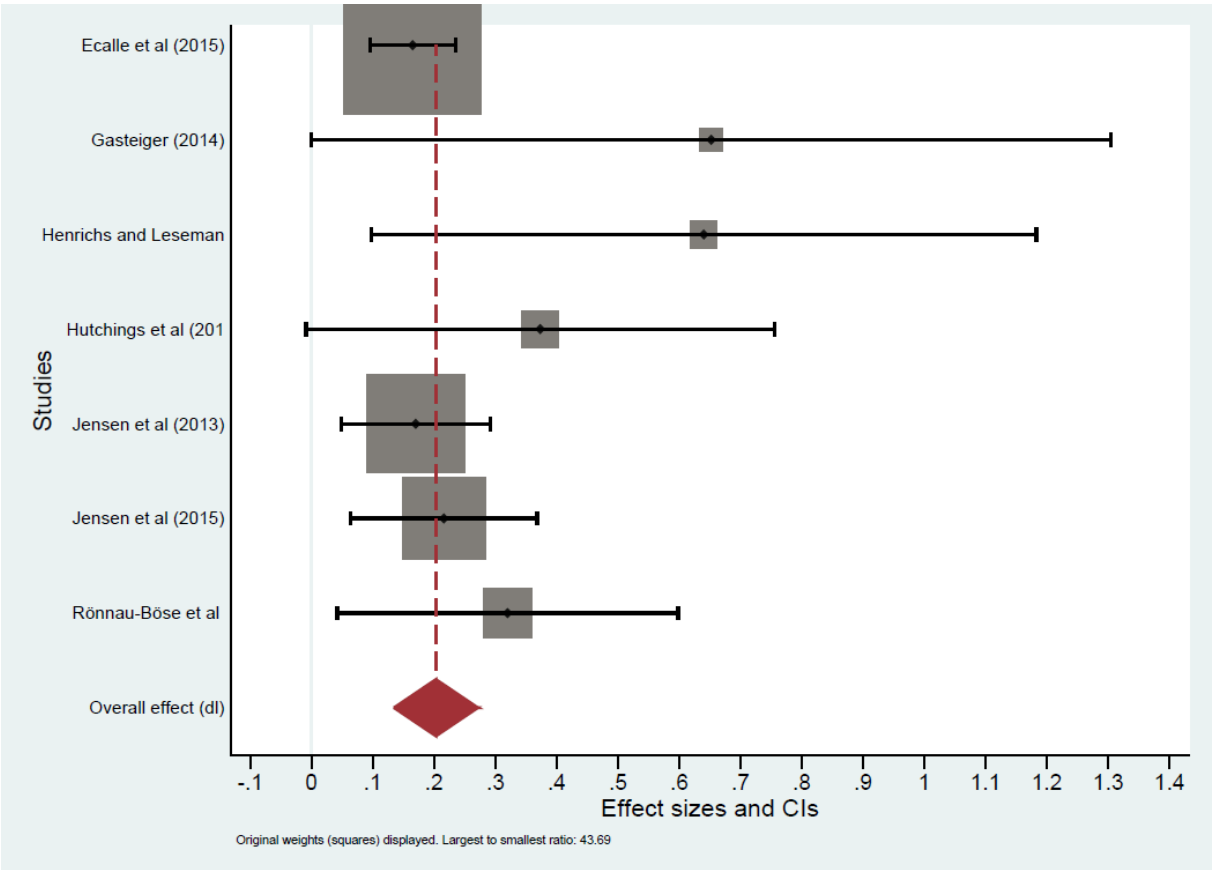
Heterogeneity Measures

Cochrane Q = 7.00 df=6 p-value=0.321

	[95% Conf.	Interval]
I^2 (%) = 14.23	0.00	74.96
H^2=1.17	0.34	3.99

tau^2 est(dl) = 0.001

Figure 3. Forest plot for meta-analysis with only 7 studies



Another study where there is some ambiguity in how to select the main effect size for the meta-analysis is the study by Henrichs and Leseman (2014). In this study, two different tasks were used as basis for the experiment. Both tasks were performed by the same teachers and children, but in the final meta-analysis reported in the previous section, we have only used the effect size from the first task (which is the one with the most successful results). We might instead replace this effect size with the average of the effect sizes from both tasks to more fully reflect the outcome of the intervention. In that case, the overall effect size from the meta-analysis decreases slightly from 0.35 to 0.34 (with a 95 % confidence interval from 0.19 to 0.50). Hence, this change has almost no influence on the overall effect size, reflecting that the results are not very sensitive to the input from one very small study.

As mentioned in the previous section, an alternative to using the average effect size from studies where more than one outcome was measured is to use one specific effect size from each study. One possible way to do this is by 'cherry picking', i.e. to select the effect sizes that are in some sense most favorable. One typical way to do this is by selecting the largest effect size from each study. To assess the sensitivity of our results to such a strategy, we have rerun the final meta-analysis of the previous section with the cherry-picked effect sizes from each study. The full results and the forest plot for this exercise is shown in Appendix B. The overall effect size increases to 0.39 (with a 95 % confidence interval from 0.26 to 0.53).

Even though such a cherry-picking strategy is not in general an advisable research strategy, it may have some merit in the sense that it can give an idea about the maximum achievable effect based on the studies at hand. If we imagine that the largest effect size could be realized for each outcome in each study, then the overall effect size from this 'cherry picking' exercise would reflect the potential impact of PD on the child outcomes.

Finally, we perform a sensitivity analysis where we use the naïve method of including all possible effect sizes (from all outcomes) in the meta-analysis. Clearly, the effect sizes are dependent within studies, since the effect of the intervention is evaluated on multiple outcome measures for the same sample of children. In the final meta-analysis of the previous section, we dealt with this dependence by aggregating the effect sizes within each study, typically by taking averages. Using the naïve method, we ignore this dependence, which may result in biased statistical inferences. By treating the effect sizes as independent, standard errors are underestimated and the resulting overall effect size might appear more precisely estimated than it really is (see Van den Noortgate et al., 2015). A more appropriate way to perform the analysis with all possible effect sizes would be by multivariate or multilevel meta-analysis, but standard software in many cases requires that the correlation between outcome measures are known (or that the same number of outcome measures are observed for each study). Information on correlations between outcome measures is not known in the included studies and the number of outcome measures is not the same across studies. Hence, we have not been able to perform this kind of analysis using the available software (such as the add-on command mvmeta in STATA), and this will therefore be left for future research.

In the naïve approach we include a total of 36 effect sizes from the 8 included studies. The result from the random-effects model in this case is that the overall effect size is estimated to 0.25 (with a 95 % confidence interval from 0.19 to 0.31). We observe two different changes here compared to the results from the final specification of the previous section. First, the overall effect size decreases compared to the final meta-analysis of the previous section, which happens because the studies with many outcome measures are those studies with low effect sizes and large sample sizes. Second, as mentioned above, the standard error of the overall effect size is underestimated, resulting in a narrow confidence interval. However, it should be pointed out that both of these changes are spurious and are consequences of ignoring the dependence between effect sizes from the same study.

Both the ‘cherry picking’ and the naïve method sensitivity analysis have also been run excluding the results from the study by Cviko et al. (2014). In both cases, the overall effect size drops as we saw above, again simply reflecting that the study by Cviko et al. (2014) finds a higher effect size than the other studies. The drop is, however, smaller than in the first sensitivity analysis, since the influence of this single study is smaller when either ´cherry picking’ or the naïve method is applied.

To summarize, the sensitivity analyses show that the results from the meta-analysis are somewhat sensitive to the various approaches and specifications, but in most cases the differences in overall effect sizes are very small and not of economic significance or the changes can be dismissed based on methodological considerations for the cases of the ‘cherry picking’ and the naïve method. All in all, the final meta-analysis of the previous section remains our preferred specification.

DISCUSSION/CONCLUSION

This systematic review of the impact of PD in ECEC in Europe has showed that there is very little research on the topic. The search strategy employed in this review is quite broad and covered more than 8,500 studies. However, after focusing only on studies fulfilling the inclusion criteria and having a clear PD intervention or training involved, we have less than 10 studies at hand. Thus, the planned moderator analysis has to be restricted to a discussion of the various aspects of each included study. Even a moderator analysis with quite few moderators is impossible due to a lack of degrees of freedom.

In the US in particular, many studies focusing on the effect of professional development have been carried out. Many of these studies are based on randomized controlled trials in relation to Head Start (e.g. the coaching intervention described in Wasik and Hindman, 2010). These studies offer great inspiration to potentially interesting initiatives, which would also be promising in a European formal pre-school or ECEC setting. More specifically, the coaching interventions studied in detail in Powell and Diamond (2010) are worth further investigation.

The observation that the vast majority of the research consists of studies from North America (USA and Canada) is not new. A number of existing reviews make the same observation, e.g. Blok et al. (2005) review the effectiveness of early intervention programs and out of 34 experimental comparisons, only 3 are for Europe (and all 3 are for the Netherlands).

More recently, Schachter (2015) has reviewed the literature on professional development in early childhood education. A quick look through the references for the included 73 studies reveals that an overwhelming majority of these studies are from the US (and Canada) – in fact, we did not succeed in finding any study from Europe. Similarly, in a recent meta-analysis of the effects of language and emergent literacy PD on educator's and children's outcomes (Markussen-Brown et al., 2015) all but one of the included studies are conducted in either Canada or the US.

Previous reviews (e.g. Fukkink and Lont, 2007; Zaslow et al., 2010) have concluded that PD has a positive impact on child outcomes, but their evidence is primarily based on US studies. Fukkink and Lont (2007) performed a meta-analysis on 4 studies reporting effects at the child level (with a total of 15 effect sizes derived from these studies) and they report that the overall effect size is 0.55, although this is not statistically significant (at 5% level). This is slightly higher than our estimated overall effect size of 0.35 from European studies, but illustrates that the European studies do not necessarily deviate much from the US studies. Zaslow et al. (2010) survey a large number of studies in great details, but do not provide effect sizes or a meta-analysis. Nonetheless, their conclusions also point in the direction that PD interventions have positive impacts on child outcomes for three different sets of outcomes: language and literacy, mathematics, and social behavior. A recent meta-analysis of the effects of PD on child outcomes in language and emergent literacy (Markussen-Brown et al., 2015, also relying on studies from North America) finds statistically positive effects for different outcomes, with overall effect sizes in the range of 0.18-0.46.

More research in a European context is clearly needed in order to shed more light on the question of the effect of PD on child outcomes. The institutional settings in relation to European pre-schools are sometimes very different from the institutional settings in the US and therefore it is very important that we also have research-based evidence from Europe. Having this evidence, we can compare to the lessons learned from the US context in more detail. Our systematic review revealed that the available research on European experiences is quite recent. Thus, it is likely that the use of RCTs in research on PD is on the rise in Europe and that such comparisons will be possible in the near future.

Despite the few studies based on European data on PD, the results from the current meta-analysis point in the direction that professional development has a clear positive impact on child outcomes. The size of the effect is somewhat sensitive to the method applied for extracting effect sizes from each of the studies. Furthermore, the exact mechanism(s) need to be further explored, but it can definitely be concluded that a greater focus on PD will be beneficial for children across Europe.

A C K N O W L E D G E M E N T S

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. The usual disclaimer applies.

R E F E R E N C E S

Abreu-Lima, I., Leal, T., Cadima, J., & Gamelas, A. (2013). Predicting child outcomes from preschool quality in Portugal. *European Journal of Psychology of Education - EJPE (Springer Science & Business Media B.V.), 28*(2), 399-420.

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.

Anders, Y., & Ulferts, H. (2015). *Meta-analysis of ECEC effects* (preliminary title). CARE project; Curriculum Quality Analysis and Impact Review of European Early Childhood Education and Care (ECEC). Forthcoming.

Bauchmüller, R., Gørtz, M., & Rasmussen, A. W. (2014). Long-run benefits from universal high-quality preschooling. *Early Childhood Research Quarterly, 29*(4), 457-470.

Blok, H., Fukkink, R., Gebhardt, E., & Leseman, P. (2005). The relevance of delivery mode and other programme characteristics for the effectiveness of early childhood intervention. *International Journal of Behavioral Development, 29*(1), 35-47.

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). *Intepreting the Evidence on Life Cycle Skill Formation*, in Hanushek and 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.

Cviko, A., McKenney, S., & Voogt, J. (2015). Teachers as co-designers of technology-rich learning activities for early literacy. *Technology, Pedagogy and Education, 24*(4), 443-459.

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.

Elango, S., Garcia, J. L., Heckman, J. J., & Hojman, A. (2015). Early Childhood Education. Manuscript prepared for *Means-Tested Transfer Programs in the United States*, Volume II, edited by Robert A. Moffitt. Available at **https://cehd.uchicago.edu/ECE-US**

Farran, D.C. and 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 Edition: Routledge.

Fröhlich-Gildhoff, K., Rönnau, M., Dörner, T., Kraus-Gruner, G., & Engel, E. M. (2008). Promotion of resilience in institutions of early childhood education (kindergarten) -first results of a prevention project. *Praxis der Kinderpsychologie und Kinderpsychiatrie, 57*(2), 98-116.

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. (2012). Mathematics education in natural learning situations: evaluation of a professional development program for early childhood educators. *Proceedings of the 36th Conference of the International Group for Psychology of Mathematics Education, Vol. 2: Opportunities to Learn in Mathematics Education*.

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 Kortenkamp, U., Brandt, B., Benz, Ch., Krummheuer, G., Ladel, S. & Vogel, R. (Eds.). Early Mathematics Learning. Selected Papers of the POEM 2012 Conference (pp.275-290). New York: Springer.

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

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 Publications.

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., Jensen, P., & Rasmussen, A. W. (2015). Does Professional Development of Preschool Teachers Improve Child Socio-Emotional Outcomes? IZA Discussion Paper No. 8957.

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

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

Lerkkanen, M.-K., Kiuru, N., Pakarinen, E., Viljaranta, J., Poikkeus, A.-M., Rasku-Puttonen, H., . . . Nurmi, J.-E. (2012). The role of teaching practices in the development of children's interest in reading and mathematics in kindergarten. *Contemporary Educational Psychology, 37*(4), 266-279.

Markussen-Brown, J., Juhl, C. B., Piasta, S. B., Bleses, D., Højen, A., & Justice, L. M. (2015). The effects of language and emergent literacy professional development for teachers and children: A meta-analysis. Manuscript, University of Southern Denmark.

Melhuish, E., Ereky-Stevens, K., Petrogiannis, K., Ariescu, A., Penderi, E., Rentzou, K., Tawell, A., Slot, P., Broekhuizen, M., & 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). Available at **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.

Neuman, S. B., & Dickinson, D. K. (2010). *Handbook of Early Literacy Research. Volume 3*: Guilford Publications.

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. Volume 3: Guilford Publications.

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 Edition: 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*.

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.

Saracho, O. N., & Spodek, B. (2013). *Handbook of Research on the Education of Young Children. 3rd Edition*: Routledge.

Sylva, K., Siraj-Blatchford, I., Taggart, B., Sammons, P., Melhuish, E., Elliot, K., & Totsika, V. (2006). Capturing quality in early childhood through environmental rating scales. *Early Childhood Research Quarterly, 21*(1), 76-92.

Sylva, K. et al. (2010). Emergent Literacy Environments: Home and Preschool Influences on Children's Literacy Development. Ch. 7 in Neuman and Dickinson (2010).

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

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. Ch. 21 in Neuman and Dickinson (2010).

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.* U.S. Department of Education, Washington, D.C.

APPENDIX A

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

APPENDIX B

Table B1. Results from meta-analysis of 'cherry-picked' effect sizes

Study	Effect	[95% Conf. Interval]		% Weight
Cviko et al (2014)	1.001	0.511	1.491	5.77
Ecalle et al (2015)	0.410	0.341	0.479	24.46
Gasteiger (2014)	0.705	0.050	1.360	3.57
Henrichs and Leseman (2014)	0.800	0.250	1.350	4.80
Hutchings et al (2013)	0.480	0.096	0.864	8.24
Jensen et al (2013)	0.202	0.080	0.324	21.51
Jensen et al (2015)	0.255	0.101	0.409	19.42
Rönnau-Böse et al (2009)	0.320	0.042	0.598	12.23
Overall effect (dl)	0.393	0.260	0.527	100.00

Heterogeneity Measures

Cochrane Q = 20.81 df=7 p-value=0.004

	[95% Conf. Interval]	
I^2 (%) = 66.36	28.76	84.12
H^2=2.97	1.40	6.30
tau^2 est(dl) = 0.018		

Figure B1. Forest plot for meta-analysis of 'cherry-picked' effect sizes