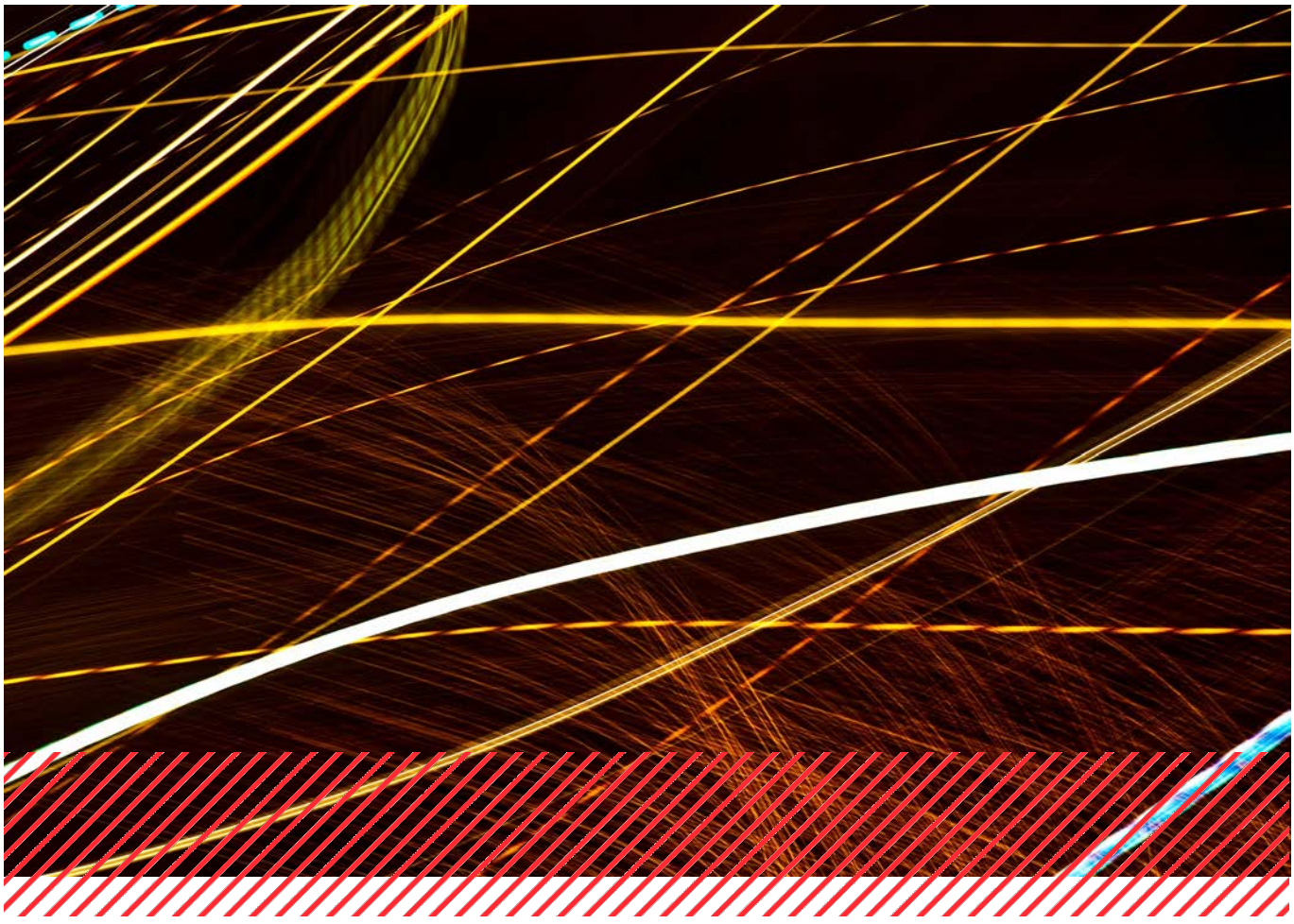
TrygFonden

Durham University

VIVE THE DANISH CENTER FOR SOCIAL SCIENCE RESEARCH

Report

A systematic review of the effective continuing professional development training of welfare professionals



Carole Torgerson, Chantal Nielsen, Louise Gascoine, Trine Filges, Ian Moore, Bjørn Nielsen & Freja Jørgensen

THE DANISH CENTER FOR SOCIAL SCIENCE RESEARCH

A systematic review of the effective continuing professional development training of welfare professionals

e-ISBN: 978-87-93626-96-6

Project: 100700

VIVE – Viden til Velfærd
The Danish Center for Social Science Research
Herluf Trolles Gade 11, 1052 København K
www.vive.dk

VIVE was established on 1 July 2017 following a merger of KORA and SFI. The Centre is an independent government institution that is to provide knowledge that contributes to developing the welfare state and the public sector. The subject areas and types of tasks of VIVE are the same as those of the two former organisations.

Table of Contents

1 List of Tables

2 List of Figures

Acknowledgements

This systematic review is the result of a collaboration between the School of Education, Durham University and VIVE – The Danish Center for Social Science Research (prior to July 1, 2017, SFI – The Danish Research Centre for Social Research). Professor Carole Torgerson (Durham University) and Senior Research Fellow Chantal Nielsen (VIVE), who led the project jointly.

The review team would like to express their gratitude to TrygFonden for awarding funding to both Durham University and VIVE (SFI). We would also like to acknowledge the co-funding provided by Durham University and VIVE to each of their staff members in the research team. Finally, we wish to acknowledge the valuable comments, queries and suggestions provided by two peer reviewers on an earlier draft. These have been very useful in finalising the report.

As a follow-up and extension of the systematic review presented in this report, a Campbell article is being prepared by Trine Filges, Carole Torgerson, Louise Gascoine, Jens Dietrichson and Chantal Nielsen. This forthcoming work is being funded by VIVE and is expected to be published in 2019.

3 Executive summary

3.1 Introduction and background

The quality of the professional development (PD) of education and welfare professionals working with children and young people is of key importance to policy makers and practitioners in these fields. In order to inform education and welfare professionals about the nature and effectiveness of a diversity of approaches to continuing professional development, a systematic review (SR) of the international literature has been undertaken.

In Denmark, there is an increasing acknowledgement of the value of working with evidence-informed approaches and methods. Therefore, the results of this SR are of utmost relevance for the Danish context.

The review aimed to systematically search for, locate, quality appraise and synthesise all the available effectiveness studies that evaluate relevant interventions using rigorous designs.

3.2 Review questions

The research questions were:

- What are the effects of continuing professional development approaches for education and welfare practitioners on: educational and social outcomes for children and young people; and outcomes for practitioners
- What empirical evidence is there on the external validity of specific PD approaches across cultures, across professions/service-deliverer types, across organisations and across service-receiver types.

3.3 Design and methods

The design of the review is a full systematic review (with mapping and in-depth review stages). The methods for each stage of the SR were outlined in a protocol, which was developed before the search for potentially relevant studies began and outlined *a priori* the inclusion and exclusion criteria. Studies that can adequately address the primary research question (which is an effectiveness question) are high-quality evaluations of continuing professional development interventions to improve educational and social outcomes for children and young people and professional practice outcomes for practitioners using experimental designs: randomised controlled trials (RCTs), quasi-randomised trials and studies of quasi-experimental designs (QEDs).

3.4 Results and discussion

The electronic searches identified 3647 potentially relevant studies for screening of titles, abstracts and full papers, using the inclusion/exclusion criteria. After three stages of independent double screening, 33 studies were included in the mapping stage of the review: all were in the area of education. The studies focused on PD in six topic areas, although most of them (30) were in two overarching topic areas: 20 in PD in language and literacy development interventions and outcomes;

and 10 in social development interventions and outcomes. The remaining three studies looked at PD in other topics, i.e. mathematics development; teaching quality; and stress reduction.

Twenty-seven studies were undertaken in the United States (US), and only one study was undertaken in each of the following countries: New Zealand, Germany, the United Kingdom (UK), Netherlands, Indonesia and Denmark. The dominance of the US as the main country in which PD interventions met our criteria clearly limits the generalisability of the findings of the mapping stage for the Danish context. The professional participants in the evaluations of PD interventions were exclusively preschool teachers ('pedagogues') and teachers. The other participants were exclusively children and young people attending preschool (including day care), kindergarten (nursery school) or school, with the majority of studies focusing on outcomes for very young, at-risk children (low socio-economic status or struggling to acquire language and literacy skills) attending preschool (from age 0) or kindergarten (23 studies). The remaining 10 studies focused on interventions evaluated in elementary (primary/middle) school settings and one in a lower secondary school setting.

Common features of the *language and literacy* PD focused on: developing teachers' knowledge and understanding in the substantive fields of reading and writing development (in two cases explicitly using evidence from research). Common features of the *social development* PD focused on: developing teachers' language use, emotional support and positive behaviour-management strategies in the classroom; strengthening teachers' interactions with the children; individualising responses to children and improving teacher-child interactions; improving classroom management skills and creating positive, supportive learning environments; and generally developing teachers' abilities to raise their expectations of children and young people.

A wide range of interventions were evaluated. In both the *language and literacy* topic area and the *social development* topic area, a number of 'branded' interventions were evaluated. Also evaluated were specific generic approaches in delivering PD, such as video recording of classroom interactions, and feedback and evidence-based strategies to raise teachers' expectations of students.

Professional outcomes: In the *language and literacy* topic area, improvements in the quality of language and literacy environment, teaching and instruction were measured. In the *social development* topic area, the following were assessed: the quality of the classroom environment; teacher-child interactions; and teacher efficacy.

Child outcomes: in the *language and literacy* topic area, standardised measures of language and literacy were used to observe improvement. In the *social development* topic area children's socio-emotional skills, socio-emotional and behavioural outcomes and, in some cases, academic outcomes were assessed using standardised measures. Also included in the mapping stage are three studies exploring various 'other' topics: mathematics development; stress reduction among teachers; and teaching quality. Each topic was evaluated by only one study (two using an RCT design and one using QED). Interventions were delivery approaches to PD and ranged from workshops and monthly meetings to summer institutes and follow-up institutes.

All of the included studies met a minimum threshold for quality due to the inclusion criterion for this review specifying a design with either a randomised control or comparison group or a non-randomised control or comparison group with baseline equivalence on primary predictor of outcome, or on all pre-test measures. Five studies used a quasi-experimental design (QED) (with baseline equivalence); one included a combination of quasi-experimental and randomised designs; and the remainder (27) used the most robust design: randomised controlled trial (RCT) design. However, the minimal quality appraisal undertaken for each study at the mapping stage in terms of noting any strengths and limitations revealed varying quality in the included studies.

The in-depth review focused on the social development topic area containing four sub-topics that had been evaluated: social and emotional development; academic and pro-social development; classroom management; and teacher-child interactions. Seven studies were undertaken in the US; and one study was undertaken in each of the following countries: New Zealand, the Netherlands and Denmark. The settings ranged from preschool through elementary schools to secondary schools, with most being early childhood settings. Participants were teaching professionals and children and young people in these settings. Although there was some individual variation in the delivery models of the professional development, the basic components were very similar across all 10 trials and included the following: workshop-based training with resources, personalised coaching/consultation using feedback on observations or video recordings of classroom practice, feedback and reflection. Control conditions were also very similar and tended to comprise business-as-usual PD.

All 10 included studies were randomised controlled trials. Overall, the included studies varied in terms of risk-of-bias judgments, and no single study could be characterised as a robust RCT with a low risk of bias on all assessed risk of bias items.

All 10 studies reported either student or teacher outcomes that enabled the calculation of a standardised mean difference and standard error approximately by the end of the intervention. Three studies reported results on student academic outcomes, and two studies reported outcomes on students' social competences. Seven studies reported outcome measures of teachers; five studies reported various measures of The Classroom Assessment Scoring System (CLASS), and two studies reported other measures of teacher outcomes.

Due to the homogeneity of PD approaches evaluated in the 10 trials, we used homogeneity of professional and student outcomes as the basis of the meta-analyses. We combined individual study outcomes to obtain an 'overall' effect size estimate of the interventions where possible. Three studies reported results on student academic outcomes. The meta-analysis showed that, despite the studies having some differences in their pedagogical approaches and students, the underlying effect of the interventions is similar; positive though very small. However, given that there are relatively few studies, one should exercise caution in assuming that there is a single true effect from PD on student academic outcomes. The remaining two studies reporting outcomes on student social competences were too different to combine.

A total of eight studies reported various and selected items of The Classroom Assessment Scoring System (CLASS) measurement tool. Four of these studies were included in the in-depth review. However, one study in the in-depth review did not have a business-as-usual control group, instead evaluating two forms of a novel PD intervention (Pianta, Mashburn, Downer, Hamre, & Justice, 2008) and so was excluded from the meta-analysis. Two studies reported CLASS summary measures. However, it was unclear whether the same CLASS items were included in the summary measures, so we decided not to combine them. It was possible to combine the results reported in the studies using selected CLASS items in meta-analyses of three items from the possible 10 items in the measurement tool: positive climate, negative climate and behavioural management. Furthermore, it was possible to combine the results from two studies in a meta-analysis of the CLASS item teacher intensity. The weighted average effects of all teacher outcomes were positive, but only positive climate and teacher sensitivity were statistically significant; the weighted averages of negative climate and behaviour management were statistically non-significant. Moreover, it should be noted that there was a high degree of heterogeneity among the studies in the analysis of negative climate.

Given that there are relatively few studies reporting teacher outcomes measured by the CLASS measurement tool, some caution should be exercised in assuming that there is a single true effect from PD on any of these teacher outcomes.

The two remaining studies reporting other measures of teacher outcomes were too different to combine.

In summary, there seems to be a positive but very small effect of PD on student academic outcomes, and there seem to be positive effects on teacher outcomes measured by CLASS. It should, however, be kept in mind that the studies only reported on selected CLASS items (positive climate, negative climate, behavioural management and teacher intensity) and that only two of the combined effects were statistically significant.

At most, the results from three individual studies could be combined in a single meta-analysis. The results of the meta-analyses should therefore be interpreted with great caution due to the very limited number of studies and the selection of measures on teacher outcomes.

3.5 Conclusions

A moderate body of experimental evidence exists in relation to the effect of PD in the topic area of education; similar evidence does not appear to exist in the topic areas of social welfare and crime and justice. A moderate number of experimental evaluations of PD in language and literacy have been undertaken, mainly in the US.

A small body of evidence exists in relation to the effect of PD in social development interventions on students and teachers. The few available studies of effectiveness have varied in terms of methods used to assess the effects. The majority of studies do not report on student outcomes, while the teacher outcomes reported are, with the exception of CLASS measures, too different to be combined.

In short, the result of the in-depth review is that there is currently insufficient evidence for conclusions to be drawn. The small number of available studies reporting similar outcomes precludes any conclusions concerning effectiveness or ineffectiveness of PD. Moreover, the limited number of studies prevented an analysis of specific PD approaches across cultures, across professions/service-deliverer types, across organisations and across service-receiver types.

For the studies in the in-depth review, *relevance of the intervention for the Danish context* was judged based on aspects such as topic, setting, sample, intervention characteristics, method of data collection etc. Interventions in three of the studies were considered to be highly relevant to, and feasible in terms of adaptation to the Danish context. None of these three studies, however, was considered to be of overall high quality in our risk-of-bias assessment. Given the limited number of rigorous studies available at this time, it would be natural to consider conducting a trial of a relevant PD intervention in the Danish context.

3.6 Recommendations

We recommend that a large randomised controlled trial (or a series of large RCTs) evaluating the effectiveness of a PD intervention in the topic area of socio-emotional development be funded and undertaken in the Danish context. The effectiveness of the intervention (compared with business-as-usual PD) on professional and student outcomes should be powered to show the pooled difference to be statistically significant. The intervention condition could be developed from one of the three studies judged to be most relevant to the Danish context, and the outcomes should be *all* items

from the CLASS measure and standardised measures of student outcomes (both social and academic outcomes). We recommend that the trial be designed, conducted and reported according to methodological criteria for rigour with respect to internal and external validity in order to achieve robust results. Particular attention would have to be paid to stringency in terms of designing and conducting a high quality RCT to minimize the risk of bias and to ensure sufficient power in the evaluation. Furthermore, our recommendation would be to identify student outcomes as the primary outcomes of interest.

To underpin this recommendation, this review concludes with a discussion of what characterises a 'receptive environment' when conducting and evaluating new interventions. It is stressed that supportive school management plays a crucial role in fostering a positive attitude among staff members, ensuring sufficient time, resources and suitable structures to enable staff collaboration during and after the intervention. Finally, support from and supervision by researchers involved in such an evaluation would also be important.

4 Introduction and background

The quality of the professional development of education and welfare professionals working with children and young people (for example, preschool teachers or 'pedagogues', school teachers, social workers, psychologists, police officers etc.) is of key importance to policy makers and practitioners in these fields. The general wellbeing of a country's citizens and the provision of better opportunities in terms of educational and social welfare outcomes (for example, participation in higher education and reduction of anti-social behaviour) have been linked to the quality of teaching and, by implication, the quality of continuing professional development (CPD). Conversely, a potential barrier to achieving these education and welfare aspirations is the variable quality of the professional training delivered to educational and/or welfare practitioners, which could mean that the education and training of these groups of professionals may, sometimes, be less than optimal.

A systematic review of the international literature was undertaken, in order to inform education and welfare professionals – policy makers and practitioners – about the nature and effectiveness of a diversity of approaches to continuing professional development. Professional development of these groups of professionals includes delivery strategies, such as: focused supervision; feedback; team work or other kinds of training/CPD approaches that are specifically focused on core teaching skills, such as language and literacy professional development.

4.1 Literature contextualization

Two previous 'tertiary' reviews – or reviews of reviews – in the field of professional development of educators have been undertaken: Dunst, Bruder, & Hamby (2015) and Cordingley et al. (2015).

In their meta-synthesis of 15 reviews, Dunst et al. (2015) looked at the features of PD (in terms of delivery, pedagogy etc.) that were associated with positive teacher and student outcomes in the included SRs and concluded that a range of key PD characteristics led to positive outcomes. However, most of the reviews in this meta-synthesis did not meet our criteria for inclusion based on key items reported in the article. This was due to a variety of factors: the review not using SR or meta-analytic design or not focusing on PD as we defined it, for example focusing on induction for beginning teachers. When a SR included in this meta-synthesis was found relevant to our review, this was subsequently citation searched for relevant empirical studies.

In their 'umbrella' review, Cordingley et al. (2015) included nine reviews from the international literature looking at effective professional development and relating the findings from the reviews to standards of rigour. One review (Timperley, Wilson, Barrar, & Fung, 2007) was found to be consistently robust in all aspects of methodology, and this review has been citation searched for our SR.

4.2 Danish context: Education system

Denmark has an elaborate early childhood education and care system offering a variety of day care options to families with children from the age of 26 weeks and until they start school at the age of 6. These options range from day care in private homes with three-four children per caretaker to nursery institutions (ages 0-3) and kindergartens (ages 3-6). Compulsory school in Denmark covers grades 0-9, i.e. ages 6-16.

The first year of compulsory school in Denmark is termed 'kindergarten class' or 'grade 0' and aims to ensure a smooth transition from kindergarten to school. The Danish *Folkeskole* is a comprehensive school system covering what in other countries is termed primary school and lower secondary schooling. Families are free to choose between a public (i.e. publically funded) school and a private school. In Denmark, there is a long-held tradition for private schools, which receive some financial support from the government, meaning that they are accessible for families who, for various reasons, want their children to attend a private school (e.g. small independent schools in rural districts, religious schools, progressive schools, schools with particular pedagogical approaches, e.g. Rudolph Steiner schools).

After completing grade 9, students can apply for admittance to upper secondary education, which is divided into two overall categories: general upper secondary education, which qualifies students for tertiary education; and vocational or technical education, which qualifies students for access to the labour market. Upper secondary schooling is for adolescents and young adults aged 16-19.

Staff who take care of the youngest children in institutional settings in Denmark are either trained pedagogues or work as pedagogue assistants. To become a pedagogue, one must complete a 3½-year professional bachelor's degree at a University College. Admission requires a high school degree. Pedagogue assistants have less education. To become a pedagogue assistant, one can access vocational training directly after the 9th or 10th grade. This takes up to 3 years and consists of courses at school and practical work in a day care institution.

Pedagogues in Denmark do not only care for and educate children under the age of 6. They are also employed in schools (primary and lower-secondary levels) and in after-school facilities. Pedagogues work in schools as teachers in kindergarten class (grade 0) and have specific tasks working with children at higher grade levels. This can be, for example, sessions focusing on promoting well-being in a class, collaboration among students and physical exercise.

Teachers in the Danish *Folkeskole* typically have a bachelor's degree in teaching, which is a four-year professional bachelor's degree taken at a University College. As with the pedagogue education, admission to study to become a teacher at a University College requires a high school degree. Apart from acquiring competencies in pedagogy and general teaching skills, teachers specialise in three core subjects (e.g. Danish, Mathematics, English or Science).

In Denmark, teachers at the upper secondary level have a master's degree from a University, usually in two subjects. Newly educated teachers are to complete postgraduate training in pedagogy during their first years of teaching in order to teach at this level. For various reasons, however, this does not always happen in practice.

4.3 Danish context: CPD

The recently introduced reform of the *Folkeskole* has set a specific goal of ensuring that, by the year 2020, 95% of lessons taught in (primary and lower secondary) school are to be taught by teachers with documented teaching competencies in the subjects that they teach. Such documented competencies can have been acquired as part of initial teacher training or through relevant PD. Financial resources have been set aside to achieve this goal. The reform also includes an ambition of increasing the competencies of pedagogues working in schools, and financial resources have been set aside to achieve this goal also.

In recent years, municipalities have increasingly worked systematically to ensure professional development for teachers and pedagogues employed not only in schools, but also in day care institutions (i.e. nurseries and kindergartens). Statistics from 2016 show that, on average, 59% of staff in Danish day care institutions are qualified pedagogues. The municipality with the highest share of educated pedagogues has 73%, while the lowest has 40%. The Danish Union of Early Childhood and Youth Educators (BUPL) has an ambition of increasing this share to 80% (BUPL, 2016).

A recent report shows that Danish municipalities have placed great emphasis on increasing the coverage of subject-specific competencies among teachers, when determining what types of PD schools should engage in (Bjørnholt et al., 2017). This differs from the situation in 2013, where inclusion was the main focus. Other topics of importance for municipalities as regards their requirements or expectations for PD focus in their schools are: training of counsellors and other resource staff, PD in goal-oriented teaching and PD in specific subjects, such as Danish and mathematics. PD in other topics such as inclusion, classroom management, relational competencies, team collaboration and cross-curricular collaboration are not as high on the agenda at the municipality level. However, interviews conducted in the study by Bjørnholt et al. (2017) indicate that these are topics left for the individual school to prioritise or not.

4.4 Aim of this review

The review aimed to systematically search for, locate, quality appraise and synthesise all the available effectiveness studies that evaluated relevant interventions using rigorous designs. By 'rigorous designs' we mean those research designs that can establish a causal link between continuing professional development interventions and outcomes for professionals themselves, children and young people. Therefore, we included: systematic review (SR) and meta-analytic designs, 'true' experiments (randomised controlled trials or RCTs), quasi-experiments (with baseline equivalence as demonstrated by pre-tests in the outcomes of interest, but excluding studies using an instrumental variable approach), including studies using regression discontinuity design. We searched substantively for studies in the fields of education, social welfare and crime and justice.

An initial scoping search on one database was undertaken, using the following search strategy:

> *TI (teacher OR social worker OR police OR psychologist) AND TI (professional development OR continuing professional development OR CPD OR in service training OR professional learning OR teacher learning OR training) AND AB (experiment* OR quasi experiment* OR QED OR control OR allocat* OR randomi#ed controlled trial OR RCT OR regression discontinuity OR RDD)*

This scoping search produced 470 potentially relevant hits, which, after screening using preliminary inclusion criteria, indicated that a range of potentially relevant studies, mainly in the field of education but also in other areas of social welfare and policing, were available to be systematically assembled. We were also aware of a recently published meta-analysis in the specific area of professional development among professionals working with children's early language and literacy development (Markussen-Brown et al., 2017). This meta-analysis formed part of the basis of our electronic and citation searching in the field of education. Note that our search covered the entire field of education and was not limited to studies on language and literacy development.

The review was completed using systematic review design and explicit methods that are open to scrutiny (Torgerson, 2003), as this minimises bias and increases confidence in the results.

4.5 Review questions

The research questions were:

- What are the effects of continuing professional development approaches for education and welfare practitioners on: educational and social outcomes for children and young people; and outcomes for practitioners
- What empirical evidence is there on the external validity of specific PD approaches across cultures, across professions/service-deliverer types, across organisations, across service-receiver types etc.

4.6 Definitions

For the purpose of this review, we have adopted the following definitions, inspired by Buysse et al. (2010):

4.6.1 Continuing professional development

- CPD encompasses facilitated learning opportunities for education and welfare professionals that have completed their ordinary (basic) training at an (higher) education institute relevant to their professional degree. This (previous) degree can be at varying ISECD levels (e.g. diploma, BA, MSc, PhD).
- CPD includes all types of facilitated learning opportunities. Some types of CPD will be shorter term, informal and situated in practice and will not lead to credits, diplomas or degrees. Other types of CPD will be longer term, involve formal coursework and take place at teachers' colleges or universities and will lead to credits, diplomas or degrees.
- The aim of CPD should be to enhance the professionals' knowledge and skills in ways that are relevant for application in practice, i.e. to serve the ultimate beneficiaries of the intervention, i.e. the children and young people with/for whom the education and welfare professionals work.
- CPD can be delivered by public or private professional development and professional training entities.

CPD can be delivered in many more or less formal ways, including coaching, mentoring, consultations and established communities or teams of practice. In such cases, the CPD must have explicitly formulated content and goals. Note that (informal) allocation of a mentor for the purpose of general collegial support is not included in this definition of CPD.

4.6.2 Education and welfare professionals

- Education and welfare professionals are employees working directly or indirectly with and for children and young people with the explicit purpose of enhancing their cognitive and non-cognitive development
- This includes, but is not limited to, education and welfare employees working towards these goals in settings such as nurseries, day care and other child care institutions, preschools and schools at different levels
- Education and welfare professionals can be either publically or privately employed, they receive salary for their work, which may be full-time or part-time

- Education[1] and welfare professionals have completed general (basic) training at a higher education institute relevant for their professional degree. This degree can be at varying ISECD[2] levels (e.g. diploma, BA, MSc, PhD)
- Education and welfare professionals are recipients of the CPD activities and interventions that are being evaluated
- Examples of education and welfare professionals include teachers, teaching assistants, preschool teachers (pedagogues), care providers, social workers, paraprofessionals, psychologists, police officers, family support providers, disability specialists and inclusion specialists
- The roles of education and welfare professionals include planning, developing, delivering and evaluating learning and development opportunities for children and young people.

[1] Teaching assistants (TAs) are not included in this, although we acknowledge the relevance of their work in terms of welfare. In the UK context, the role of TA is not a degree-level profession, although it is likely that there are many TAs working who have a degree.

[2] ISCED: International Standard Classification of Education.

5 Design and methods

The design of the review is a full systematic review (with mapping and in-depth review stages). The design and methods of the review were informed by the Campbell Collaboration policy briefs (Campbell Collaboration, 2018); 'Systematic reviews: CRD's guidance for undertaking reviews in health care' (University of York, Centre for Reviews and Dissemination, 2009); the 'Cochrane Collaboration Handbook' (Higgins & Green, 2011); the *Handbook of Research Synthesis* (Hedges & Cooper, 1994); '*Systematic Reviews' (Torgerson, 2003)*. The design and methods for each stage of the SR were outlined in a Protocol that was developed before the search for potentially relevant studies began and outlined, *a priori,* the inclusion and exclusion criteria. The protocol was published as a note at The Danish National Centre for Social Research (SFI) (Torgerson et al., 2017) following approval from TrygFonden (main funder of the review).

The reporting of each stage of the systematic review process was guided by the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) statement (Moher, Liberati, Tetzlaff, & Altman, 2009) to ensure transparency.

Studies that can adequately address the primary research question (which is an effectiveness question) are high-quality evaluations of continuing professional development interventions to improve educational and social outcomes for children and young people and professional practice outcomes for practitioners using experimental designs: randomised controlled trials, quasi-randomised trials and quasi-experiments. We included only study designs that employ a treatment-control or a treatment-comparison group design. A control group is defined as a non-treatment condition, while a comparison group receives an alternative treatment. Studies using single group pre-post-comparisons were not included; in order to establish causality (i.e. to be able to state that a specific professional development intervention *causes* an improvement in the outcomes stated above), study designs which can adequately control for all other known and unknown variables that could affect outcome are required (Cook, Campbell, & Boston, 1979; Shadish, Cook, Campbell, & Boston, 2002).

1. Randomised and quasi-randomised controlled trials (allocated at either the individual level or cluster level, e.g. class/school/social worker/geographical area etc.).
2. Quasi-experimental studies (including regression discontinuity design, but excluding studies using an instrumental variable approach – see Appendix A for our rationale for excluding studies of these designs). We also included only QED studies that demonstrated baseline equivalence in the main outcomes of interest. A further requirement was that these studies were able to identify an intervention effect. Studies where, for example, the treatment was given to teachers in one school only and the comparison group was teachers at another school (or more schools for that matter) were not able to separate the treatment effect from the school effect.

This review focuses on research evidence from academic journals and other published research from the last 20 years (as this provides the most up-to-date evidence for policy makers, practitioners and funders on effective practices, strategies and interventions). In order to limit the possibility of publication bias, research from difficult-to-locate 'grey' literature was searched for and included. Our approach to the search for 'grey' literature is described in a separate section below.

Studies in which at least one of the groups received a CPD intervention compared to either standard practice ('business-as-usual') or an alternative CPD intervention were included.

5.1 Inclusion criteria

Table 1: Inclusion and exclusion criteria

INCLUDED	EXCLUDED
Date: 1997 to present (last 20 years)	Date: Pre-1997
Publication status: Published or unpublished but in the public domain	
Nature of research: Empirical research or review of empirical research	Nature of research: Non-empirical research or review of non-empirical research
Study design: RCT; quasi-experiment (with baseline equivalence), including RDD	Study design: Study using IV approach; non-experimental study designs (i.e. studies without a control or comparison group)
Topic: Education; social welfare; crime and justice	Topic: Not education; social welfare; crime and justice
Participants: Welfare professional (preschool teacher, pedagogue, school teacher, social worker, psychologist, police officer[3])	Participants: Not welfare professionals (e.g. volunteers) or welfare professionals in a school-based role that does not require a professional degree (e.g. teaching assistants/TAs)
Participants: Target group (children and young people between the ages of 0 and 18 years)	Participants: Aged 19 years and over (adults)
Intervention: Intervention in continuing professional development (CPD) in the three topic areas. CPD includes, but is not restricted to: focused supervision; feedback; team work or other kinds of training/PD approaches; literacy and language teaching skills, problem-solving teaching skills, socio-emotional development skills and other CPD content	Intervention: Does not have a CPD component; initial training intervention/PD (e.g. initial teacher training)
Outcomes: Primary: Educational, social welfare and crime and justice outcomes for children and young people; Secondary: any intermediate outcomes on children and young people, such as at-risk behaviours; family outcomes; any outcomes for practitioners that are focused on improving any aspect of professional practice. Studies were included only if they included at least one valid and reliable outcome that had been standardised on a different population *[and was 'objective', i.e. not 'experimenter-designed' and not self-reported].	Outcomes not related to education, social welfare and crime and justice. Practitioner outcomes not focused on improving professional practice, e.g. higher job satisfaction *['Experimenter-designed' outcomes] *[Self-reported outcomes]

*Note: Inclusion and exclusion criteria specifically relating to outcomes (experimenter-designed and self-reported) were added as a variation to the Protocol in the third stage of screening.

5.2 Systematic reviews and meta-analyses

Systematic reviews, meta-analyses and narrative reviews were included at the mapping stage and for citation searching (see below). Identification of relevant systematic-reviews and meta-analyses was integrated into the general search and citation strategy.

5.3 Electronic searching

We conducted initial scoping searches in key databases (e.g. ERIC, PsycINFO, SocIndex, Web of Knowledge). We then developed search strategies in an iterative process and, once finalised, conducted all the systematic electronic searches in the following seven databases:

- ERIC (searched through EBSCOhost)

[3] Graduate entry routes into the police in the UK context have been established.

- PsycINFO (searched through EBSCOhost)
- SocIndex (searched through EBSCOhost)
- Academic Search Premier (searched through EBSCOhost)
- Teacher Reference Center (searched through EBSCOhost)
- Web of Knowledge (Social Science Citation Index & Science Citation Index/SSCI + SCI) (searched via Thomson Reuters)
- ASSIA (searched through ProQuest).

The results of all of the electronic searches were combined into a master database in a software database specifically designed for processing studies in a SR: EPPI-Reviewer 4 (Thomas, Brunton, & Graziosi, 2010). The search strings for each database can be seen in Appendix C1.

5.4 Grey literature search strategy

In order to identify relevant grey literature for the review (reports, academic theses, working papers etc.) different strategies were utilised. We searched specific targeted relevant online repositories, such as the Danish and US clearing houses for educational research (https://ies.ed.gov/ncee/wwc/WhoWeAre).

Furthermore, we searched general research repositories (such as Social Care Online) and national research portals such as Forskningsdatabasen (Danish National Research Database), SwePub (Academic content from Swedish universities) and NORA (Norwegian Open Research Archive). Searches on Google Scholar for grey literature were also developed (see Appendix C1).

5.5 Citation searching

Due to the time restraints of the review process, we prioritised citation tracking of the most relevant identified studies. We performed citation searching on systematic reviews and meta-analyses that were included after the second stage (full-text) screening. In general, the citation tracking was retrospective, i.e. we searched the bibliography of the relevant studies. We made a judgement to prioritise exhaustive searching and therefore used systematic citation searching to supplement the primary strategy (namely systematic electronic searching).

5.6 Screening at 1st, 2nd and 3rd stages

Once de-duplicated, a random sample of studies was independently triple screened in EPPI at the first stage (titles and abstracts only) by the three reviewers using the inclusion/exclusion criteria (Table 1) by way of quality assurance. The database was then split into equal thirds, and each third was double screened by two reviewers. Any disagreements were resolved through discussion, with arbitration where necessary by a third reviewer. If necessary, a fourth reviewer was available to provide confirmation of inclusion/exclusion. Potentially relevant studies (i.e. studies remaining after title and abstract screening) were located and retrieved. Once retrieved, all full papers were double screened at the second stage, with arbitration (where necessary) as described above. All included studies were re-screened at the third stage. This stage of screening was added as a variation of the Protocol to exclude studies that only used experimenter-designed or self-reported outcomes, as these kinds of outcomes are susceptible to the introduction of bias.

5.7 Mapping: Data extraction and quality appraisal – synthesis

All studies included at the third stage were independently double data extracted by at least two reviewers working in pairs, a bespoke data extraction tool being devised for this purpose. The pairs of reviewers then reconciled their data extractions, and a completed data extraction for each study was finalised in EPPI-Reviewer. Key, basic data about: bibliographic information (author, country); participants; settings; details of intervention condition (CPD); details of control or comparison condition(s); and outcome measures were extracted and tabulated. A number of quality items were also extracted. A minimal model for quality appraisal was developed, which included design (RCT, individual or cluster); allocation sequence method; and attrition. We checked the following: whether clustering was taken into account in the analysis for cluster-randomised controlled trials; the extent of baseline equivalence (for RCTs and QEDs); the sample size and level of attrition of professionals and children; whether there was blinding in outcome ascertainment; and whether an intention to treat analysis was undertaken.

The mapping tables focused on key characteristics (interventions and outcomes) of the included studies: PD interventions targeting specific groups of professionals and outcomes targeting both the professionals and the children/young people (PD and wider outcomes).

5.8 In-depth review: Additional data extraction and quality appraisal – synthesis

After consultation with TrygFonden, a research question for an in-depth review narrowed down the focus to a specific area of interest to the funder based on the overall results of the mapping in terms of topic and sub-topic coverage. Two main topic areas emerged in the map (language and literacy development; and social development) with an additional four minor topic areas also present, and a decision was made to narrow the focus to social development interventions and outcomes for the in-depth review.

More detailed data extraction of the studies included in the in-depth review was undertaken, including information about participants, settings, intervention, control or comparison conditions, outcomes (professionals and children/young people) and results. Information extracted included the following:

- Eligibility criteria specified
- Method to generate allocation to groups
- Participants and outcome assessor
- Presentation of estimate of effect size and its precision.

In addition, data were extracted on topic/focus; setting – structural aspects; sample; PD participation; intervention – practical aspects; content; cultural aspects; teacher autonomy; data collection; and outcomes in focus. These data were used to compile contextualisation tables for individual studies in the in-depth review, with a judgement of the extent to which each of the individual items were relevant to the Danish context and an overall judgement using the following algorithm:

- High: at least eight items high and no low or moderate to low items
- Moderate to high: at least eight items moderate to high or higher and no low items
- Moderate: at least eight items moderate or higher
- Moderate to low: at least eight items moderate to low
- Low: at least eight items low.

A modified version of the risk-of-bias model developed by Professor Barnaby Reeves in association with the Cochrane Non-Randomized Studies Method group (Reeves, Deeks, Higgins, & Wells, 2011) was used in order to develop a tool to assess the risk of bias in the included randomised, quasi-randomised and quasi-experimental studies included in the in-depth review. This model, an extension of the Cochrane Collaboration's risk-of-bias tool, covers risk of bias both in RCTs and non-randomised studies with a well-defined control or comparison group.

The modified version of this model should address the following nine risk-of-bias judgement items:

5.9 Risk-of-bias judgement items

- Sequence generation (judged on a low/high risk/unclear scale)
- Allocation concealment (judged on a low/high risk/unclear scale)
- Confounders (judged on a 5-point scale/unclear)
- Blinding (judged on a 5-point scale/unclear)
- Incomplete outcome data (judged on a 5-point scale/unclear)
- Selective outcome reporting (judged on a 5-point scale/unclear)
- Other potential threats to validity (judged on a 5-point scale/unclear)
- *A priori* protocol (judged on a yes/no/unclear scale)
- *A priori* analysis plan (judged on a yes/no/unclear scale).

On a 5-point scale, 1 corresponds to low risk of bias and 5 to a high risk of bias. A score of 5 on any of the items assessed on the 5-point scale translates to a risk of bias so high that the findings were not considered in the data synthesis, because they are more likely to mislead than inform.

Quality appraisal of the included studies preceded any declaration of results. Use of the aforementioned modified risk-of-bias tool potentially excluded studies with too high a risk of bias.

As different computational methods may produce effect sizes that are not comparable, we have chosen full transparency regarding all methods used in the primary studies (research design and statistical analysis strategies) and exercised caution when synthesising effect sizes[4].

The synthesis for the in-depth review combined the results meta-analytically, focusing on outcomes targeting specific groups of participants (professionals and students) in the topics of social development and language and literacy development respectively.

Quality assurance

Each stage of the SR was undertaken by at least two reviewers working independently and then reconciling and reaching agreement on all decisions, with arbitration by a third or fourth reviewer where necessary. In the early stages of both screening and data extraction, 'training' in consistency of applying decisions and judgement was checked through screening and data extraction of a sample of studies by all four reviewers. At each stage of the review, expert methodological advice was sought by a fifth reviewer (TF), when the methodological complexity of a study necessitated her

[4] Special caution was intended to be exercised concerning studies using regression discontinuity (RD) to estimate a local average treatment effect (LATE). These were to be included, but subject to a separate analysis depending on the comparability of the LATEs and the effects from other studies. We intended to check the sensitivity of our results to the inclusion of RD studies. In addition, we intended to discuss the limitation in generalisation of results obtained from these types of studies. However, no studies employing a RD were included in the mapping and the in-depth review.

advice and judgement. Data extraction, risk-of-bias assessment and extraction of numerical data for effect size calculation and pooling of effect sizes in the meta-analyses were all undertaken by two reviewers, working as a pair.

6 Results and discussion

6.1 Systematic searches

The electronic searches were completed in seven databases. Additionally, grey literature was searched for in seven different locations. All searches were performed in April and May 2017. The searches identified a total of 4328 records, which was reduced to 4245 records due to automatic de-duplication processes in the databases. De-duplication was performed in Mendeley prior to bibliographical files for the 4245 records being imported into EPPI-Reviewer 4 for management and screening. Table 2 indicates both the number of records retrieved from each database searched and the number remaining after de-duplication in EPPI-Reviewer[5]. After de-duplication in EPPI-Reviewer, 3647 records remained for first-stage screening.

[5] Additional de-duplication was required in EPPI as several duplicates still remained after de-duplication in Mendeley.

Table 2: Databases searched and number of records retrieved

Database	Date of search	Platform	Number of records found	Number of records exported into biblio-graphical files	Number of records after de-duplication in EPPI
Systematic searches					
Web of Science (SSCI + SCI)	01/05/2017	Thompson Reuters	544	544	468
Academic Search Premier (ASP)	17/04/2017	EBSCO (searched simultaneously)	536	528	397
Education Resources Information Center (ERIC)			1423	1385	1305
PsycINFO			807	791	706
SocIndex			217	215	100
Teacher Reference Center (TRC)			156	156	67
Applied Social Sciences Index and Abstracts (ASSIA)	04/05/2017	ProQuest	236	223	205
Grey literature searches					
Forskningsdatabasen (Danish National Research Database)	01/05/2017	Web based	177	171	170
SwePub (Academic content from Swedish universities)	01/05/2017	Web based	65	65	64
Norwegian Open Research Archive (NORA)	01/05/2017	Web based	4	4	4
Google Scholar	02/05/2017	Google	34	34	33
Social Care Online (SCO)	01/05/2017	N/A	129	129	128
US Clearing house for educational research	03/05/2017	N/A	0#	-	-
Danish Clearing house for educational research	03/05/2017	N/A	0**	-	-
TOTAL	**N/A**	**N/A**	**4328**	**4245**	**3647**

Note: # No records were found that had not already been found in the systematic and previous grey literature searches.

6.2 Citation searches

Upon completion of the second stage of screening, eight systematic reviews, 'tertiary' reviews (review of reviews) or meta-analyses remained. (Dunst et al., 2015; Gaudin & Chalies, 2015; Hwang, Bartlett, Greben, & Hand, 2017; Kelcey & Phelps, 2013; Lander, Eather, Morgan, Salmon, & Barnett, 2017; Markussen-Brown et al., 2017; Snell, Dowsell Forston, Stanton-Chapman, & Walker, 2013; Yoon, Duncan, Lee, Scarloss, & Shapley, 2007). Four reviews were added from the EPPI publication page found at https://eppi.ioe.ac.uk/cms/Default.aspx?tabid=274. The four reviews (Cordingley, Bell, Isham, Evans, & Firth, 2007; Cordingley, Bell, Evans, & Firth, 2005; Cordingley, Bell, Rundell, Evans, & Curtis, 2003; Cordingley, Bell, Thomason, & Firth, 2005) were coded in the EPPI SR library as professional development reviews and were therefore of direct relevance to our review. To extract relevant studies, each record (in PDF format) was searched using keywords, including *'control' 'experimental'* and *'randomised', in order* to locate relevant references. The results section for each review was also scrutinised to maximise the reach of the citation searching.

The citation searches (conducted by Ian Moore) of the 12 records above found 41 papers that were not already located in the original systematic searches. Each paper was screened using the criteria used for the first and second screening stages of the results of the electronic searches. Included papers (after first-stage screening) were collated in a table with reasons for inclusion or exclusion. This was shared with two other reviewers (LG and CT), who each reviewed half of the papers. In cases of doubt, the records were checked by a third reviewer before a final decision was made. Twenty-two records were included from citation searches (after screening at the first and second stages) and moved on to third-stage screening (in EPPI-Reviewer). Nineteen (of the 41) records were excluded in the first and second stages, predominantly due to reasons of inappropriate outcome measures as per the exclusion criteria in Table 1. An expert in the field identified four systematic reviews (Basma & Savage, 2017; Blank & de las Alas, 2009; Timperley et al., 2007; Zaslow et al., 2010) and one 'tertiary' review (Cordingley et al., 2015). Basma and Savage (2017) was not screened, as it was identified too late in the process. It will be included in any update of this systematic review. Screening of the four reviews led to one additional study being identified for inclusion in our systematic review.

6.3 Screening in the first, second and third stages[6]

Figure 1 shows the flow of records through the systematic review process using a PRISMA flow diagram (Moher, Liberati, Tetzlaff, & Altman, 2009). Inter-coder agreement at first-stage screening (title and abstract) was over 90% in all pairings of reviewers (range: 90-97%). In the first stage screening, 3410 records were excluded, leaving 237 records eligible for full-text screening, two of which were not available. Thus, 235 records were screened for inclusion in the second stage (full text).

In the second stage screening, full texts were assessed for inclusion based on the criteria set out in Table 1 (p. 19). Inter-rater reliability at this stage (include/exclude only) was lower than in the first-stage screening, but all disagreements were resolved by a third reviewer, and all parties agreed before coding was finalised. In total, 132 records were excluded in the second stage (for reasons, see Table 3), 2 were unavailable, which left 103. Table 4 shows that the remaining records were predominantly empirical and focused on continuing professional development in education. In Table 4, it can be seen that there were 84 empirical studies (in the 103 records remaining after the second-stage screening). These were combined with 22 empirical studies from citation searching, meaning that a total 106 records were carried forward to screening in the third stage prior to data extraction, which led to 74 additional studies

[6] Appendix C3 shows the numbers of records for each database searched that were included and excluded in all stages (first, second and third stages).

being excluded (see Table 5 for reasons) and 33 studies being included (Table 4). All 33 studies were coded as having an 'education' focus.

Table 3: Reasons for exclusion in the second stage

Reason for exclusion in full-text screening stage	Number of records excluded
Date	0
Nature of research	4
Study design	40
Topic	2
Participants (profession)	6
Participants (target age group)	0
Exclude on intervention	83
Exclude on outcomes	4
TOTAL:	132 (139*)

Note: * Seven records were excluded for more than one reason, thus a total of 132 records were excluded.

Table 4: Included records (type and focus) after second and third-stage screening (including citation searches from the 12 SR/MA)

Stage of screening	Total number of empirical records	Topic	Record type		
			Empirical	Systematic Review/Meta-analysis/'Tertiary' review	Review (e.g. narrative review or report)
Records remaining after second stage screening	84	Education	81	12	11
		Social Welfare	3	-	-
		Crime and Justice	-	-	-
Records from citation searches (added before third stage screening*)	22	Education	22	-	-
Empirical records screened at third stage	106	Education	103	-	-
		Social Welfare	3	-	-
Records remaining after third stage screening	32	Education	32	-	-
		Social Welfare	0	-	-
		Crime and Justice	-	-	-
Systematic reviews/meta-analysis/'Tertiary' review identified by expert reviewer for screening and citation searching	-	Education	-	5	-
Empirical records added	1	Education	1	-	-
TOTAL number of empirical studies included in mapping	33	Education	33	-	-

Note: * From the 8 SR/MA as above, plus 4 reviews from the EPPI website (12 citations searched in total). 41 records were first and second stage screened manually, before uploading the remaining 22 to EPPI for third stage.

The most striking result of the process of searching and screening to inclusion in the third stage is that all of the 33 included empirical studies were in the area of education. This was despite exhaustive searches to include any relevant studies in all three areas. It is possible that empirical studies have been undertaken to evaluate the effectiveness of CPD interventions in the areas of social welfare and crime and justice, but that they did not meet our strict inclusion criteria. So, for example, they could have used a research design without an appropriate control or comparison group, or they could have used experimenter-designed or non-validated outcome measures.

Figure 1: PRISMA flow diagram showing the flow of records to the mapping stage, based on Moher, Liberati, Tetzlaff, & Altman (2009)

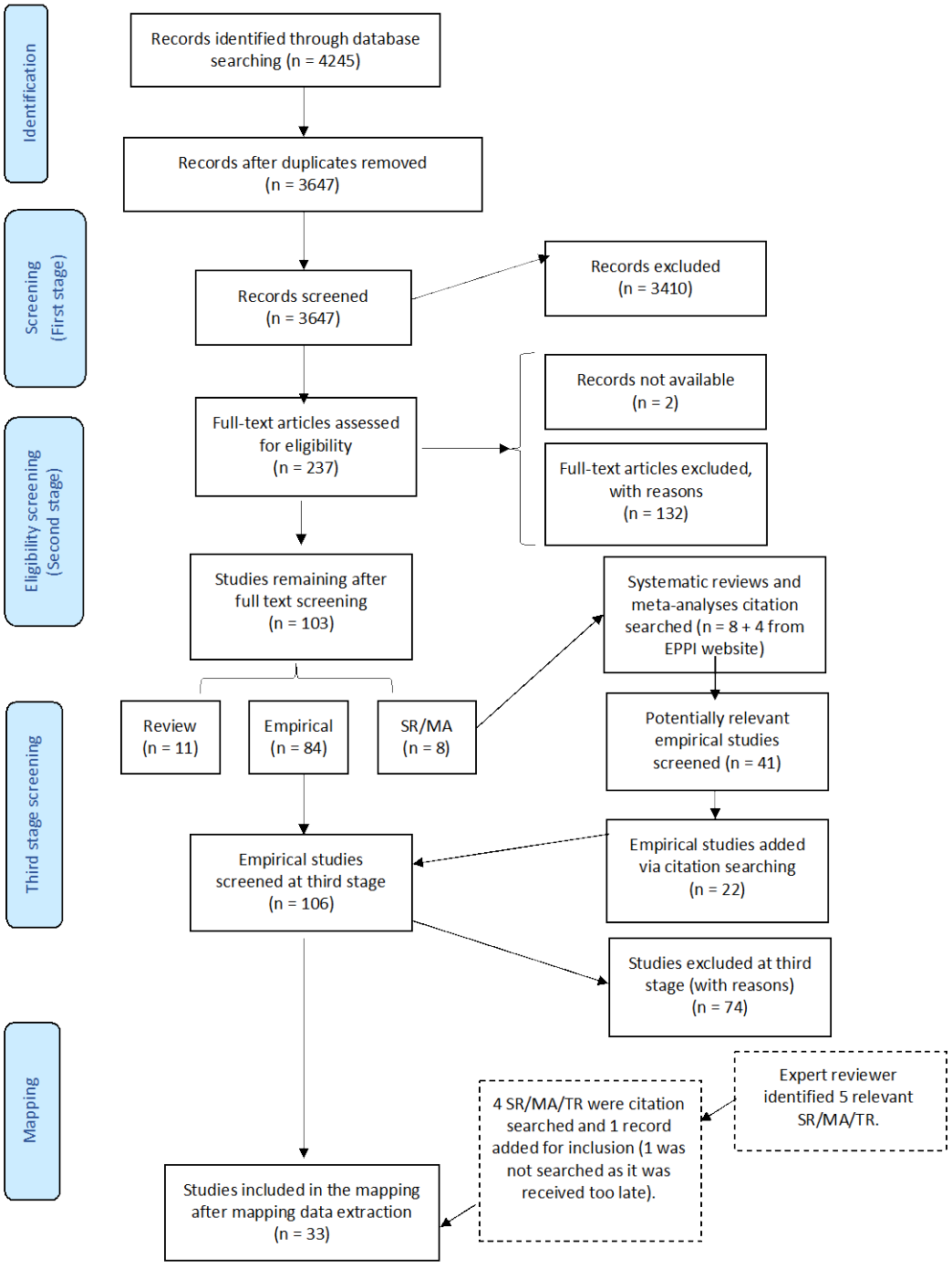


Table 5: Reasons for exclusion in the third stage

Reasons for exclusion in third-stage screening	Number of records excluded
Lack of clarity in reporting results or results not reported (e.g. trial protocol)	9
Lack of clarity in describing control condition or control group absent	4
Intervention (does not fit stated definition of PD)	7
Lack of baseline equivalence	6
Exclude on topic (e.g. focusses entirely on health) or focus (e.g. teacher burnout, motivation) as per protocol	6
Exclude on study design	4
Exclude on outcome measures:	
Experimenter-designed or adjusted outcome measures	17
Outcome measures not validated	3
Self-report outcome measures only	6
Other reason for exclusion based on outcome measure	12
TOTAL:	74

Experimenter designed outcome measures designed by the author(s) have typically been developed for the specific study and have not been validated or standardised against another sample. In some cases, the instruments have been pilot tested, but this is not adequate in terms of providing full confidence in the quality and validity of the outcome measure. In other cases, the authors have combined existing instruments with experimenter-designed items, and these can thus be thought of as *experimenter adjusted outcome measures*. The use of *self-reported outcome measures* is also quite widespread in many of the studies found in the early screening for this review – typically alongside other more objective and reliable outcome measures. The problem here is of course – by definition – risk of self-reporting bias – typically in the direction of over-estimating a possible effect of the intervention. We have excluded studies that rely exclusively on self-reported outcome measures, and which are thus not based on validated assessment tools.

6.4 Results and discussion: Mapping

Table 6 presents the bibliographic details, topics, aims of professional development and design, interventions and outcomes of the 33 included empirical studies. The table in Appendix F provides additional, detailed information from the records about the specifics of each intervention presented in Table 6.

Table 6: Data extraction (mapping)

Author, date, country, title	**Design**	**Topic (stage of schooling, participant characteristics)**	**Aims of PD**	**Intervention** (name and **description of PD format**)	**Outcome and outcome measures (professional and child)**
OVERARCHING TOPIC: LANGUAGE AND LITERACY DEVELOPMENT					
Al Otaiba et al., 2011, US Assessment data-informed guidance to individualize kindergarten reading: Findings from a Cluster-Randomized Control Field Trial	RCT	LANGUAGE AND LITERACY DEVELOPMENT (KINDERGARTEN)	To develop teachers' abilities to differentiate or individualise instruction – based on ongoing assessments of students' language and literacy skills - to ensure that most students learn to read	Individualized Student Instruction for Kindergarten (ISI-K) **Professional development and in class support – baseline workshop, coaching and classroom support**	Professional outcomes excluded (not validated) Child outcomes: Vocabulary: Picture Vocabulary subtest of the WJ-III (Woodcock) test. Word reading skills: WJ-II Letter Word Identification (Woodcock) test. Ability to decode pseudo-words: Word Attack subtest of the WJ-III (Woodcock) test DIBELS Nonsense Word Fluency (NWF) and Phoneme Segmenting Fluency (PSF) tasks
Bos et al., 1999, US Interactive, collaborative professional development in early literacy instruction: Supporting the balancing act	QED	LANGUAGE AND LITERACY DEVELOPMENT (EARLY ELEMENTARY AND SPECIAL EDUCATION)	To support early elementary and special education teachers in integrating explicit instruction (into their curricula) for children at risk of reading failure	Project RIME (Reading Instructional Methods of Efficacy) **Professional development course and in-class support/collaboration**	Professional outcomes excluded (experimenter designed and self reported) Child outcomes: Three measures from the Woodcock-Johnson Tests of Achievement III
Buysse et al., 2010, US Effects of a professional development program on classroom practices and outcomes for Latino dual language learners	RCT	LANGUAGE AND LITERACY DEVELOPMENT (DUAL LANGUAGE LEARNERS: PRE-KINDERGARTEN)	To provide specific instructional strategies to support teachers in addressing language and literacy skills of DLLs	Nuestros Niños program **Professional development: individual and collaborative support**	Professional: The Early Language and Literacy Classroom Observation (ELLCO) Toolkit (Education Development Center, 2002) Child outcomes: Woodcock Language Proficiency Battery-Revised: English and Spanish Forms; Peabody Picture Vocabulary Test English and Spanish; Phonological Awareness Tasks; Naming Letters (National Center for Early Development & Learning, 2003); and Where's My Teddy Story and Print Concepts (FACES: The Head Start Child and Family Experiences Survey, 2003).

Author, date, country, title	**Design**	**Topic (stage of schooling, participant characteristics)**	**Aims of PD**	**Intervention** (name and **description of PD format**)	**Outcome and outcome measures (professional and child)**
Cabell et al., 2011, US The Impact of Teacher Responsivity Education on Preschoolers' Language and Literacy Skills	RCT	LANGUAGE AND LITERACY DEVELOPMENT (PRESCHOOLERS; LOW SES)	To train educators in how to improve their responsively oriented interactions with children, in particular in the importance of talking to children about print during literacy related activities	Adapted from Learning Language and Loving it. **Professional development: training**	Professional: None Child outcomes: Children's language skills: Grammar using "composite of two subtests from the standardized, norm-referenced CELF Preschool-2 (Wiig et al., 2004)", The Word Structure subtest and the Sentence Structure Subtest. Receptive vocabulary – Peabody Picture Vocabulary Test-III (Dunn and Dunn, 1997) Expressive Vocabulary subtest of the CELF Preschool-2 Emergent literacy skills (including print-concept knowledge and alphabet knowledge. Print concept knowledge – 14-item Preschool Print and Word Awareness test (Justice, Bowles, & Skibbe, 2006). Upper-Case Alphabet Knowledge and the Lower-Case Alphabet Knowledge tasks of the Phonological Awareness Literacy Screening for Preschool (Invernizzi, Sullivan, Meier, & Swank, 2004)

Author, date, country, title	**Design**	**Topic (stage of schooling, participant characteristics)**	**Aims of PD**	**Intervention** (name and **description of PD format**)	**Outcome and outcome measures (professional and child)**
Dickinson & Caswell, 2007, US Building support for language and early literacy in preschool classrooms through in-service professional development: Effects of the Literacy Environment Enrichment Program (LEEP)	QED	LANGUAGE AND LITERACY DEVELOPMENT (PRESCHOOLERS; LOW SES)	To help teachers improve the quality of support that they provide for children's language and literacy development	LEEP Effects of the Literacy Environment Enrichment Program **Professional development: course and in-class support – lectures, video recordings, analysis of work samples**	Professional: "Early Language and Literacy Observation (ELLCO) Toolkit (Smith et al., 2002). The ELLCO was developed by the same team of researchers that created LEEP, but it was a separate endeavour. ELLCO reflected the same theoretical understanding of literacy that was communicated to teachers taking LEEP, but it was not specifically tailored to assess the effects of LEEP on classroom practices. This tool is comprised of three overall components, each of which has two subscales. The three components are: (1) the Classroom Observation and Teacher Interview, a classroom observation section that uses anchored rating scales to evaluate classrooms along 14 holistic dimensions; (2) the Literacy Environment Checklist, a checklist for rating materials and classroom spaces; (3) the Literacy Activities Rating Scale, a scale for coding literacy-related activities that occur while the observation is carried out." (p. 248). Child outcomes: None
Gallagher et al., 2011, US, An evaluation of the individualized learning intervention: A mentoring program for early childhood teachers.	RCT	LANGUAGE AND LITERACY DEVELOPMENT (PRESCHOOL) LEARNING BEHAVIOUR (PRESCHOOL)	To provide ongoing support through mentoring, thus getting higher quality teachers	Individualized learning intervention **Professional development: mentoring (seminars and on-going in classrooms)**	Professional outcomes: None Child outcomes: Collectively termed as FACES outcomes: Peabody Picture Vocabulary Test (3rd ed.), Woodcock-Johnson Psychoeducational Battery, Leiter-R Attention Sustain Scale, McCarthy Dra a design scale, Story and Print concepts. In addition to this: The Preschool Learning Behaviours Scale (PLBS), The Language and Emergent Literacy Assessment (LELA)

Author, date, country, title	**Design**	**Topic (stage of schooling, participant characteristics)**	**Aims of PD**	**Intervention** (name and **description of PD format**)	**Outcome and outcome measures (professional and child)**
Garet et al., 2008, US The Impact of Two Professional Development Interventions on Early Reading Instruction and Achievement.	RCT	LANGUAGE AND LITERACY DEVELOPMENT (PRIMARY)	To improve the knowledge and practice of teachers and the reading achievement of their students in high-poverty schools	Language Essentials for Teachers of Reading and Spelling (LETRS) by Louisa Moats (2005) **Professional development: Institutes and seminars and resources; coaching**	Professional outcomes excluded (not validated) Child outcomes: "Students' reading achievement. The key measure was the standardised total reading score obtained from the district assessments." p. 18.
Hindman & Wasik 2012, US Unpacking an Effective Language and Literacy Coaching Intervention in Head Start: Following Teachers' Learning over Two Years of Training	RCT	LANGUAGE AND LITERACY DEVELOPMENT (PRE-KINDERGARTEN)	To develop teachers' skills in language and literacy development and instruction	Exceptional Coaching for Early Language and Literacy (ExCELL) **Professional development: training by expert coaches over 2 years**	Professional outcomes: The quality of the language/literacy environment was measured using the Literacy Environment Checklist of the Early Language and Literacy Classroom Observation (ELLCO; Smith, Dickinson, Sangeorge, & Anastasopoulos, 2002) Classroom interaction quality was observed using the Instructional Support subscale of the Classroom Assessment Scoring System (CLASS; Pianta, La Paro, & Amre, 2006) Child outcomes: The Peabody Picture Vocabulary Test-III (PPVT; Dunn & Dunn, 1998) Among 4-year-old children, letter recognition was measured using the Uppercase Alphabet subtest of the Phonological Awareness Literacy Screening (PALS-PreK; Invernizzi, Sullivan, Meier, & Swank, 2004) Phonological skills were assessed using the Rhyme and Beginning Sound Awareness subtests of the PALS-PreK
Howlin et al., 2007, UK The effectiveness of Picture Exchange Communication System (PECS) training for teachers of children with autism: a pragmatic, group randomised controlled trial	RCT	LANGUAGE AND LITERACY DEVELOPMENT (AUTISTIC CHILDREN)	To train teachers in the use of the Picture Exchange Communication System and help in dealing with autistic children	**Training in Picture Exchange Communication (PECS)** **Professional development: workshop with expert consultants and feedback and monitoring**	Professional outcomes: None Others: Scores on formal language tests: expressive and receptive language

Author, date, country, title	**Design**	**Topic (stage of schooling, participant characteristics)**	**Aims of PD**	**Intervention** (name and **description of PD format**)	**Outcome and outcome measures (professional and child)**
Kammermeyer et al., 2016, Germany Promotion of literacy and numeracy in pyramid classrooms in Germany	RCT	LANGUAGE AND LITERACY DEVELOPMENT (PRESCHOOL)	To improve teachers' abilities to improve students' learning in literacy and maths	The Pyramid Approach The Kindergarten of the Future in Bavaria Approach The Letter & Number World Approach **Professional development: training and materials**	Professional outcomes: None Child outcomes: To measure literacy and numeracy achievement wortgewandt & zahlenstark ('eloquent & mathematically strong') von Moser and Berweger (2007) was used: two measures of literacy skills, phonological awareness (58 items) and letter knowledge and first reading (66 items), as well as one measure of maths skills, numeric-mathematical skills (120 items). To measure vocabulary, the wortgewandt ('eloquent') part of the test of Moser and Berweger (2007) with 37 items was used. This subtest is limited to expressive vocabulary and tests nouns, verbs and adjectives. Non-verbal intelligence was measured with the CPM (Becker, Schaller, & Schmidtke, 1980; Raven, Raven, & Court, 2010), a paper-and-pencil test in the German version, and standardisation from Bulheller and Häcker (Raven et al., 2010) served as a control variable
Landry et al., 2009, US Effectiveness of Comprehensive Professional Development for Teachers of At-Risk Preschoolers	RCT	LANGUAGE AND LITERACY (PROGRESS MONITORING, PRESCHOOLERS, AT RISK)	To promote effective instructional practices in order to improve children's linguistic skills	'Comprehensive Professional Development' **Professional development: mentoring and feedback**	Professional: Excluded (experimenter designed) Other: P. 453: "The Expressive One-Word Picture Vocabulary Test (EOWPVT; Brownell, 2000) was used to measure children's expressive vocabulary." P. 454: "The Auditory Comprehension subtests of the English and Spanish versions of the Preschool Language Scale – fourth edition (PLS-4; Zimmerman, Steiner, & Pond, 2002) were used to assess complex receptive language." English and Spanish versions of the Developing Skills Checklist. P. 454: "The Preschool Comprehensive Test of Phonological Processing (PCTOPPP)" Rashotte, 1999).

Author, date, country, title	**Design**	**Topic (stage of schooling, participant characteristics)**	**Aims of PD**	**Intervention** (name and **description of PD format**)	**Outcome and outcome measures (professional and child)**
McCollum et al., 2013, US Coaching Teachers for Emergent Literacy Instruction Using Performance-Based Feedback	RCT	LANGUAGE AND LITERACY DEVELOPMENT (PRESCHOOL)	To help teachers use research-based instruction to teach emergent literacy skills to young children	Performance Based Feedback (Coaching) **Professional development: coaching and classroom support**	Professional: ELLCO: "This research version of the ELLCO includes four sections: a Learning Environment Checklist (LEC); two sets of one to five qualitative rating items grouped into General Classroom Environment (GCE) and Language, Literacy, and Curriculum (LLC); and a Literacy Activities Rating Scale (LARS) to record the frequency and duration of nine literacy behaviors." (p. 33). Child outcomes: none
Neuman & Cunningham, 2009, US The Impact of Professional Development and Coaching on Early Language and Literacy Instructional Practices	RCT	LANGUAGE AND LITERACY DEVELOPMENT (KINDERGARTEN)	To improve early childhood educators' language and literacy instructional practices and child outcomes in high-priority urban areas, serving the very poorest children in Michigan's poorest cities	Language and literacy course **Professional development: year-long coaching intervention**	Professional outcomes: "Teacher Practice: We used two measures to assess the quality of language and literacy practices in center-based and home-based care settings: The ELLCO (Smith & Dickinson, 2002) and the Child/Home Early Language and Literacy Observation (CHELLO; Neuman, Dwyer, & Koh, 2007), which was specially developed to measure home-based practices. Both measures were based on the theoretical assumptions of ecological psychology (Bronfenbrenner, 1979), which attribute children's learning to the influences of the physical and the instructional supports in their environments." (p. 545). Child outcomes: none
Piasta et al., 2012, US Impact of professional development on preschool teachers' conversational responsivity and children's linguistic productivity and complexity	RCT	LANGUAGE AND LITERACY DEVELOPMENT (PRESCHOOL)	To promote the conversational responsivity of preschool educators working with children at risk of developmental and academic difficulties	Learning Language and Loving It approach **Professional development: workshop, coaching, feedback and self-reflection**	Professional outcomes: excluded (experimenter designed) Child outcomes: children's linguistic productivity and complexity using systematic analysis of language transcripts (SALT)

Author, date, country, title	**Design**	**Topic (stage of schooling, participant characteristics)**	**Aims of PD**	**Intervention** (name and **description of PD format**)	**Outcome and outcome measures (professional and child)**
Powell et al., 2010, US Effects of an early literacy professional development intervention on head start teachers and children	RCT	LANGUAGE AND LITERACY DEVELOPMENT (PRE-KINDERGARTEN)	To improve teachers' use of evidence-based literacy instruction leading to improvements in children's literacy achievement	Classroom Links to Early Literacy **Professional development: workshop and expert coaching**	Professional outcomes: Early Language and Literacy Classroom Observation (ELLCO). The General Classroom Environment (six items: e.g. organisation, management) and the Language, Literacy, and Curriculum (five items, e.g. oral language facilitation, approaches to book reading) subscales were used Child outcomes: Peabody Picture Vocabulary Test – Third Edition (PPVTIII). Woodcock-Johnson III Tests of Achievement – Letter Word Identification. Concepts About Print (Clay, 1985) measure. Letter Naming assessment employed in Head Start's Family and Child Experiences Survey. Writing (of child's name): (not included in d/e) Prepublication version of the Test of Preschool Early Literacy (Lonigan, Wagner, & Torgesen, 2007). The Alliteration Individual Growth and Development Indicator from Get it, Got it, Go! (Center for Early Education and Development, University of Minnesota, 2005)
Saraniero Patricia et al., 2014, US Unlocking My Creativity": Teacher Learning in Arts Integration Professional Development	RCT	LANGUAGE AND LITERACY DEVELOPMENT (PRIMARY)	To increase teacher confidence, use and frequency of arts integration	DREAM Developing Reading Education with Arts Methods **Professional development: summer institute with or without instructional coaching**	Professional: Teacher Participant Pre-/Post-Survey – demographic info survey. Post-institute survey: assessed teacher knowledge and confidence in arts instruction, arts state standards, arts integration and institute feedback. Lesson plan work samples: Arts integration lesson used in reading instruction. Scored with project-designed rubric by project evaluator. Focus groups and Mid-year Survey-examined impact of intervention on teaching and learning. Other: California Standards Test Language Arts test: state-administered normed test in English language arts

Author, date, country, title	**Design**	**Topic (stage of schooling, participant characteristics)**	**Aims of PD**	**Intervention** (name and **description of PD format**)	**Outcome and outcome measures (professional and child)**
Scanlon et al., 2008, US Reducing the incidence of early reading difficulties: Professional Development for classroom teachers versus direct interventions for children	RCT	LANGUAGE AND LITERACY DEVELOPMENT (KINDERGARTEN)	To develop teachers' knowledge in order to enable them to more fully understand their students' needs	The Interactive Strategies Approach (ISA) **Professional development: training**	Professional: Classroom language arts systematic sampling and instructional coding (CLASSIC) system Other: PALS-K version of phonological awareness and literacy screening battery Basic reading skills cluster (pp. 348-9)
Schwanenflugel et al., 2010, US PAVEd for Success: An Evaluation of a Comprehensive Preliteracy Program for Four-Year-Old Children	QED (RCT)	LANGUAGE AND LITERACY DEVELOPMENT (PRESCHOOL)	To support the teachers' development of preliteracy skills in 4-year-olds	PAVEd for Success **Professional development: institute; workshops; classroom-based support**	Professional outcomes: General classroom quality using the Revised Early Childhood Environmental Rating Scale (ECERS-R; Harms, Clifford, & Cryer, 1998), a global assessment of overall quality in child care and early education environments; Early Language and Literacy Classroom Observation (ELLCO; Smith & Dickinson, 2002). Child outcomes: Early literacy indicators and early decoding skills Kindergarten DIBELS. The kindergarten DIBELS (Good & Kaminski, 2002; Kaminski & Good, 1998) included Letter Naming, Phoneme Segmentation and Nonsense Word Fluency outcome assessments
Vernon-Feagans et al., 2013, US Live Webcam Coaching to Help Early Elementary Classroom Teachers Provide Effective Literacy Instruction for Struggling Readers: The Targeted Reading Intervention	RCT	LANGUAGE AND LITERACY DEVELOPMENT (TECHNOLOGY-BASED LITERACY TRAINING: KINDERGARTEN/PRIMARY)	To provide teachers with feedback in their classrooms as the teachers instruct struggling readers to improve the teachers' literacy teaching	Targeted Reading Intervention (TRI) **Professional development: webcam coaching and website materials; coaching sessions**	Professional outcomes: none Other: Word Attack, Letter Word Identification, Passage Comprehension, Spelling of Sounds and PPVT (Peabody Picture Vocabulary Test)
Wasik & Hindman, 2011, US Improving Vocabulary and Pre-Literacy Skills of At-Risk Preschoolers Through Teacher Professional Development	RCT	LANGUAGE AND LITERACY DEVELOPMENT (PRESCHOOL)	To provide teachers with conceptual knowledge and instructional strategies to enable them to support young children's development of vocabulary, alphabet	Exceptional Coaching for Early Language and Literacy (ExCELL) **Professional development: intensive and ongoing; resources**	Professional outcomes: Early Language and Literacy Classroom Observation (ELLCO; Smith, Dickinson, Sangeorge, & Anastasopoulos, 2002), Classroom Assessment Scoring System (CLASS; Pianta, La Paro, & Amre, 2007)

Author, date, country, title	Design	Topic (stage of schooling, participant characteristics)	Aims of PD	Intervention (name and description of PD format)	Outcome and outcome measures (professional and child)
			knowledge and phonological sensitivity		Child outcomes: Vocabulary: The Peabody Picture Vocabulary Test-III (PPVT; Dunn & Dunn, 1997) measures receptive language skills with strong psychometric properties. Preliteracy skills. Both alphabet knowledge and phonological skills were assessed using subtests of the Phonological Awareness Literacy Screening (PALS-PreK; Invernizzi, Sullivan, Meier, & Swank, 2004). Phonological awareness skills were measured by the Rhyme Awareness and the Beginning Sound Awareness subtests of the PALS-PreK. Alphabet knowledge. For pre-kindergarteners, letter recognition was measured using the upper and lowercase alphabet subtests of the PALS-PreK.
OVERARCHING TOPIC: SOCIAL DEVELOPMENT					
Allen et al., 2011, An interaction-based approach to enhancing secondary school instruction and student achievement	RCT	TEACHER-CHILD INTERACTIONS (ADOLESCENTS)	To improve teacher student interactions in the classroom by enhancing the fit between teacher-student interactions and adolescents' developmental, intellectual and social needs	The My Teaching Partner – Secondary program (MTP-S) **Professional development: integrates initial workshop-based training, an annotated video library and a year of personalized coaching followed by a brief booster workshop**	Professional: excluded Other: "the Virginia state standards assessment instrument applicable to the course being taught (20, 21)" (p. 3)
Domitrovich et al., 2009, US Fostering High-Quality Teaching With an Enriched Curriculum and Professional Development Support: The Head Start REDI Program	RCT	SOCIAL DEVELOPMENT (PRE-K)	To improve the quality of teachers' language use, emotional support, and positive behaviour-management strategies in the classroom	Head Start REDI (Research-based Developmentally Informed) **Professional development: resources; workshops and mentoring; coaching**	Professional: *Classroom Assessment Scoring System* (CLASS; La Paro & Pianta, 2003) Other: "the Virginia state standards assessment instrument applicable to the course being taught (20, 21)" (p. 3)
Early et al., 2017, US Improving teacher-child interactions: A randomized controlled trial of Making the Most of Classroom Interactions and My Teaching Partner professional development models	RCT	SOCIAL DEVELOPMENT (TEACHER-CHILD INTERACTIONS PRESCHOOL)	To strengthen teacher-child interactions in preschool classrooms	Making the Most of Classroom Interactions and My Teaching Partner **Professional development: teaching instruction and support or coaching (video recording and feedback)**	Professional: Teacher-child interactions: Classroom Assessment Scoring System (CLASS; Pianta, La Paro, & Hamre, 2008): 3 measures

Author, date, country, title	**Design**	**Topic (stage of schooling, participant characteristics)**	**Aims of PD**	**Intervention** (name and **description of PD format**)	**Outcome and outcome measures (professional and child)**
Fukkink & Tavecchio, 2010, Netherlands Effects of Video Interaction Guidance on early childhood teachers	RCT	TEACHER-CHILD INTERACTIONS (PRE-KINDERGARTEN)	To improve interaction skills of early educators Feedback via discussion about videos	Video Interaction Guidance training **Professional development: video interaction guidance**	Professional: Caregiver interaction scale (Arnett, 1989): stimulating and authoritarian caregiving behaviour – validated and delivered by assessors. "The 'Job Resources' scale was selected from a work-satisfaction instrument that was developed by Curbow et al. (2000). These measures focus primarily on 'emotional fulfilment from the job due to relationships with the children and parents, seeing the growth in children, and feeling like the work is supported and is important" (p. 1655). VIG job satisfaction scale: experimenter designed. Other: none
Jensen et al., 2015, Denmark Does Professional Development of Preschool Teachers Improve Child Socio-Emotional Outcomes?	RCT	SOCIO-EMOTIONAL DEVELOPMENT (KINDERGARTEN)	To develop the teachers' provision of learning opportunities for the individual child in ways that match the child's skill level	VIDA.2* **Professional development: training**	Professional: Other: Children's socioemotional skills measured by preschool teachers' assessment of each child using the Strengths and Difficulties Questionnaire (SDQ) (Goodman, 1997). 1) emotional symptoms, 2) conduct problems, 3) hyperactivity, 4) peer relationship problems and 5) prosocial behaviour. In addition, the total SDQ score, which is obtained as the sum of the first four subscales, and the SDQ impact score, which is obtained from the impact supplement of SDQ. This latter score reflects whether the preschool teacher thinks the child has any emotional or behavioural difficulties and, if so, to what extent these difficulties result in distress and social impairment. High reliability and validity "that the psychometric properties of the SDQ are strong, particularly for the teacher version." (p. 29).

Author, date, country, title	**Design**	**Topic (stage of schooling, participant characteristics)**	**Aims of PD**	**Intervention** (name and **description of PD format**)	**Outcome and outcome measures (professional and child)**
Murray et al., 2012, US Grade Level Effects of the Incredible Years Teacher Training Program on Emotion Regulation and Attention	RCT	ACADEMIC AND PRO-SOCIAL DEVELOPMENT (K-2 STUDENTS)	To improve academic and socio-emotional outcomes for K-2 grade students	IYT (Incredible Years Teacher Classroom Management Programme) **Professional development: training; video modelling, behavioural rehearsal of key skills through numerous role plays, classroom practice assignments, and teacher goal setting and self-monitoring.**	Professional: Observational change in teacher practices was assessed with the CLASS (Pianta & Hamre, 2005) and the Teacher Coder Impressions Inventory (TCI; Webster-Stratton, Reid, & Hammond, 2001). The CLASS is a multi-dimensional standardised instrument that has been widely used in early education classrooms and has been associated with variation in students' achievement and social adjustment. Subscales examined for this study include Positive Climate, Negative Climate, and Behaviour Management. Other: student socio-emotional and behavioural outcomes 'Inattention' (Revised Teacher Social Competence scale (R-TSC; Conduct Problems Prevention Research Group, 1995)) Conners' DSM-IV Inattention scale (Conners, 2001). Academic outcomes: Academic Competence subscale of the R-TSC and the Star Early Literacy/Reading and Math computerized assessment (STAR) Academic Competence scale
Ottmar et al., 2013, US Does the Responsive Classroom Approach Affect the Use of Standards-Based Mathematics Teaching Practices?: Results from a Randomized Controlled Trial	RCT	SOCIO-EMOTIONAL DEVELOPMENT (PRIMARY SCHOOL)	To help teachers create a supportive and safe environment for learning	Responsive Classroom (RC) approach **Professional development: training; coaching**	Professional: Mathematics Teacher Efficacy Beliefs Inventory MTEBI Mathematical Knowledge for Teaching Assessment (MKTA). The other teacher measures are author developed (M-Scan, C-POM and FOI (fidelity of implementation)). Other: none
Pianta et al., 2008b, US Effects of web-mediated professional development resources on	RCT	TEACHER-CHILD INTERACTIONS (PRESCHOOL)	To provide feedback to teachers about their emo-	MyTeachingPartner	Professional: Classroom Assessment Scoring System (CLASS; Pianta, La Paro, & Hamre, 2008)

Author, date, country, title	**Design**	**Topic (stage of schooling, participant characteristics)**	**Aims of PD**	**Intervention** (name and **description of PD format**)	**Outcome and outcome measures (professional and child)**
teacher-child interactions in pre-kindergarten classrooms			tional, organizational and instructional interactions with students	**Professional development: video exemplars and consultation process**	Other: Pre-K version of the Phonological Awareness Literacy Screening (PALS; Invernizzi, Sullivan, Meier, & Swank, 2004)
Raver et al., 2008, US Improving preschool classroom processes: Preliminary findings from a randomized trial implemented in Head Start settings	RCT	CLASSROOM MANAGEMENT (PRESCHOOL)	To improve teachers' ability to provide positive emotional support and well-structured classroom management to their classrooms	Chicago School Readiness Project (CSRP) **Professional development: Training and coaching**	Professional: "Classroom Assessment Scoring System (CLASS; La Paro, Pianta, & Stuhlman, 2004) and the Early Childhood Environment Rating Scale, revised edition (ECERS-R; Harms, Clifford, & Cryer, 2003)." (p. 8). Other: none
Rubie-Davies & Rosenthal, 2016, New Zealand Intervening in teachers' expectations: A random effects meta-analytic approach to examining the effectiveness of an intervention	RCT	TEACHER-CHILD INTERACTIONS (GRADES 2-10)	To offer teachers strategies and technique to improve teachers' expectations of students to ensure that all teachers have high expectations to their students.	Complex intervention **Professional development: workshops**	Professional: Other: Student achievement "the e-asTTle programme is an online assessment tool that has been developed primarily for the assessment of students in Grades 2-10. Teachers use the program to create tests specific to their needs." (p. 86).
OVERARCHING TOPIC: OTHER					
Azkiyah et al., 2014, Indonesia The effects of two intervention programs on teaching quality and student achievement	RCT	TEACHING QUALITY (SECONDARY)	To use education standards and a teacher improvement intervention to improve teaching quality	Improving teacher quality **Professional development: course**	Professional outcomes: Teacher quality: classroom observations 52 items scored on a 5-point Likert scale to indicate frequency and quality of items observed Child outcomes: Pre- and post-tests in reading comprehension test developed from the standards test (modified). 20 items

Author, date, country, title	**Design**	**Topic (stage of schooling, participant characteristics)**	**Aims of PD**	**Intervention** (name and **description of PD format**)	**Outcome and outcome measures (professional and child)**
Flook et al., 2013, US Mindfulness for Teachers: A Pilot Study to Assess Effects on Stress, Burnout, and Teaching Efficacy	RCT	STRESS REDUCTION (TEACHERS)	To reduce stress in teachers	Mindfulness-Based Stress Reduction course **Professional development: course**	Professional: Teacher Classroom Behaviour – An observational coding system was used to assess teachers' behaviour in the classroom. The Classroom Assessment Scoring System (CLASS; LaParo, Pianta, & Stuhlman, 2004) measures three dimensions of teaching behaviour in the classroom: emotional support, classroom organisation and instructional support. Cortisol Measurement – Saliva samples were collected to measure cortisol (nmol/L) over a span of three consecutive working days. Neuropsychological and attentional tasks – The Cambridge Neuropsychological Test Automated Battery (CANTAB, 1999) is a computerised neurocognitive battery that was used to assess sustained attention and affective attentional bias Other: none
McMeeking et al., 2012, US Effects of a Teacher Professional Development Program on the Mathematics Achievement of Middle School Students	QED	MATHEMATICS DEVELOPMENT (MIDDLE SCHOOL)	To improve multiple facets of teachers' teaching competencies, e.g. problem-solving skills, new techniques for enquiry-based learning and building critical thinking skills	The Rocky Mountain Middle School Math and Science Partnership (RMMSMSP) **Professional development: summer institutes; follow-up institutes**	Professional outcomes: none Child outcomes: Colorado Student Assessment Program (CSAP)

Note * "VIDA is a Danish acronym for knowledge-based efforts for socially disadvantaged children in daycare" (Jensen et al., 2015: 27)

The studies focused on PD in a total of six topic areas, although most (29) were in two overarching topic areas: 20 in PD in language and literacy development interventions and outcomes, including one with a focus on just language and one with a focus on technology-based literacy development (Al Otaiba et al., 2011; Bos et al., 1999; Buysse et al., 2010; Cabell et al., 2011; Dickinson & Caswell, 2007; Gallagher et al., 2011; Garet et al., 2008; Hindman & Wasi, 2012; Howlin et al., 2007; Kammermeyer et al., 2016; Landry et al., 2009; McCollum et al., 2013; Neuman & Cunningham, 2009; Piasta et al., 2012; Powell et al., 2010; Saraniero et al., 2014; Scanlon et al., 2008; Schwanenflugel et al., 2010; Vernon-Feagans et al., 2013; Wasik & Hindman, 2011); and 10 in PD in social development interventions and outcomes (Allen et al., 2011; Domitrovich et al., 2009; Early et al., 2017; Fukkink & Tavecchio, 2010; Jensen et al., 2015; Murray et al., 2012; Ottmar et al., 2013; Pianta et al., 2008; Raver et al., 2008; Rubie-Davies & Rosenthal, 2016). Four areas in social development were evaluated: social and emotional development; academic and pro-social development; classroom management; and teacher-child interactions. The remaining three studies looked at PD in other topics: mathematics development; teaching quality; progress monitoring; and stress reduction (Azkiyah et al., 2014; Flook et al., 2013; Sample et al., 2012) (see Table 7).

Table 7: Topics of studies included in the mapping

Topic	Number of studies
LANGUAGE AND LITERACY DEVELOPMENT	
Language and literacy development	18
Technology-based literacy development	1
Language development	1
SUBTOTAL:	20
SOCIAL DEVELOPMENT	
Socio-emotional development	3
Academic and pro-social development	1
Classroom management	1
Teacher-child interactions	5
SUBTOTAL:	10
OTHER TOPICS	
Mathematics development	1
Teaching quality	1
Stress reduction	1
SUBTOTAL:	3
OVERALL TOTAL	**33**

As Table 6 shows, 27 studies were undertaken in the US, and only one was study undertaken in each of the following countries: New-Zealand, Germany, the UK, Netherlands, Indonesia and Denmark. The dominance of the US as the main country in which PD interventions meeting our criteria have been evaluated using rigorous methods, and within our specific parameters, clearly limits the generalisability of the findings of the mapping for the Danish context. Research demonstrating (some degree of) effectiveness in the setting in which the intervention was developed, tested and evaluated cannot necessarily be generalised to other contexts. According to Gardner, Montgomery, & Knerr (2016), there is a growing literature on the topic of transferability of effective interventions from one cultural and structural context to another. Cultural norms, family and societal values, educational structures and political priorities will all influence the acceptability and effectiveness of attempts to 're-plant' specific interventions in a context other than the one in which they were originally 'grown'. Such differences are important, when considering the relevance of and potential for transferring interventions from one setting to another; yet

it is also important to look for commonalities that may indeed facilitate the process Gardner, Montgomery, & Knerr (2016).

The professional participants in the evaluations of PD interventions were exclusively preschool teachers ('pedagogues') and teachers. The other participants were exclusively children and young people attending preschool (including day care), kindergarten (nursery) or school, with the majority of studies focusing on outcomes for very young, at-risk children (low socio-economic status or struggling to acquire language and literacy skills) attending preschool (from age 0) or kindergarten (22 studies). The remaining 10 studies focused on interventions evaluated in elementary (primary/middle) school settings and one in a lower secondary school setting.

When judging the relevance of the settings in which the interventions have taken place in the Danish context, it should be noted that the level of education and the staff composition may differ somewhat. In Denmark, pedagogues and teachers have professional bachelor's degrees, and it is relatively uncommon to have employees with a master's degree or a PhD in day care or school settings. Moreover, in Danish day care institutions 6 out of 10 staff members are pedagogues, while the remaining staff members are unskilled assistants. Another structural difference is that, in Denmark, it is pedagogues who teach school children in their first (reception) year at school (termed 'kindergarten class'). Moreover, as mentioned earlier, the reform of the Danish *Folkeskole* has given pedagogues an even more significant role in schools at higher grade levels also.

Studies focusing on very young, at-risk children have high relevance to the Danish context in that there is a focus on early identification and early intervention, in Denmark. However, having low socio-economic status in the US differs greatly from having low socio-economic status in Denmark. In the US studies, children from low-socioeconomic status families are typically identified based on whether or not they are eligible for free or reduced-price meals at school. Regarding identification of students struggling with acquiring language and literacy skills, there is a language skills screening programme in Denmark that is offered to all 3-year-olds. Additional attention is paid to children for whom Danish is not their mother tongue, in terms of assessing whether a specific intervention is necessary or whether a close follow-up is sufficient.

Another obvious difference between Denmark and the other country settings is the ethnic minority composition of the children and youth populations. In the US, attention is paid to the racial/ethnic composition (i.e. Caucasian, African-American and Hispanic). In the New Zealand study, the focus is on the Maori population and the ethnic minorities originating from the Pacific Islands and Asia. In Denmark, the ethnic minorities of students in grade 9 are mainly from Turkey, Iraq and Lebanon (1st and 2nd generation immigrants). Other minority populations are from countries such as Pakistan, Somalia, Afghanistan and the former Yugoslavia.

6.4.1 Aims of PD

Common features of the *language and literacy* PD focused on: developing teachers' knowledge and understanding in the substantive fields of reading and writing development (in two cases explicitly using evidence from research). Specifically, PD aimed to develop teachers' instructional strategies, methods and techniques (in the substantive area); teachers' abilities to differentiate or individualise instruction; teachers' abilities to support children generally in their language and literacy development; teachers' confidence and their abilities to interact responsively with the children; and, finally, to fill in the gaps in teachers' conceptual knowledge and understanding.

Common features of the *social development* PD focused on: developing teachers' language use, emotional support and positive behaviour-management strategies in the classroom; strengthening teachers'

interactions with the children; individualising responses to children and improving teacher-child interactions; improving classroom management skills and creating positive, supportive learning environments; and generally developing teachers' abilities to increase their expectations of children and young people.

Other areas of PD covered in single studies aimed at reducing teacher stress[7], increasing teacher quality and enabling teachers to develop critical thinking skills in the children and young people.

The common focus of the *language and literacy* PD studies on developing teachers' competencies related to the substantive fields of reading and writing is relevant for the Danish context. Embedded in the agreement from 2014 to reform the public school was an agreement to ensure that *all* students are taught by teachers with subject-specific teaching competencies. This goal is to be met by the year 2020, and hence teachers without the necessary competencies will be required to participate in relevant PD. However, a recent overview of the degree of competency coverage by subject shows that the greatest gap is not in the subject Danish (language and literacy) (Bjørnholt et al., 2017). Gaps are substantially larger in other subjects, such as Christianity (religion), Science (primary levels) and History.

PD focused on generally supporting children in their language and literacy development is clearly relevant for the Danish context – particularly in day care settings (age 0-6) and in areas where there are many families with ethnic minority background[8] or families able to support their children in their language and literacy development. Interventions that are more generally oriented towards filling in gaps in teachers' conceptual knowledge and understanding would be relevant for the Danish context, as long as such gaps are adequately identified and subsequently addressed by the offered PD.

Focus on PD aimed at enhancing teachers' abilities to differentiate or individualise instruction is clearly relevant for the Danish context. This is an aspect that Danish teachers find particularly challenging, and the task has become more complex in recent years, the political aim being to include as many children with special/additional needs in mainstream schools as possible. Consequently, inclusion and individualised/differentiated teaching are both a key focus in the professional development agreed on under the 2014 reform for the *Folkeskole*.

Also relevant for the Danish context is the common focus of the *social development* PD studies on teachers' interactions with the children, emotional support and the development of supporting learning environments. One of the main aims of the 2014 public school reform is to reduce the influence of problematic family backgrounds on students' academic performance. One mechanism through which this goal could be achieved is to enhance teachers' abilities to interact positively with *all* children. This includes having high expectations of all students, irrespective of their family background. Social development PD may also be relevant, since several studies find an increasing prevalence of teenagers with poor mental well-being. Creating a positive class environment/climate may help promote positive mental health.

In Denmark, a lot of attention has been paid to disruptive behaviour in classrooms. Hence, PD focused on increasing teachers' classroom management skills and improving their abilities to use positive behaviour management strategies is highly relevant.

6.4.2 Interventions

A wide range of interventions were evaluated. In the *language and literacy* topic area, sixteen studies evaluated a number of 'branded' interventions, for example: Project RIME; Learning Language and

[7] Teacher stress falls under our included outcomes of interest, since it can adversely affect teacher performance and student outcomes, and is distinct from aspects related to, for example, teacher job satisfaction, the outcome of which is not of interest for this particular SR.

[8] In Denmark, the performance gap between ethnic Danish students and 2nd generation immigrants at age 15 is larger the average OECD performance gap, and there does not seem to have been substantial reductions in this gap over the recent years (Greve & Kressel, 2017).

Loving It (two studies); LEEP Effects of the Literacy Environment Enrichment Program; Exceptional Coaching for Early Language and Literacy (ExCELL) (two studies); PAVEd for Success (see Table 6). Similarly, in the social development topic area a number of 'branded' interventions were evaluated, including: Head Start (classroom management strategy development); Making the Most of Classroom Interactions and My Teaching Partner; Incredible Years Teacher Classroom Management Programme; Responsive Classroom and Chicago School Readiness Project (CSRP). Also evaluated were video recording of classroom interactions and feedback and evidence-based strategies to improve teacher expectations of students.

In response to a questionnaire[9], about professional development among teachers and pedagogues in Denmark, 41% of teachers and 54% of pedagogues responded that the most recent PD activity in which they had participated concerned a pedagogical theme, e.g. inclusion, relational competencies or children's crying (Danmarks Evalueringsinstitut, 2013). For both groups of professionals, training in specific methods and 'branded' interventions, such as PALS, the LP-model[10] and courses on cooperative learning, was also common (26% and 25% of teachers and pedagogues, respectively). Training to take on a specific function, such as behaviour-counsellor ('AKT-vejleder') or language-pedagogue, was what 18% and 24% of teachers and pedagogues received as their most recent PD. While 27% of teachers had received subject-specific PD and didactics, this was only the case for 7% of pedagogues. PD related to collaboration (e.g. team collaboration) had been attended by 12% of teachers and 18% of pedagogues. The share of teachers and pedagogues that had attended other types of PD was 13% and 14%, respectively.

6.4.3 Outcomes

Professional outcomes: In the language and literacy topic area, improvement in the quality of language and literacy environment, teaching and instruction was measured using: The Early Language and Literacy Classroom Observation (ELLCO) Toolkit (Smith & Dickinson, 2002) (7 studies); Classroom interaction quality was observed using the Instructional Support subscale of the Classroom Assessment Scoring System (CLASS; Pianta, La Paro, and Hamre (2006) (3 studies); Classroom language arts systematic sampling and instructional coding (CLASSIC) system (one study); and in one study only a global assessment of overall quality in the early education environment was made, using the Revised Early Childhood Environmental Rating Scale (ECERS-R; Harmes, Clifford, and Cryer (1998). Many of the studies in this topic area either did not assess professionals' outcomes at all or they did so using experimenter-designed or non-standardised outcomes. In the social development topic area, the quality of the classroom environment was assessed using the Classroom Assessment Scoring System (CLASS; La Paro and Pianta (2003)), five studies used CLASS (Domitrovich et al., 2009; Early et al., 2017; Murray et al., 2012; Pianta et al., 2008; Raver et al., 2008); teacher-child interactions were measured using CLASS (one study: Early et al., 217); and teacher efficacy in mathematics were assessed in one study (Ottmar et al., 2013) using the teacher Efficacy Beliefs Inventory (MTEBI) (Enochs, Smith, & Huinker, 2000) Mathematical Knowledge for Teaching Assessment (MKTA) (Hill, Schilling, & Ball, 2004).

The outcome measures used in these studies are relevant for the Danish context, yet it is important to acknowledge that the creation of a teacher-evaluation culture is relatively new in Denmark. Danish teachers increasingly work together in teams (subjects, classes) – not only to coordinate practical matters, but also with a view to coordinating, evaluating and improving their teaching practices. One way of doing this is by having colleagues or the school principal observe and provide feedback on teachers' teaching. Also, in the topic of social development the outcome measures are relevant for the Danish

[9] Note that respondents could tick more than one answer, so the percentages do not add up to 100%.

[10] The LP model is short for "Learning environment and Pædagogical analysis". It is an analytical model developed by the Norwegian professor Thomas Nordahl.

context. The CLASS outcome measure is already used by researchers in Denmark, but it would be important to ensure understanding among teachers of what this kind of measurement contributes. Educational research in Denmark tends to be dominated by qualitative measures, and there is scepticism – among some – towards quantifying what could be termed 'soft' outcomes.

Child outcomes: In the language and literacy topic area, standardised measures of language and literacy were used to observe improvement, for example the Woodcock Johnson tests of achievement III (two studies); Woodcock Language Proficiency Battery-Revised: English and Spanish Forms (WLPB-R; (Woodcock, 1991; Woodcock & Muńoz-Sandoval., 1995)). Peabody Picture Vocabulary Test (PPVT-III; Dunn and Dunn (1997)) (see Table 6). In the social development topic area, children's socio-emotional skills were measured by preschool teachers' assessment of each child, using the Strengths and Difficulties Questionnaire (SDQ) (Goodman, 1997) in one study, and student socio-emotional and behavioural outcomes were measured in one study using teacher ratings of emotion regulation, pro-social behaviour and inattention on the Revised Teacher Social Competence scale (R-TSC) (Conduct Problems Prevention Research Group, 1995) and the Conners' DSM-IV Inattention scale (Conners, 2001). In all other studies, children's socio-emotional outcomes were not assessed using standardised measures. However, in some cases academic outcomes of the child participants were assessed using standardised measures.

In Denmark, the SDQ is a well-known and widely used instrument for assessing socio-emotional problems among children at all ages. It is both used by professionals and practitioners to identify children in need of support, to evaluate interventions that have been initiated, and it is also used by researchers. In fact, there is a non-profit secretariat in Denmark supporting the use of SDQ in the Danish setting (Danish Secretariate for SDQ and DAWBA in Denmark, 2014). The other child outcome measures would have to go through a translation, adaptation and validation process to ensure relevant use in Denmark, if this has not yet been done.

6.4.4 Other topic areas

Also included in the map, are three studies exploring diverse 'other' topics: mathematics development; stress reduction among teachers; and teaching quality. Each topic was evaluated by only one study (two of RCT design and one QED). Interventions ranged from workshops and monthly meetings to summer and follow-up institutes. Although the CLASS outcome measure was used to assess professional development in one of the RCTs, the other three studies employed measures that were unique to their study (with respect to this review). Therefore, we have recorded the mapping data extraction for these three studies in Table 6 but make no further comment due to the limitations of one study exploring each topic.

6.4.5 Quality

All of the included studies met a minimum threshold for quality, due to the inclusion criterion for this review specifying a design with either a randomised control or comparison group or a non-randomised control or comparison group with baseline equivalence on the primary predictor of outcome or on all pre-test measures. Three studies used a quasi-experimental design (QED); one included a combination of quasi-experimental and randomised designs; and the remainder (29) used the most robust design: randomised controlled trial (RCT) design.

However, the minimal quality appraisal undertaken for each study at mapping, in terms of noting any strengths and limitations, revealed variable quality in the included studies. For individual RCTs, sample size at allocation for teachers ranged from 13 teachers (McCollum et al., 2013) to 116 teachers (Saraniero et al., 2014), while for children the sample size ranged from 88 children (Howlin et al., 2007) to 1278 children (Rubie-Davies & Rosenthal, 2016). For cluster RCTs, the sample size for teachers ranged

from 30 (Wasik & Hindman, 2011) to 486 teachers (Early et al., 2017). For children, the allocation sample sizes ranged from 193 (Buysse et al., 2010) to 5,530 (Garet et al., 2008). As above, in some cases the sample size was very small. In RCTs, small sample sizes mean that chance imbalances are more likely to have an impact on the results. In some of the trials, the potential for bias in the results could not be ruled out due to post-randomisation or post-allocation bias being introduced by various factors. In addition, in some cases attrition was high to the extent that a potential for bias would have been introduced. Attrition affected both professionals and children. For example, attrition of professionals ranged from 1 participant (e.g. Howlin et al., 2007) to 99 participants (Garet et al., 2008); and attrition of children ranged from 1 participant (Saraniero, 2014) to 564 participants (McMeeking et al., 2012). Some lack of baseline equivalence was seen in some of the RCTs, and blinding for outcome ascertainment was not always stated or achieved. Conversely, some high-quality conduct was noted, for example clustering being taken into account in the analyses of cluster RCTs.

6.4.6 Results: Quality assurance (Mapping)

Screening at all stages was completed independently by two reviewers, and agreement to include or exclude was high. In cases of disagreement, agreement was achieved through discussion. Referring back to Table 3, we can see that 74 studies were excluded in third-stage (full-text) screening by way of quality assurance – 51% of the studies excluded in the third stage were excluded for reasons pertaining to the outcome measures reported. All four reviewers agreed on these exclusions. Data extraction for the 33 remaining studies included in the mapping was independent. Agreement was very high, and any differences were resolved by discussion and with occasional reference to a third reviewer. The 'run comparison' feature in EPPI was used to examine the independent data extractions for similarity and to ensure completeness of and sufficient detail in the data extraction. The final data extraction was agreed on by both reviewers before being marked as complete in EPPI-Reviewer.

7 Results and discussion: In-depth review on social and emotional development

7.1 Data extraction (study characteristics)

Table 8 and Table 9 present the study characteristics of the 9 out of 10 studies included in the in-depth review. Pianta et al. (2008) was not extracted for the in-depth review, as it did not contain a 'business as usual' control group. Six of the 9 studies were undertaken in the US, and one study was undertaken in each of the following countries: New Zealand, the Netherlands and Denmark. The settings ranged from preschool through elementary and secondary schools, with most being early childhood settings. Participants were teaching professionals and children and young people in these settings. Although there was some individual variation in the delivery models of the professional development (specifically in relation to dosage and timing), the basic components were very similar across all 9 trials and included the following components: workshop-based training with resources, personalised coaching/consultation using feedback on observations or video recordings of classroom practice, feedback and reflection. Control conditions were also very similar and comprised business-as-usual PD (sometimes with waiting list design). Note: Pianta et al. (2008) evaluated two slightly different PD interventions and had no business-as-usual control – see above.

Table 8: Study characteristics (included studies 1-5)

Study		1 – Rubie-Davies and Rosenthal (2016)	2 – Allen et al. (2011)	3 – Fukkink & Tavecchio (2010)	4 – Early et al. (2017)	5 – Murray et al. (2014)
Participants	**Country**	New Zealand	US	Netherlands	US	US
	Professional	Elementary school teachers	Secondary school teachers	Early childhood education and care teachers	Pre-K teachers	Early elementary teachers of grades K-2
	Other (e.g. students)	Elementary school students, grades 2-7	Secondary school students, aged 11-18	Children in day care centres (but not focus of study)	Preschool children, age 4	Early elementary students, grades K-2
Eligibility	**Professional**	Teachers working in recruited schools	Teachers working in recruited schools and able to select a focal course for study (p. 2 in SOM, Supplementary Online Material).	Not clear, but apparently teachers in 2 major day care providers SKON and KOREIN, p. 1654.	Preschool teachers in selected schools and classes, who were (i) NOT in their 1st year as a Georgia Pre-K teacher, (ii) would NOT be absent most of the year due to illness or pregnancy (p. 61).	Teachers working in selected schools. No additional details provided
	Other (e.g. students)	Students attending elementary schools at various socioeconomic levels in the city in which the study took place	Students in selected courses, whose parents had provided written consent. Students also provided written consent.	Children in aforementioned day care centres	Preschool children in the state of Georgia's Pre-K program.	Children in selected schools. No additional details provided
Intervention	**Setting**	12 elementary schools from a suburban area of a city in Auckland, New Zealand	12 secondary schools	Day care centres	Universal Pre-K programme, open to all 4-year-olds, in variety of settings (private child care, local schools, Head Start centres, etc.)	Elementary schools in rural school districts selected due to below average school resources and difficulty recruiting and retaining highly qualified teachers.
	Dosage	4 workshops: mid-March, end-March, mid-April, mid-May. After the workshops, researchers met with intervention teachers 3 further times to provide support and answer queries.	About twice a month, teachers sent in video recordings of class sessions. Trained teacher consultants reviewed these and followed up with 20-30 min phone conferences.	Typically four sessions of 10 min videotaping, followed by session with discussion of video clips selected by trainer.	Each teacher participated for 1 academic year. MMCI: 10 two-and-a-half-hour workshops delivered over five full-day sessions, spread across five months (Oct./Nov.-Feb./Mar.). One training day/month. Between ses-	4-6 monthly full-day workshops (5 in the present study). Average number of workshop training hours across the 5 training days was 34. Teachers also had two brief consultation visits (avg. 44 min. total).

Study		1 – Rubie-Davies and Rosenthal (2016)	2 – Allen et al. (2011)	3 – Fukkink & Tavecchio (2010)	4 – Early et al. (2017)	5 – Murray et al. (2014)
					sions, homework assignments. MTP: Sept.-Apr. Cycles of video, review, feedback takes 2 weeks. No pre-specified goal for number of sessions. Average was 7.57.	
	Length	One school year	One school year (13 months)	Not clear, only mention of retention measurement 3 months after intervention. Mention of 3 measurements in all: pre-, post-, retention (only for intervention group).	3 years (2011-12, 2012-13, 0213-14). New cohorts of teachers selected each year	1 year
	Description	Teacher expectation intervention. Through PD workshops teachers were trained in the practices of teachers who have high expectations for all students. Three key areas: grouping and learning experiences, class climate, and goal setting.	The My Teaching Partner–Secondary programme (MTP-S), a web-mediated approach, "targets the motivational and instructional qualities of teachers' ongoing, daily interactions with students." p. 2. Workshop-based training, an annotated video library, and 1 year of personalised coaching followed by a booster workshop.	"In the training, teachers were videotaped by their trainer for approximately 10 min while they are working with their groups. The trainer watched the video subsequently and selected a number of video fragments for review. In a next session, the trainer and the teacher subsequently engaged in a detailed discussion of these video clips." (p. 1654).	Two intervention conditions: 1) Making the Most of Classroom Interactions (MMCI): face-to-face PD, where small groups of teachers meet for instruction and support. Focus is to learn to identify and analyse effective interactions; 2) My Teaching Partner (MTP): one-to-one remote coaching model, where teachers work with coach using cycles of video-recorded observations of teaching, review, and feedback. Teachers receive specific feedback about emotional climate, organizational structure, instructional support based on videos through, for instance, conference	IYT is a teacher training programme that is part of a comprehensive series of interventions including parent, child and teacher training components that were designed to prevent and treat aggressive behaviour and conduct problems in young children aged 3-8 years. Its approach includes validated training methods such as video-modelling, behavioural rehearsal of key skills through numerous role plays, classroom practice assignments, and teacher goal setting and self-monitoring. Workshops cover building positive relationships with students and parents, proactive classroom management strategies, effective use of incentives, “coaching” students’ social and emotional

Study		1 – Rubie-Davies and Rosenthal (2016)	2 – Allen et al. (2011)	3 – Fukkink & Tavecchio (2010)	4 – Early et al. (2017)	5 – Murray et al. (2014)
					calls.	development, teaching calm-down and problem solving, and positive discipline techniques, such as redirection, ignoring and time out. Workshops are led by two trained co-leaders with approximately 12-15 teachers in each group
Control	**Description**	Regular professional development programme of their school. They also attended 3 workshops related to this programme (p. 78). Control group teachers were waitlisted for full participation in the following year.	Business-as-usual PD (SOM p. 2). Both I and C teachers participated in 3-hour workshop prior to beginning of school year and also prior to allocation.	Control stated but no details, assume business as usual	During first year of study, control group teachers had access to same online library of video clips demonstrating best practices in teacher-child interactions as MMCI and MTP teachers (no data about usage). In 2nd and 3rd years, control group teachers participated in same 15-hour PD required of all Georgia's Pre-K teachers.	Business as usual, but not explicated.

Table 9: Study characteristics (included studies 6-9)

Study		6 – Jensen et al. (2015)	7 – Domitrovich et al. (2009)	8 – Ottmar et al. (2013)	9 – Raver et al. (2008)
Participants	**Country**	Denmark	US	US	US
	Professional	Trained preschool staff (BA educated), also untrained preschool staff	Teachers and assistant teachers	Third grade teachers who taught mathematics in the 2008/9 school year	Teachers and teachers' aides in preschool Head Start settings
	Other (e.g. students)	Children in preschools	N/A – focus solely on professionals	Third grade students	Children in the preschool Head Start settings
Eligibility	**Professional**	Preschool staff working in the recruited settings	Working in preschools participating in Head Start in three counties in one state	Third grade teachers in schools willing to receive training in RC approach and participate in a research study	Teachers and teachers' aides working at eligible Head Start sites
	Other (e.g. students)	3-5-year-old children, focus on disadvantaged children	N/A – focus solely on professionals	Third-grade students in the participating schools	Children in preschool classes
Intervention	**Setting**	Preschools	Preschools (participating in Head Start programs)	Primary school (in one mid-Atlantic school district)	Preschools
	Dosage	17 days training across the two years	3-day workshop, weekly mentoring (on average 3 hours per week of observation too)	1 x 35 hour RC training institute, 2008-9 school year – one-day workshop, three consultations and classroom visits by RC coach	5 training sessions each lasting 6 hours; coaching
	Length	2 years	2 years	1 school year	1 school year (2 successive cohorts)
	Description	VIDA intervention. Three key elements – knowledge (provided by training programme), reflection (about practices) and activities (via reflection leading to new activities). Intervention trained teachers to work on an evidence-based platform. A focus point of the training was to improve children's skills by being responsive to their needs	REDI curriculum – Preschool PATHS curriculum – to promote children's socio-emotional skills. 3 day workshop – approx. 5 day general, 1 day language and literacy and 1 day with socio-emotional emphasis. Half way through the year – 1 day booster – brief review of the REDI curriculum. Weekly mentoring support for intervention teachers by local education consultants. Approx. 3 hours a week of REDI trainers observing and 1-	RC (Responsive Classroom) approach – to help "create a supportive and safe environment for learning" (p. 437). RC approach aiming to foster prosocial skills in the classroom. Focus on the process of learning (as opposed to only the product), critical social skills and teachers knowing individual characteristics and families of the children that they teach. Also, encouraging support from other adults in the community.	Treatment group teachers received: (1) training, (2) Mental Health Consultation (MHC) (p. 7.) A behaviourally and evidence-based teacher training package was selected and purchased, and a seasoned trainer with Licensed Clinical Social Worker (LCSW) qualifications delivered the 30 hours of teacher training during autumn and winter, adapting the Incredible Years teacher training module (Webster-Stratton, Reid, & Hammond, 2004).

Study		6 – Jensen et al. (2015)	7 – Domitrovich et al. (2009)	8 – Ottmar et al. (2013)	9 – Raver et al. (2008)
		and potentials. Principals also offered 2-day course and workshop about facilitating organizational learning processes in the preschools.	hour meeting per week with lead and assistant teacher.	Summer school with additional workshop and consultations and classroom visits (by RC coaches).	
Control	**Description**	Control condition not described, only "did not receive any training" (p. 26)	Head Start business as usual (i.e. without REDI) Including in-service training. In addition, each teacher worked with an assigned supervisor or mentor, who visited the classroom on a monthly basis and provided teachers with individual feedback." (p. 578). Summary of differences between REDI intervention and ordinary Head Start programme (p. 578-579).	P. 440: Business-as-usual approaches to social and emotional learning and/or classroom management. No training on RC approaches received by control teachers. Information was gathered by the researchers about what this comprised (including textbooks used, testing and guides for direct instruction).	P. 8: Business as usual plus teachers were given classroom support (teachers' aides) to ensure similar ratios of staff to children across intervention and control conditions.

7.2 Contextualisation

Considerations about the relevance and feasibility of specific aspects of the interventions in each of the eight studies included in the in-depth review (excluding the study undertaken in Denmark) are presented in Appendix D. Relevance has been judged on eleven different aspects, to ascertain the feasibility of conducting such a trial in Denmark. An overall judgment of relevance for the Danish context has been provided for each study according to the algorithm described in the design section. Three studies were judged to be of 'high' overall relevance to the Danish context (Allen et al., 2011; Domitrovich et al., 2009; Murray et al., 2012); two were judged to be of 'moderate-high' relevance (Early et al., 2017; Fukkink & Tavecchio, 2010), and three were judged to be of overall 'moderate' relevance to the Danish context (Ottmar et al., 2013; Raver et al., 2008; Rubie-Davies & Rosenthal, 2016).

In all cases but one, the topic relevance of the PD intervention for the Danish context was judged as 'high', focusing on improving teachers' skills with respect to creating positive teacher-student/child interactions (including having high expectations of all students), good classroom management, developing behaviour management skills and fostering a positive socio-emotional climate in classrooms. The topic of the study by Ottmar et al. (2013) was judged as being of 'moderate' relevance due to its focus on teachers' use of standards-based mathematics teaching practices, which could be seen as somewhat restrictive. This is because teachers and pedagogues in Demark generally enjoy a high degree of autonomy with regard to the methods they use in their work. Consequently, interventions that are strongly manual-based or otherwise leave limited flexibility for teachers/educators have been judged to be of 'moderate' (Ottmar et al., 2013) or 'moderate-high' relevance (Allen et al., 2011; Domitrovich et al., 2009; Early et al., 2017; Raver et al., 2008). By contrast, interventions which engage teachers by enabling them to contribute with their best situational judgment, are seen to be more relevant in working cultures with high degrees of teacher autonomy (Funkkink & Tavecchio, 2010; Murray et al., 2014; Rubie-Davies and Rosenthal, 2016).

The content of the PD in the in-depth studies has been judged as being of 'high' relevance for the Danish context in nearly all studies, since they are focused on training that has clear links to practice, e.g. teacher training packages with a focus on learning and implementing strategies for direct use in the classroom. The content of one study (Rubie-Davies and Rosenthal, 2016) has been judged as 'moderate-high' in terms of relevance due to its rather heavy focus on research-based knowledge as a significant part of the training. As Danish teachers are trained at university colleges, and their basic training is a professional BA degree, they may find this focus too academic/theoretical.

Apart from such differences in the level of education of teachers and pedagogues in Denmark compared with teachers and educators in the US, the Netherlands and New Zealand, differences also exist in the settings in which the trials were conducted. In studies where participating schools are a random sample with no specific characteristics, the main issue of relevance and feasibility concerns structural/systemic differences (e.g. Allen et al., 2011; Early et al., 2017) or ethnic composition (e.g. Rubie-Davies and Rosenthal, 2016). In other cases, the study is conducted in specific settings, such as Head Start classrooms in the US (e.g. Domitrovich et al., 2009; Raver et al., 2009). In the latter cases, adaptation would mean testing the intervention in similar settings, for instance low-income school districts in larger cities such as Copenhagen, Aarhus and Odense.

Whether the *PD participation* is mandatory or voluntary is also an aspect to be considered in terms of contextualisation. In Denmark, municipalities often set overall aims for the type of PD in which teachers and pedagogues are expected to participate – in some cases at the topic level, in others with regard to specific methods. However, school management is almost always closely involved in the final decision

regarding the types of PD activities in which staff at their particular school/day care institution are engaged. Staff members are typically also involved in a discussion of which types of PD would be relevant for them and the school/institution. Three of the studies (Murray et al., 2014; Ottmar et al., 2013; Raver et al., 2009) invited teachers to participate. In three other studies, the decision to participate was made at the school level, and teacher participation was mandatory (Allen et al., 2011; Early et al., 2017; Rubie-Davies & Rosenthal, 2016), while in Domitrovich et al. (2009) and Funkkik & Tavecchio (2010) information on participation was unclear.

Practical aspects of conducting the intervention trials were judged as being 'high' in relevance, with two exceptions. The intensive training involved in the intervention studied by Ottmar et al. (2013) might be a barrier. The scheduling of trainings in the weekends (with financial compensation) in the Raver et al. (2009) study is not in line with common practice in Denmark, where PD typically takes place during ordinary working hours, with teachers receiving no financial compensation.

Data collection methods and *outcomes in focus* in the studies are generally judged as being of 'high' relevance for the Danish trial context, since teachers would be asked to make video recordings of their own teaching, with a view to receiving feedback, fill in questionnaires and contribute to observation data collected by external researchers/staff. Although it would be more common practice in Denmark to have colleagues or school management conduct classroom observations, in a trial context use of trained external observers would be perfectly acceptable. Indeed, Danish schools and institutions are increasingly involved in action research and hence are becoming used to contributing actively to research.

7.3 Three studies judged as highly relevant in a Danish context

In the following, we turn our attention to the three studies for which the overall judgment for relevance and feasibility was 'high'. The study by Allen et al. (2011) focuses on methods to improve the motivational and instructional qualities of teachers' interactions with their secondary students (age 11-18). A PD intervention of this kind would be relevant to evaluate in three different overall settings in Denmark and would therefore have to be adapted to different types of teachers and students. Students aged 11-15/16 would still be in compulsory school (a public school (*Folkeskole*) or a state-subsidised private school) and taught by teachers with a professional Bachelor's degree. Students aged 15/16-18 would be in either general high school (taught by teachers with a Master's degree) or in vocational training (taught by teachers with other types of training). The intervention is structured as workshop-based training supplemented by one year of personalised coaching/feedback based on video recordings of class sessions. This set-up seems to be relevant for all types of teachers who would be involved in such a trial in Denmark, though the use of the CLASS-S instrument to collect data would probably have to be well-motivated to ensure acceptability among all types of teachers, since they would differ in terms of their experience in using such instruments.

The study by Domitrovich et al. (2009) focuses on improving teachers' language use, emotional support and positive behaviour management strategies for use in preschool classrooms. The focus is on Head Start classrooms targeting children from disadvantaged families in both rural and urban areas. The main need for adaptation to the Danish context would concern identifying similar settings, i.e. kindergarten settings with children from disadvantaged families. The intervention in this trial is based on a specific curriculum targeting children's emergent language/literacy skills and socio-emotional development. Obviously, this would require translation and cultural adaptation in order to provide pedagogues in Denmark with relevant material.

The PD intervention evaluated by Murray et al. (2014) is the Incredible Years Teacher (IYT) classroom management programme. The relevance and feasibility of testing this in Denmark is high, since IYT is a well-known programme in Denmark. There is a strong belief among Danish teachers that students'

socio-emotional wellbeing in school is of key importance, and some would say that it is indeed a pre-requisite for academic learning, so there would most probably be an interest in finding out whether this programme can document effects on students' academic performance. In Denmark, students in the public *Folkeskole* are tested regularly (starting in grade 2 for the mandatory national tests, but earlier using other tests, which may or may not be mandatory, depending on school/municipality policy) in the main subjects, such as Danish language, mathematics. Hence, one necessary adaptation would be to decide which test(s) to use as relevant outcome measures. The intervention centres on video modelling, rehearsal of key skills though role play, classroom practice assignment, teacher goal setting and self-monitoring. Workshops take place in groups of 12-15 teachers. Since many teachers in Danish schools work in teams (subjects, grade), it may be useful to consider testing this intervention in such teams.

7.4 Numerical values

In the following, we present the numerical values extracted from all nine studies in the in-depth review, with a view to presenting the results of the subsequent meta-analysis. A risk-of-bias assessment of each of the studies is also presented and discussed as part of the overall quality assessment of the studies. We then return to the three studies mentioned above-, in which the interventions being evaluated were judged to be highly relevant and suited for possible trial in a Danish context. Together with the findings presented in the following sections, we will discuss to what extent these three interventions are promising and whether they should be considered for a trial in Denmark.

Table 10 and Table 11 present the numerical values for the nine studies that were extracted in preparation for the series of meta-analyses. All studies reported either student or teacher outcomes that enabled the calculation of a standardised mean difference and standard error approximately by the end of the intervention. Hedges' *g* was used to estimate the SMD, and we applied the small *N* correction (Lipsey & Wilson, 2001, pp. 47-49)

Three studies reported only selected items on student academic outcomes (Allen et al., 2011; Murray et al., 2014; Rubie et al., 2015), and two studies reported outcomes on students' social competences (Jensen et al., 2015; Murray et al., 2014). Six studies reported outcome measures for teachers; four studies (Domitrovich et al., 2009; Early et al., 2017; Murray et al., 2014; Raver et al., 2008) reported various measures of The Classroom Assessment Scoring System (CLASS), and two studies (Fukkink & Tavecchio, 2010; Ottmar et al., 2013) reported other measures of teacher outcomes.

Table 10: Numeric data (studies 1-5)

Study	1 – Rubie-Davies and Rosenthal (2016)	2 – Allen et al. (2011)	3 – Fukkink & Tavecchio (2010)	4 – Early et al. (2017)	5 – Murray et al. (2014)
Type of outcome	Continuous, test scores range from approx. 1100-1900	Continuous, test scores standardised on a 200-600-point scale.	Four-point scale	10 dimensions, each measured on a 7-point scale, with 1-2 low, 3-5 mid-range, 6-7 high. The 10 dimensions are organised in 3 domains.	CLASS scales measured on a 7-point scale. STAR standard scores used in analyses reflect students' ability levels on a continuous vertical scale spanning grade levels, which measure absolute growth. Subject/ grade limitations (p. 3).
Professional or other?	Other (children)	Other (children)	Professional	Professional	Professional and students
Outcome (there may be more than one, record them all)	Mathematics and reading achievement	Student achievement in various courses: maths/science, language, arts/social studies, (p. 2). Teachers: observations based on videos that are analysed according to CLASS-S, but this is part of the intervention and hence a mediating factor rather than an outcome as such.	Caregiver interaction scale (Arnett, 1989), author-constructed scales to capture sensitive responsivity and verbal simulation. VIG-specific measures. Job resource scale and VIG job satisfaction scale. Not relevant outcomes according to review protocol.	Teacher-child interactions using validated CLASS, 3 domains: (i) Emotional Support; (ii) Classroom Organisation; (iii) Instructional Support. Additional teacher outcomes (pp. 61-62)	Observational change in teacher practices based on CLASS, of which subscales were used for this study: Positive Climate, Negative Climate and Behaviour Management. TCI seems to have at least two subscales: Harsh and Competent (see note in Table 1 of paper).
Time Point (s) (record the exact time, there may be more than one, record them all)	Beginning (mid-end February), middle (mid-end June) and end (end-November) of school year. Academic year in New Zealand runs from beginning February to mid-December.	Baseline (similar course taken by the student the year before), end of intervention year and end of post-intervention year (in course of teachers' choice according to study instructions).	Pre-test, post-test, retention (only intervention)	Fall and spring observations. Each observation included 6 30-min observation cycles. Measured by trained observer. Scores are average of all 6 cycles conducted. Average of 194 calendar days between pre- and post-test observations, slightly more for MTP than MMCI since the former takes longer time by design (p. 63).	Autumn and spring observations of approx. 2 hours of instructional time (approx. five months between observations)

Study	1 – Rubie-Davies and Rosenthal (2016)	2 – Allen et al. (2011)	3 – Fukkink & Tavecchio (2010)	4 – Early et al. (2017)	5 – Murray et al. (2014)
Source (questionnaire, admin data, other (specify) or unclear)	Mathematics and reading achievement assessed using e-asTTle, an online assessment tool. Curriculum strands selected by first author for this study.	Commonwealth of Virginia Standards of Learning (SOL) testing system.	Caregiver interaction scale (Arnett, 1989): stimulating and authoritarian caregiving behaviour – validated and delivered by trained assessors. Other scales were constructed so not used here.	Classroom Assessment Scoring System (CLASS; Pianta et al., 2008)	CLASS (Pianta & Hamre, 2005) and the Teacher Coder Impressions Inventory (TCI; Webster-Stratton, Reid, & Hammond, 2001). Revised Teacher Social Competence (R-TSC) scale (teacher-rated); Conners' DSM_IC Inattention scale (Conners, 2001); Academic competence subscale of R-TSC; Star Early Literacy/Reading and Math (STAR) (nationally normed computerised adaptive test).
Valid Ns (only applicable for continuous outcome data). Mention treatment and comparison	84 teachers randomised: 43 intervention; 41 control. Students: 1241 intervention, 1167 control	78 teachers; Intervention year: 1267 students in 76 classrooms. Post-intervention year: 970 additional students in 61 classrooms (61 teachers).	95 teachers in total. n = 52 for the experimental group, n = 43 for the control group	Final sample (p.60): Total (over the three years): 486 teachers in 336 schools/centre at pre-test and who had pre- and post-test CLASS observations. MMCI: 175; MTP: 151; Control: 160.	97 teachers, 47 allocated to intervention group, of which 45 participated. Student data available for analysis: 598 intervention, 560 control
Method of estimation	Test score means	End-of-year student achievement test scores I vs. C, but unclear where these are raw scores or whether they have been adjusted for achievement test scores from previous year and teacher and student demographic characteristics, etc. See paper p. 3. and note at bottom of Table S1, p. 2 in SOM.	Pre-test and post-test means and std. dev. Effect sizes (using corrected means). Since authors mention that ES are calculated using corrected means, should it be assumed that the reported means and SD are raw?	Pre-test and post-test means, std. dev., range.	Unadjusted (raw) means and SDs as well as predicted changes in outcomes taking account of nesting of students within teacher, grade level and school

Study	1 – Rubie-Davies and Rosenthal (2016)	2 – Allen et al. (2011)	3 – Fukkink & Tavecchio (2010)	4 – Early et al. (2017)	5 – Murray et al. (2014)
Statistics (risk ratio, odds ratio, standard error, 95 cf, DF, p-value, chi2)	Suggest using T1 and T3	Post-intervention year (2-year study period by design): End of Year Achievement Test Scores, MEAN (SD) Intervention: 488.2 (72.0), Control: 482.2 (65.9)	Stimulating caregiving: ES_corr = .61* and Authoritarian caregiving: ES_corr = .37	Raw means and SD. Effect sizes while controlling for pre-test scores. (Predicted odds of reaching predefined cut points at post-test also available)	Raw means and SD. Predicted changes, in which various background characteristics are taken into account. Main effects mentioned in text on p. 4: "Results indicate a significant main effect of the intervention on Positive Climate (b = .77, p = .007), reflecting that teachers in the intervention group were rated as having significantly more positive change in classroom climate than control teachers over time." "Student outcome analyses identified no main intervention effects."
Page numbers and notes	Table 2, p. 80: N, means and std. dev. T1 (beginning), T2 (middle), T3 (end) of school year.	Supplementary Online Material, Table S1. Achievement Test Scores I vs. C in Intervention Year (p. 1) and Post-Intervention Year (p. 2). Means and std. dev. NB: Note at bottom of table states that "Analyses used hierarchical linear models". Not clear whether results in S1 are raw or adjusted scores. Table S2., regression results	Table 2. Only use results for "Stimulating caregiving" and "Authoritarian caregiving" since they are from the validated Arnett scale. Note that a decrease in the latter is a favourable development. See note about corrected means used for ES estimation	Table 2, p. 62, use only the 3 CLASS outcomes. Table 3, p. 65 has results that take account of pre-test scores and clustering. Table 4, p. 66, has predicted odds of reaching cut points at post-test.	Table 1, p. B-1: Teacher outcomes: CLASS and TCI (Note that only TCI Competent has been reported. According to note at bottom of table, there is also a TCI Harsh component. Reason for non-reporting is not provided.) Figure 3: Predicted change in STAR for 1st and 2nd grade, separately.
Level of aggregation	Student level results	Student level results, HLM models take account of nesting of children within teachers and teachers within schools	Teacher	Teacher	Teachers and students, since there are outcomes for both

Study	1 – Rubie-Davies and Rosenthal (2016)	2 – Allen et al. (2011)	3 – Fukkink & Tavecchio (2010)	4 – Early et al. (2017)	5 – Murray et al. (2014)
Notes	No apparent clustering within schools or classrooms	Used post-intervention year means and prior year SDs (in order to use SDs from the same students as post-means). Number of students T/C: 419/551	ES as reported, number of teachers T/C: 52/43	Cluster RCT: schools/centres in Cohort 1 (first year) blocked by region and randomly allocated to one of 3 arms. In years two and three (cohorts 2 and 3) classrooms (not teachers) were randomly allocated (no blocking), i.e. allowing different classrooms within same school/centre to have same prob. for participation in each of the 3 conditions.	Slight baseline differences, see top of page 4, which the raw score comparisons – by definition – cannot take into account.
Used in meta-analysis?	Used means from T3 and total SD from T1. Number of students T/C: Maths: 831/808; Reading: 852/707. Simple average of maths and read ESs	Used post-intervention year means and prior year SDs (in order to use SDs from the same students as post-means). Number of students T/C: 419/551	Could not be pooled with other studies but two effect sizes are shown in a forest plot by themselves.	Report ESs. Number of teachers used: MMCI: 175; MTP: 151; Control: 160	Post-means and pre-SDs. Number of teachers T/C: 45/50. Number of students T/C: 598/560. Student outcome used a simple average of maths, reading and early literacy.

Table 11: Numeric data (included studies 6-9)

Study	6 – Jensen et al. (2015)	7 – Domitrovich et al. (2009)	8 – Ottmar et al. (2013)	9 – Raver et al. (2008)
Type of outcome	Strengths and Difficulties Questionnaire (SDQ), five subdomains, total SDQ score, SDQ impact score. All SDQ scores are normalised by pre-test std. dev. in control group	CLASS and TSRS – continuous; CLASS: 7-point scale, TSRS: 5-point scale. CLEO: Four components are counts/min, one is a richness-sensitivity rating on 5-point scale	M-Scan: eight dimensions measured on a 7-point scale; MKTA: scale from -2.0 to 2.0	Classroom management
Professional or other?	Other (children)	Professional (teacher)	First two – professional, latter – child	Professional
Outcome (there may be more than one, record them all)	Socioemotional skills of the children (emotional symptoms, conduct problems, hyperactivity, peer relationship problems, prosocial behaviour)	The Classroom Assessment Scoring System (CLASS; La Paro and Pianta, 2003) – 10 dimensions of teaching quality; The Teaching Style Rating Scale (TSRS; Domitrovich, Cortes & Greenberg, 2000) – teacher behaviour; The Classroom Language and Literacy Environment Observation (CLEO; Holland-Coviello, 2005) – child-directed talk from a specific teacher	Mathematics Scan (M Scan; Berry et al., 2010); Mathematical Knowledge for Teaching Assessment (MKTA); Stanford Achievement Test, 10th edition (Stanford-10; Harcourt Educational Measurement, 2002) – to look at average classroom achievement, not at individual child level (aggregated by teacher ID); abbreviated version of Mathematics teacher Efficacy Beliefs Inventory (MTEBI)	CLASS (classroom quality): 4sub-scales: Positive climate Negative climate Teacher sensitivity Behaviour management
Time Point(s) (record the exact time, there may be more than one, record them all)	P. 29: Pre- (March 2011) and post- (March 2013), also collected mid-way in March 2012 but not used in main estimations (rather to do robustness checks on results and check validity of teacher-reported SDQ.	Pre-intervention and post-test	2008-2009 school year, Three 3-month windows of observation were established: autumn (September to November), winter (December to February), and spring (March to May)."; MKT - Spring 2009 (follow up); Stanford Test – Spring of second grade (baseline only)	Spring
Source (questionnaire, admin data, other (specify) or unclear)	Strengths and Difficulties Questionnaire (SDQ) – teacher-completed	CLASS: observational; TSRS: observational; CLEO: observational	Observation; online assessment; standardised achievement test	Observations using 7-point Likert scales
Valid Ns (only applicable for continuous outcome data). Mention treatment and comparison.	Total = 58 preschools; 29 control; 29 intervention; 686 children – 396 intervention and 290 control	84 (but, pre-test scores not used as a covariate in the model, as they were only available for 57 teachers)	24 schools in total (13 intervention, 11 control schools); 94 teachers out of a possible 100 in the participating schools	N = 18 sites: 9 Intervention and 9 Control

Study	6 – Jensen et al. (2015)	7 – Domitrovich et al. (2009)	8 – Ottmar et al. (2013)	9 – Raver et al. (2008)
Method of estimation	Regression model (pre-test not included – Table 5 (p. 35), and other factors, e.g. gender	P. 581 multiple regression models (pre-test not included). For the latter two outcome measures – random intercept models.	P. 445: "the unconditional model (Model 1) tested whether a two-level hierarchical linear modelling (HLM) model or ordinary least-squares (OLS) regression were more suitable. The intra-class correlation coefficient (ICC 0.13), coupled with the nesting of teach ers in schools, led us to use two-level HLM (Raudenbusch, Bryk, Cheong, Congdon, & du Toit, 2004) to analyse our outcomes. Model 2 tested the impact of the RC approach on the use of standards-based practices and included teacher-level variables (MKT, PMTE, MTOE, teacher experience, and average classroom achievement) and school-level variables (school assignment [RC vs. control], AYP status and Title 1 status). Model 3 included the FOI variable. All level 1 variables were grand-mean centred, while all three level 2 (school) predictors were uncentred."	HLM Difference in means after adjusting for baseline predictors
Statistics (risk ratio, odds ratio, standard error, 95 cf, DF, p-value, chi2)	-0.216*** (*** p < 0.01) SE 0.075	EMOTIONAL SUPPORT: TSRS Positive Emotional Climate Scale – significant positive intervention impact (p = 0.05); CLASS – individual item (positive climate) (d = .61, p = .04); BEHAVIOURAL SUPPORT: three-item TSRS (p = .002), proactive-preventive classroom management (p = .001); COGNITIVE LINGUISTIC TEACHING QUALITY: intervention teachers more statements (p = .001) and ask more questions (p < .001)	(p. 446). 0.39** SE 0.18	0.89 0.98 (coeff.) .20 SE d = 0.89 0.60 minus 0.72 (co-eff) .19 SE d = 0.64 0.54 0.54 (coeff.) .20 SE d = 0.53 0.55 0.55 (coeff.) .29 SE d = 0.52

Study	6 – Jensen et al. (2015)	7 – Domitrovich et al. (2009)	8 – Ottmar et al. (2013)	9 – Raver et al. (2008)
Page numbers and notes	Table 5, p. 35. Total SDQ score. Sub-groups are available	Table 2, p. 583: M and SD for measures of teaching quality (end of year); Table 3 – Effects of HS REDI on teaching quality emotional behavioural support; Table 4: Effects of HS REDI on teaching quality cognitive linguistic support; Table 5: Effects of HS REDI on teaching quality: linguistic support in three classroom settings	Table 2, p. 447: "Table 2. Results from Two-Level HLM Model Examining the Impact of the RC Approach and the Contribution of MKT, Teacher Self-Efficacy, FOI, and School Contextual Factors on Mathematics Teaching Practices"	Tables 1, 2 and 3, p. 12 and pp. 23-25 HLM to adjust for clustering
Level of aggregation	Child level, but adjusted for clustering	Teacher level	MSCAN – teacher level, MKTA – teacher level, Stanford Achievement Test (assessment of child but aggregated at teacher level in this study)	Class and site level using HLM to account for clustering at two levels
Notes	Adjusted for clustering (p. 35) Table 5: Preschools randomised but analysis on child level	P. 581 – all TSRS variables deviated 'more substantially from normality'.		
Used in meta-analysis?	ES and SE reported	Number of classrooms (for CLASS as one measurement per classroom) T/C: 22/22. Number of teachers (for TSRS) T/C: 42/42. Post-means and post-SDs (do not report other than post-SDs). NB: TSRS NOT USED as it deviates from normal.	Mean difference coefficient from model 2 and pre-SD from Table 1. Number of teachers T/C 43/45.	ES reported. Number of teachers T/C: 41/41 (assuming an equal number in both groups: (87-4)/2

7.5 Risk-of-bias assessment

All nine included studies were randomised controlled trials. Overall, the included studies varied on risk-of-bias judgments, and no single study could be characterised as a robust RCT with low risk of bias on all assessed risk-of-bias items, although the study by Rubie-Davies and Rosenthal (2016) had only minor problems. In this respect, this study had the highest quality overall. The ratings of each study in relation to the nine domains in the risk-of-bias tool, as well as the descriptions used for the assessments, are shown in Appendix E: Risk-of-Bias Assessment. The risk-of-bias judgments are based on pre-specified questions and a 5-point scale with ratings of 1 = low risk and 5 = high risk. Further details on risk of bias are provided in the design and methods section.

Two studies reported the use of appropriate randomisation methods (Raver et al., 2009; Rubie-Davies and Rosenthal, 2016); the remaining studies did not report the method of randomisation. One study was rated high on sequence generation and allocation concealment, even though the sequence generation method was not reported (Early et al., 2017). However, it was reported that classrooms were randomised and teachers were assigned to the selected classrooms, to determine whether she or he was eligible for participation, any classes being replaced when a teacher was ineligible. This is not proper randomisation, as randomisation occurred too soon. As is common in social intervention, it is generally impossible to blind participants or those delivering the interventions. Five studies clearly stated that outcome assessors were blinded to allocation status (Early et al., 2017; Fukkink & Tavecchio, 2010; Ottmar et al., 2013; Raver et al., 2009; Rubie-Davies and Rosenthal, 2016). Overall attrition levels were not high; only two studies had relatively high levels of attrition (Allen et al., 2011; Jensen et al., 2015), and one study did not report the level of attrition (Fukkink & Tavecchio, 2010). None of the studies dealt with missing data.

Only one study was free of selective reporting bias (Rubie-Davies and Rosenthal, 2016). Four of the studies had serious problems of various kinds, rated 4 on the 'other risk of bias' item (Domitrovich et al., 2009; Fukkink & Tavecchio, 2010; Jensen et al., 2015; Murray et al., 2014). We could not locate a protocol for any of the studies. We had reason to expect that an *a priori* analysis plan was present in one study (Allen et al., 2011).

Confounding was not relevant in the in-depth review of risk-of-bias assessment since we did not find any non-randomised studies to include in the in-depth review.

7.6 Meta-analyses

Due to the homogeneity of PD approaches evaluated in the nine trials (see above), we used professional and student outcomes as the basis of the series of meta-analyses presented below. This was possible due to fact that a number of the trials used the same outcome measures for observation of effects of the interventions.

We report the results of a series of meta-analyses below, where individual studies with homogeneity of outcome are combined to obtain an 'overall' effect size estimate of the interventions where possible. All meta-analyses are inverse variance weighted, using random effects statistical models that incorporate both the sampling variance and between-study variance components into the study-level weights. Random effects weighted mean effect sizes are calculated using 95% confidence intervals.

If outcomes are too different to combine in a meta-analysis, the study-level effect sizes are shown.

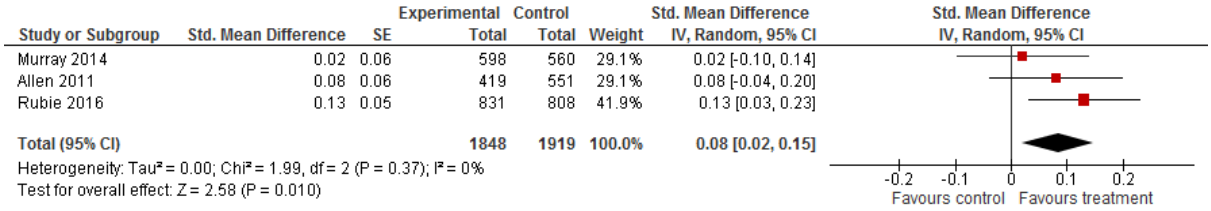
Analysis was conducted in RevMan5 (Informatics, 2016) and results displayed graphically in forest plots.

7.7 Student outcomes

Three studies reported results on student academic outcomes (Allen et al., 2011; Murray et al., 2012; Rubie-Davies & Rosenthal, 2016) as displayed in Figure 2.

The meta-analysis of the studies showed no evidence of statistical heterogeneity with an I^2 value of 0%, and the estimated τ^2 is 0.00, which suggests that, despite the studies having some differences in their pedagogical approaches and students, the underlying effect of the interventions is similar. However, given that there are relatively few studies, some caution should be exercised in assuming that there is a single true effect from PD on student academic outcomes.

Figure 2: Students' academic scores



An extensive sensitivity analysis was also undertaken adjusting for clustering using an ICC of 0.05, 0.1 and 0.22[11]. While ignoring clustering will not produce biased estimates of intervention effects, it will bias the standard errors and make something appear statistically significant, when in truth the observed difference could be largely due to chance. The resulting forest plots (

Figure 3 to Figure 5) show that with a small amount of clustering (i.e. an ICC of 0.05) the results are no longer statistically significant at p = 0.05. This suggests that, although the overall effect on student academic outcomes remains positive, we need to show caution in attributing a treatment effect, as this could still result from chance. Consequently, one recommendation would be to undertake a large trial with sufficient power to show the pooled difference to be statistically significant.

[11] Note: Correcting the effect size and standard error using an intra-cluster correlation (ρ) of 0.05, 0.1 and 0.22, we used the following formulas (see Hedges, 2007, p. 349):

$$d = \left(\frac{MD}{SD}\right)\sqrt{1 - \frac{2(n-1)\rho}{N-2}}$$

$$SE = \sqrt{\left(\frac{N^T + N^C}{N^T N^C}\right)(1 + (n-1)\rho) + d^2\left(\frac{(N-2)(1-\rho)^2 + n(N-2n)\rho^2 + 2(N-2n)\rho(1-\rho)}{2(N-2)[(N-2) - 2(n-1)\rho]}\right)}$$

where n is cluster size and N^T, N^C are treatment and control group sample sizes, and N is total sample size

Figure 3: Students' academic scores, adjusting for clustering ICC = 0.05

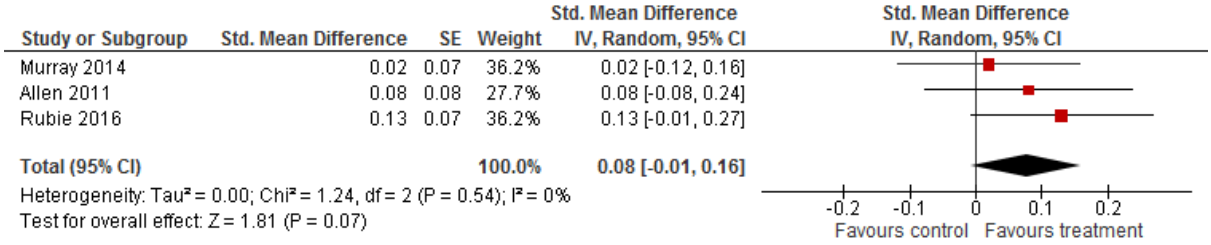


Figure 4: Students' academic scores, adjusting for clustering ICC = 0.1

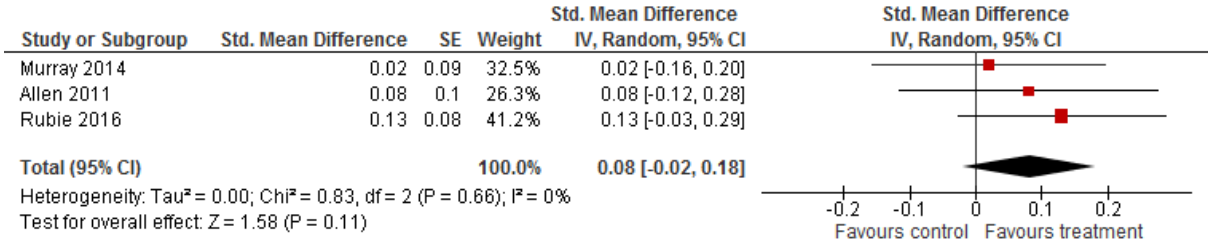
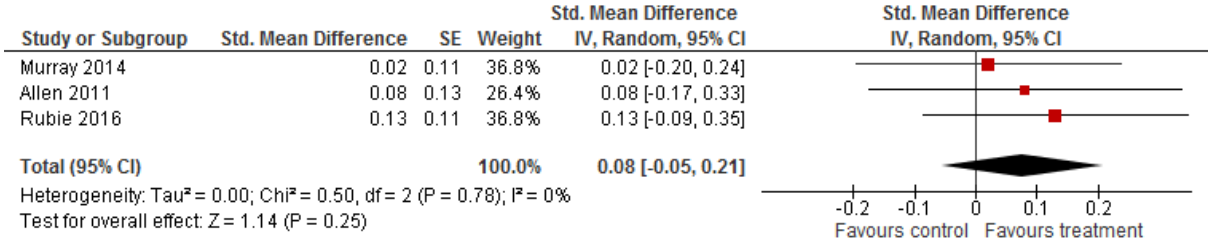


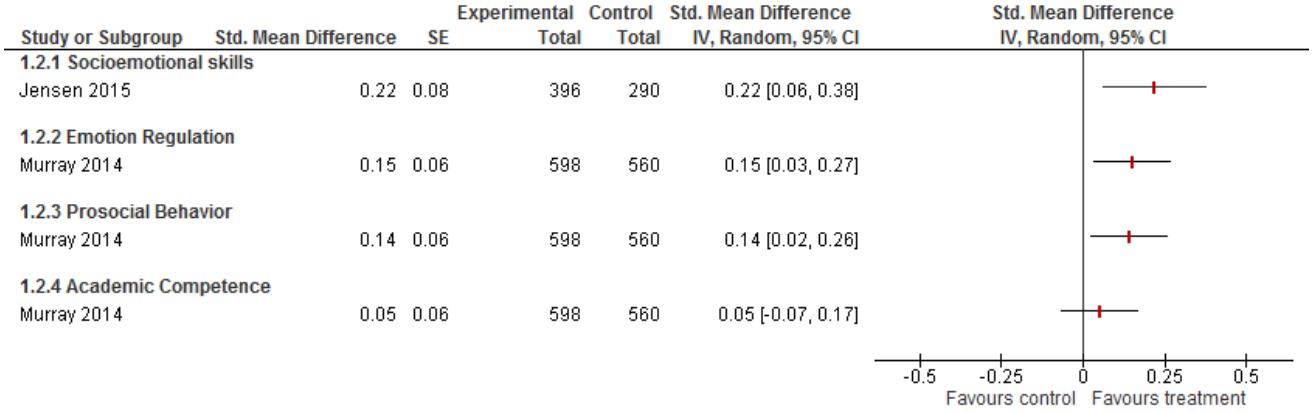
Figure 5: Students' academic scores, adjusting for clustering ICC = 0.22



Two studies reported outcomes on student social competences (Jensen et al., 2015 – one outcome; Murray et al., 2014 – three outcomes), although the measures were too different to be combined. The reported results from the two studies are displayed in Figure 6 below.

All results indicated a positive effect with study-level effect sizes varying between 0.05 and 0.22.

Figure 6: Students' social competences



The analysis in the study by Jensen et al. (2015) took into account clustering. Thus, in the sensitivity analysis only the results reported in Murray et al. (2014) were adjusted for clustering. The resulting forest plots (Figure 7 to Figure 9) show that with a small amount of clustering (i.e. an ICC of 0.1) the results are no longer statistically significant at a p = 0.05, while the result reported in Jensen et al. (2015) is unchanged.

Figure 7: Students' social competences ICC = 0.05

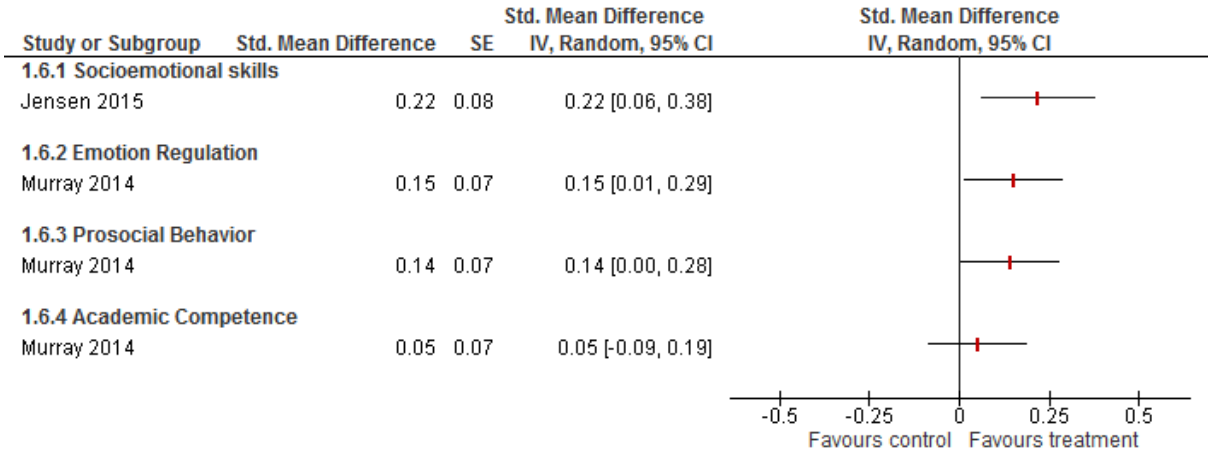


Figure 8: Students' social competences ICC = 0.1

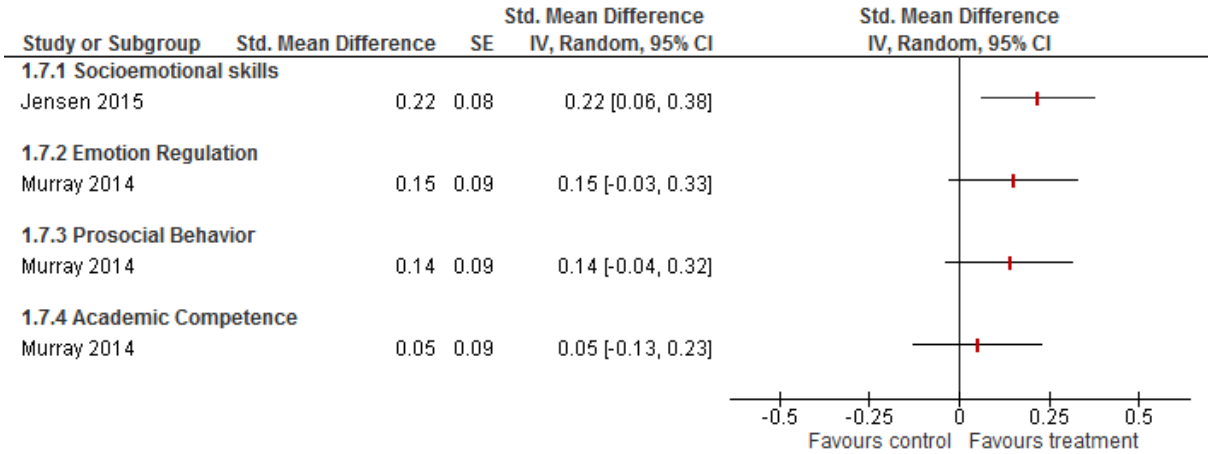
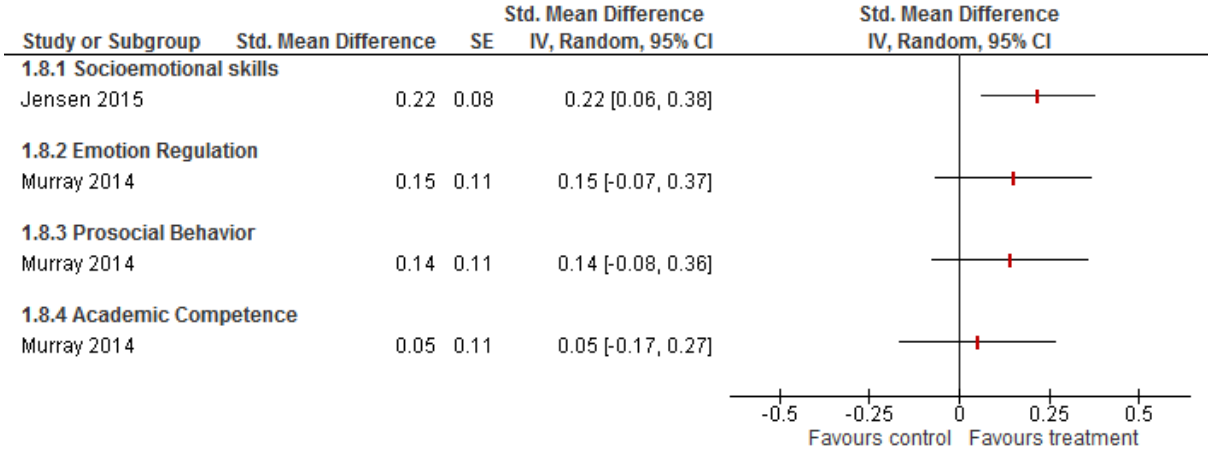


Figure 9: Student social competences ICC = 0.22



7.8 Teacher outcomes

Three studies (Domitrovich et al., 2009; Murray et al., 2014; Raver et al., 2008) reported various measures of The Classroom Assessment Scoring System (CLASS), and two studies (Domitrovich et al., 2009; Early et al., 2017) reported CLASS summary measures.

CLASS has ten items, but only nine items are reported in Domitrovich et al. (2009), three are reported in Murray et al. (2014), and four are reported in Raver et al. (2008). We decided to combine the results reported in Domitrovich et al. (2009), Murray et al. (2014) and Raver et al. (2008) in analyses of positive climate, negative climate and behavioural management, and to use the results reported in Murray et al. (2014) and Raver et al. (2008) in an analysis of teacher intensity. The results of these four analyses are shown in Figure 10 to

Figure 13. The weighted average effect of all teacher outcomes are positive, but only positive climate and teacher sensitivity are statistically significant; the weighted average of negative climate and behaviour management are statistically non-significant.

There is a high degree of heterogeneity between the studies in the analysis of negative climate as indicated by the values of I^2 and τ^2. There is some degree of heterogeneity in the analyses of positive climate and behaviour management, but, as indicated by the value of I^2, it may not be of high practical importance, even though the values of τ^2 may seem relatively high. However, values of τ^2 should be interpreted with caution, as the DerSimonian and Laird estimate of τ^2 is overestimated, on average, and the bias can be substantial when the number of studies is small (Borenstein, Hedges, Higgins, & Rothstein, 2010).

Although the p-value of the Q-statistic is notoriously underpowered to detect heterogeneity in small meta-analyses, the estimated τ^2 is 0.00, and I^2 is 0% in the analysis of teacher sensitivity, implying that heterogeneity between the two studies used in this analysis is not present. Given there are relatively few studies reporting teacher outcomes measured by CLASS, some caution is needed in making an assumption that there is a single true effect from PD on any of these teacher outcomes.

Regarding the two studies (Domitrovich et al., 2009; Early et al., 2017) reporting CLASS summary measures, we decided not to combine them in a meta-analysis as it was unclear whether the same CLASS items were included in the summary measures of the two studies. Early et al. (2017) reports on three summary measures (for both of the interventions analysed in this study) but does not report which items are included in which summary. Domitrovich et al., 2009 reports on two summary measures and

reports which items are used. However, the items used in the summary measures are different from the ones reported in La Paro and Pianta (2003) (which Domitrovich et al. 2009 refers to). In

Figure 14, the individual study results are shown for the summary measures. All of the results indicate a positive effect, although they are not all statistically significant.

Figure 10: Positive climate

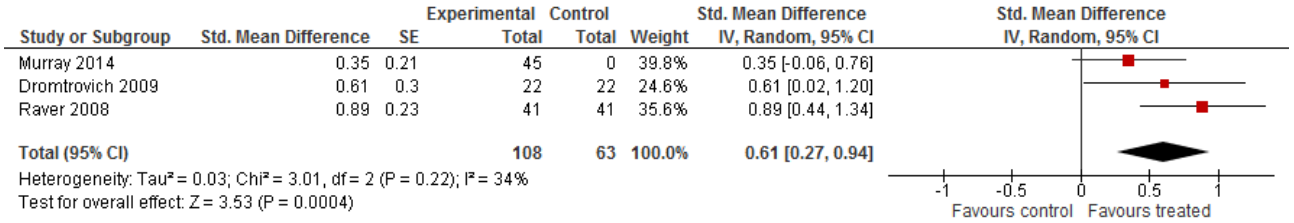


Figure 11: Negative climate

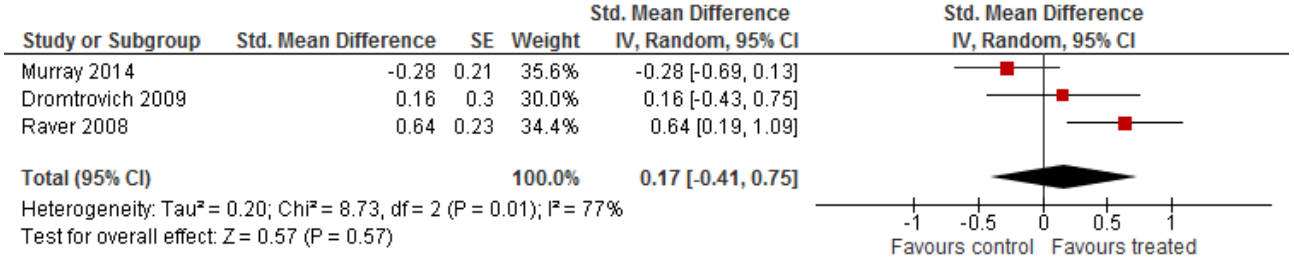


Figure 12: Behaviour management

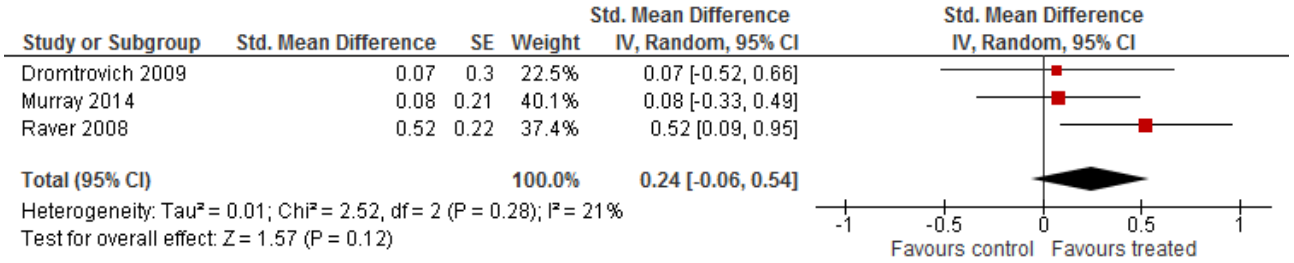


Figure 13: Teacher sensitivity

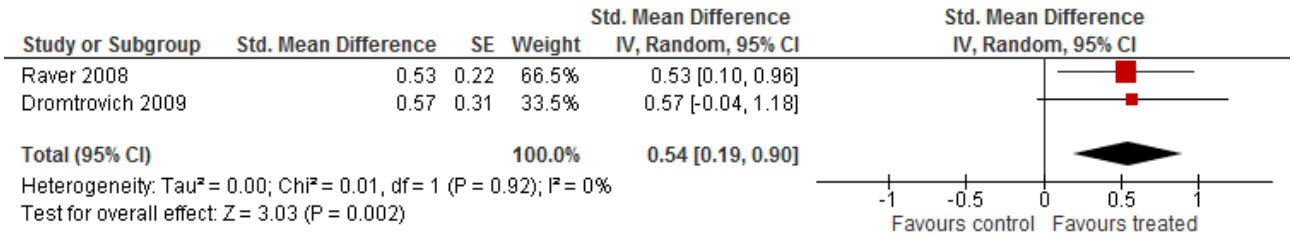
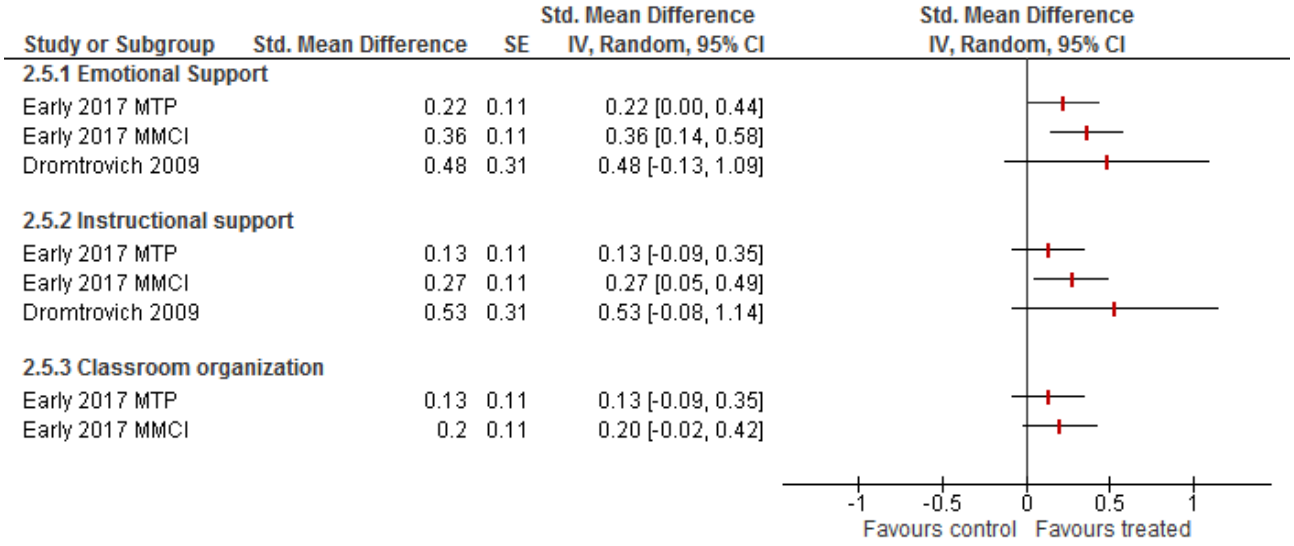


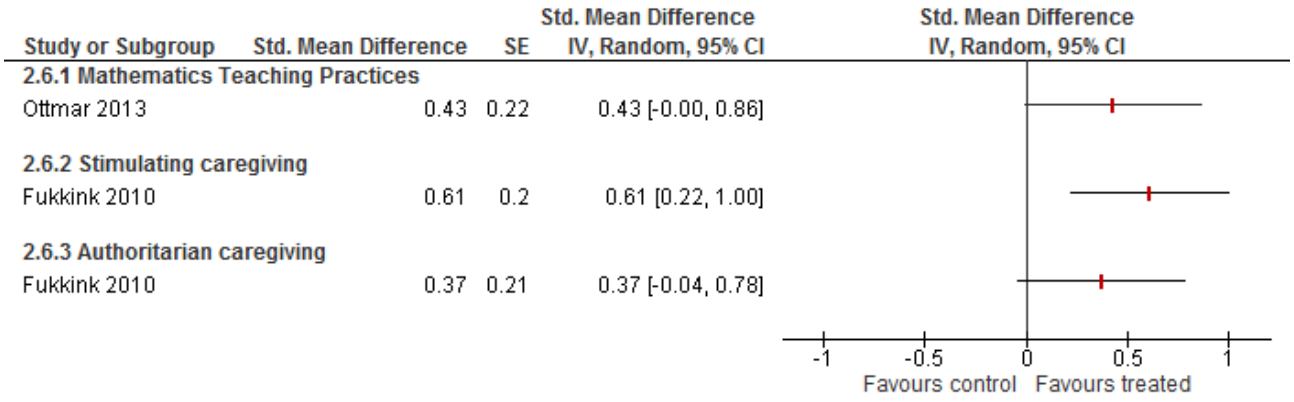
Figure 14: Summary CLASS



Two studies (Fukkink & Tavecchio, 2010; Ottmar et al., 2013) reported other outcome measures of teachers. Fukkink & Tavecchio (2010) reported two measures from the Caregiver interaction scale (Arnett, 1989), and Ottmar et al. (2013) reported on Mathematics teaching practices. The single-study effect sizes are shown in

Figure 15. All of the results indicate a positive effect, although they are not all statistically significant.

Figure 15: Other



Conclusions

A moderate body of experimental evidence exists in relation to the effect of PD in the topic area of education, but similar evidence does not appear to exist in the topic areas of social welfare and crime and justice. A moderate number of experimental evaluations of PD in language and literacy have been undertaken, mainly in the US.

A small body of evidence exists in relation to the effect of PD in social development interventions for students and teachers. The few available studies of effectiveness have varied methods of assessing the effects. The majority of studies do not report on student outcomes, while, with the exception of CLASS measures, the teacher outcomes reported are too different to be combined.

Conclusions: meta-analysis

In summary, there seems to be a positive although very small effect on student academic outcomes, and the statistical significance is questionable when clustering is taken into account.

There seem to be positive effects on teacher outcomes measured by CLASS, although the studies only reported on selected CLASS items (positive climate, negative climate, behavioural management and teacher intensity), and only two of the combined effects were statistically significant.

At most, the results from three individual studies could be combined in a single meta-analysis. The results of the meta-analyses should therefore be interpreted with great caution, due to the very limited number of studies and selection of measures on teacher outcomes.

In short, the result of this review is that there is currently insufficient evidence for conclusions to be drawn. The small number of available studies reporting similar outcomes precludes any conclusions concerning effectiveness or ineffectiveness of PD. Moreover, the limited number of studies prevented an analysis of specific PD approaches across cultures, across professions/service-deliverer types, across organisations, across service-receiver types etc.

Results: Quality assurance

Data extraction for the in-depth review, risk-of-bias assessment and extraction of numerical data were undertaken by two reviewers. Agreement was initially quite good, and full consensus was achieved through discussion.

8 Discussion: relevance and feasibility in a Danish context

As in several other western societies, there is political pressure in Denmark to promote the use of evidence-informed interventions, i.e. ones that have been proven to be effective according to the highest possible levels of effectiveness research standards. This is true of interventions in the broader social sector as well as in the narrower sector of schooling and education.

None of the three studies, which were judged as 'high' in terms of relevance and feasibility in the contextualisation section, were considered as being of high quality according to our risk-of-bias assessment. One study had a relatively high level of attrition, while the two others suffered from other serious risk-of-bias problems. Moreover, the three studies did not report all relevant student and teacher outcomes.

Summing up, though the three studies each suffered from various problems they were judged to be highly relevant and feasible for the Danish context. Hence, we suggest that the interventions presented in these three studies (Allen et al., 2011, Domitrovich et al., 2009; Murray et al., 2014) could be considered for rigorous testing trials in Denmark. Specific attention would then have to be paid to stringency by conducting a well-designed RCT with low risk of bias as well as ensuring that the sample sizes are large enough to provide sufficient power. Moreover, it should then be considered which types of outcomes are the most relevant. Our recommendation would be to have student outcomes as the primary outcomes (e.g. academic achievement, socio-emotional and behavioural outcomes). The reason for this is that the ultimate goal of any teacher PD ought to be to have a positive impact on students' well-being and academic progress in school. Teacher outcomes would then be considered as secondary outcomes in the sense that they are important, but mainly as intermediate factors working toward the ultimate goal of improving student outcomes. In this way, such adapted trials in Denmark would have the potential to make useful contributions to the PD effectiveness literature, if due consideration was made to the strengths and weaknesses of the studies found in this review.

8.1 Practical considerations regarding new trials in Denmark

In all the three cases mentioned above, where the interventions have been identified as having a certain potential for trial in Denmark, implementation can be helped along if one thinks carefully about how they might be received by the professionals for whom they are intended. This aspect, i.e. the 'receiving environment', is closely related to how use of research in general is received and used by professionals.

Dyssegaard and Egelund (2017) have conducted a systematic review of what promotes and hinders the use of research in school settings. Their review includes studies from the EU, Switzerland, Norway, USA, Canada and New Zealand. Based on their review, Dyssegaard & Egelund (2017) have written a concise knowledge brief to make the results available and relevant for stakeholders in and around schools in Denmark. They also discuss the importance of taking account of the culture and organisation of the particular day care institution or school in which a given intervention is to be implemented. This includes considering previous experiences with development of new practices as well as the level and type of competencies already present among the staff. Finally, the staff must have a clear vision, understanding and appreciation of the aim and relevance of the proposed change (Dyssegaard & Egelund, 2017).

One of their main conclusions of their review was that *school management* plays an important role in terms of creating a positive attitude among staff members regarding, for instance, implementation of a

new initiative. This can be achieved by involving the staff in the process, ensuring that they have the necessary time and resources available to prepare, plan for and participate in the new initiative.

Another of their findings is that *collaboration among staff members working in teams* is valuable, since this enables the sharing of new knowledge and experiences. Dyssegaard & Egelund (2017) find that *support and supervision from researchers* is a good idea in terms of assisting teachers in how to use the instruments that are being introduced/evaluated (e.g. observation guides) or how to interpret data (e.g. from tests). In general, their review shows that it is important to have *resource staff and support systems available both during and after implementation* of an initiative. In Denmark, such supportive 'anchoring' could be used with the 'learning consultants' employed in the Ministry of Education or the 'pedagogical consultants' employed by the local municipalities. Other options would be to have the follow-up support provided by appointed supervisors at the Psychological Pedagogical Advisory (PPR) units at the municipal or school level, in addition to continuing support from the school management.

Several of the characteristics of the PD interventions evaluated in our in-depth review do indeed address some of the points highlighted by Dyssegaard & Egelund (2017). The need for qualified support during the implementation of a new PD initiative is part of the PD evaluated in the form of personalised or remote coaching, mentoring sessions, one-to-one consultations etc. by trained, external staff (e.g. Allen et al., 2011; Domitrovich et al. 2009, Ottmar et al. 2013, Rubie-Davies and Rosenthal, 2016). Qualified support can also be given through presence in the classroom (e.g. Raver et al. 2009, where mental health consultants assist teachers on a weekly basis). Other interventions in our in-depth review contain organizational elements that are highlighted by Dyssegaard & Egelund, 2017 as being important, e.g. encouraging collaboration among teachers by providing the PD in groups that meet face to face with instructors (Early et al., 2017; Murray et al., 2012). Finally, the important role of school management being actively supportive and involved in the new PD initiative is an element of the Jensen et al. (2017) study.

8.2 Main recommendation

We recommend that a large randomised controlled trial evaluating the effectiveness of a PD intervention in the topic area of socio-emotional development should be funded and undertaken in the Danish context. The effectiveness of the intervention (compared with business-as-usual PD) on professional and student outcomes should be powered to show the pooled difference to be statistically significant. The intervention condition would ideally be developed from one of the three most relevant studies with respect to the Danish context, and the outcomes should be *all* items from the CLASS measure (not selected items) and standardised measures of student outcomes (social and academic outcomes). We recommend that the trial be designed, conducted and reported according to methodological criteria for rigour with regard to internal and external validity in order to achieve robust results.

9 References

(NB * indicates the 33 records that were included in the mapping)

* Al Otaiba, S., Connor, C. M., Folsom, J. S., Greulich, L., Meadows, J., & Li, Z. (2011). Assessment Data-Informed Guidance to Individualize Kindergarten Reading Instruction: Findings from a Cluster-Randomized Control Field Trial. The *Elementary Schol Journal, 111*(4), 535-560. doi: 10.1086/659031

* Allen, J. P., Pianta, R. C., Gregory, A., Mikami, A. Y., Lun, J. (2011). An interaction-based approach to enhancing secondary school instruction and student achievement. *Science*, 333, 1034–1037.

Arnett, J. (1989). Caregiver interaction scale. *Princeton, NJ: Educational Testing Service*.

* Azkiyah, S. N., Doolaard, S., Creemers B., P. M., Van Der, Werf, & M, P. C. (2014). The effects of two intervention programs on teaching quality and student achievement. *Journal of Classroom Interaction, 49*(1), 4-11.

Basma, B., & Savage, R. (2017). Teacher Professional Development and Student Literacy Growth: a Systematic Review and Meta-analysis: Springer.

Bjørnholt, B., Kristensen, K., Jacobsen, R.H., Bæk, T.A., Iversen, K., Skov, P.R., . . . Jensen, J.M. (2017). Kompetenceudvikling og kompetencedækning i folkeskolen ["Development of teachers' comptencies and competency coverage in the Danish Folkeskole"]. Retrieved December 2017, from https://www.kora.dk/media/7284851/11358-kompetenceudvikling-folkeskolen.pdf

Blank, R. K. , & de las Alas, N. (2009). Effects of teacher professional development on gains in student achievement: How meta-analysis provides evidence useful to education leaders. Washington DC, USA.

Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2010). A basic introduction to fixed-effect and random-effects models for meta-analysis. *Res Synth Methods, 1*(2), 97-111. doi: 10.1002/jrsm.12

* Bos, C. S., Mather, N., Narr, R.F., & Babur, N. (1999). Interactive, Collaborative Professional Development in Early Literacy Instruction: Supporting the Balancing Act. *Learning Disabilities Research and Practice, 14*(4), 227-238. doi: 10.1207/sldrp1404_4

BUPL. (2016). Alt for få uddannede pædagoger i daginstitutionerne, artikel på www.bupl.dk den 23.11.2016.

* Buysse, V., Castro, D. C., & Peisner-Feinberg, E. (2010). Effects of a professional development program on classroom practices and outcomes for Latino dual language learners. *Early Childhood Research Quarterly, 25*(2), 194-206. doi: 10.1016/j.ecresq.2009.10.001

* Cabell, S. Q., Justice, L, M., Piasta, S, B., Curenton, S, M., Wiggins, A., Turnbull, K. P., & Petscher, Y. (2011). The Impact of Teacher Responsivity Education on Preschoolers' Language and Literacy Skills. *American Journal of Speech-Language Pathology, 20*(4), 315-330.

Campbell Collaboration. (2018). Campbell Collaboration Policy Briefs Series. Retrieved 14 January 2018, 2018, from https://www.campbellcollaboration.org/better-evidence/policy-briefs.html

Conduct Problems Prevention Research Group. (1995). Psychometric properties of the social competence scale-teacher and parent ratings. *Fast Track Project Technical Report). University Park, PA: Pennsylvania State University*.

Conners, C. K. (2001). *Conners' rating scales revised*: Multi-Health Systems, Incorporated.

Cook, T. D., Campbell, D., & Boston, M. A. (1979). Quasi-Experimentation: Design and Analysis Issues for Field Settings. *HoughtonMifflin*.

Cordingley, P, Bell, M., Isham, C., Evans, D., & Firth, A. (2007). What do specialists do in CPD programmes for which there is evidence of positive outcomes for pupils and teachers *Research Evidence in Education Library*. London: EPPI-Centre, Social Science Research Unit, Institute of Education, University of London.

Cordingley, P, Higgins, S, Greany, T, Buckler, N, Coles-Jordan, D, Crisp, B, . . . Coe, R. (2015). Developing great teaching: lessons from the international reviews into effective professional development. London: Teacher Development Trust.

Cordingley, P., Bell, M., Evans, D., & Firth, A. (2005). The impact of collaborative CPD on classroom teaching and learning. Review: What do teacher impact data tell us about collaborative CPD? *Research Evidence in Education Library*. London: EPPI-Centre, Social Science Research Unit, Institute of Education, University of London.

Cordingley, P., Bell, M., Rundell, B., Evans, D., & Curtis, A. (2003). How does collaborative Continuing Professional Development (CPD) for teachers of the 5-16 age range affect teaching and learning *Research Evidence in Education Library* (pp. 205-216). London: EPPI-Centre, Social Science Research Unit, Institute of Education, University of London.

Cordingley, P., Bell, M., Thomason, S., & Firth, A. (2005). The impact of collaborative continuing professional development (CPD) on classroom teaching and learning. Review: How do collaborative and sustained CPD and sustained but not collaborative CPD affect teaching and learning *Research Evidence in Education Library*. London: EPPI-Centre, Social Science Research Unit, Institute of Education, University of London.

Danish Secretariate for SDQ and DAWBA in Denmark. (2014). Danish Secretariate for SDQ and DAWBA in Denmark. Retrieved December 2017, from http://sdq-dawba.dk

Danmarks Evalueringsinstitut. (2013). Strategier for læreres og pædagogers kompetenceudvikling, ("Strategies for the development of competencies among teachers and pedagogues") Retrieved December 2017, from https://www.eva.dk/sites/eva/files/2017-07/Strategier%20for%20laereres%20og%20paedagogers%20kompetenceudvikling_web.pdf

* Dickinson, D. K., & Caswell, L. (2007). Building support for language and early literacy in preschool classrooms through in-service professional development: Effects of the Literacy Environment Enrichment Program (LEEP). *Early Childhood Research Quarterly, 22*(2), 243-260. doi: 10.1016/j.ecresq.2007.03.001

* Domitrovich, C. E., Gest, S. D., Gill, S., Bierman, K. L., Welsh, J., & Jones, D. (2009). Fostering High-Quality Teaching with an Enriched Curriculum and Professional Development Support: The Head Start REDI Program. *Am Educ Res J, 46*(2), 567-597. doi: 10.3102/0002831208328089

Dunn, L. M., & Dunn, L. M. (1997). *PPVT-III: Peabody picture vocabulary test*: American Guidance Service.

Dunst, C. J., Bruder, M. B., & Hamby, D. W. (2015). Metasynthesis of in-service professional development research: Features associated with positive educator and student outcomes. *Educational Research and Reviews, 10*(12), 1731.

Dyssegaard, C. B., & Egelund, N. (2017). Vidensnotat om anvendelse af forskning i grundskolen. Retrieved January 2018, from http://edu.au.dk/fileadmin/edu/Udgivelser/Clearinghouse/Vidensnotat-om-anvendelse-af-forskning-i-grundskolen.pdf

* Early, D. M., Maxwell, K. L., Ponder, B. D., & Pan, Y. (2017). Improving teacher-child interactions: A randomized controlled trial of Making the Most of Classroom Interactions and My Teaching Partner professional development models. *Early Childhood Research Quarterly, 38*, 57-70.

Enochs, L. G, Smith, P. L, & Huinker, D. (2000). Establishing factorial validity of the mathematics teaching efficacy beliefs instrument. *School Science and Mathematics, 100*(4), 194-202.

* Flook, L., Goldberg, S. B., Pinger, L., Bonus, K., & Davidson, R. J. (2013). Mindfulness for teachers: A pilot study to assess effects on stress, burnout and teaching efficacy. *Mind Brain Educ, 7*(3), 182-195. doi: 10.1111/mbe.12026

* Fukkink, R. G., & Tavecchio, L. W. C. (2010). Effects of Video Interaction Guidance on early childhood teachers. *Teaching & Teacher Education, 26*(8), 1652-1659.

* Gallagher, P. A., Abbott-Shim, M., & VandeWiele, L. (2011). An evaluation of the individualized learning intervention: A mentoring program for early childhood teachers. *Early Child Development and Care, 14*(2), 57-74.

Gardner, F., Montgomery, P. & Knerr, W. (2016) , Transporting Evidence-Based Parenting Programs for Child Problem Behavior (Age) Between Countries: Systematic Review and Meta-Analyses, Journal of Clinical Child & Adolescent Psychology, : 749-762. *45*(6 SRC - GoogleScholar), 3-10.

* Garet, M. S., Cronen, S., Eaton, M., Kurki, A., Ludwig, M., Jones, W., . . . Doolittle, F. (2008). The Impact of Two Professional Development Interventions on Early Reading Instruction and Achievement. NCEE 2008-4030. *National Center for Education Evaluation and Regional Assistance, U.S. Department of Education*.

Gaudin, C., & Chalies, S. (2015). Video viewing in teacher education and professional development: A literature review. *Educational Research Review, 16*, 41-67. doi: 10.1016/j.edurev.2015.06.001

Goodman, R. (1997). The Strengths and Difficulties Questionnaire: a research note. *Journal of child psychology and psychiatry, 38*(5), 581-586.

Greve, J., & Kressel, K. F. (2017). PISA Etnisk. Hvordan elever med indvandrerbaggrund klarer sig i PISA-testen og deres holdninger og forventninger til naturvidenskab. *København: KORA – Det Nationale Institut for Kommuners og Regioners Analyse og Forskning*.

Harmes, T., Clifford, R., & Cryer, D. (1998). Early childhood environment rating scale-revised: New York: Teachers College Press.

Hedges, L. V., & Cooper, H. M. (Eds.). (1994). *The handbook of research synthesis*: Russell Sage Foundation.

Higgins, J. P. T., & Green, S. (editors). (2011). *Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011],* The Cochrane Collaboration, Retrieved from www.handbook.cochrane.org

Hill, H. C., Schilling, S. G., & Ball, D. L. . (2004). Developing measures of teachers' mathematics knowledge for teaching. *The Elementary School Journal, 105*(1), 11-30.

* Hindman, A. H., & Wasi, B. A. (2012). Unpacking an Effective Language and Literacy Coaching Intervention in Head Start: Following Teachers' Learning over Two Years of Training. *Elementary School Journal, 113*(1), 131-154.

* Howlin, P., Gordon, R. K., Pasco, G., Wade, A., & Charman, T. (2007). The effectiveness of Picture Exchange Communication System (PECS) training for teachers of children with autism: a pragmatic, group randomised controlled trial. *J Child Psychol Psychiatry, 48*(5), 473-481. doi: 10.1111/j.1469-7610.2006.01707.x

Hwang, Y. S., Bartlett, B., Greben, M., & Hand, K. (2017). A systematic review of mindfulness interventions for in-service teachers: A tool to enhance teacher wellbeing and performance. *Teaching & Teacher Education, 64*, 26-42.

Informatics, Cochrane. (2016). Knowledge Management Department. RevMan 5 download and installation.

* Jensen, B., Jensen, P., & Rasmussen, A. W. (2015). Does Professional Development of Preschool Teachers Improve Child Socio-Emotional Outcomes? Bonn: IZA. Institute for the Study of Labor / Forschungsinstitut zur Zukunft der Arbeit.

* Kammermeyer, G., Stuck, A., & Roux, S. (2016). Promotion of literacy and numeracy in pyramid classrooms in Germany. *Early Child Development and Care, 186*(1), 153-172. doi: 10.1080/03004430.2015.1052424

Kelcey, B., & Phelps, G. (2013). Considerations for Designing Group Randomized Trials of Professional Development With Teacher Knowledge Outcomes. *Educational Evaluation and Policy Analysis, 35*(3), 370-390. doi: 10.3102/0162373713482766

La Paro, K.M., & Pianta, R.C. (2003). CLASS: Classroom assessment scoring system. *Charlottesville: University of Virginia*.

Lander, N, Eather, N., Morgan, P. J., Salmon, J., & Barnett, P. A. (2017). Characteristics of Teacher Training in School-Based Physical Education Interventions to Improve Fundamental Movement Skills and/or Physical Activity: A Systematic Review. *SPORTS MEDICINE, 47*(1), 135-161. doi: 10.1007/s40279-016-0561-6

* Landry, S. H., Anthony, J. L., Swank, P. R., & Monseque-Bailey, P. (2009). Effectiveness of Comprehensive Professional Development for Teachers of At-Risk Preschoolers. *Journal of Educational Psychology, 101*(2), 448-465. doi: 10.1037/a0013842

Lipsey, M. W., & Wilson, D. B. (2001). Practical meta-analysis. *Applied Social Research Methods Series v 49*. Thousand Oaks, Calif.: SAGE Publications, Inc.

Markussen-Brown, J., Carsten, B., Shayne, B., Anders, H., Laura, M., & Dorthe, Piasta. (2017). The effects of language-and literacy-focused professional development on early educators and children: A best-evidence meta-analysis Early Childhood Research Quarterly: *38*(1), 97-115.

* McCollum, J. A., Hemmeter, M. L., & Hsieh, W. Y. (2013). Coaching Teachers for Emergent Literacy Instruction Using Performance-Based Feedback. *Topics in Early Childhood Special Education, 33*(1), 28-37. doi: 10.1177/0271121411431003

* McMeeking, L.B., Orsi, R., & Cobb, R.B. (2012). Effects of a Teacher Professional Development Program on the Mathematics Achievement of Middle School Students. *Journal for Research in Mathematics Education, 43*(2), 159-181.

Moher, D., Liberati, A., Tetzlaff, J., & Altman, D.G. (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *PLoS Med, 6*(7), e1000097. doi: doi:10.1371/journal.pmed.1000097

* Murray, D. W., Rabiner, D. L., & Carrig, M. M. (2012). Grade Level Effects of the Incredible Years Teacher Training Program on Emotion Regulation and Attention: Society for Research on Educational Effectiveness.

* Neuman, S. B., & Cunningham, L. (2009). The Impact of Professional Development and Coaching on Early Language and Literacy Instructional Practices. *American Educational Research Journal, 46*(2), 532-566. doi: 10.3102/0002831208328088

* Ottmar, E. R., Rimm-Kaufman, S. E., Berry Robert, Q., & Larsen Ross, A. (2013). Does the Responsive Classroom Approach Affect the Use of Standards-Based Mathematics Teaching Practices?: Results from a Randomized Controlled Trial. *Elementary School Journal, 113*(3), 434-457.

Pianta, R. C., La Paro, K. M., & Hamre, B. K. (2006). Classroom assessment scoring system: Manual k-3 version. *Charlottesville, VA: Center for Advanced Study of Teaching and Learning, University of Virginia*.

* Pianta, R. C., Mashburn, A.J., Downer, J.T., Hamre, B.K., & Justice, L. (2008). Effects of web-mediated professional development resources on teacher–child interactions in pre-kindergarten classrooms. *Early Childhood Research Quarterly, 23*(4), 431-451.

* Piasta, S. B., Justice, L.M., Cabell, S.Q., Wiggins, A.K., Turnbull, K.P., & Curenton, S.M. (2012). Impact of professional development on preschool teachers' conversational responsivity and children's linguistic productivity and complexity. *Early Childhood Research Quarterly, 27*(3), 387-400. doi: 10.1016/j.ecresq.2012.01.001

* Powell, D. R., Diamond, K.E., Burchinal, M.R., & Koehler, M.J. (2010). Effects of an early literacy professional development intervention on head start teachers and children. *Journal of Educational Psychology, 102*(2), 299-312. doi: 10.1037/a0017763

* Raver, C. C., Jones, S. M., Li-Grining, C. P., Metzger, M., Smallwood, K., & Sardin, L. (2008). Improving Preschool Classroom Processes: Preliminary Findings from a Randomized Trial Implemented in Head Start Settings. *Early Child Res Q, 63*(3), 253-255. doi: 10.1016/j.ecresq.2007.09.001

Reeves, B. C., Deeks, J. J., Higgins, J. P. T., & Wells, G. A. (2011). *Including non-randomized studies*. Chapter 13 In: Higgins, J.P.T. & Green, S. (eds.), Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 (updated March 2011). The Cochrane Collaboration.

* Rubie-Davies, C. M., & Rosenthal, R. (2016). Intervening in teachers' expectations: A random effects meta-analytic approach to examining the effectiveness of an intervention. *Learning and Individual Differences, 50*, 83-92. doi: 10.1016/j.lindif.2016.07.014

* Saraniero, P., Goldberg, M. R., & Hall, B. (2014). "Unlocking My Creativity": Teacher Learning in Arts Integration Professional Development. *Journal for Learning Through the Arts, 10*(1).

* Scanlon, D. M., Gelzheiser, L. M., Vellutino, F. R., Schatschneider, C., & Sweeney, J. M. (2008). Reducing the Incidence of Early Reading Difficulties: Professional Development for Classroom Teachers vs. Direct Interventions for Children. *Learn Individ Differ, 18*(3), 346-359. doi: 10.1016/j.lindif.2008.05.002

* Schwanenflugel, P. J., Hamilton, C. E., Neuharth-Pritchett, S., Restrepo, M. A., Bradley, B. A., & Webb, M. Y. (2010). PAVEd for Success: An Evaluation of a Comprehensive Preliteracy Program for Four-Year-Old Children. *Journal of Literacy Research, 42*(3), 227-275. doi: 10.1080/1086296x.2010.503551

Shadish, W. R., Cook, T. D., Campbell, T. D., & Boston, M. A. (2002). Experimental and Quasi-experimental Designs for Generalized Causal Inference. Boston & New York: *Houghton Mifflin*.

Smith, M. W., & Dickinson, D. K. (2002). Early Language & Literacy Classroom Observation (ELLCO) Toolkit, Research Edition [with] User's Guide.

Snell, M. E., Dowsell Forston, L., Stanton-Chapman, T. L., & Walker, V. L. (2013). A review of 20 years of research on professional development interventions for preschool teachers and staff. *Early Child Development and Care, 183*(7), 857-873.

Thomas, J., Brunton, J., & Graziosi, S. (2010). EPPI-Reviewer software for research synthesis. *EPPICentre Software London Social Science Research Unit of Education, 4 SRC - GoogleScholar*.

Timperley, H., Wilson, A., Barrar, H., & Fung, I. (2007). Teacher professional learning and development: Best evidence synthesis iteration. Wellington, New Zealand: Ministry of Education.

Torgerson, C. (2003). *Systematic Reviews*. London: Continuum.

Torgerson, C., Nielsen, C., Gascoine, L., Filges, T., Moore, I. , & Viinhold Nielsen, B.C. (2017). A Systematic Review (SR) of the Effective Continuing Professional Development (CPD) Training of Welfare Professionals. Retrieved November 2017, from https://en.sfi.dk/publications/a-systematic-review-sr-of-the-effective-continuing-professional-development-cpd-training-of-welfare-professionals-13324/

University of York. Centre for Reviews and Dissemination. (2009). *Systematic reviews: CRD's guidance for undertaking reviews in health care* Retrieved from https://www.york.ac.uk/crd/SysRev/!SSL!/WebHelp/SysRev3.htm

* Vernon-Feagans, L., Kainz, K., Hedrick, A., Ginsberg, M., & Amendum, S. (2013). Live Webcam Coaching to Help Early Elementary Classroom Teachers Provide Effective Literacy Instruction for Struggling Readers: The Targeted Reading Intervention. *J.ournal of Educational Psychology, 105*(4), 1175-1187. doi: 10.1037/a0032143

* Wasik, B. A., & Hindman, A. H. (2011). Improving Vocabulary and Pre-Literacy Skills of At-Risk Preschoolers Through Teacher Professional Development. *Journal of Educational Psychology, 103*(2), 455-469. doi: 10.1037/a0023067

Woodcock, R. W. (1991). *Woodcock Language Proficiency Battery-Revised: English and Spanish Forms: Examiner's manual.* Itasca, IL : Riverside Publishing.

Woodcock, R. W., & Muńoz-Sandoval. (1995). *Woodcock Language Proficiency Battery-Revised: Spanish Form*. Itastca, IL: Riverside.

Yoon, K. S., Duncan, T., Lee, S. W., Scarloss, B., & Shapley, K. L. . (2007). Reviewing the Evidence on How Teacher Professional Development Affects Student Achievement. Issues & Answers. REL 2007-No. 033. *Regional Educational Laboratory Southwest (NJ1)*.

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. *Office of Planning, Evaluation and Policy Development, US Department of Education*.

Appendices

Appendix A: Justification of exclusion of studies using an instrumental variable (IV) approach

Studies using instrument variables (IV) for causal inference will not be included, as the interpretation of IV estimates is challenging. IV only provides an estimate for a specific group, namely people whose behaviour changes due to changes in the particular instrument used. It provides no information about effects on never-takers and always-takers because the instrument does not affect their treatment status. The estimated effect is thus applicable only to the subpopulation whose treatment status is affected by the instrument. As a consequence, the effects differ for different IVs, and care has to be taken as to whether they provide useful information. The effect is interesting when the instrument on which it is based is interesting, in the sense that it corresponds to a policy instrument of interest. Furthermore, if those affected by the instrument are not affected in the same way the IV estimate is an average of the impacts of changing treatment status in both directions and cannot be interpreted as a treatment effect. To turn the IV estimate into a LATE requires a monotonicity assumption. The movements induced by the instrument go in one direction only, from no treatment to treatment. The IV estimate, interpreted as a LATE, is only applicable to the complier population, those that are affected by the instrument in the 'right way'. It is not possible to characterise the complier population, as an observation's subpopulation cannot be determined and defiers do not exist by assumption.

In the binary-treatment-binary-instrument context, the IV estimate can, given monotonicity, be interpreted as a LATE, i.e. the average treatment effect for the subpopulation of compliers. If treatment or instruments are not binary, interpretation becomes more complicated. In the binary-treatment–multivalued-instrument (ordered to take values from 0 to J) context, the IV estimate, given monotonicity, is a weighted average of pairwise LATE parameters (comparing subgroup j with subgroup j-1). The IV estimate can thus be interpreted as the weighted average of average treatment effects in each of the J subgroups of compliers. In the multivalued-treatment (ordered to take values from 0 to T)–multivalued-instrument (ordered to take values from 0 to J) context, the IV estimate for each pair of instrument values, given monotonicity, is a weighted average of the effects from going from t-1 to t for persons induced by the change in the value of the instrument to move from any level below t to the level t or any level above. Persons can be counted multiple times in forming the weights.

Bibliography

Angrist, J.D., & Pischke, J.S. (2009*). Mostly Harmless Econometrics: An Empiricist's Companion.* Princeton, NJ: Princeton University Press.

Buyesse, V. & P.J. Winton (2009). Reaching Consensus on a Definition of Professional Development for the Early Childhood Field. *Topics in Early Childhood Special Education*, 28 (4), 235-243.

Heckman, J.J. & Urzúa, S. (2010). Comparing IV with structural models: What simple IV can and cannot identify. *Journal of Econometrics, 156*, 27-37.

Heckman, J.J., Urzúa, S. & Vytlacil, E. (2006). Understanding instrumental variables in models with essential heterogeneity. *The Review of Economics and Statistics, 88*(3), 389-432.

Appendix B: Results of initial scoping search

Al Otaiba, S., Connor, C. M., Folsom, J. S., Greulich, L., Meadows, J., & Li, Z. (2011). Assessment Data-Informed Guidance to Individualize Kindergarten Reading Instruction: Findings from a Cluster-Randomized Control Field Trial. *Elem Sch J, 111*(4), 535-560. doi: 10.1086/659031

Álvarez, Pedro, De La Fuente, Emilia I, Perales, F Javier, & García, Juan. (2002). Analysis of a quasi-experimental design based on environmental problem solving for the initial training of future teachers of environmental education. *The Journal of Environmental Education, 33*(2), 19-21.

Antoniou, P., & Kyriakides, L. (2011). The impact of a dynamic approach to professional development on teacher instruction and student learning: results from an experimental study. *School Effectiveness and School Improvement, 22*(3), 291-311. doi: 10.1080/09243453.2011.577078

Arbolino, L. A., & Learning, U. S. (2007). *In-service training on child abuse for classroom teachers: What is the effectiveness of mandated training? (67), ProQuest Information & Retrieved from http://search.ebscohost.com/login.aspx?direct=true&db=psyh&AN=-034&site=ehost-live Available from EBSCOhost psyh database*.

Aronin, S. A., & Learning, U. S. (2010). *Integrating universal design for learning through content video with preservice teachers*.

Barr, D. J., Boulay, B., Selman, R. L., McCormick, R., Lowenstein, E., Gamse, B., . . . Learning, U. S. (2013). *A randomized controlled trial of professional development for interdisciplinary civic education: Impacts on humanities teachers and their students* (Vol. 117).

Butler, Michelyn Cynthia. (2012). *Implementation of Evidence-Based Book-Reading Strategies by Head Start Teachers: Benefits of Professional Development and Effect on Children's Literacy Outcomes*: ERIC.

Campanaro, M. (2009). *The effects of a professional development training protocol on teacher implementation of comprehension strategy instruction*.

Davis, C. A., & Learning, U. S. (2005). *Effects of in-service training on teachers' knowledge and practices regarding identifying and making a focus of concern students exhibiting internalizing problems*.

Ellis, H. A. (2014). Effects of a Crisis Intervention Team (CIT) training program upon police officers before and after Crisis Intervention Team training. *Arch Psychiatr Nurs, 28*(1), 10-16. doi: 10.1016/j.apnu.2013.10.003

Ellis-Mills, P. E., & Learning, U. S. (1996). *Benefits of training sixth-grade Bahamian teachers to use conflict resolution techniques in the classroom*.

Filer, A. F., & Learning, U. S. (2016). *Investigating the use of traditional and online instruction for teachers of children with Autism Spectrum Disorder: A case for blending training models*.

Foltz, S. P. (2000). *The effect of police academy training on self-esteem and locus of control in law enforcement recruits*.

Frank, G. (2007). *The effect of elementary teachers' professional development activities intended to increase the number of migrant student nominations for gifted and talented programs*.

Lane, Cristianne, Prokop, Mary Jo Surges, Johnson, Evelyn, Podhajski, Blanche, & Nathan, Jane. (2013). Promoting early literacy through the professional development of preschool teachers. *Early Years, 34*(1), 67-80. doi: 10.1080/09575146.2013.827157

Powell, Ed. (2005). Conceptualising and facilitating active learning: teachers' video-stimulated reflective dialogues. *Reflective Practice, 6*(3), 407-418. doi: 10.1080/14623940500220202

Roeser, R. W., Schonert-Reichl, K. A., Jha, A., Cullen, M., Wallace, L., Wilensky, R., . . . Harrison, J. (2013). Mindfulness Training and Reductions in Teacher Stress and Burnout: Results From Two Randomized, Waitlist-Control Field Trials. *JOURNAL OF EDUCATIONAL PSYCHOLOGY, 105*(3), 787-804. doi: 10.1037/a0032093

Skoretz, Y. M., & Learning, U. S. (2011). *A study of the impact of a school-based, job-embedded professional development program on elementary and middle school teacher efficacy for technology integration*.

Sluijsmans, Dominique M. A., Brand-Gruwel, Saskia, & van Merriënboer, Jeroen J. G. (2002). Peer Assessment Training in Teacher Education: Effects on performance and perceptions. *Assessment & Evaluation in Higher Education, 27*(5), 443-454. doi: 10.1080/0260293022000009311

Tao, X., Chongde, L., & Jiliang, S. (1999). Effect of cognitive self-instruction training on the improvement of teachers' teaching-regulated ability. *Psychological Science China, 22*(1 SRC - GoogleScholar), 5-9.

Wasik, B. A., & Hindman, A. H. (2011). Improving Vocabulary and Pre-Literacy Skills of At-Risk Preschoolers Through Teacher Professional Development. *JOURNAL OF EDUCATIONAL PSYCHOLOGY, 103*(2), 455-469. doi: 10.1037/a0023067

Appendix C: Final search strategies and tracking of records from each database searched

C1: Final search strategies

Web of Science (Social Science Citation Index & Science Citation Index)

1997-2017. Search performed 1/5/2017.

Search	Terms	Results
S7	#6 AND #5 AND #2 AND #1 *Indexes=SCI-EXPANDED, SSCI Timespan=1997-2017*	544
S6	#4 OR #3 *Indexes=SCI-EXPANDED, SSCI Timespan=1997-2017*	5,490,876
S5	TI=(teacher* OR pedagogue* OR school counsellor* OR social worker* OR police* OR psychologist* OR probation officer* OR family support OR support worker*) *Indexes=SCI-EXPANDED, SSCI Timespan=1997-2017*	46,106
S4	TS=(profess* OR develop* OR learn* OR train* OR supervis* OR feedback OR team work OR education*) *Indexes=SCI-EXPANDED, SSCI Timespan=1997-2017*	5,490,876
S3	TI=(profess* OR develop* OR learn* OR train* OR supervis* OR feedback OR team work OR education*) *Indexes=SCI-EXPANDED, SSCI Timespan=1997-2017*	1,055,729
S2	TI=(random* control* trial* OR rct* OR trial* OR review* OR intervent* OR meta analys*) *Indexes=SCI-EXPANDED, SSCI Timespan=1997-2017*	694,515
S1	TS=(child* OR student* OR adolescent* OR teen* OR preschool*) *Indexes=SCI-EXPANDED, SSCI Timespan=1997-2017*	1,419,905

References imported into Mendeley's database: 544

References imported to EPPI-reviewer after duplication in Mendeley: 544

EBSCO databases (ERIC, Academic Search Premier, PsycINFO, Teacher Reference Center, SocIndex)

1997-2017. Search performed 17/4/2017

Search	Terms	Results
S6	S1 AND S2 AND S3 AND S4 Limiters – Published date: 19970101-20171231	3,139
S5	S1 AND S2 AND S3 AND S4	4,626
S4	SU (child* OR student* OR adolescent* OR teen* OR preschool*)	2,981,315
S3	SU ((random* control* trial* OR rct* OR trial* OR review* OR intervent*) OR TI (random* control* trial* OR rct* OR trial* OR review* OR intervent*))	2,395,696
S2	SU ((continu* OR proffes* OR in service OR teacher*) AND (professional OR development OR learning OR training OR supervis* OR feedback OR team work OR education*))	1,101,911
S1	SU ((teacher* OR classroom* OR para* OR auxiliary) AND (teacher* OR assist* OR support* OR aid*) OR pedagogue* OR school counsellor* OR social worker* OR police* OR psychologist* OR probation officer* OR family support OR support worker*))	1,249,212

The individual results (after duplication) from each EBSCO database are listed below:

- *Academic Search Premier* – 528 imported to EPPI-reviewer after de-duplication in Mendeley
- *ERIC* –1385 imported to EPPI-reviewer after de-duplication in Mendeley
- *PsycINFO* – 791 imported to EPPI-reviewer after de-duplication in Mendeley
- *SocIndex* – 215 imported to EPPI-reviewer after de-duplication in Mendeley
- *Teacher Reference Center* – 156 imported to EPPI-reviewer after de-duplication in Mendeley.

ASSIA

1997-2017

Searches performed 4/5/2017

Search	Terms	Results
S6	1 AND 2 AND 3 AND 4 AND pd (19970101-20171231)	236
S5	1 AND 2 AND 3 AND 4	243
S4	su(((teacher* OR classroom* OR para* OR auxiliary (teacher* OR assist* OR support* OR aid*)) OR pedagogue* OR school counsellor* OR social worker* OR police* OR psychologist* OR probation officer* OR family support OR support worker*))	36, 690
S3	su(((continu* OR proffes* OR in service OR teacher* (professional OR development OR learning OR training OR supervis* OR feedback OR team work OR education*))))	9,347
S2	su(random* control* trial* OR rct* OR trial* OR review* OR intervent*) OR ti(random* control* trial* OR rct* OR trial* OR review* OR intervent*)	88,826
S1	su((child* OR student* OR adolescent* OR teen* OR preschool*))	198,896

References imported into Mendeley: 236

References imported to EPPI-reviewer after duplication in Mendeley: 223

C2: Grey Literature Search Strategies

Forskningsdatabasen (Danish National Research Database)

Searches performed 01/05/2017

professional develop training effect – 75 hits.
professional develop training random – 7 hits.
professional develop train control – 29 hits.
continued professional develop effect – 60 hits.
continued professional develop random – 9 hits.
continued professional develop control – 46 hits.
continued professional train effect – 25 hits.
continued professional train random – 3 hits.
continued professional train control – 8 hits.
"CPD" effect – 13 hits.
"CPD" trial – 2 hits.

References imported into Mendeley's database: 177

References imported to EPPI-reviewer after duplication in Mendeley: 171

SwePub (Academic content from Swedish universities)

Searches performed 01/05/2017

tit:(profess* develop* train*) – 14 hits.
tit:(profess* develop* effect*) – 3 hits.
tit:(profess* develop* control*) – 1 hits.
tit:(continu* profess* develop*) – 42 hits.
tit:(continu* profess* random*) – 1 hits.
tit:(continu* profess* control*) – 1 hits.
tit:(continu* profess* train*) – 1 hits.
Tit: ("CPD") – 10 hits.

References imported into Mendeley's database: 65

References imported to EPPI-reviewer after duplication in Mendeley: 65

NORA (Norwegian Open Research Archive)

Searches performed 01/05/2017

"professional development" AND effect – 12 hits.
professional development effect – 88 hits.
professional development trial – 10 hits.
"professional development" AND training – 11 hits.
"continued professional development" – 4 hits.
continued professional training – 14 hits.

References imported into Mendeley's database: 4

References imported to EPPI-reviewer after duplication in Mendeley: 4

Social Care Online

1997- 2017. Searches performed 01/05/2017

TITLE: continued professional OR professional*
AND
TITLE: train* OR learn* OR develop*
AND
AB: random* control* trial* OR rct* OR trial* OR review* OR intervent* OR effect*

References imported to EPPI-reviewer after duplication in Mendeley: 129

Google Scholar

1997-2017. Searches performed 02/05/2017

alleititel: professional effect develop OR learning OR training – 182 hits.
alleititel: continued professional develop OR learning OR training – 23 hits.
alleititel: professional intervention develop OR learning OR training – 81 hits.
alleititel: social work effect develop OR learning OR training – 10 hits.
alleititel: police effect develop OR learning OR training – 31 hits.
alleititel: continued professional develop OR learning OR training OR effect – 24 hits.
alleititel: continued professional development teacher OR psychologist OR police OR social OR worker – 5 hits.

References imported into EPPI (after external screening in Mendeley): 34

US Clearinghouse for educational research & Danish Clearinghouse for educational research

Searched performed 3/5/2017

We manually checked relevant repositories. Identified studies were already found in the database search, so none were imported into EPPI.

C3: Records included and excluded at all stages by database (* = grey literature)

Database	No. of records after de-duplication in EPPI	No. of records excluded at 1st stage screening	No. of records screened at full-text stage	No. of records not available	No. of records excluded at full-text screening (with reasons)	No. of records remaining after full-text screening	No. of *empirical* records forward to 3rd stage screening	No. of empirical records excluded in 3rd stage screening (with reasons)	No. of empirical records included in mapping data extraction
Web of Science	468	399	69	-	38	31	23	19	4
ASP	397	366	31	-	19	12	9	6	3
ERIC	1305	1251	54	-	24	30	26	19	7
PsycINFO	706	660	46	-	32	14	11	8	3
SocIndex	100	99	1	-	-	1	1	1	0
TRC	67	67	0	-	-	-	-	-	-
Forskningsdata-basen*	170	168	2	-	-	2	2	1	1
SwePub*	64	62	2	-	-	2	2	2	0
NORA*	4	4	0	-	-	-	-	-	-
ASSIA	205	183	22	-	12	10	9	7	2
Google Scholar*	33	27	6	2	3	1	1	1	0
SCO*	128	124	4	-	4	0	-	-	-
TOTAL (systematic searches)	**3647**	**3410**	**237**	**2**	**132**	**103**	**84**	64	20
TOTAL (citation searching from included SR/MA)	-	-	-	-	-	-	**22**	10	12
Expert reviewer SR/MA	-	-	-	-	-	2 (SR/MA)	**1**	-	1
TOTAL (ALL)	**3647**	**3410**	**237**	**2**	**132**	**103**	**106**	74	33

Appendix D: Contextualisation: Relevance and feasibility of interventions from the studies in the in-depth review in the Danish context

	Allen (US)	**Considerations about relevance and feasibility in the Danish context**	**Relevance**
Topic/focus	Motivation and engagement among secondary students through positive student-teacher interactions with a view to improving students' academic achievements	Relevant for DK, particularly since general upper secondary school (age 15-18) has become broader, and hence lack of motivation and disengagement are more prevalent	High
Setting **– Structural aspects**	Secondary schools, students age 11-18	Students age 11-15/16 would be in grades 5-9/10 in primary school setting. Age groups 15/16-18 would be in either general high school or vocational training	High
Sample	Secondary teachers teaching, 35% BA, 65% MA, 83% white, 8% African-American, 6% mixed, 3% other	Teachers of students in primary school settings hold a professional bachelor's degree, while teachers in general secondary school would hold a master's degree	Moderate-high
PD participation	Mandatory among participating schools subject to having chosen a focal course subject	Municipality level influence on what types of PD teachers participate in, but school management is also involved in these decisions	Moderate-high
Intervention	Workshop-based training, video library, brief booster workshop and 1 year of personalised coaching based on teacher sending video recordings of class sessions for review and feedback from trained teacher consultant, 20-30 min phone conference where consultant speaks to teacher on how to enhance interactions using CLASS.	Relevant	High
– Practical aspects	Participation in programme for 13 months	Relevant	High
– Content	Motivational and instructional qualities of teachers' daily interactions with students	Relevant	High
– Cultural aspects	The "My Teaching Partner -Secondary" programme involves coaching from an external "partner"	Relevant	High
– Teacher autonomy	Not much room for flexible implementation, e.g. videotaping had to follow a standard protocol	High degree of teacher autonomy in DK, although also some use of standardised methods	Moderate-high
Data collection	Teachers had to facilitate collection of student self-report data, video recordings of own teaching to be coded by external coders	Relevant	High
Outcomes in focus	CLASS-S (secondary): Emotional Support, Classroom Organization, and Instructional Support, state testing system to capture academic achievement data	Relevant	High
Overall judgement: High			

	Domitrovich (US)	Considerations about relevance and feasibility in the Danish context	Relevance
Topic/focus	Fostering high-quality teaching with enriched curriculum and professional development support, school readiness of children from disadvantaged families	Relevant for DK	High
Setting **– Structural aspects**	Head Start classrooms target children from disadvantaged families in Pennsylvania, rural and urban, most centres small (1-2 classrooms) but 4 larger (3-5 classrooms)	Head Start classrooms settings in DK obviously differ from those in the US, but could be relevant in certain school districts in larger cities	Moderate
Sample	Lead and assistant teachers, > 80% Caucasian, > 95% English as primary language, most lead teachers had a 4-year degree, most assistant teachers had high school or some post-high school education, most lead teachers had Child Development Associate credential or teaching certificate.	Level of education similar to DK, although differences with regards to certification traditions	High
PD participation	Not clearly stated	-	-
Intervention	4 days of workshop training, weekly in-class support from a mentor teacher, Head Start REDI (Research-based Developmentally Informed)	Relevant	High
– Practical aspects	3-day training workshop, midway 1-day "booster workshop, weekly mentoring sessions from REDI trainers, teachers received detailed manuals and kits containing all materials needed to implement intervention	Relevant	High
– Content	Specific curriculum targeting children's emergent language/literacy skills (vocabulary, syntax, phonological awareness, print awareness), and socio-emotional development, PD focuses on improving teacher's language use, emotional support and positive behaviour-management strategies, teachers received mentoring in use of "language coaching" strategies; Preschool PATHS Curriculum to promote children's socio-emotional skills (e.g. friendship skills, emotional understanding and expression, self-control, conflict-solving)	Relevant, particularly in certain school districts in larger cities	High
– Cultural aspects	Focus on high quality early childhood care, particularly for children from disadvantaged families, with the intention of narrowing the achievement gap that is associated with differences in family background	Relevant	High

	Domitrovich (US)	Considerations about relevance and feasibility in the Danish context	Relevance
– Teacher autonomy	Limited flexibility for teachers, very specific activities in programme, e.g. use a 10-15 min Sound Game activity at least 3 times/week	High degree of autonomy among pedagogues in DK, although also some use of standardised methods	Moderate-high
Data collection	REDI implementation quality monitored through teacher self-reports, REDI trainers observed curriculum components at least once a month and rated fidelity	Relevant	High
Outcomes in focus	Teaching quality assessed using Classroom Assessment Scoring System (CLASS), the Teaching Style Rating Scale (TSRS) and the Classroom Language and Literacy Environment Observation (CLEO)	Relevant	High
Overall judgement: High			

	Early (US)	Considerations about relevance and feasibility in the Danish context	Relevance
Topic/focus	General teacher-child interactions, without focus on any specific content area	Relevant	High
Setting **– Structural aspects**	Preschool for 4-year-olds (Pre-K) in a variety of settings, incl. private child care, local schools, Head Start centres, military bases, non-profit programmes, universal, yet 55% of children are from low-income families, more than half of the classrooms were in private settings	4-year-olds in DK are in kindergartens (for children aged 3-6) or in integrated day-care facilities (for age 0-6), public/private (independent) structure will differ from US structure, no equivalent to Head Start centres in DK, but could be relevant for institutions in low-income districts	Moderate-high
Sample	Pre-K teachers at least in their second year of teaching, 65.2% BA/BS, 26.1% MA/MS, Ph.D., 8.7% less than BA/BS	Level of education among staff in Danish day care institutions serving 4-year-olds is lower, i.e. more unskilled, virtually none with MA/MS, Ph.D.	Moderate-high
PD participation	Mandatory for this trial (random selection at school/centre level)	Municipality level influence on what types of PD teachers participate in, but school management is also involved in these decisions	Moderate-high
Intervention	2 PD models that combine skills training with in-service coaching or consultation: My Teaching Partner (MTP) and Making the Most of Classroom Interactions (MMCI)	Relevant	High

	Early (US)	Considerations about relevance and feasibility in the Danish context	Relevance
– Practical aspects	MMCI: Groups of teachers meet face-to-face with instructors, 5 full-day sessions (10 work-shops), access to print and web-based resources, home-work assignments; MTP: one-to-one consultations, remote coaching, sends self-video re-cordings to coach, receives spe-cific feedback Instruction/coach-ing delivered by the state Pre-K consultants (not researchers or teachers); Participation for one academic year	Relevant	High
– Content	Focus is on instructional support as well as classroom climate and management	Relevant	High
– Cultural aspects	Authors suggest that the finding that MMCI was more effective among less educated teachers is due to novelty and openness to change.	Staff in DK day care are less educated, but have less experience with research-based knowledge, since their initial train-ing is not at a university	Moderate
– Teacher autonomy	Limited flexibility for teachers to influence content during inter-vention	High degree of teacher autonomy in DK, although also some use of standardised methods	Moderate-high
Data collection	Independent data collectors conducted CLASS observation in classroom, 6 x 30 minutes	Relevant	High
Outcomes in focus	CLASS: Emotional Support, Classroom Organisation, and Instructional Support	Relevant	High
Overall judgement: Moderate-high			

	Fukkink (Netherlands)	Considerations about relevance and feasibility in the Danish context	Relevance
Topic / focus	Interaction skills of early childhood educators and care teachers	Relevant	High
Setting **– Structural aspects**	Childcare providers that use the Video Interaction Guidance (VIG) method on a large scale in their organisation	Relevant in a trial setting, but may require further introduction to method and practicalities, if institution is not used to using VIG	Moderate-high
Sample	Caregivers/teachers in day care centres and other childcare centres, staff (average age 28 years), work experience (5 years), education level not mentioned	59% of staff in day care institutions are trained pedagogues, the rest are unskilled	Uncertain
PD participation	Not mentioned	-	-
Intervention	Video recordings of approx. 4 x 10-15 minutes of teacher working with children, trainer-selected fragments from which feedback is discussed with teacher, unclear who trainer should be	Relevant	High
– Practical aspects	Short filming sessions, which are comparable in duration to other video feedback interventions	If institution is not accustomed to regularly using VIG, could be disturbing for staff and children	Moderate
– Content	Training teachers to act in ways that show they recognise and react to children's contact initiatives in positive verbal and non-verbal ways, i.e. primarily sensitivity and emotional support, but also verbal simulation	Relevant	High
– Cultural aspects	Focus in VIG training model: interactions are successful when initiatives of the child are received and responded to in a positive manner	In line with strong focus on promoting autonomy of young children and positive teacher-child relations	High
– Teacher autonomy	After videotaping sessions: short questionnaire for teacher (and person filming) about representativeness of video clip	Relevant to respect teacher's possibility of self-rating representativeness	High
Data collection	Teachers contribute by allowing videotaping and filling in short questionnaires (representativeness of situation video recorded and job satisfaction)	Impression that it is not very common for PD to take place through VIG, but probably feasible in a trial setting, responding to questionnaires seems relevant and acceptable	Moderate-high
Outcomes in focus	Behaviour of caregivers and their interactions with children rated on a four-point scale: stimulating and authoritarian caregiving (+ other intermediate outcomes and final outcomes not relevant for SR)	Relevant	High
Overall judgement: Moderate-high			

	Murray (USA)	Considerations about relevance and feasibility in the Danish context	Relevance
Topic/focus	Effects of Incredible Years Teacher (IYT) classroom management programme on students' emotion regulation, attention and academic achievement	Relevant, IYT is used in DK	High
Setting **– Structural aspects**	Schools in rural south-eastern US with lower than average level of school resources with difficulty recruiting and retaining qualified teachers, highly diverse with regards to race/ethnicity (54% Caucasian, 22% Hispanic, 17% Afro-Amer.), poverty, achievement levels, class size range 16-24	Some regions are challenged in terms of recruiting and retaining qualified staff, public school structure is moving in direction of larger schools, at the same time smaller private schools are emerging, general finding in DK that there is no systematic connection between level of school resources and academic achievement, very different ethnic/racial structure	Low
Sample	97 K-2 teachers, primarily female and Caucasian, 30% hold a master's degree	Teachers in DK have a professional bachelor's degree and are predominantly female and of native Danish ethnicity	Low-moderate
PD participation	Participation upon teacher consent and randomisation	Municipalities level influence on what types of PD teachers participate in, IYT being one of them	High
Intervention	Video-modelling, behavioural rehearsal of key skills through role play, classroom practice assignments, teacher goal-setting and self-monitoring	Relevant approach, positive self-modelling in line with high degree of autonomy of pedagogical staff in DK	High
– Practical aspects	5 full-day workshops led by 2 trained co-leaders with 12-15 teachers in each group, teachers received 2 brief consultation visits in classroom, regular emails to support implementation. Trained researcher assistants observed each teacher's classroom for approx. 2 hrs.	Feasible mode of implementation	High
– Content	The building of positive relationships with students and parents, proactive classroom management strategies, use of incentives, "coaching" students social and emotional development, teaching calm-down and problem-solving, positive discipline techniques	Relevant for DK, where focus in past years has been – and continues to be – on effective classroom management to ensure good quality learning environments for children	High
– Cultural aspects	IYT is an acknowledge classroom management programme, which has demonstrated impact on young children's social emotional impact, natural to study possible impact on academic achievement	IYT well-known programme in DK, strong belief among teachers that socio-emotional well-being is a prerequisite for academic learning	High
– Teacher autonomy	Approach builds upon high degree of teacher autonomy	Relevant, consideration should be given to the time spent on self-monitoring	High
Data collection	Teacher ratings of students' emotion regulation, pro-social behaviour, inattention, teacher ratings of students' academic competence	Relevant, positive to engage teachers in rating students themselves in addition to academic achievement tests	High

	Murray (USA)	Considerations about relevance and feasibility in the Danish context	Relevance
Outcomes in focus	CLASS used to observe teacher practice, teacher ratings of students (see above), STAR Early Literacy/Reading and Math computerized assessment (nationally normed, adaptive test)	Social and emotional skills are highly valued alongside academic development, national tests	High
Overall judgement: High			

	Ottmar (US)	Considerations about relevance and feasibility in the Danish context	Relevance
Topic/focus	Impact of the Responsive Classroom (RC) approach and RC practices on classroom social interactions and teachers' use of standards-based mathematics teaching practice	Relevant, yet common goals in DK are primarily directional and intended as guidelines	Moderate
Setting – Structural aspects	Objectives in the Common Core State Standards Initiative (CCSSI) reflect both the mathematics content and the necessary social and self-regulatory interactions that contribute to students' mathematics learning	Common goals formulated by the Ministry of Education that stipulate the overall competency goals that students are expected to meet at each grade level in each subject, the common goals are intended as a tool and a guide for the teacher	Moderate-high
Sample	Teachers of 3rd grade students, racial composition of teachers (83% White, 6.8% African American, 1.1% Hispanic, 9.1% other), 30.7% BA, 69.3% MA or other graduates	Slightly different ethnic composition among teachers in DK, more substantial is apparent difference in level of education	Low-moderate
PD participation	Teachers (*n*=100) at participating schools were invited to participate, 94% accepted	Relevant to invite teachers	High
Intervention	Social and emotional learning intervention designs to promote pro-social skills and assertion in the classroom	Relevant	High
– Practical aspects	Intensive training, 1-week (35 hours), 1-day workshop, 3 training sessions with RC coach, observations in classroom, recommendations, email and phone communication, coach also conducted a lesson in teacher's class, held debriefing sessions and mini-workshops	Feasible mode of implementation, but time requirement could be a barrier	Moderate
– Content	Learn strategies for implementing key practices of RC approach, e.g. morning meeting, rule creation, interactive modelling, positive teacher language, logical consequences	Relevant	High
– Cultural aspects	Focus on promoting the cognitive, social and organizational skills needed to implement standards-based mathematics teaching practices, focus on the process of learning (not only the product), focus on the importance of social interactions to support cognitive development	Relevant focus for DK, although it would not be common to consider this only in mathematics, but more generally	Moderate-high

	Ottmar (US)	Considerations about relevance and feasibility in the Danish context	Relevance
– Teacher autonomy	Manuals-based, very specific intervention, not much flexibility for teachers	Moderately relevant, high degree of teacher autonomy in DK, although some use of manual-based approaches	Moderate
Data collection	Observational data regarding teacher's instructional practice collected by trained coders, fidelity of implementation, teacher efficacy and mathematical knowledge. Plus, online teacher self-report questionnaires regarding demographic and teaching beliefs information.	Observation of teacher practice would typically be formative and by colleagues, supervisors or school principals	Moderate
Outcomes in focus	Teachers' use of standards-based mathematics teaching practices using Mathematics Scan (M-Scan), 8 specific dimensions	Since common goals in DK are primarily directional and intended as guidelines, it would be questionable whether observations of teachers' adherence to these common goals would be relevant in their own right. Could instead be framed as a research-based approach to generating greater learning of mathematics among students	Low-moderate
Overall judgement: Moderate			

	Raver (USA)	Considerations about relevance and feasibility in the Danish context	Relevance
Topic/focus	Emotional climate in preschool classrooms as a prerequisite for young children's socioemotional development and early learning	Relevant for DK where this view on child development and learning is shared	High
Setting **– Structural aspects**	Head Start settings in Chicago for children 3-5 years old, site selection criteria incl. receipt of HS funding, high-poverty neighbourhoods, described as an intervention in a community-based context with limited financial resources available to support disadvantaged families	Head Start classrooms settings in Chicago obviously different from deprived neighbourhoods in DK, but could be relevant in certain districts of larger cities	Moderate
Sample	70% of teachers were African American, 20% were Latino, and 10% were European American; approx. ½ of teachers had an associate's degree, > ¼ had a high school degree or some college experience, < ¼ had BA or higher	Very different ethnic composition of teachers and somewhat higher level of education, although also use of unskilled staff in kindergartens	Low
PD participation	Self-nomination at site level, teachers at site invited to participate – contingent on giving their consent to being observed, to "host" a CSRP intervention staff member (mental health consultant or teacher's aide) in their classroom for entire school year	Relevant	High

	Raver (USA)	Considerations about relevance and feasibility in the Danish context	Relevance
Intervention	Chicago School Readiness Project (CSRP): teacher training paired with intensive, on-site mental health consultation with social workers providing capacity-building for teachers and mental health services for children, mental health consultant (MHC) present in classrooms 1 morning/week, MHC work manual-based as coach for teacher, but also free to provide child-focused consultation with small number of children	All schools in DK have (general) health workers, whom students have consultations with at regular intervals during their time in school, there are also school psychologists, and some schools also have a social worker affiliated with the school	High
– Practical aspects	5 training sessions on Saturdays, each lasting 6 hours, financial reimbursement per hour, 75% of teachers participated in at least 1 training session, 63% participated in more than half of the offered trainings	PD typically takes place during ordinary working hours, receiving ordinary pay	Low-moderate
– Content	Behaviourally and evidence-based teacher training package, adaptation of Incredible Years teacher training module	Relevant, Incredible Years is a familiar programme in DK	High
– Cultural aspects	Need to address children's disruptive behaviour and teachers' ability to create emotionally positive classroom climates	Relevant, focus on inclusion in DK	High
– Teacher autonomy	Manual-based, MHC act as coaches for teachers, relatively low degree of flexibility for teachers	High degree of autonomy among pedagogues in DK, although also some use of standardised methods	Moderate-high
Data collection	12 trained blind observers collected classroom-level data using Classroom Assessment Scoring System (CLASS) and the Early Childhood Environment Rating Scale (ECERS-R)	Relevant	High
Outcomes in focus	Teachers' classroom management practices	Relevant	High
Overall judgement: Moderate			

	Rubie-Davies (New Zealand)	Considerations about relevance and feasibility in the Danish context	Relevance
Topic/focus	High teacher expectations to both low and high achievers	Relevant, aim of school system reform (2014) is to challenge all students, so that they achieve the best they can	High
Setting – Structural aspects	Elementary teachers typically group students (within class) by ability in reading and mathematics	Grouping by ability is not common practice, yet large degree of flexibility in grouping	Moderate-high
Sample	66% of teachers at elementary school (yrs. 1-6) and 33% at middle school (yrs. 7-8). Student age range 6-13. Ethnic: Maori, Pacific Isl., Asian	1st-7th grade (age 6/7-12/13) Ethnic: Turkish, Pakistani, Somali etc.	Low-moderate
PD participation	Required annually, school decides which type of training	Municipality level influence on what types of PD teachers participate in	Moderate-high
Intervention	Self-analysis of video while teaching aspects from workshop, individually or in groups	Relevant	High
– Practical aspects	Workshops, meeting with researchers, network with teachers from other schools	Relevant	High
– Content	Research-based knowledge about influence of teacher expectations on student achievement, flexible grouping vs. ability grouping, goal setting through test-based feedback, ways to enhance class climate to enhance well-being	Relevant	Moderate-high
– Cultural aspects	Goal setting, e.g. through use of online assessment tool	Relevant, but might need to consider which assessment tool is most meaningful for teachers to use	Moderate-high
– Teacher autonomy	Teachers allowed to implement changes they felt comfortable with, not a 'scripted' intervention	Relevant	High
Data collection	Voluntary use of online assessment tool by teachers to collect data on student performance, teachers are used to using the tool to create tests specific to their needs	Use of online assessment tools is increasing, but varies a lot across schools and teachers, there are mandatory national tests but in practice theses are used primarily for progress monitoring rather than as pedagogical tool	Moderate-high
Outcomes in focus	Student achievement in reading and mathematics, teachers learn to measure and depict class climate in sociogram but not used as outcome	Relevant	High
Overall judgement: Moderate			

Appendix E: Risk-of-Bias Assessment

Author	Allen et al.	Domitrovich et al.	Early et al.
Year	2011	2009	2017
Country	USA	USA	USA
Sequence generation (Judgement)	Unclear	Unclear	High
Sequence generation (Description, quote from paper or describe key information)	Not described	Stratified (by location, length of programme student proportion of minority and Spanish-speaking children and centre size) groups, p. 573. Centres are randomised. No details given about the sequence generation	Method not described. P. 61 (2 and 3-year assignment) Classes, rather than teachers, were selected for participation because often teachers were not assigned to classrooms until very close to the start of the academic year, and occasionally teachers were not assigned until after the school year had begun. Thus, the final step in the random selection and assignment process involved finding out which teacher was assigned to the selected classroom, determining whether she or he was eligible for participation and replacing any classes where the teacher was ineligible. This is not true randomisation; randomisation occurs too soon. Not sure it could/should be classified as a RCT for these two years (and as they do not show results separately for the first year this concerns the whole trial)
Allocation concealment (Judgement)	Unclear	Unclear	High
Allocation concealment (Description, quote from paper or describe key information)	Allocation concealment is not clearly described. (pp. 2-3 in SOM)	Not mentioned	
Blinding (Judgement)	3	3	2
Blinding (Description, quote from paper or describe key information)	Not entirely clear, but since the testing system used is part of an official state testing system, it should be free from potential teacher bias.	"no overlap between staff working on the research team collecting observations of teachers and those working on the intervention team" (p. 574)	Independent data collectors conducted CLASS observations at start and end of school year. Different data collectors for autumn and spring obs. Pp. 62-63. Data collectors were unaware of project design, blind to teachers' PD conditions.
Incomplete outcome data addressed (Judgement)	3	2	2

Author	Allen et al.	Domitrovich et al.	Early et al.
Incomplete outcome data addressed (Description, quote from paper or describe key information)	Attrition: According to the text, 78 teachers participated in the intervention year, i.e. were randomised. In Table S1, total number is 76. In post-intervention year (study is designed to measure effects one year after intervention has been completed and with new/different students), 61 teachers had both fully participated in evaluation and used end-of-year achievement tests for students. (SOM p. 8), i.e. attrition is 15 teachers out of 76. Reasons for attrition at this stage includes teaching a course with no end-of-year test, residential moves, declining further participation due to other tasks. Authors claim no selective attrition at this stage, see SOM p. 8.	P. 574: pre-intervention baseline observations only available for 57 (68%) of the 84 initially participating teachers due to staff turnover. Final analysis was not restricted to these 57 teachers, end-of-year observations were available for all classrooms and 84 teachers.	27 teachers, or 5.3% of the original sample, left the study between the pre- and post-test. Of these, 8 had been assigned to MMCI (4.4% of the original MMCI sample); 8 to MTP (5.0% of the original MTP sample), and 11 to control (6.4% of the original control sample). Thus, the differential attrition rate was 2.0% (6.4 minus 4.4), p. 61. Low attrition. Reasons for attrition are stated on p. 61 (most stopped teaching in Georgia's Pre-K during that year), and it is further stated that no differences were found in any of the three CLASS domains on pre-test scores. Results not shown, however.
Free of selective reporting (Judgement)	3	2	3
Free of selective reporting (Description, quote from paper or describe key information)	Intention-to-treat analysis mentioned at top of p. 10 in SOM, but results not reported	Teachers recruited over 2 consecutive years (p. 574). No data shown separated by cohort, yet cohort is included in the analyses (along with setting) but no coefficients are reported.	No descriptive of attritors, and results on differences in pre-test scores not shown (only mentioned that there were no differences). A sensitivity check in which the main impact analyses were repeated including only the teachers in the second and third year of the project. Not all results reported, and results separated by each year should have been reported.
Free of other bias (Judgement)	1	4	2
Free of other bias (Description, quote from paper or describe key information)		Centres are randomised, and all the classrooms in the same centre are allocated to the same group. Teachers recruited over 2 consecutive years (p. 574). There was a staff turnover of at least 27 teachers. Newly recruited teachers may have known the treatment status of the centre.	P. 64: In the first year of the study, teachers in the control group (n = 51) had access to the same online library of video clips demonstrating best practices in various aspects of teacher-child interactions as the MMCI and MTP teachers. No data are available regarding how much those teachers accessed the library, but anecdotal evidence suggested that it was used very little. In the second and third years, teachers in the control group (n = 109) participated in the same 15 hours of professional development required of all Georgia's Pre-K teachers. Topics varied, but included behaviour management, child assessment, outdoor learning and others. Teachers in the control group did not receive direct training related to the CLASS, although some of the professional development opportunities may have been aligned with CLASS concepts.

Author	Allen et al.	Domitrovich et al.	Early et al.
A priori protocol (Judgement)	Unclear	Unclear	Unclear
A priori protocol (Description, quote from paper or describe key information)	Not mentioned	Not mentioned	Not mentioned
A priori analysis plan (Judgement)	Yes	Unclear	Unclear
A priori analysis plan (Description, quote from paper or describe key information)	Authors expected changes to accumulate over the course of the year during which teachers were exposed to the intervention, and they thus focused their evaluation on whether changes in student achievement would be observed in the 2nd year of the study, with a new class of students and no further coaching of the teacher. Two-year evaluation period.	Not mentioned	Not mentioned
Confounders	N/A	N/A	N/A

Author	Fukkink & Tavecchio	Jensen et al.	Murray et al.
Year	2010	2015	2014
Country	Netherlands	Denmark	USA
Sequence generation (Judgement)	Unclear	Unclear	Unclear
Sequence generation (Description, quote from paper or describe key information)	Not described.	Unclear (p. 28), preschools are randomised (cluster randomisation)	Within school randomisation of grades (K-2). Method not described
Allocation concealment (Judgement)	Unclear	Unclear	Unclear
Allocation concealment (Description, quote from paper or describe key information)	Not mentioned	Not mentioned	Not mentioned
Blinding (Judgement)	3	4	3
Blinding (Description, quote from paper or describe key information)	“The rating (or scoring) of the video clips was carried out following a “blind procedure,” with assessors who had not been informed of the time of measurement (pre-test, post-test or retention measurement) or condition (VIG or control)” (p. 1656). Different assessors used to rate individual teachers.	No blinding of outcomes (teacher administered SDQ)	Trained research assistants blind to randomisation status observed teachers
Incomplete outcome data addressed (Judgement)	Unclear	3	1
Incomplete outcome data addressed (Description, quote from paper or describe key information)	Only explicit mention of attrition from post-test to retention measurement, i.e. for the intervention group (9 out of 52). Not clear whether this means that attrition from pre-to post-test is zero or not. No numbers provided in tables. They do not report how many centres are randomised	P. 30 "the overall attrition rate was 22%, but with an uneven distribution across the intervention and control groups with significantly different attrition rates of 16% and 28%, respectively." Appendix Table A1 (ref. p. 30) – attrition rates, but only p values reported. Children with weaker outcomes more likely to attrit (p. 30).	Attrition of n = 2 teachers in intervention group and 0 in the control group (total 2%). Very low attrition
Free of selective reporting (Judgement)	3	3	3

Author	Fukkink & Tavecchio	Jensen et al.	Murray et al.
Free of selective reporting (Description, quote from paper or describe key information)	Not clear why control group was not measured 3 months after the intervention stopped, only intervention group.	Attrition, reasons stated (residential moves etc.) – more attrition in control group. Table A.1 in appendix – means not reported for attrition, only p values. Not ITT analysis. Only give standardised means in tables (no raw scores in Table 2). Sensitivity analysis. Teacher attrition not reported.	According to note at bottom of Table 1 there is also a TCI Harsh component. Reason for non-reporting is not provided.
Free of other bias (Judgement)	4	4	4
Free of other bias (Description, quote from paper or describe key information)	P. 1653: The VIG method, which is widely used in the Netherlands, is an integral part of a masters course School-Video Interaction Guidance (S-VIG) in the Netherlands. Unclear whether the intervention is VIG training or is VIG in itself. Unclear whether all involved teachers use VIG and some then get training. If the intervention is not training, it is possible that centres using VIG and centres not using VIG have been chosen (perhaps randomly), in which case it is not a RCT.	"Originally, the intervention was also implemented in a fourth municipality. However, the data on child outcomes for the control group in this municipality showed unexpected high SDQ-scores. This amounts to relatively few children in the final sample of children present for the full period and thus could not be detected in the original balancing tests. The available data did not allow us to investigate the reasons behind this further, and the current study therefore only includes data from three municipalities." (p. 28). Late entrants not in sample. Analytic sample consists of children present for the full period of two years.	Not reported whether parental consent to use student data was obtained before or after randomisation. Treated/control distribution of grades not reported. Not reported whether some teaches teach more than one grade. Worst case, a teacher may teach a class in a treated grade *and* a control grade.
A priori protocol (Judgement)	Unclear	Unclear	Unclear
A priori protocol (Description, quote from paper or describe key information)	Not mentioned	Not mentioned	Not mentioned
A priori analysis plan (Judgement)	No	Unclear	No
A priori analysis plan (Description, quote from paper or describe key information)	Not mentioned	Not mentioned	Not mentioned
Confounders	N/A	N/A	N/A

Author	Ottmar et al.	Raver et al	Rubie-Davies and Rosenthal
Year	2013	2009	2016
Country	USA	US	New Zealand
Sequence generation (Judgement)	Unclear	Low	Low
Sequence generation (Description, quote from paper or describe key information)	Schools are matched and randomised. Method not reported.	Random numbers generator (p.7) to assign sites (n = 18) stratified by cohort and in matched pairs to I and C	P. 76 in Rubie-Davies and Rosenthal (2016): a research assistant not involved in the study was blindfolded and drew names from a container for each school, and the assignment was recorded by a further research assistant. When there were an uneven number of teachers participating in a school, more teachers were randomly assigned to the intervention group than to the control group.
Allocation concealment (Judgement)	Unclear	Low	Low
Allocation concealment (Description, quote from paper or describe key information)	Not mentioned	Matlab uniform random numbers generator. It is non-sequential and therefore concealed.	Non-sequential and therefore low risk of bias
Blinding (Judgement)	3	3	2
Blinding (Description, quote from paper or describe key information)	Videos of observations were coded blindly (sent away to coders who were blind to assignment) p. 442.	A cadre of 12 trained observers (blind to intervention status of each site as well as to the approaches taken by training and MHCs) collected class-room level data. (p. 8).	Teachers administered tests, albeit according to a protocol. Researchers entered paper-and-pencil responses online. Positive aspect: Statistical analyses of data were conducted centrally.
Incomplete outcome data addressed (Judgement)	2	2	2
Incomplete outcome data addressed (Description, quote from paper or describe key information)	Not attrition per se – 100 teachers approached and 94 consented to participate. Attrition is present at follow up – Table 2 p. 447 (results) – teachers n = 88 (94 teachers in baseline sample) – reasons not given for this attrition. Stated on p. 441 that an additional six teachers were excluded because they taught mathematics in a foreign language.	Attrition: teachers (4/87); children (88/543). 54 children entered late. Attrition analyses limited to children due to low number of teacher attritors: comparison of 455 children at baseline with 93 children who left (no sig. differences); comparison of 455 children at baseline with 59 who entered late (only difference children who entered late were younger). No sig. differences between exit and entry status and I and C on 8 demographic variables. More girls entered late. Page 10	Attrition: 7% left the study during the year due to personal reasons, promotion, retirement.
Free of selective reporting (Judgement)	3	3	1

Author	Ottmar et al.	Raver et al	Rubie-Davies and Rosenthal
Free of selective reporting (Description, quote from paper or describe key information)	Descriptive statistics not shown other than overall means (treated and control pooled) and correlated with treatment status.	Results for CLASS (4 subscales) reported in Tables 1, 2 and 3. No raw means for either pre- or post-intervention are shown, only covariate adjusted means and standard errors.	
Free of other bias (Judgement)	1	2	2
Free of other bias (Description, quote from paper or describe key information)		Evidence from descriptive statistics suggests substantial variability in classroom quality among sites. Page 11: Note that of the scores mentioned here only the ECERS-R is a baseline score, the rest are post-intervention (mention that the CLASS positive climate is from March, which is post, does not explicitly mention month for the three remaining, but they are probably also March scores). With all that variability, pre- as well as post-, it is even more worrying that they do not show raw means by group (only covariate-adjusted means). The overall variability post-intervention could be a result of the treatment (if the variation is mainly between groups).	First, author created the tests using the e-asTTle tool. It was then reviewed by deputy principals of involved schools.
A priori protocol (Judgement)	Unclear	Unclear	Unclear
A priori protocol (Description, quote from paper or describe key information)	Not mentioned	Not mentioned	Not mentioned
A priori analysis plan (Judgement)	Unclear	Unclear	Unclear
A priori analysis plan (Description, quote from paper or describe key information)	Not mentioned	Not mentioned	Not mentioned
Confounders	N/A	N/A	N/A

Note: The RoB tool is provided in the Methods section. In the 5-point scale, 1 corresponds to Low risk of bias and 5 corresponds to High risk of bias.

Appendix F: Additional detailed intervention description for the 33 studies included in mapping

Author, date	Topic (stage of schooling, participant characteristics)	Intervention name	Detailed description of intervention
OVERARCHING TOPIC: LANGUAGE AND LITERACY DEVELOPMENT			
Al Otaiba et al., 2011	LANGUAGE AND LITERACY DEVELOPMENT (KINDERGARTEN)	Individualised Student Instruction for Kindergarten (ISI-K)	"…targeted professional guidance to provide teachers with the necessary tools to differentiate or individualize instruction..." (p. 536). First-author delivered workshops. Research partners delivered coaching (biweekly) (p. 544). Teachers met once a month for 8 months. "During their biweekly visits, research partners reinforced the professional development, assisted if needed with technology, modelled small-group strategies and often led a center activity or read aloud to students." (p. 544).
Bos et al., 1999	LANGUAGE AND LITERACY DEVELOPMENT (EARLY ELEMENTARY AND SPECIAL EDUCATION)	Project RIME (Reading Instructional Methods of Efficacy)	COMPONENT 1: COURSE FOR TEACHERS – "Content on (a) factors that affect early reading and spelling development, (b) types of assessment that may be used to detect children with early reading and spelling difficulties, and (c) teaching strategies, methods, and techniques for addressing and reducing the effects of these difficulties. Teachers kept journals and were asked to develop a plan and procedures for integrating early intervention strategies into their existing curriculum." (p. 228). COMPONENT 2: SCHOOL COLLABORATION – "Classroom collaboration in which project staff observed and collaborated with teachers as they taught + monthly school study and support meetings." (p. 229). Course lasted for 1 month, "the course was taught in June for 2 1/2 weeks/ Teachers attended sessions that were 3 1/2 hours in length", (p. 229). Coaching was led by Project RIME staff members with various backgrounds incl. expertise in speech and language pathology, early literacy and reading disabilities, bilingual special education.

Author, date	Topic (stage of schooling, participant characteristics)	Intervention name	Detailed description of intervention
Buysse et al., 2010	LANGUAGE AND LITERACY DEVELOPMENT (DUAL LANGUAGE LEARNERS: PRE-KINDERGARTEN)	Nuestros Niños program	"(1) professional development institutes to promote teachers' acquisition of core content knowledge and skills, (2) individualized consultation sessions to support teachers in implementing new instructional strategies in the classroom, and (3) community of practice meetings to provide participating teachers with opportunities for feedback, reflection, and collaborative problem-solving." p. 197. "Teachers participated in a three-day institute at the beginning of the school year conducted over a period of several weeks. Following the institutes, teachers were assigned to work with one of two bilingual Latina consultants (who were also responsible for facilitating the three-day institutes and community of practice meetings). Each consultant worked with a group of 6-8 teachers for eight weeks, visiting individual teachers once every other week and conducting community of practice meetings with her assigned group of teachers on alternate weeks. The consultants had graduate degrees in an early childhood-related field. They participated in intensive professional development and once-a-week reflective supervision meetings, both facilitated by the research team, to support their consultation practice and facilitation of the community of practice meetings." (p. 198).
Cabell et al., 2011	LANGUAGE AND LITERACY DEVELOPMENT (PRESCHOOLERS; LOW SES)	Adapted from Learning Language and Loving it.	Professional development focused on responsivity as adapted from Learning Language and Loving It. The training focused on a set of strategies designed to promote children's engagement and participation in extended conversational interactions across the school day (pp.320-1). "The professional development package for teachers in the intervention condition contained two components: (a) direct training designed to increase teachers' conversational responsivity in the classroom and (b) access to a consultant who provided off-site coaching throughout the academic year." (p. 320).

Author, date	Topic (stage of schooling, participant characteristics)	Intervention name	Detailed description of intervention
Dickinson & Caswell, 2007	LANGUAGE AND LITERACY DEVELOPMENT (PRESCHOOLERS; LOW SES)	LEEP Effects of the Literacy Environment Enrichment Program	"LEEP was given as a 45-hour course (…) delivered in two 3-day intensive sessions with support for application of new learning provided between sessions (…) to help teachers build (…) knowledge about literacy development (…) and to learn and gain facility using appropriate classroom strategies. Each session included lectures, videotapes of classroom activity and work samples that participants analyzed, and opportunities for participants to break into smaller groups to discuss concepts and relate them to classroom practices." p. 246. The intervention was delivered throughout New England by the regional Head Start Training and Technical Assistance system that was based at EDC. The intervention lasted one year for each course/teacher. Total intervention was spread out over a two-year period (1998–1999 and 1999–2000).
Gallagher et al., 2011	LANGUAGE AND LITERACY DEVELOPMENT (PRESCHOOL) LEARNING BEHAVIOUR (PRESCHOOL)	Individualised learning intervention	3 components (including Mentor Teacher Seminar, ongoing mentoring and support for the mentors via the Mentor Coordinator). 50-hour Mentor Teacher Seminar (2-day sessions over approx. 4 months). Then, mentors worked with protégées to provide classroom with learning environments that were responsive to individual children (and to promote learning outcomes).
Garet et al., 2008,	LANGUAGE AND LITERACY DEVELOPMENT (PRIMARY)	Language Essentials for Teachers of Reading and Spelling (LETRS)	Teacher Institute Series "(...) based on Language Essentials for Teachers of Reading and Spelling (LETRS) by Louisa Moats (2005). The full LETRS series consists of 12 modules that cover content intended to be consistent with the recommendations of the National Reading Panel NRP for reading instruction in grades K-6. Treatment group A and B teachers who attended the institutes and seminars received copies of the participant books for each of these modules in addition to other materials that addressed the NRP findings, vocabulary development, fluency-building strategies, and differentiated instruction." (p. 28). Coaching "treatment B added a half-time coach to each participating school to work with second-grade teachers in applying the content learned in the institute series within the context of implementing their core reading program. All second-grade teachers in treatment group B schools participated in the coaching, which began in early fall 2005 and continued until the end of the school year in spring 2006." (p. 32). "It was expected that teachers would receive, on average, 60 hours of coaching during the school year" (p. xvii).

Author, date	Topic (stage of schooling, participant characteristics)	Intervention name	Detailed description of intervention
Hindman & Wasik, 2012	LANGUAGE AND LITERACY DEVELOPMENT (PRE-KINDERGARTEN)	Exceptional Coaching for Early Language and Literacy (ExCELL)	In ExCELL, expert coaches visit Head Start teachers' classrooms for 3 hours per week over 2 full academic years, modelling best practices, observing teachers as they try out new behaviours, and providing individualized feedback. ExCELL provides HS teachers with 2 years of coaching on language and literacy development and instruction. Teachers are trained on five modules, including (a) oral language, (b) sound awareness, (c) alphabet knowledge, (d) emergent writing and (e) book reading, a technique that can advance all of these other skill sets. Begins with a 2-day summer workshop. There were four coaches altogether. "Teachers participated in the program for 2 full academic years. In the first year, coaches presented all five modules, and in the second year they revisited each topic to help teachers refine their knowledge." (p. 6).
Howlin et al., 2007,	LANGUAGE AND LITERACY DEVELOPMENT (AUTISTIC CHILDREN)	PECS	A 2-day PECS workshop for teachers plus 6 half-day, school-based training sessions with expert consultants. "The consultants recommended and demonstrated strategies for advancing children's use of PECS in the classroom, monitored teachers' progress and provided systematic feedback on their implementation of PECS. Following each visit, class teachers were provided with written summaries, agreed action points and future goals. The PECS consultants encouraged teachers to facilitate children's use of PECS in various sessions during the school day, according to the principles outlined in the PECS manual." (p. 475). The intervention lasted for 5 months.

Author, date	Topic (stage of schooling, participant characteristics)	Intervention name	Detailed description of intervention
Kammermeyer et al., 2016	LANGUAGE AND LITERACY DEVELOPMENT (PRESCHOOL)	The Pyramid Approach, The Kindergarten of the Future in Bavaria Approach, The Letter & Number World Approach	(1) The 'Pyramid Approach' developed by Jef van Kuyk (2006). The curriculum (van Kuyk, 2009) follows a holistic approach with a sequential framework as a structuring principle based on the dynamic system theory, which goes beyond Piaget and Vygotsky (van Kuyk, 2011): the method creates physical and psychological space for self-regulation through play and initiative learning. (2) A domain-specific training programme focusing on literacy and numeracy: an approach with many ideas and suggestions for early learning in mathematics and literacy in natural learning situations that are present in the everyday life of kindergarten. (3) A domain-specific Developmentally Appropriate Approach focusing on literacy and numeracy: the approach is clearly defined in terms of school subjects and teacher control. It focuses on numbers and operations and is characterised by lessons in which teachers introduce either numbers or letters, one after the other, based on detailed instructions. The training contains special materials to present the 'number & letter world' to the children, including puppets for every number and letter. "The teachers took part in a two-day professional training course on using the selected approach. They received special materials for the approach free of charge and instructions on how to use them. The teachers in the Pyramid treatment did not receive any special professional training in promoting literacy and mathematics because supporting these domains is part of their certification. The teachers must declare themselves willing to implement the approach in accordance with its guidelines. The professional training course for the 'Letter & Number World' training (TRAIN) was provided by the head of the 'Institute for preschool education' (Institut für vorschulisches Lernen), which markets the approach commercially. The professional training course for the Developmentally Appropriate Approach 'Kindergarten of the Future in Bavaria' (DAA) was conducted by researchers who were involved in developing this approach." (p. 162).

Author, date	Topic (stage of schooling, participant characteristics)	Intervention name	Detailed description of intervention
Landry et al., 2009	LANGUAGE AND LITERACY (PROGRESS MONITORING, PRESCHOOLERS, AT RISK)	'Comprehensive Professional Development'	A 2x2 design was used for cross-mentoring and progress-monitoring of conditions among the 4 PD programs. Specifically, some teachers received both in-classroom mentoring and detailed, instructionally linked feedback concerning children's progress in language and literacy. Some teachers received no mentoring but did receive the detailed, instructionally linked feedback concerning children's progress. Some teachers received in-classroom mentoring but only limited feedback on children's progress, which was not linked to curricular activities. Finally, some teachers received no mentoring and only limited feedback concerning children's progress. All 4 PD conditions included the same year-long, facilitated online course that emphasised language and literacy instruction, practice of learned material in one's classroom and participation in online message boards with fellow teachers. Intervention lasted 2 years. Classrooms in Ohio, Maryland and Florida participated in the project during the 2004-2005 school year. Classrooms in Texas participated during the 2005-2006 school year. "During initial classroom visits, the facilitators assessed the instruction being used in the classroom and evaluated the classroom environment using the Teacher Behavior Rating Scale (TBRS; Landry, Crawford, Gunnewig, & Swank, 2001). Subsequent mentoring visits consisted of helping teachers with classroom arrangement, instructional lessons, and instructional planning. Facilitators also provided teachers with reflective follow-up designed to help the facilitator and teacher discuss positive instructional pieces and those that needed attention. Written feedback was provided using the CIRCLE Glows and Grows Mentoring Tool (Tuyman, Aston, & Gunnewig, 2001). Facilitators mentored teachers twice a month for 2 hrs each visit." (p. 452).
McCollum et al., 2013	LANGUAGE AND LITERACY DEVELOPMENT (PRESCHOOLERS)	Performance-Based Feedback (Coaching)	Teachers met for two days for orientation (10 hrs total) regarding important areas of emergent literacy, focusing on 3 clusters of teaching skills: A. Vocabulary, information comprehension, narrative structure; B. Phonological awareness, alphabetic principle; C: Print concepts, written language. Each teacher then selected one activity setting in which to practice and received coaching on each cluster. Three brief additional group meetings (1.5 hrs each) were held during the year to share experiences; information on useful websites and ways to facilitate parent involvement in early literacy learning. Coaching (biweekly) followed a clinical supervision cycle (Krajewski, 1993). Each session included (a) a brief pre-observation discussion to re-establish the focus and context of the observation and to review percentage data from the previous session, (b) teaching-observing, (c) meeting to view and discuss the data and compare it with previous observations and with the 80% criterion and (d) looking ahead to what the teacher could do differently on the next visit.

Author, date	Topic (stage of schooling, participant characteristics)	Intervention name	Detailed description of intervention
Neuman & Cunningham, 2009	LANGUAGE AND LITERACY DEVELOPMENT (KINDERGARTEN)	Language and literacy course	"The intervention constituted a 45-hour, three-credit course in language and literacy held at one of four local community colleges, closest in proximity to the child-care site. For those randomly selected, a yearlong coaching intervention occurred in addition to the professional development course." (p. 540). “Coaches engaged teachers in reflection and goal setting; the coaches helped to identify desired outcomes and strategies to achieve these outcomes; collaboratively, they developed an action plan for the implementation of new practices the following week, which became the source of further reflection and action. Sessions were weekly, one on one, and on site, for approximately 1 to 1-1/2 hours. Coaching sessions for the first 15 weeks were designed to align with the professional development course.” (p. 543).
Piasta et al., 2012	LANGUAGE AND LITERACY DEVELOPMENT (PRESCHOOL)	Learning Language and Loving It approach	Learning Language and Loving It approach: LLLI training provides participating educators with a rationale that supports the relevance and importance of new or modified practices, opportunities for practice and self-reflection, and coaching and feedback over the course of an intensive 20-hr workshop featuring regularly-scheduled, small, evening group meetings (n = 8) and biweekly on-site coaching with certified speech-language pathologists serving as LLLI facilitators. "The training consisted of eight distinct sessions that focused on strategies that set up and promoted extended conversational exchanges with children (i.e. communication-facilitating strategies) as well as ways to enrich children’s language within conversational exchanges by providing advanced language models (i.e. language-developing strategies). In August, treatment teachers attended a 3-day training institute (approximately 13 h of PD in total)”. (pp. 389-391).
Powell et al., 2010	LANGUAGE AND LITERACY DEVELOPMENT (PRE-KINDERGARTEN)	Classroom Links to Early Literacy	"(…) classroom strategies to improve children’s oral language skills and code-focused skills, particularly phonological awareness and letter knowledge." (p. 302) and, "(…) instructional practices to improve children’s knowledge of print concepts (Clay, 1979)." (p. 303). 2-day workshop (16 hours) followed by expert coaching (to provide individualised feedback). "The literacy coaches led most of the workshop and met in small groups with assigned teachers to learn about each classroom. Teachers assigned to the remote condition received training in uses of the equipment (see below) while teachers assigned to the on-site condition prepared a predetermined literacy artefact (stick puppet) that also was made available to teachers in the remote condition. There were no differences across semesters in content coverage, presenters, and the amount of time devoted to each content area. The main purpose of coaching was to provide individualized feedback to teachers in improving the implementation of evidence-based practices emphasized in the intervention.” (p. 303).

Author, date	Topic (stage of schooling, participant characteristics)	Intervention name	Detailed description of intervention
Saraniero et al., 2014,	LANGUAGE AND LITERACY DEVELOPMENT (PRIMARY)	DREAM Developing Reading Education with Arts Methods	2 approaches to teacher professional development – a summer institute model and a model combining the summer institute with instructional coaching. The intervention trained third and fourth-grade teachers to integrate visual arts and theatre into the reading curriculum. Summer institute: 30 hours of PD in arts integration. Teachers received standards-based instruction in both art forms. Coaching: Teachers who received coaching spent 25 hours with the coach over the school year in a variety of activities, including lesson planning, observing each other teach, co-teaching and reflecting together. Coaching Group: Up to 25 teachers were assigned to this group each year, and they attended the summer institute. They also received 25 hours, on average, of instructional arts coaching during the school year. Teachers received a stipend for attending the summer institute, but not for the coaching. B. Institute-only Group: Up to 25 teachers were assigned to this group each year, and they attended the summer institute but received no coaching support during the school year. Teachers received a stipend for attending the summer institute.
Scanlon et al., 2008	LANGUAGE AND LITERACY DEVELOPMENT (KINDERGARTEN)	The Interactive Strategies Approach (ISA)	Professional development for classroom teachers, which served as a Tier 1 intervention, small group supplemental intervention for children, which served as a Tier 2 intervention, or both professional development for teachers and direct intervention for children (pp. 350-2). The intervention lasted 1 year and was led by Early Literacy Leaders (ELL): 3 classroom teachers; 3 reading teachers; 1 speech and language pathologist; 3 building administrators who held reading certification.

Author, date	Topic (stage of schooling, participant characteristics)	Intervention name	Detailed description of intervention
Schwanenflugel et al., 2010	LANGUAGE AND LITERACY DEVELOPMENT (PRESCHOOL)	PAVEd for Success	The PAVEd for Success programme contains guidelines that encourage the development of foundational alphabetic, phonological and semantic knowledge necessary for early decoding. All experimental teachers received professional development on all these elements, but for some teachers, this was the extent of their professional development from us (called Universal Quality Literacy Practices or UQLP). Other experimental teachers received additional professional development on phonological awareness activities (UQLP C PA), others on explicit vocabulary enhancement practices (UQLP C VE) and still others on all of these (UQLP C PAVE). The experimental teachers received professional development through (a) a three-day institute held immediately prior to the beginning of school, (b) a two-hour, on-site, follow-up workshop conducted approximately two weeks after the start of the intervention, and (c) on-site development, as classroom-based support throughout the intervention phase by preliteracy specialists who visited the classrooms every two to three weeks. During the three-day institute, teachers were divided into groups based on their assignment to condition. The structures of the professional development within intervention components were similar, though teachers received instruction only on the components relevant to their condition.
Vernon-Feagans et al., 2013	LANGUAGE AND LITERACY DEVELOPMENT (TECHNOLOGY-BASED LITERACY TRAINING: KINDERGARTEN/PRIMARY)	Targeted Reading Intervention (TRI)	"The Targeted Reading Intervention (TRI) provides teachers with professional development for struggling readers through state-of-the-art webcam coaching that allows literacy coaches thousands of miles away to provide real-time feedback to teachers in their classrooms, as the teachers instruct struggling readers. The programme also provides extensive website materials for instruction, webcam workshops and webcam team/grade level meetings, as well as e-mail correspondence between teacher and coach." (p. 2). “Teachers in experimental schools used the TRI in one-on-one sessions with 1 struggling reader in the regular classroom for 15 min a day until that struggler made rapid reading progress. Teachers then moved on to another struggling reader until all 5 struggling readers in the class received the TRI during the year. Biweekly webcam coaching sessions between the coach and teacher allowed the coach to see and hear the teacher as she instructed a struggling reader in a TRI session, and the teacher and child could see and hear the coach. In this way, the classroom teacher was able to receive real-time feedback from the coach." (Abstract)

Author, date	Topic (stage of schooling, participant characteristics)	Intervention name	Detailed description of intervention
Wasik & Hindman, 2011	LANGUAGE AND LITERACY DEVELOPMENT (PRESCHOOL)	Exceptional Coaching for Early Language and Literacy (ExCELL)	ExCELL has two main components: (a) intensive and ongoing staff development and (b) books, materials and lesson plans that support the development of children's language and literacy. The PD intervention provided teachers with conceptual knowledge and instructional strategies that support young children's development of vocabulary, alphabet knowledge and phonological sensitivity: systematic training in all five areas of language and preliteracy development, with a continued focus on book reading as a hub of language and literacy instruction. "All intervention teachers participated in a summer literacy institute over 4 half-days prior to children's entry into their HS programs. The purpose of this training was to familiarize the teachers with the goals of the project and to thoroughly explain the training and coaching procedures." (p. 458). "Thereafter, coaches provided 9 months of 3- to 4-week training cycles. Each cycle began with coaches providing a 3-hr group training for the teachers. Several days later, coaches modeled target behaviors in each teacher's classroom, and 1 week later coaches observed teachers' use of the new strategy." (p. 458).
OVERARCHING TOPIC: SOCIAL DEVELOPMENT			
Allen et al., 2011	TEACHER-CHILD INTERACTIONS (ADOLESCENTS)	The My Teaching Partner-Secondary program (MTP-S)	"The MTP-S intervention integrates initial workshop-based training, an annotated video library, and a year of personalized coaching followed by a brief booster workshop. During the school year, teachers send in video recordings of class sessions in which they are delivering a lesson. Trained teacher consultants review recordings that teachers submit and select brief segments that illustrate either positive teacher interactions or areas for growth in one of the dimensions in the CLASS-S. These are posted on a private, password-protected web site, and each teacher is asked to observe his or her behavior and student reactions and to respond to consultant prompts by noting the connection between the two. This is followed by a 20- to 30-minute phone conference in which the consultant strategizes with the teacher about ways to enhance interactions using the CLASS-S system. This cycle repeats about twice a month for the duration of the school year." (p. 2).

Author, date	Topic (stage of schooling, participant characteristics)	Intervention name	Detailed description of intervention
Domitrovich et al., 2009	SOCIAL DEVELOPMENT (PRE-K)	Head Start REDI (Research-based Developmentally Informed)	"The intervention was delivered by classroom teachers and integrated into existing curricula (i.e. Creative Curriculum and High/Scope) used by the Head Start centers. It included explicit curriculum-based lessons, center-based extension activities, and “teaching strategies” designed to be used throughout the day to generalize key intervention concepts. Teachers received detailed manuals and kits containing all materials needed to implement the intervention (...) teachers received mentoring in the use of language coaching strategies, including vocabulary support, expansions and grammatical recasts, and decontextualized talk to provide a general scaffold for language development in the classroom (Dickinson & Smith, l994)." (p. 575).
Early et al., 2017	SOCIAL DEVELOPMENT (TEACHER-CHILD INTERACTIONS PRESCHOOLERS)	Making the Most of Classroom Interactions and My Teaching Partner	Two intervention conditions: 1) Making the Most of Classroom Interactions (MMCI), a cohort-model where small groups of teachers met for five days of instruction and support; 2) My Teaching Partner (MTP), in which teachers worked one-on-one with a coach, using cycles of video-recorded observations of teaching, review and feedback. "Making the Most of Classroom Interactions MMCI is a face-to-face professional development model in which a group of teachers meets regularly with trained instructors to learn to identify and analyze effective interactions in classrooms and discuss ways to interact intentionally to increase children’s learning." (pp. 57-8). "My Teaching Partner MTP is a one-to-one, remote coaching model that provides specific feedback to teachers about emotional climate, organizational structure, and instructional support using a standardized coaching cycle format. During each cycle, the participating pre-k teacher makes a video recording of her or himself interacting with children in the classroom and sends it to the coach, who then reviews the video and posts feedback and questions about the interactions with children to a secure website for the teacher to review." (p. 58).
Fukkink & Tavecchio, 2010	TEACHER-CHILD INTERACTIONS (PRE-KINDERGARTEN)	Video Interaction Guidance training	"In the training, teachers were videotaped by their trainer for approximately 10 min while they are working with their groups. The trainer watched the video subsequently and selected a number of video fragments for review. In a next session, the trainer and the teacher subsequently engaged in a detailed discussion of these video clips. The training is usually comprised of four sessions (average number of sessions: 3.83, sd 1⁄4 0.94)." (p. 1654).

Author, date	Topic (stage of schooling, participant characteristics)	Intervention name	Detailed description of intervention
Jensen et al., 2015	SOCIO-EMOTIONAL DEVELOPMENT (KINDERGARTEN)	VIDA.2	The full intervention consists of three elements: knowledge, reflection and activities. Knowledge is provided by the training programme, then knowledge forms the basis for reflection about practices, and reflection subsequently leads to the development of new activities. These three elements alternate during the two-year period of the intervention. "During a total of 17 days distributed across two years, preschool teachers are trained to improve their practices, refresh and update their pedagogical knowledge and expertise, and thereby improve child outcomes. The intervention trains preschool teachers to work on an evidence-based platform. The training sessions focus on theoretical aspects, and they further include educational training targeting the interplay between theoretical knowledge and practice. As an important part of the implementation into practice, the teachers learn to reflect on how to transfer the theoretical knowledge into practice. This final part is crucial to ensure that the teachers are able to implement the pedagogical methods and tools into renewed practices. A focus point is to improve every child's skills by being responsive to the child's needs and potentials, working with the individual child and peer groups (inclusion), and thus enhance the well-being and learning of all children, but hopefully with a relatively larger effect for socially disadvantaged children." (p. 27). In addition, every year each of the principals is offered a two-day course and a workshop on facilitating the organizational learning processes in the preschools. This set-up ensures support and training for the participants to implement their new knowledge into practice in the daily work at the preschool, which is known to strengthen the effect of the intervention.
Murray et al., 2012	ACADEMIC AND PRO-SOCIAL DEVELOPMENT (K-2 STUDENTS)	IYT (Incredible Years Teacher Classroom Management Programme)	"IYT is a teacher training program that is part of a comprehensive series of interventions including parent, child, and teacher training components that were designed to prevent and treat aggressive behavior and conduct problems in young children aged 3-8 years. Its approach includes validated training methods such as video-modelling, behavioral rehearsal of key skills through numerous role plays, classroom practice assignments, and teacher goal setting and self-monitoring. IYT is provided in 4-6 monthly full-day workshops (5 in the present study) that cover building positive relationships with students and parents, proactive classroom management strategies, effective use of incentives, “coaching” students’ social and emotional development, teaching calm-down and problem-solving, and positive discipline techniques such as redirection, ignoring, and time out. Workshops are led by two trained co-leaders with approximately 12-15 teachers in each group" (p. 2).

Author, date	Topic (stage of schooling, participant characteristics)	Intervention name	Detailed description of intervention
Ottmar et al., 2013	SOCIO-EMOTIONAL DEVELOPMENT (PRIMARY SCHOOL)	Responsive Classroom (RC) approach	"Prior to data collection (summer 2008), all third-grade teachers assigned to the RC condition completed a one-week RC1 training institute constituting 35 hours of instruction and professional development. (...) During this training, participants learned strategies for implementing key practices of the RC approach, including morning meeting, rule creation, interactive modelling, positive teacher language, and logical consequences. In addition, during the 2008-2009 school year, teachers received one-day workshops, three consultations and classroom visits by their RC coach, as well as email and phone communication with their coaches. During the coaching visits, teachers were observed and given feedback about their use of RC practices. In addition, the coach conducted a lesson in their class, held debriefing sessions and mini-workshops, and led meetings with teachers and administrators." (pp. 441-442).
Pianta et al., 2008	TEACHER-CHILD INTERACTIONS (PRE-SCHOOLERS)	MyTeachingPartner (MTP)	Two components: access to video content – examples of high-quality teacher-child interactions tied to specific dimensions of the CLASS and (b) a consultation process that provides regular, multi-modal, ongoing, targeted feedback to pre-k teachers One group received WebOnly access – "provided teachers with exemplars of other teachers' interactions in the form of 1-2 min video clips pertaining to a specific dimension of interaction accompanied by a text description of that teacher's behavior" (p. 437).
Raver et al., 2008	CLASSROOM MANAGEMENT (PRE-SCHOOL)	Chicago School Readiness Project (CSRP)	Treatment group teachers received: (1) training and (2) Mental Health Consultation (MHC). "Training – All treatment-assigned teachers (including lead teachers and assistant teachers) were invited to participate in five trainings on Saturdays, each lasting six hours. A behaviourally- and evidence-based teacher training package was selected and purchased, and a seasoned trainer with Licensed Clinical Social Worker (LCSW) qualifications delivered the 30 hours of teacher training over fall and winter, adapting the Incredible Years teacher training module (Webster-Stratton, Reid, & Hammond, 2004). Teachers were reimbursed $15 per hour for their participation. Examination of rates of participation suggests that 75% of teachers participated in at least one training and 63% of teachers participated in more than half of the trainings. Each teacher spent an average of 18 hours (SD = 12) in training from September through March, with each classroom receiving an average of 50 training hours (SD = 26) across teachers in each classroom." (p. 7).

Author, date	Topic (stage of schooling, participant characteristics)	Intervention name	Detailed description of intervention
Rubie-Davies & Rosenthal, 2016	TEACHER-CHILD INTERACTIONS (GRADES 2-10)	Complex intervention	4 workshops in relation to expectancy research and the relationship between high expectation of teachers in relation to achievement. “At the first workshop, teachers were introduced to the expectancy research so that they understood the background to the study and how the field of teacher expectation research had developed” (p. 87). “At the second workshop, teachers learned how high expectation teachers worked with flexible groups rather than keeping students in fixed within-class ability groups for mathematics, how they allowed students to make choices in relation to the learning activities that they completed, and how they encouraged students to work with a variety of peers to complete their tasks” (Ibid.). “In the third workshop, teachers learned about ways that high expectation teachers enhanced the class climate by, for example, promoting peer support and interactions, using positive behavior management strategies, being encouraging of all students, and showing that they cared for their students” (Ibid.). “In the final workshop, the focus for the intervention teachers was on student motivation, engagement, and autonomy, and the importance of teacher evaluation and feedback” (Ibid.). Researchers met with the participants 3 more times in that year (to offer support and answer questions).
OVERARCHING TOPIC: OTHER			
Azkiyah et al., 2014	TEACHING QUALITY (SECONDARY)	(No specific name is mentioned for the intervention)	Two interventions: Group 1 – use of education standards (elaborated standards document). Group 2 – combination of education standards with a teacher improvement programme (document containing characteristics of effective teachers) (pp. 6-7 and Table 1). Teachers in both experimental groups attended a one-day workshop to discuss the elaborated document. Teachers in the second intervention group also attended a one-day workshop on effective teaching and six half-day monthly meetings to discuss and advise teachers about the two documents and how to use them.
Flook et al., 2013	STRESS REDUCTION (TEACHERS)	Mindfulness-Based Stress Reduction course	"The course (MBSR) was offered during the academic year in Fall 2011, which allowed teachers to directly apply the skills they learned within the context of their classroom and actual teaching. Outside of class, teachers were encouraged to practice between 15-45 minutes per day for six days/week and were provided with guided recordings to support their practice. The course lasted eight weeks, 2.5 hours/week, plus a day-long immersion (6 hours), totalling approximately 26 hours of group practice and instruction." (p. 5).

Author, date	Topic (stage of schooling, participant characteristics)	Intervention name	Detailed description of intervention
McMeeking et al., 2012	MATHEMATICS DEVELOPMENT (MIDDLE SCHOOL)	The Rocky Mountain Middle School Math and Science Partnership (RM-MSMSP)	"These courses were designed to contain approximately 80% content knowledge and 20% pedagogical content knowledge and were taught in 2- to 3-week institutes offered in the summer. The content from these courses could be situated in Hill and Ball's (2009) Mathematical Knowledge of Teaching model as contributing to teachers' common and specialized content knowledge. In addition, structured follow-up (SFU) units associated with each content course were taught across 4 Saturdays during 1 academic semester. These were focused 80% on pedagogical content knowledge and a review of the content from the summer (based on data from teacher content inventories, evaluation feed-back, or both) and 20% on general pedagogy, the latter including work focused on English language learners, integration of literacy within the mathematics classroom, and differentiation. The SFUs required that teachers implement a "teaching innovation" in the classroom during the academic semester and share those practices and their students' learning results during the class. In addition, some of the content courses were taught in academic-year intensive sessions, which delivered the content and the SFU simultaneously in one course." (p. 164).