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

613318

CARE

Curriculum Quality Analysis and Impact Review of European ECEC

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

D4.2:
Effects of ECEC on academic outcomes in literacy and mathematics: Meta-analysis of European longitudinal studies

DUE DATE OF DELIVERABLE: 31 DECEMBER 2015

ACTUAL SUBMISSION DATE: 30 DECEMBER 2015

Start date of project: 01-01-2014

Duration: 36 Months

CARE contractor: Utrecht University

„As educational researchers, we find ourselves in the mildly embarrassing position of knowing less than we have proven. The proofs reside in a vast literature that is often superciliously scorned and insufficiently respected. Extracting knowledge from accumulated studies is a complex and important methodological problem to which I commend your attention.“

(Glass, 1976, S. 8)

Title: D4.2: Effects of ECEC on academic outcomes in literacy and mathematics: Meta-analysis of European longitudinal studies

Organisation: Freie Universität Berlin and University of Oxford

Authors (main authors in bold):

Hannah Ulferts (Germany)

Yvonne Anders (Germany)

Main contributors to developing design and analysis:

Hannah Ulferts (Germany)

Yvonne Anders (Germany)

Paul Leseman (Netherlands)

Edward Melhuish (UK)

Contributing researchers:

UK:	Edward Melhuish
Finland:	Marja-Kristiina Lerkkanen, Gintautas Silinskas
Germany:	Hannah Ulferts, Yvonne Anders, Ana Susac
Netherlands:	Paul Leseman, Pauline Slot, Martine Broekhuizen
Portugal:	Joana Cadima

Email: Hannah.Ulferts@fu-berlin.de; Yvonne.Anders@fu-berlin.de;

Number of PM:

Dissemination Level:

Version	date	Authors	Status	changes
0.1				
0.2				
0.3				

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

ACKNOWLEDGEMENTS

We are grateful to the European Commission for funding the project CARE (Curriculum Quality Analysis and Impact Review of European ECEC) and to our colleagues in the CARE project, especially the partners, who undertook new analyses of primary data for this study, who supported us with information about the studies and helped with interpreting the study findings. We are also grateful to the extremely engaged coders of included studies and the CARE partners that have provided us with valuable feedback at the meetings in Berlin (Germany), Tønsberg (Norway), and Reggio Emilia (Italy). We would also like to thank the members of the Advisory Committee for their valuable comments and information on longitudinal findings in Europe. All feedback has helped to shape the study design and the interpretation of the findings.

Executive summary

Early **C**hildhood **E**ducation and **C**are (ECEC) increasingly receives attention in Europe. One main public interest lies in the potential beneficial effects of ECEC on children's development and later educational careers, especially for vulnerable children and children who grow up in disadvantaged families. We present the results of a meta-analysis of longitudinal studies in Europe regarding the relative impact of variations in ECEC experience and outcomes in two central academic domains: mathematics and literacy. Our meta-analysis adds additional evidence to previous research syntheses in the field. It picks up various shortcomings of previous analyses. By creating a compressed knowledge basis about evidence on the developmental impact of European ECEC, this report aims at contributing to the overall objective of CARE to create an evidence-based and culture-sensitive framework of European ECEC.

This report aggregates findings four core aspects, commonly used to describe ECEC experiences across countries, across types of ECEC, across different programmes and across pedagogical approaches. *ECEC quantity* refers to variations in children's exposure to ECEC (the "dose") and can be further categorized into the comparison of no ECEC vs. some ECEC experience (i.e., the absolute effect of ECEC quantity), as well as differences in duration, intensity and age of entry (i.e., the relative effect of ECEC quantity). *Structural quality* refers to aspects such as class size, teacher-child ratio, formal staff qualifications, and group size in the setting. Structural quality can be subject to regulation by policy and funding. It is the concept of process quality that describes the nature of the interactions between preschool teachers and children, the interactions among children and the interaction of children with space and materials. Different conceptualizations of process quality include *global process quality* (such as warm climate or child-appropriate behaviour, commonly assessed by observational measures like ECERS-R, CIS or CLASS) as well as the extent *of pre-academic promotion* relating to the promotion of learning in areas such as literacy, emerging mathematics and science. It is hypothesized that process quality has direct effects on children's learning and development, while structural quality has indirect effects through its influences on process quality.

After a thorough and systematic search, selection and coding procedure, we included 226 separate findings of 22 European longitudinal studies, thereby, gathering knowledge about the developmental impact of ECEC on developmental outcomes for over 43,000 children in Europe. Evidence spans different phases of the educational career from pre-school to secondary school. Using three-level longitudinal meta-analysis, we aggregated findings to four overall effects (i.e., global process quality, extent of pre-academic promotion, structural quality, and quantity). The included studies differ, for example, in location and its ECEC system, in design and sample characteristics, or in the assessment measures for outcomes and ECEC. Besides an overview of important study characteristics, this report investigated if findings varied within and between studies. Differences in study characteristics are expected to be linked to variations in findings, which was tested for some important study characteristics with mixed-effects model. This moderator analysis studied if ECEC effects varied between the two outcome domains (literacy or mathematics), if effects were persistent across ages and different phases of the academic career, and if different measures of ECEC vary in their effects. Additionally, we reviewed existing European longitudinal evidence on differential effects for disadvantage children.

Our meta-analysis confirmed the claim of other meta-analyses and reviews in the field of ECEC: the different experiences children gather within childcare are important and they have developmental impact on academic outcomes. To our knowledge, this meta-analysis is the first synthesis which studied systematically if ECEC effects differ between literacy and mathematics. Our overall results imply that children benefit from higher global process quality (*ES* = .11), more pre-academic promotion (*ES* = .10), and from a greater amount of ECEC experience (*ES* = .12). Transforming our correlational measures into the commonly used measure of effect size, Cohen's *d*, results in effect sizes in the range .20 to .24. Surprisingly, these transformed aggregated results comparing variation in ECEC are only slightly lower than results of other meta-analyses that entail findings for specific ECEC interventions, and contrasts for outcomes for children with and without ECEC experience.

Observed overall effects vary substantially between and within studies, and variations can partly be explained by different moderators. All of the quality effects vary by outcome domain and by the measure of ECEC. Global process quality seems to be more strongly related to literacy outcomes, whereas pre-academic promotion is more strongly related to mathematics outcomes. We did not find indications for a decline of ECEC effects with age, except that quantity had a stronger association with outcomes measured in the ECEC period than to outcomes in a later phase of children's educational career. The available research uses various measures to assess the four ECEC aspects, and the moderator analysis suggests that the choice of measure relates to the strength of observed relationships to child outcomes. Interaction-focused measures tend to be more strongly associated to child outcomes than those including an evaluation of material surroundings in overall quality ratings. Also for structural quality, we found only the variations in staff qualification, and not variations in environmental arrangements, to relate to child outcomes. No differences between absolute effects of ECEC versus effects of relative variations in ECEC quantity were apparent.

A review of differential findings for disadvantaged children reveals that research evidence is sparse in Europe, and studies address this question in different ways. Overall findings show that disadvantaged children benefit from earlier enrolment and higher quality of educational processes, but that they need additional and specific support for themselves as well as for their families to exploit learning opportunities in ECEC, if they are to catch up with their peers.

Our results imply that substantial gains are to be expected by improving the quality and quantity of regular ECEC provision in Europe. Enhancing quality and quantity of regular ECEC is beneficial for all children, including the disadvantaged. Effects are persistent across different ages and phases of the academic career. Conclusions about the benefits of improving structural quality are less straight-forward; except for the benefit of higher staff qualifications. Structural quality has an impact on child development through process quality and its effect depends on the interaction of different structural aspects and its influence on the quality of processes. Furthermore, though important for outcomes in both domains, variations in ECEC quality differ in their effects for outcomes in literacy and mathematics. Focusing on interactions when measuring process quality may be more efficient in order to assess ECEC's potential to foster academic development. Thus far existing staff questionnaires seem to capture less of the relevant aspects of ECEC processes than observational measures do.

Recommendations

1) Enhancing the quality of pedagogical processes and providing an extended ECEC service can be an effective and sustainable approach to increase academic benefits for children of various backgrounds across countries and across varying ECEC systems.

2) To assure that promotion and stimulation in *various* academic domains happens on a regular basis is one of the main challenges in regular provision

3) Regulations of structural quality which address environmental arrangements are necessary prerequisites for high process quality, but investments in better environmental arrangements are not sufficient to foster children's learning.

4) Improvements to staff qualifications and efforts to enhance the professional skills of teachers and thus improve pedagogical practice are more promising. Staff qualifications and professional development are key components of structural quality for improving process quality, and thereby, child outcomes.

5) Quality monitoring should focus on pedagogical interactions and processes.

6) Measuring interactions captures beneficial pedagogical processes, measuring material surroundings captures the pedagogical opportunity structure.

7) The effects of ECEC quantity are also a question of relative amount (not just absolute effect).

8) Disadvantaged children need intensive and high quality support, including parental support.

9) Decisions on research funding should be based on considerations regarding the proposed research design including: reliability and validity of measures for child outcomes and ECEC aspects, balanced research questions with regards to effects of quantity and quality and their interactions, representativeness and size of the sample, consideration of family and child characteristics in studying ECEC effects, potential to study differential effects for disadvantaged children.

Table of Contents

Introduction

Early **C**hildhood **E**ducation and **C**are (ECEC) increasingly receives attention in Europe and worldwide. This includes financial investments driven by the hope of beneficial effects of ECEC on children's development and later educational careers, especially for vulnerable children and children who grow up in disadvantaged families. Thus, the impact of ECEC on children's developmental outcomes is a pressing research question addressed in this report. We present the results of a meta-analysis of longitudinal studies in Europe regarding the relations between ECEC experiences and outcomes in two central academic domains: mathematics and literacy. ECEC experiences are commonly characterized by aspects of ECEC quantity and quality (Anders, 2013; Burger, 2010). While quantity refers to the "dose" of ECEC experience, quality characteristics capture the variation between ECEC settings. The meta-analysis looks at variations in ECEC quantity and quality in their relation to child outcomes. For quality aspects, we studied the developmental impact of structural aspects, differences in general pedagogical processes, and aspects of pre-academic promotion. For quantity aspects the report covers evidence for the developmental impact of differences in age of entry, duration and intensity of ECEC, as well as the impact of ECEC attendance versus no ECEC experience at all. The meta-analysis provides information about the relation of these aspects to developmental outcomes. Thus, these findings have important implications for policy making. The strength of effects for outcomes may vary across studies, domains, and different phases of the individual's academic career leading to the heterogeneity of previously reported findings. The meta-analysis quantifies the extent to which findings vary and studies potential reasons for variations. By studying moderators of the effects, including developmental domain, age and phase of academic career, and measures of ECEC experience, the meta-analysis allows for an estimation of the generalizability of findings, e.g. across different phases of the lifespan, and domains. These differential analyses affect implications for policy and future research in ECEC, and provide direction for targeted funding. This report is part of the project *Curriculum Quality Analysis and Impact Review of European Early Childhood Education and Care (CARE)*, funded by the European Union's 7th Framework program (THEME [SSH.2013.3.2-2] Early childhood education and care: promoting quality for individual, social and economic benefits). The task reported on here is part of WP4, entitled *Impact of ECEC in short, medium & long-term*.

Our objectives directly relate to the CARE-project's overall aims of working towards an evidence-based, comprehensive and culturally sensitive European framework for evaluating and monitoring ECEC quality and child wellbeing, including indicators of ECEC quality and child wellbeing for use in policy making. Within the CARE-project, WP4 specifically focuses on assessing the impact of ECEC on children's outcomes on several time scales, identifying potential moderators of this impact, and examining the potential links between children's outcomes and countries' monitoring and quality assurance systems. The tasks of WP 4 comprise: (1) an updated review of research into the impact of ECEC on child development (completed: Melhuish et al., 2015; deliverable D4.1), (2) a meta-analysis of European studies on the impact of ECEC on child development (deliverable D4.2) and (3) a comparative review of quality monitoring and assurance systems across Europe (deliverable D4.3). Already completed studies of the CARE-project linked to this task are: (2) an overview of European ECEC curricula and curriculum template (Sylva, Ereky-Stevens, & Ariescu, 2015; deliverable D2.1), (3) a secondary analyses of large scale studies in five countries assessing relations between structural and process quality in European ECEC (Slot, Lerkkanen, & Leseman, 2015a; deliverable D2.2), (4) a comparative review of approaches to ECEC staff professionalization in Europe (Jensen et al., 2015; deliverable D3.1), (5) a literature review on the effectiveness of different types of funding and

governance of ECEC (Akgündüz, Ünver, Plantenga, & Nicaise, 2015; deliverable D5.1), (6) an initial framework for evaluating and monitoring ECEC quality and wellbeing (Moser et al., 2014; deliverable D6.1), and (7) a first report on the values, beliefs and concerns of parents regarding ECEC services in nine European countries (Broekhuizen, Leseman, Moser, & van Trijp, 2015; deliverable D6.2).

Additional ongoing studies within CARE address (7) the cultural interpretations of quality and the cultural factors that shape the implemented curriculum in ECEC as observed in different countries, and provide (8) a meta-analysis on the impact of professional development and (9) an economic analysis of the costs and benefits of ECEC. Finally, to provide a general framework of shared concepts and a basic model of ECEC services as embedded in wider local, regional and national contexts that can integrate separate studies within CARE, a document was developed including the goals of the project and definitions of core concepts in ECEC (Moser et al., 2014; D6.1).

This report synthesizes findings of European longitudinal studies and quantifies the impact of various ECEC aspects. By creating a compressed knowledge basis about evidence of the developmental impact of ECEC, this report aims at contributing to the overall objective of CARE to create an evidence-based and culture-sensitive framework for European ECEC.

Theoretical background

Attendance rates of ECEC throughout Europe have been persistently growing in recent years, with currently about 95% of children over the age of 4 attending some sort of professional centre-based care (European Commission/EACEA/Eurydice/Eurostat, 2014). Hence, ECEC nowadays represents an almost universal experience for children in Europe. Apart from the aim to maximize parental labour market participation, ECEC's aim is to foster children's development. ECEC is seen as the first step for children into a phase of institutional learning (Roßbach & Weinert, 2008). Besides caring and assuring security and wellbeing of children, learning and development in different domains should form an essential part of ECEC experience (Sylva et al., 2015).

There is growing evidence for the potential of ECEC to foster children's developments in various (pre-)academic domains (Anders, 2013; Melhuish et al., 2015). Research also underpins that ECEC can serve as an effective means to combat educational inequalities (Eurydice, 2009): children from disadvantaged background are at risk of falling behind their peers in their development before entering school (Arnold & Doctoroff, 2003). Different indicators of disadvantage are linked to impoverished learning environments and educational opportunities at home (Anders et al., 2012; Eurydice, 2009; Kluczniok, Lehrl, Kuger, & Roßbach, 2013; Niklas & Schneider, 2013). ECEC institutions that provide rich learning environments and engage in fostering early learning can compensate for this disadvantage, thereby helping disadvantaged children catch up and providing all children with a good start in school (Anders, 2013; Melhuish et al., 2015). It has been argued that investing in early education programmes could have large long-term monetary and non-monetary benefits (Heckman, 2006; Knudsen, Heckman, Cameron & Shonkoff, 2006). In 2000, the *World Declaration on Education for All* called for "expanding and improving comprehensive early childhood care and education, especially for the most vulnerable", and over recent years one could observe increased financial support for early childhood care and education (ECCE) programmes in many countries. Besides quantitative expansion of ECEC to maximise coverage, several European countries implemented new policies and governmental initiatives to optimize the quality of provision such as

legislation regarding structural, safety and qualification standards and teacher training. However, concerns about the low general quality of ECEC services in many countries and unanswered questions regarding the appropriate age for entering ECEC and the appropriate amount of ECEC give reason to doubt if ECEC is realising its potential (Anders, 2013; Kuger & Kluczniok, 2008).

First insights into ECEC's potential for children's development came from intervention studies in the US, e.g. Head Start or the Perry-Preschool Project, and several syntheses of their findings (Anders, 2013; Kuger, Sechtig & Anders, 2012; Melhuish et al., 2015). These interventions started in the 1960s. The NICHD-study which started in the U.S. in the 1990s demonstrated potential beneficial effects of high quality regular care. Research activity in Europe was lower for many years. At the same time the differences in ECEC provision between many European countries and the US-system are obvious, and the generalizability of US findings for a European context can be doubted (Burger, 2010; Kuger, Sechtig & Anders, 2012). At the beginning of the 21st century, the Organisation for Economic Co-operation and Development [OECD] published results from large-scale international assessments of student's achievements around the world (OECD, 2004) leading to growing interest in large-scale investigations of educational institutions and systems including ECEC in Europe. Guided by previous scientific insights and the need to channel investments and efforts in ECEC development, European large-scale projects started to investigate the impact of ECEC on child development with the intention of capturing more accurately the typical (real-world) experiences of children in European childcare (Anders, 2013; Burger, 2010). A central goal of these projects is to go beyond the simple question of "Can ECEC have an impact on child development?" and to answer the question "Which specific aspects of ECEC have an impact on child development?". This meta-analysis aims at providing additional knowledge related to this question by synthesizing currently available European evidence in order to provide a strong knowledge basis of effective aspects in regular provision for European ECEC policy, investments and developments, and to guide future research.

The report starts with a brief overview of the conceptualization of ECEC in terms of quality and quantity. Following this, the main research findings on ECEC impact on child development in countries around the world with a strong focus on European findings are summarized. The next section highlights findings and shortcomings of existing research syntheses. Finally the rationale and the specific research questions of the meta-analysis are outlined.

Conceptualization of differences in ECEC experiences in terms of quality and quantity

The ECEC experiences of children across countries, types of ECEC, and pedagogical approaches, can be considered in terms of aspects of ECEC quantity and ECEC quality (Anders, 2013).

ECEC quantity refers to variations of the amount of exposure to ECEC (the "dose") and can be further categorized into the comparison of no ECEC vs. some ECEC experience, as well as differences in duration, intensity, age of entry. Relevant indicators differ between countries, depending on the characteristics of their ECEC system. Some studies include children who experienced some form of childcare and children without ECEC experience (sometimes referred to as the "home group" or "home sample"). Variations in duration are usually measured by the total or absolute number of years or months children spent in ECEC. This is closely connected to the age of entry, because a younger entry age normally results in a longer duration, but it is also dependent on country-specific policies for enrolment into primary school. Intensity covers differences in number of hours per week,

month or year or the differentiation of children in full- or part-time provision. Relations of ECEC quantity to child outcomes, i.e. effects of ECEC quantity, therefore, resemble dosage-effects as commonly described in clinical or intervention studies. We will use the terms interchangeably within this report.

One dimension that shows variation between and within ECEC types as well across countries in Europe, is the quality of ECEC (see Figure 1). The quality of preschool learning is seen as a multidimensional concept covering various characteristics of quality including structural characteristics and process quality (NICHD ECCRN, 2002a; 2002b; Pianta et al., 2005; Kluczniok & Roßbach, 2014). Structural quality refers to aspects such as class size, teacher-child-ratio, formal staff qualification levels, provided materials and group size in the setting. Structural quality can be subject to regulation by policy and funding. Many European ECEC systems share a child-oriented view of pedagogy in ECEC, putting much emphasis on free play and the children's right to choose activities and material (Anders, 2014). In this case the creation of learning opportunities may be more dependent on structural characteristics such as learning materials or the spatial setting. School effectiveness research sometimes refers to factors that are relevant to create supportive learning environments, which can be seen as prerequisites for high quality pedagogical interactions (Ditton, 2000; Scheerens & Creemers, 1989). It is the concept of process quality that describes the nature of interactions between preschool teachers and children, the interactions among children and the interaction of children with space and materials. Some approaches also highlight the quality of interactions between staff and parents (e.g. Lamb-Parker et al., 2001; Reynolds, Mavrogenes, Bezruczko, & Hagemann, 1996). It is hypothesized that process quality has direct effects on children's learning and development, while structural quality has indirect effects through its influences on process quality (Pianta et al., 2005).

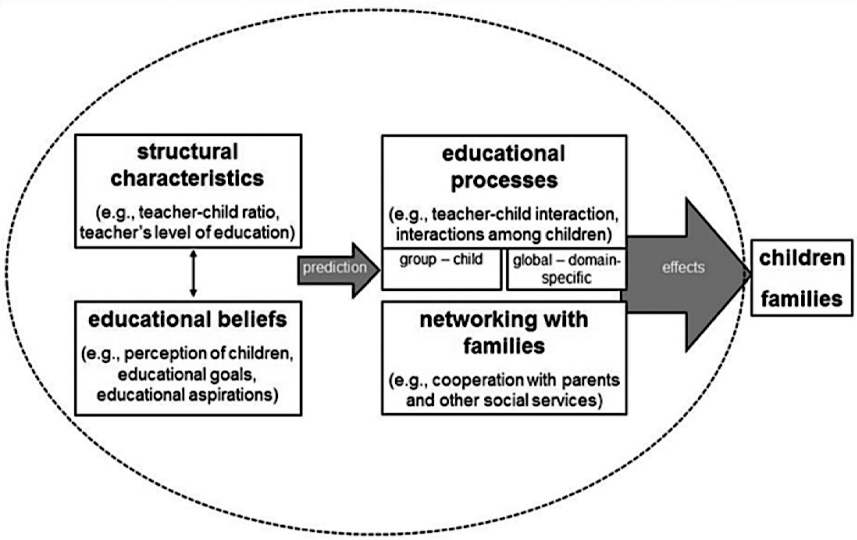

Figure 1 Structure-process model of educational quality in ECEC (Kluczniok & Roßbach, 2014; based on Roux & Tietze, 2007; Tietze et al., 1998).

Different theories also highlight the importance of certain educational beliefs and orientations of teachers (e.g. their definition of their professional role, their educational aspirations and goals, their view of the child) and treat them as a quality dimension with direct impact on process quality and indirect effects on children's development. Some conceptualizations of ECEC quality cover further

dimensions (OECD, 2006), but most concepts include the dimensions of structural quality and process quality.

With regard to the question of impact of ECEC on children's development, process quality with its direct link to the child's outcomes is the key concept. Different conceptualizations of process quality cover global aspects (such as warm climate or child-appropriate behaviour; e.g. Harms, Clifford & Cryer, 1998) as well as promotion and stimulation in learning areas such as literacy, emerging mathematics and science (Kuger & Kluczniok, 2008; Sylva, Siraj-Blatchford & Taggart, 2003). Specific measures of the extent of promotion in various domains evolved as a consequence of a growing emphasis of ECEC science, practice and policy of domain-specific learning. Children need to acquire domain-specific knowledge and skills to succeed in academic areas like mathematics, science and literacy (Roßbach & Weinert, 2008). As learning takes places by domain-specific learning processes, it needs to be supported accordingly, i.e. by domain-specific promotion.

It is shared understanding that the process quality of ECEC can be best measured by observation (Sylva et al., 2006; Mashburn et al., 2008). Commonly, it is doubted if questionnaires are equally appropriate to capture the nature of the provision and the interactions in the setting. Standardized observational instruments exist that have been used in different studies in various countries to investigate preschool quality and its effects on children's learning (see also method section). Well established and widely used are the *Early Childhood Environment Rating Scale-Revised Edition* (ECERS-R; Harms et al. 1998) measuring global aspects of process quality and its extension ECERS-E (Sylva et al., 2006), which measure the process quality in areas of mathematics, literacy, science and diversity. Both the ECERS-R and ECERS-E involve observation to produce quality ratings. The scores range from 1 to 7, with 1 indicating *inadequate quality*, 3 *minimal quality*, 5 *good quality*, and 7 *excellent quality*. Both instruments include ratings of the material surrounding into their ratings of overall quality.

Other instruments focus on interactions, for example the widely-used Caregiver Interaction Scale (CIS; Arnett, 1989) assesses aspects of positive relationships between children and caregivers, as well as punitiveness, permissiveness and detachment of the caregivers. Another standardized instrument measuring preschool quality at group level with an interaction focus is the *Classroom Assessment Scoring System* (CLASS; cf. Pianta, La Paro & Hamre, 2008; Pianta & Hamre 2009). The CLASS is also a well-established, reliable, and valid observational instrument to assess teacher-child interactions in three broad domains: emotional support, classroom organization, and instructional support. These domains are based on developmental theory and research suggesting that interactions between children and adults are the primary mechanism of child development and learning (cf. Pianta et al., 2008). Scaling of interactions with CLASS resembles the ECERS-R ratings ranging from 1 to 7: 1 or 2 indicating *low quality*; 3, 4, or 5 indicating *mid-range of quality*; and 6 or 7 indicating *high quality*. The CLASS focuses on aspects boosting the effectiveness of teacher-child interactions, which are associated with important achievement gains for children from preschool through secondary school (cf. Mashburn et al., 2008). Other instruments focus on educational processes at child level (e.g., *Emerging Academic Snapshot*, EAS; cf. Ritchie, Howes, Kraft-Sayre & Weiser, 2001; *Observation of Activities in Preschools*, OAP; cf. Palacios & Lera 1991; *Target Child Observation/Zielkindbeobachtung*, ZiKiB; cf. Kuger, Pflieger & Roßbach, 2006a, 2006b; Smidt, 2012). These measures cover important details regarding the dynamic everyday activities and interactions of individual children with their peers and preschool teachers. To carry out observations is extremely time- and cost-intensive. Furthermore, it has been argued that larger observation intervals would be necessary to ensure that

the nature of ECEC provision is captured in a representative way (Kane & Staiger, 2012; Praetorius, Pauli, Reusser, Rakoczy, & Klieme, 2014). As a consequence recent attempts have been undertaken in different research projects to develop and evaluate questionnaire measures of process quality (e.g. Anders & Ballaschk, 2014; Bäumer, Außmann, von Maurice & Blossfeld, 2013; Camehl, Schober, & Spieß, 2012; Slot, Leseman, Verhagen, & Mulder, 2014). Not only large-scale research projects profit from reliable and valid questionnaire measures but also ECEC systems which rely on continuous quality assessments to monitor and assure quality, like the ECEC systems in Scotland or England (Bradshaw, Lewis, & Hughes, 2014; Sammons et al., 2002).

Research has provided insight into the nature and variability of process quality in different countries. Some European studies report only low to moderate levels of quality (Kuger & Kluczniok, 2008; Slot et al., 2015a, see also Table 4 in the Appendix). Especially promotion and stimulation in different learning domains seems to be insufficient in most ECEC systems and mainly restricted to the fostering of language development.

The following section gives insight into developments in the area of research on the impact of ECEC on children's development.

Studies on European ECEC and main differences to US studies

The most famous research findings on the impact of ECEC have been derived from U.S. research, especially from the longitudinal evaluation studies on the effects of specific programmes for disadvantaged children (Garces, Thomas & Currie, 2000; Schweinhart et al., 2005) and the NICHD study (NICHD ECCRN, 2005). However, these findings may not be generalizable to European context and similar research has also grown within Europe. These studies include small experimental studies which do not cover natural settings as well as large-scale studies on the regular provision of ECEC (Anders, 2013). Early research focused on the most basic quantitative indicators, often comparing children with and without ECEC experience, and often used retrospective designs (see Schweinhart et al., 2005; Andersson 1992, 1994). Given the high attendance rate, the number of children with no preschool experience has become very low, and thus the effect of attendance compared to no ECEC experience is not the main question of interest anymore. Therefore, research has shifted its focus from the question "does ECEC have an impact on child development" to the question of "which aspects of ECEC have an impact?" Studies have become more sophisticated, use complex designs and investigate the impact of a range of quantitative and qualitative characteristics of ECEC in a longitudinal way. Furthermore while family characteristics have been treated as control variables in early studies, recent studies also look into interaction effects of the home and preschool environments.

When comparing studies from different countries it becomes obvious that European studies often do not share such a long follow-up history compared to U.S. studies. Within Europe there are country-specific concepts of disadvantage and studies investigate different aspects of structural quality (Anders, 2013; Slot et al., 2015a; see also Table 2 in the Appendix).

In the following, we give a short overview of the main longitudinal studies in Europe which have produced most robust findings on impact recently (see also Table 1 in the Appendix):

Two large-scale longitudinal studies began in the late 1990s in the United Kingdom. The *Effective Pre-School, Primary and Secondary Education* (EPPSE) study followed over 3000 children in England from pre-school to post-compulsory education (see Sammons et al., 2002; 2004a; 2004b; 2007; 2008; 2011; 2014). It focused on identifying which aspects of ECEC have an impact on children's attainment, progress and development in cognitive, academic and socioemotional domains. It focused on the effect of quantity, structural and process quality and the type of ECEC provision. The EPPSE study is the best known longitudinal study on ECEC effects in Europe. A parallel study was conducted in Northern Ireland following over 800 children until the age of 11 years, and focusing on same outcome domains and same aspects of ECEC provision (see Melhuish et al., 2002; 2004; 2012).

A large-scale longitudinal study named *Bildungsprozesse, Kompetenzentwicklung und Selektionsentscheidungen im Vor- und Grundschulalter* (educational processes, competence development and selection decisions in preschool- and school age, BiKS) started in two German states in 2005, with a sample of 547 children (von Maurice et al., 2007). The children are assessed regularly from the age of three on cognitive and socioemotional outcomes in relation to quantity, and structural and process quality of ECEC, and the data is still being collected. At about the same time in Finland the *First Steps (Interaction and Learning within Children-Parent-Teacher Triangle)* study started with over 2000 children, who were followed from their kindergarten year to the end of fourth grade in primary school (see Pakarinen et al., 2010; Pakarinen, Kiuru, Lerkkanen, Poikkeus, Ahonen, & Nurmi, 2011; Pakarinen, Lerkkanen, Poikkeus, Siekkinen, & Nurmi, 2011). The study aimed at establishing links between children's academic skills development during the transition from kindergarten to primary school, in relation to classroom process quality, and parent-teacher interactions. Recently, a similar cohort study has started in the Netherlands, under the name *Pre-COOL (Cohortonderzoek OnderwijsLoopbanen het jonge kind / voor- en vroegschoolse educatie)*, aiming to establish the effects of early childhood care and education on children's development and school achievement (see Slot et al., 2014).

In Europe, various programmes and initiatives aim at enhancing quality of ECEC. Some of them have been evaluated and also include measures of child outcomes in addition to ECEC quality and quantity measures (Anders, 2013; Melhuish et al., 2015). Several of these evaluated programs target families and children in need of educational support. Families are motivated to enroll their children timely into ECEC and children and families get (sometimes intense) additional educational or further support (Anders, 2013; Blok, Fukkink, Gebhardt, & Leseman, 2005). This approach can be considered as multifaceted, it provides wide-ranging services. These combined programs are assumed to be of superior effectiveness, especially when parent-training is involved (Blok et al., 2005; Burger, 2010; Ramey & Ramey, 1998), but the developmental impact of regular ECEC provision alone remains unknown. There are, however, programmes focusing on improving ECEC in regular provision and reaching disadvantage children by providing high-quality ECEC without offering additional intervention components. Those evaluation studies resemble previously described longitudinal projects and can also provide important insights into the developmental impact of variations in ECEC experiences. Examples of such evaluation studies are the German studies *Kindergarten der Zukunft in Bayern* (KiDZ; Roßbach, Sechtig, & Freund, 2010) and *Stärkung der Bildungs- und Erziehungsqualität in Kindertageseinrichtungen und Grundschule – Gestaltung des Übergangs* (TransKiGS; Fried, Hoeft, Isele, Stude, & Wexeler, 2012), the federal German initiative *Early Chances* (Anders et al., 2015; Flöter, Weigel, & Schmerse, 2015), and the Dutch *Utrecht Mixed Preschool Groups* study (de Haan, Elbers, Hoofs, & Leseman, 2013).

Despite differences in details, these European longitudinal studies show great resemblance in design characteristics, research questions, as well as measures of ECEC experiences and developmental outcomes. Therefore, they form a relatively homogenous sample of primary studies of high quality covering a wide range of ECEC experiences in Europe and relationships to child outcomes.

Evidence for the impact of quantity and quality

Various studies around the world have investigated ECEC's developmental impact, and an exhaustive review is provided elsewhere (e.g., Anders, 2013; Burger, 2010; Melhuish et al., 2015). This section therefore provides a brief and general overview about main findings and studies in the field. Primary studies and syntheses of studies point to the link of ECEC to a broad range of developmental outcomes, including emotional and physical health, behavioural and socioemotional development and academic outcomes. Studies on health and behavioural outcomes often include studies on the effects of specific interventions, programmes that combine a type of ECEC provision with home-visits and/or parent-trainings or other services beyond education and care (Nores & Barnett, 2010; Blok et al., 2005). Studies focusing on ECEC mainly study impacts on a range of cognitive outcomes, academic and socioemotional outcomes. Evidence shows associations of several aspects of ECEC quantity and quality to these child outcomes.

With regard to the effects of the "dose" of ECEC participation of children under the age of three years the many US-American and European studies reported beneficial effects of an early starting age and advantages of institutional care compared to informal care settings on the development of language and cognitive skills (NICHD, 2005; Sammons et al., 2002; Loeb, Fuller, Kagan, & Carrol, 2004; Bernal & Keane, 2007; Gregg, Washbrook, Propper, & Burgess, 2005; Hansen & Hawkes, 2009; Love et al., 2003; Sylva et al., 2011b; Broberg, Hwang, Lamb, & Bookstein, 1990; Broberg, Wessels, Lamb, & Hwang, 1997). However, some large European studies also report inconsistent effects (Driessen, 2004; Sammons et al., 2008). Research evidence on socio-emotional outcomes is even more heterogeneous. A number of studies, comprising the well-known, large and comprehensive US-American NICHD-study found evidence for negative effects of early institutional care on developmental aspects such as problem behaviour and less prosocial behaviour (NICHD, 2002c, 2003a). Other authors reported null effects (Bornstein, Hahn, Gist, & Haynes, 2006; Votruba-Drzal, Coley, Maldonado- Carreño, Li-Grining, & Chase-Lansdale, 2010; Love et al., 2003; Bassok, French, Fuller, & Kagan, 2008).

The research evidence is more consistent with regard to the intensity of care in the early years and the impact of care stability. In the first two years of a child's life the use of institutional care of more than 6 hours a day or 24 hours a week seems to go hand in hand with less beneficial effects on cognitive outcomes and a higher likelihood of negative effects on socio-emotional outcomes (Anders, 2013). Furthermore the stability of care arrangements is important to avoid negative effects in the socio-emotional domain and to achieve best results in the cognitive domain. With respect to ECEC for children age 3 and older, the quantitative aspects seem to be less important. Some evidence exists that the overall duration is relevant for cognitive and language-related outcomes, especially for children from disadvantaged backgrounds (Sylva, Melhuish, Sammons, Siraj-Blatchford, & Taggart, 2004; Bos et al., 2003; Bassok et al., 2008). But overall, half-day programmes seem to be as effective as full-day programmes (Gormley Jr., Gayer, Phillips, & Dawson, 2005; Sammons, 2010).

Studies that analysed the effects of structural aspects of quality produced mixed findings. Often the expected link to a setting's process quality could not be established empirically (see Slot et al., 2015a). At the same time, studies could not show the expected indirect effects of structural quality on children's development (Anders, 2013). A possible explanation for the inconsistent relationships and effects can be that, due to quality regulations within countries, the range of structural and process quality within countries (and study regions) is also restricted (Love et al., 2003). Slot and colleagues (2015a) argued that a further reason for the inconsistency of effects may be that potential interaction effects are not sufficiently considered in most analyses. For example, beneficial effects of a higher level of teacher qualification or a smaller ratio of children to staff may be dependent on the composition of the group of children.

But with respect to process quality, various studies from different countries have proven that higher process quality predicts children's learning gains and development in different developmental areas (cf. Anders et al., 2012, Anders, Große, Roßbach, Ebert, & Weinert, 2013; Belsky et al., 2007; Ebert et al., 2013; ECCE Study Group, 1999; NICHD ECCRN, 2003, 2005; Peisner-Feinberg et al., 2001; Sammons et al., 2008a; Sylva et al., 2004; Tietze, Hundertmark-Mayser, & Roßbach, 1999; Vandell et al., 2010). Effects tend to be higher and more stable for children aged 3 years and older than for younger children, also effects are higher and more consistent for cognitive and language-related outcomes compared to socio-emotional development (see Anders, 2013 for a review). Although effects decrease over time, the benefits of attending a high quality ECEC setting remain traceable years later, even when children have reached adult age (Reynolds, Ou, & Topitzes, 2004; Schweinhart et al., 2005). Having attended an ECEC setting of high process quality does not only provide children with a better start for formal schooling, but also relates to better progress while children move through primary and secondary school (Anders et al., 2013; Melhuish et al., 2008a, b; Sammons et al., 2008b, 2011; Sylva, Melhuish, Sammons, Siraj-Blatchford, & Taggart, 2011a). However, with regard to the question if disadvantaged children benefit more from higher process quality than more privileged children, research evidence is mixed.

Current Research Syntheses – Findings and Shortcomings

Several research syntheses summarize the huge amount of studies on the developmental impact of ECEC worldwide, simultaneously providing an illustrative picture of the differences in study designs and effects (Ahnert, Pinquart, & Lamb, 2006; Anders, 2013; Anderson et al., 2003; Blok et al., 2005; Camilli, Vargas, Ryan & Barnett, 2010; Chambers, Cheung, Slavin, Smith & Laurenzano, 2010; Dalli, White, Rockel, & Duhn, 2011; Gilliam & Zigler, 2001; Gorey, 2001; Keys et al., 2012; Melhuish, 2004; Melhuish, et al., 2015; Mitchell, Wylie & Carr, 2008; Nores & Barnett, 2010; Kelchen et al., 2011). Reviews and meta-analyses have repeatedly shown beneficial effects of ECEC participation on children's cognitive, academic and socioemotional developmental outcomes (Anderson et al., 2003; Burger, 2010; Camilli et al., 2010; Gorey, 2001). In general, these syntheses tend to agree that ECEC participation has potential benefits for children.

Since childcare is becoming an increasingly universal experience for children, rather than noting the beneficial effects of ECEC attendance per se, there is a need for synthesizing evidence for the developmental impact of variations in ECEC experiences. Various meta-analyses and reviews claim that the effects of ECEC on children's development depend on different aspects of quantity and quality of the ECEC experience (Anders, 2013; Blok et al., 2005; Burger, 2010; Melhuish et al. 2015).

Nowadays, the scientific literature provides a rich source for describing and measuring these differences as described above. Notwithstanding slight differences in definitions, aspects, and measures, studies use a similar core set of dimensions of ECEC quantity and quality (as previously outlined), and there seems to be general understanding about how these dimensions interact and affect child development (see Figure 2a). These dimensions have been shown to be applicable to describe and study the differences of ECEC experiences across countries, various programmes and pedagogical approaches, different types of provisions, and groups of children.

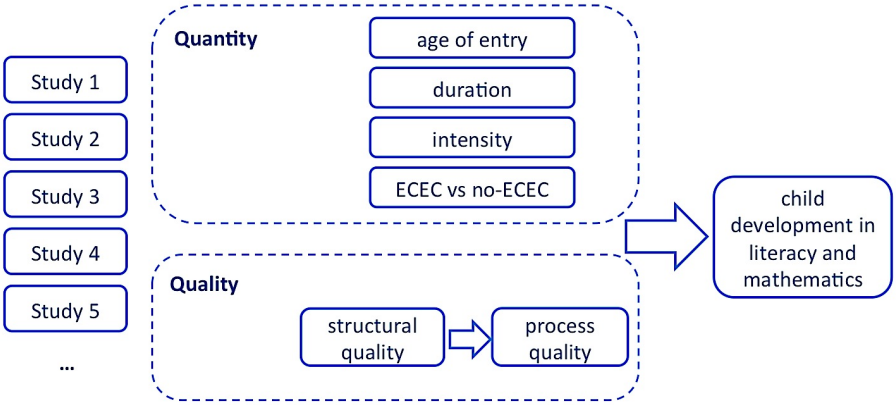

Figure 2a. Schematic overview of frequently studied quantity and quality dimensions of ECEC affecting developmental outcomes.

Attempts to synthesize scientific knowledge and evidence about core characteristics of ECEC quality in order to inform policy makers and the scientific community have been made previously (Anders, 2013; Burger, 2010, Melhuish et al., 2015).These syntheses seldom went beyond systematically reviewing the findings. Statistically integrating the findings (i.e. meta-analysis) and providing integration of quantitative results may yield more concise knowledge about the relative importance of different ECEC aspects, possibly leading to clearer policy implications (see Figure 2b).

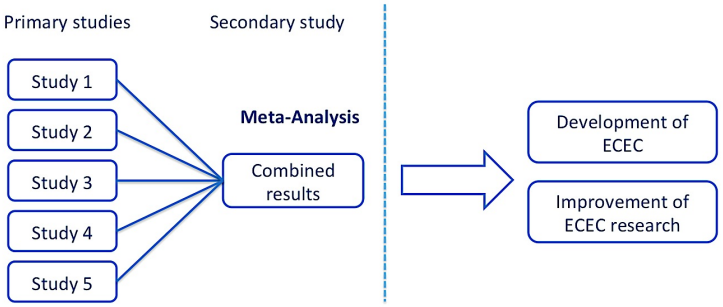

Figure 2b. Value and function of a meta-analysis on developmental impact of ECEC.

Previous meta-analyses mainly focused on the question if ECEC has developmental impact, instead of studying the relevant question of which aspects have developmental impact. They also suffer from several further shortcomings, which ultimately calls for new meta-analytical approaches:

Firstly, they were mainly based on US evidence (Gorey, 2001; Blok et al., 2005; Camilli et al., 2010), and some included studies are methodologically weak or not up to date (e.g., various studies on Perry-Preschool do not represent current reality of ECEC; Anders et al., 2013). Even when aiming for

an international scope (e.g., Nores & Barnett, 2010), previous meta-analyses largely lack important European evidence. Given important differences between ECEC systems and research in Europe and the USA (Burger, 2010), the current meta-analysis focuses on findings about the developmental impact of ECEC in Europe.

Secondly, many meta-analyses exclude relevant studies that do not adhere to a certain research design. They mainly use findings derived from group contrast designs (see Figure 2c). Often meta-analyses on the impact of an intervention outside of educational science limit their synthesis to findings from random-assignment, controlled experimental studies (Cooper, 2009; Higgins & Green, 2008). For conclusions on the impact of ECEC that means that they only use the studies that contrast the children who are exposed to a certain ECEC programme with children not exposed to any comparable programme (i.e., attending another programme or not attending ECEC at all), but are equivalent in important background characteristics (Burger, 2010). This ensures that observed differences in developmental outcomes are mainly linked to the evaluated programme. In educational science, restricting included evidence to randomized controlled studies leads to an exclusion of a large amount of relevant studies, which represents a serious threat to the validity of meta-analytical results (Card, 2012). Hence, the majority of meta-analyses also include studies that adopt quasi-experimental designs and investigate the impact of naturally occurring variations in ECEC experience, for example contrasting children attending ECEC to children without ECEC experience, or comparing children participating in different ECEC programmes without random assignment. They usually use some coefficient of standardized mean difference such as Hedge's *g*, Cohen's *d* or Glass's *g* (Card, 2012) as measures of the effect sizes. Despite this widening of selection criteria for relevant studies to quasi-experimental studies, the published meta-analyses generally exclude studies without contrast designs, which includes a number of high quality European studies, studying children in regular provision without a sample of children lacking ECEC experience, or samples of children in different ECEC programmes.

Thirdly, as other meta-analyses normally use findings from studies with group contrast designs, they entail a large proportion of findings for specific educational programmes (e.g. Head Start, Perry-Preschool), sometimes even combining ECEC with other early intervention components (e.g. home-visits, parent-trainings, nutrition programmes). A multifaceted approach and providing more wide-ranging services (Burger, 2010; Ramey & Ramey, 1998), especially when parent-training is involved (Blok et al., 2005), is assumed to be of superior effectiveness. Hence the developmental impact of regular ECEC provision alone remains unknown.

Fourthly, meta-analyses mostly studied the effects of variations in quantity and quality of ECEC indirectly, by rating the assessed programmes in these regards (Gorey, 2001; Blok et al., 2005). Generally programmes of higher quality and quantity are linked to stronger benefits. These aggregated findings do not allow for estimating the developmental gain that is to be expected by increasing the quality and/or quantity in ECEC provision.

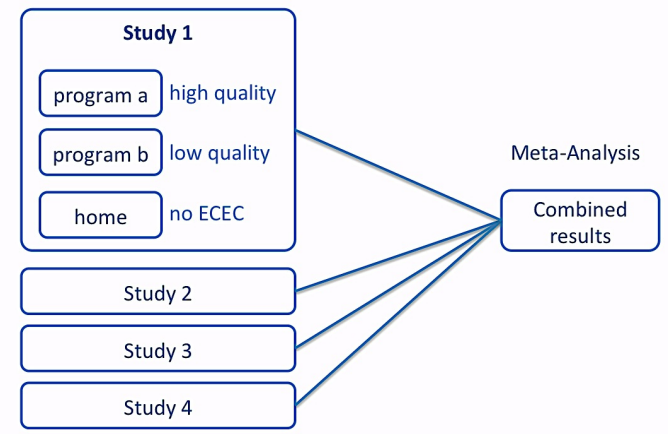

Figure 2c. Common meta-analytic approach for the impact of ECEC using studies with group comparison design and studying indirectly the effects of quality and quantity aspects.

A fifth shortcoming of the existing meta-analyses on ECEC effects is that they subsume effects for a great variety of outcomes into one overall effect for broad developmental areas such as cognitive outcomes, health outcomes, and socioemotional development. Aggregating findings for academic outcomes is a promising meta-analytical focus, as some quantitative syntheses compared effects for cognitive outcomes, including academic outcomes, to outcomes in other domains and found that results tend to be stronger and more homogenous (Blok et al., 2005; Camilli et al., 2010; Nores & Barnett, 2010). Despite comparably homogenous empirical results for cognitive outcomes, the previously subsumed findings for outcomes of the cognitive domain are based on outcome measures of extremely heterogeneous nature, e.g. IQ measures, measures of long-term and short-term memory, and other measures of cognitive functioning, as well as outcomes in literacy and mathematics (Camilli et al., 2010; Nores & Barnett, 2010). Reviewing the effects for cognitive outcomes, Burger (2010) emphasized that reading, vocabulary, writing and math scales, in contrast to IQ tests and test of basic cognitive skills, primarily measured the acquisition of what is taught in ECEC or schools. We deemed it reasonable to focus on synthesizing the outcomes that lie in the scope of ECEC's tasks. According to most curricula of ECEC in different European countries (Anders, 2014; Sylva et al., 2015), fostering development in different pre-academic domains (e.g. literacy, mathematics, science and art) is one of the core tasks of regular ECEC provision. Even though fostering socioemotional development is another core task of ECEC, the effects tend to be weaker and less homogenous (Camilli et al., 2010; Blok et al., 2005). For cognitive outcomes it is not very likely that positive effects on one specific aspect would be accompanied by negative effects on another (Nores & Barnett, 2010). On the other hand, socioemotional outcomes form a broad domain, and it is possible for ECEC to have both positive and negative effects on differing aspects of this domain (NICHD, 2002a, 2002b, 2005).

To sum up: the value of meta-analyses on the impact of ECEC lies in their potential to deduce the important implications for ECEC policy and research based on a comprehensive and systematically combined body of evidence. However, we noticed important shortcomings of previous syntheses: Firstly, most meta-analyses focus strongly on US evidence including that of weak quality, while neglecting the available European evidence. The focus on American studies, as well as on the effectiveness of specific programmes, puts in question the generalizability of the findings to the European context. Secondly, they address the effect of ECEC quantity and/or quality indirectly by

comparing various programmes differing in those regards with one another, instead of studying the developmental impact of each aspect separately and directly. Thirdly, they subsume developmental outcomes of very heterogeneous nature under one aggregated finding for outcomes in the cognitive domain.

In order to provide the evidence basis for a European ECEC framework, WP4 of the CARE project addresses these shortcomings. We provide a synthesis of relevant European longitudinal studies with high quality designs by directly estimating the aggregated effects of various aspects of ECEC quantity and quality (see Figure 2d). We aggregate effect sizes for the strength of the relationship between ECEC quality and quantity and child outcome to study directly their relative overall importance. Furthermore, we focus on the developmental impact of regular ECEC provision on developmental outcomes in two significant academic domains: literacy and mathematics. Specifically, the core question that needs answering is which aspects of ECEC are the most relevant to explain its developmental impact on the academic outcomes in European settings of regular centre-based care.

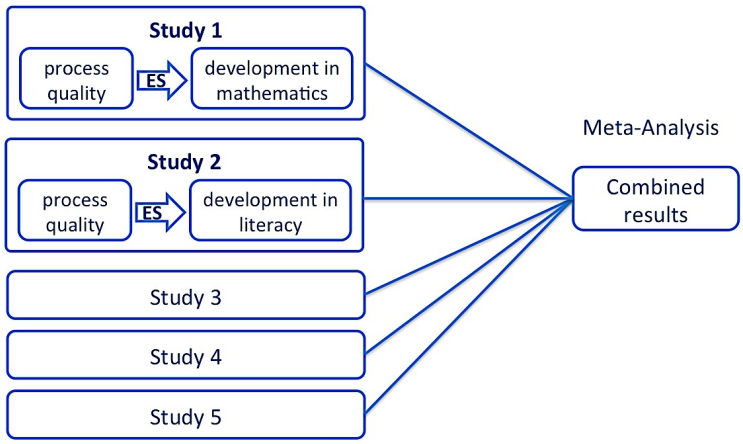

Figure 2d. Alternative meta-analytic approach for the impact of ECEC using studies without group comparison design and studying directly the effects of quality and quantity aspects.

Studying Moderators of ECEC Effects

The aim of a meta-analysis is to aggregate evidence of primary studies investigating the same research question, e.g. ECEC's impact on developmental outcomes. Therefore, primary studies and reported findings need to be comparable, but they still differ to some extent. The included studies in meta-analyses on ECEC's impact differ, for example, in studied region and its ECEC system, in designs and sample characteristics, or in the measures used to assess the outcomes or the ECEC aspects. These differences are expected to be to some degree linked to differences in reported findings. It is important to explore the heterogeneity in findings and study its potential explanations. Meta-analysis allows for studying potential sources of variations when findings are heterogeneous, i.e. a moderator analysis (Card, 2012; Cooper, 2009). Some study characteristics appear as consistent moderators across several meta-analyses on ECEC's impact, and hence should be included in every meta-analysis on this topic. Our meta-analysis differs from others in important aspects, as outlined in the previous section; we included additional moderators than have been subject of other meta-analyses.

Previous quantitative syntheses generally subsume developmental outcomes in the cognitive domain without studying the domain-specificity of effects (Camilli et al., 2010; Nores & Barnett, 2010). Current ECEC policy and scientific literature emphasize the role of domain-specific learning processes and promotion (Sylva et al., 2015; Roßbach & Weinert, 2008). So far, meta-analyses have not studied if aggregated ECEC effects diverge between domains, meaning if effects are consistent across various domains (Reynolds et al., 2014).

A recent systematic review focusing on both US and international large-scale studies, including several birth cohort studies, found strong positive effects for short-term and weaker for long-term cognitive outcomes (Burger, 2010). There seems to be a general tendency of effects in the cognitive domain to fade with age (Heckman, 2006; Camilli et al., 2010, but see Nores & Barnett, 2010). Instead of assuming a gradual decline, some primary studies on the impact of ECEC studied if the exposure to new institutional influences, like the quality and quantity of primary and secondary schools, lead to a decrease in the strength of relationships of the ECEC experiences and the outcome (Anders et al., 2013; Lehrl, Kluczniok, & Roßbach, 2015; Sammons et al., 2011). Any new institutional influences may affect the strength of relationship between ECEC experiences and child development.

To measure the same aspect of ECEC experience, scientific literature provides a rich source of different measures. By choosing different measures, studies assess slightly different aspects of ECEC experience, though sometimes labelled with the same scientific term. Choice of measure can also affect reliability and validity of assessments. Both can influence how strongly the aspects relate to the child outcomes and can explain the differences in findings across and within studies. Scientific literature needs to compare different measures conceptually and empirically. It is important to clarify which aspect of ECEC experience each measure exactly covers and if some measures are more strongly related to child outcomes. Some reviews, for example, point out important differences between concepts used in primary studies and link them to differential findings (Anders et al, 2013; Melhuish et al., 2015). There are also examples of primary studies comparing different measures: The study of Pianta et al. (2005) and the results of the secondary analysis of CARE (Slot et al., 2015a) compared measures of process quality and investigated how they relate to the structural aspects of ECEC. Recent research projects studied if staff questionnaires about processes in ECEC are equally reliable and valid as observational measures (Camehl et al., 2015; Anders & Ballaschk, 2014). The ECCE study conducted a cross-country comparison of the relationship of various aspects of structural quality to child outcomes (Tietze et al., 1999). The relationships of some aspects, e.g. child-staff-ratio and square meters per child, even differed in observed direction across countries. The EPPSE study (Sammons et al., 2008) studied how children attending ECEC differed from children without ECEC experience (i.e. the absolute effect of ECEC), but additionally found that variation in duration of the ECEC attendance was associated with child outcomes (i.e., relative effect of ECEC). Hence investigating if findings are generalizable across different ECEC measures has important scientific and practical implications.

Aims and research questions of the study

The objectives of WP4, including the study reported here, as stated in the Description of Work (DOW) were the following:

> To assess the impact (short-, medium- and long-term) of ECEC in Europe, including variations in "quality", on children's outcomes, including cognitive, language, social, emotional and educational development.
>
> 4.1 To review the effects of ECEC, including variations in the quality of ECEC, upon children's short-, medium-, and long-term developmental outcomes.
>
> 4.2 To identify possible moderators of the effects (both at individual and contextual level).
>
> 4.3 To investigate possible differential effects, including compensatory effects, of ECEC for children from disadvantaged backgrounds.

WP4 addresses these objectives by means of three tasks with different foci and methodological approaches, as outlined in the task description. The first report of WP4 is an extensive review of the evidence of international and European studies (Melhuish et al., 2015; D4.1). The meta-analysis builds on this review. It provides systematic overview and comparison of important study characteristics, and compresses the reviewed research evidence into aggregated results. By combining these two related methods of research synthesis we are able to sufficiently address the above specified complex research questions, providing the necessary basis for deriving ECEC policies. The review provided a broad and detailed description of the relevant past and current research and scientific literature worldwide. It embedded the pressing questions for European ECEC policy into the wider context. The meta-analysis focuses on determining the size and the nature of effects for selected key characteristics of ECEC in Europe exclusively. We use a focused quantitative approach to synthesize a spectrum of research into one single estimator for each of the ECEC aspects. This allows for a straightforward interpretation about their relative importance for children's developments. It also enables us to systematically compare effects across domains and across studies, and to deduce and statistically test explanations for differences in findings. To sum up, the current meta-analysis seeks to complete the review, to address the shortcoming of previous meta-analyses, and to inform ECEC policy makers about relative importance of key ECEC aspects for children's development in two important academic domains. Additionally, we review the evidence for differential effects for children from disadvantaged backgrounds. A quantitative synthesis did not seem feasible, because only few studies specifically address this question and they address it in very different ways. A thorough review of evidence can foster future research in the field, which paves the way for a quantitative synthesis of evidence on this important topic in the future.

In view of these aims, the present study addresses the following set of research questions:

RQ1: How strongly do global process quality, extent of pre-academic promotion, structural quality, and quantity of ECEC on average relate to developmental outcomes in mathematics and literacy?

RQ2: Do the reported ECEC quantity and quality effects differ substantially within and between studies?

RQ3: Do the following characteristics of studies and findings moderate the strength of relationship of ECEC aspects to developmental outcome?

3.1: *Domain of developmental outcome:* Are ECEC aspects stronger associated to developmental outcomes in literacy than mathematics, or vice versa?

3.2*: Age and phase of academic career*: Do ECEC effects fade out with age, i.e. gradually decline in strength, or do they decrease after children have left ECEC?

3.3 *Measures of ECEC aspects*: Do strength of effects vary dependent on the type of measures of ECEC aspects?

RQ4: Which European longitudinal evidence on differential effects for disadvantaged children exist?

Method

In conducting the present meta-analysis we followed the common steps of a meta-analysis as outlined by Cooper (2009; Card, 2012; see Figure 3). The method section outlines the three steps which follow the formulation of research questions: search and selection, data extraction and general analytical approach.

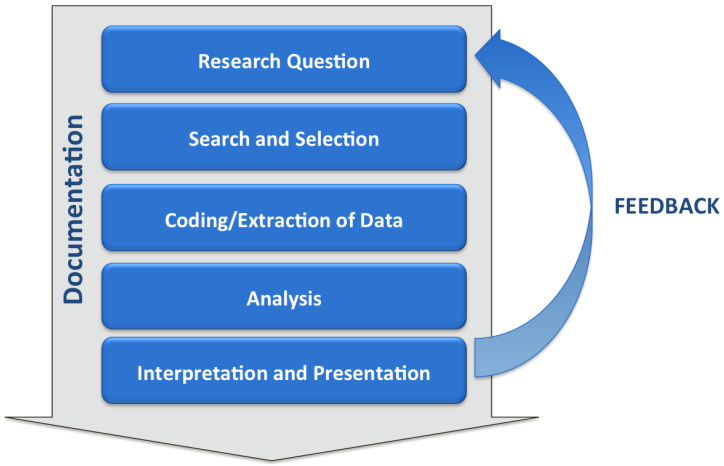

Figure 3. Steps of every meta-analysis.

Search and Selection of Studies

We conducted a systematic search and selection process by following a detailed protocol[1], which entailed the selection criteria for studies, as well as guidelines for the sources of information, search terms and their combinations.

[1] The search and selection protocols will be provided by the authors on request.

We selected studies according to the following criteria: We exclusively included longitudinal European studies, with a first measurement point collected not earlier than 1990. The studies had to assess quality of ECEC, including at least one of the following: structural quality, global process quality and pre-academic promotion, and/or quantity of ECEC, including at least one of the following: home sample, duration, age of entry and intensity. These aspects had to be analysed in relation to children's outcomes in mathematics and/or literacy. The type of care assessed had to be center-based care. Hence excluded were the studies assessing, for example, nanny or childminder care. Other exclusion criteria included using specific risk groups as sample (e.g. samples of hyperactive children or children with special needs) and evaluating a specific training programme. Studies that included intervention programme aimed at enhancing the quality of regular provision (e.g. Early Chances, KIDZ and TransKiGs in Germany, and the Utrecht Mixed Preschool Groups in the Netherlands) were included in the meta-analysis. We excluded evaluations of very specific training programmes, characterized by professionals training children in a specific set of skills over a comparably brief period of time. Those trainings do not represent children's experience in regular ECEC provision, as they are often conducted by external professionals. They normally follow a specific sequence of training sessions (relatively rigid curriculum or manual) and end after the last session.

In order to identify the publications, manuscripts and unpublished material suitable for our purpose, we conducted a search process of several steps involving different sources of information. Firstly, we contacted the CARE partners, who provided us with information about relevant research in their respective countries, especially yet unpublished work or references not published in English.

Secondly, we conducted a thorough search of electronic databases (e.g. EBSCOhost, Web of Science) and search engines (e.g. Google Scholar). We sought references that included a term related to ECEC (e.g., "childcare", "preschool", "early education") and quantity (e.g., "structure", "arrangement", "quality", or "climate") or quality aspects (e.g., "age of entry", "duration", or "home") along with a term related to the studies outcome domains (e.g., "numeracy," "math," "counting", "literacy," "reading" "vocabulary," and "language"). The search was restricted to studies in European countries (e.g., "Europe", "Germany", "France"). When needed, we further limited the search results by specifying the design of studies to longitudinal (e.g., "longitudinal", "growth" or "development").

Thirdly, we cross-checked our search database with reference lists of several reviews (e.g. Anders, 2013; Burger, 2010; Melhuish et al., 2015; Nores & Barnett, 2010) on closely-related topics identified earlier during the search process. We also checked websites of ministries and research institutes in the different EU countries for research projects of interest. As a final check of completion, we circulated our database with ECEC experts in Europe, and asked them to identify any publications or studies that were potentially missing.

Data extraction

In order to extract the necessary data from the collected manuscripts we employed a heavily structured coding scheme. Due to the complexity of study designs and analyses the coding scheme allowed for a very fine-grained extraction of different types of information ranging from information about the effects (e.g. size and significance of effects, and information about conducted analyses), applied measures for outcomes and ECEC aspect (including information about reliability and type of

instrument), and general study characteristics (e.g. timing of assessments, sample characteristics). This detailed coding was necessary to provide a complete picture about the body of evidence for the current meta-analysis, and about potential influences on the size of reported effects. Important differences in study design, adopted analysis and ECEC systems are all possible explanations for differences in effect sizes between studies. Furthermore, it is important to gather a complete picture about important study characteristics to guide improvements in the field of ECEC research. The coding scheme was developed by integrating the work of other WPs that pointed to potentially relevant moderators of ECEC effects.

To be coded, a reference had to report of at least one effect size for the relationship of a quantity or quality aspect to a developmental outcome in one domain. The effect sizes selected for coding were always part of the most controlled analysis (i.e. the model with most predictor variables), which was assumed to provide the most conservative estimation of the effect. We searched for findings for each measurement point of a study. We coded separately findings at the same measurement point for different subsamples (e.g. findings for samples of different countries in the ECCE study), individual measures of the same ECEC aspect (e.g. CIS and ECERS-R for global process quality), and different outcomes (e.g. vocabulary and pre-reading skills as literacy outcomes, or number sequence and basic arithmetic as math outcomes). Generally, we preferred findings for the total scores of the same ECEC measure. If references entailed only findings for subscales of measures, we used subscale information as an approximation of findings for the total score (e.g. findings for a subscale of the ECERS-R as a substitute for findings for its total score). Other characteristics that were extracted from the studies included descriptive information about sample sizes and characteristics, outcome assessment dates and ages of children, instruments, and details about analysis.

We had six coders extracting the data. The coders were primarily Masters or PhD students in areas such as Psychology and Human Factors. All of them possessed deep knowledge about multivariate analysis, complex and longitudinal study designs. We provided coders with the structured coding scheme, which entailed some example codings, and a coding manual with explanations of all categories[2]. They all received intensive training at the beginning and continuous support throughout the coding process.

General analysis approach

Description of effect sizes

We used standardized coefficients for the relationship of the ECEC aspects and child outcomes as the effect sizes (ES). Whenever a reference entailed a relevant finding indicating the strength of the relationship of an ECEC aspect and a developmental outcome in literacy or mathematics, we included it in our coding. A finding was relevant if it captured the association of the outcome to one of the four ECEC aspects, i.e. global process quality, pre-academic promotion, structural quality or quantity of ECEC. We could only include in our meta-analysis the effect sizes belonging to the correlation family, such as bivariate correlations or regression coefficients. Furthermore, these effects sizes had to be standardized so that they ranged from -1 to 1, with an effect size of 0 indicating no relationship between the ECEC aspect and the child outcome.

[2]The SPSS coding scheme and the manual will be provided by the authors on request.

The complexity of the designs of the included primary studies in our meta-analysis allowed for a variety of analytical approaches. Effect size measures were derived from various analytical approaches, but all represented the strength of relationship of ECEC aspect to developmental outcome and were standardized as described previously. The types of analysis ranged from bivariate correlations, standardized regression coefficients to latent growth curve modelling. Analyses also differed in the number of predictor and control variables. Most of the analyses included various child (e.g. age at assessment, gender) and family background characteristics (e.g. socioeconomic status, maternal education), as well as further variables (e.g. further ECEC aspects, quality of primary school). As longitudinal studies assessed developmental outcomes at several time points, analyses normally considered prior attainment when information was available, i.e. after the first wave of child outcome assessment. Analytical discrepancy was also apparent in further details, such as if the hierarchical structure of the ECEC system was accounted for, or if a form of missing estimation procedure was used. Analytical approaches not only varied across studies, but also within studies. For example, the effect sizes referring to growth in skills require controlling for prior attainment, which is only possible at later waves of outcome assessments. The strength of the reported findings is linked to adopted analytical approach (Becker & Wu, 2007; Bowman, 2012; Burger, 2010; Cooper, 2009).

Preparation of data for estimation of overall effects

In case a publication reported unstandardized coefficients, we standardized the coefficients post-hoc, if sufficient information was available (e.g. means and standard deviations). Otherwise we excluded it from the analysis. If information about the size of the effects was not available, we asked the authors for (standardized) results. This way we obtained unpublished results for the effects of global process quality and pre-academic promotion from the EPPSE study for some measurement points, and from the Contexts and Transition study. For the extent of pre-academic promotion we computed additional effect sizes for the NEPS study for further outcomes and measurement points using similar analyses as Durda (2015). If information on prior attainment was available, we added prior attainment as a control variable to the analysis[3]. We excluded effect sizes for vocabulary as a literacy outcome from the pre-COOL study at the first measurement point (k_{ES} = 3 for global process quality and k_{ES} = 1 for pre-academic promotion), which for the majority of children fell slightly after preschool entry. As children were exposed to the preschool environment for only a very brief period at the first measurement point, it is unlikely that preschool effects were already present, and that the effect sizes for this measurement point represent the developmental impact of Dutch preschools. Furthermore, the pre-COOL study is largely based on Dutch large-scale targeted programs Voor- en Vroegschoolse Educatie (VVE, Pre- and early primary school education; Slot, 2014). These well-structured education programs are mainly reserved for children from disadvantaged background and put strong emphasis on children's cognitive and language skills. That is why effect sizes at the first measurement point are partly biased by selective placements of disadvantaged children into these education-oriented settings.

According to the common theoretical distinction (see theoretical background) we built four separate overall ECEC effects (see Figure 4): quantity, global process quality, pre-academic promotion and structural quality, all of which include effect sizes for the relationship to various measures of literacy

[3] M*plus* outputs of the analyses will be provided by the authors on request.

and mathematics outcomes at different ages (see Table 3 in the Appendix for details). Apart from differing outcome measures, comparable but slightly different measures of the ECEC aspects contributed to the overall effects.

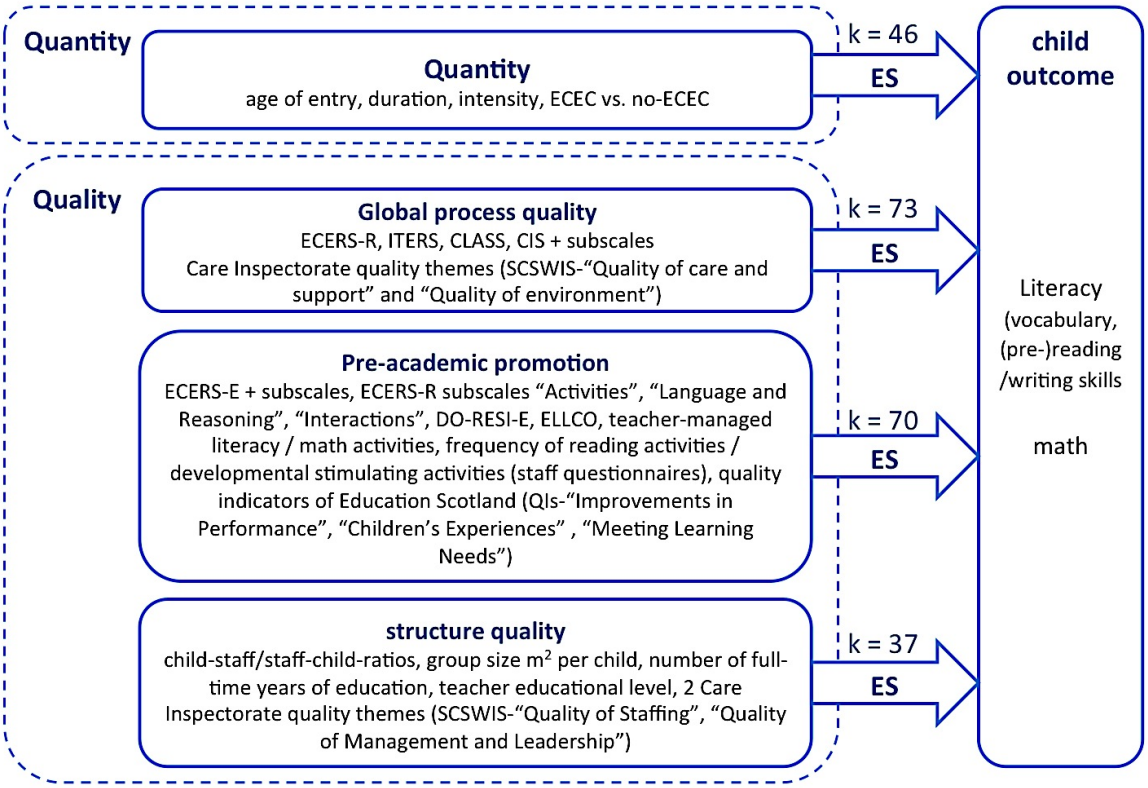

Figure 4. Grouping of separate findings to the four overall effects (k = number of included findings).

Global process quality measures assess the nature of the pedagogical interactions between preschool teachers and children, covering aspects such as warm climate and child-appropriate behaviour. Most studies used established research instruments based on observations, like ECERS-R, CLASS or CIS and their subscales. The most commonly used Early Childhood Environment Rating Scale (ECERS-R) is an observational instrument based on child centred pedagogy, assessing resources for indoor and outdoor play (Harms et al., 1998). The Infant-Toddler Environment Rating Scale (ITERS) is used to assess activities, materials, interaction and programme structure in group settings for infants and toddlers up to 30 months of age (Harms, Cryer, & Clifford, 1990). The Caregiver Interaction Scale (CIS) is used to assess aspects of positive relationships between children and caregivers, as well as punitiveness, permissiveness and detachment of the caregivers (Arnett, 1989). The Classroom Assessment Scoring System (CLASS) is a measure of classroom organization, and emotional and instructional support (Pianta et al., 2008). Furthermore country-specific inspectorate assessments were also included as instruments for global process quality, i.e. two of the self-evaluation scales of the Social Care and Social Work Improvement Scotland (SCSWIS- scales "Quality of care and support" and "Quality of environment"; Bradshaw et al., 2014).

Pre-academic promotion measures mainly the extent of pedagogical stimulation in a certain areas such as mathematics, science and literacy. The most commonly used instrument was the Extension of the Early Childhood Environment Rating Scale (ECERS-E) or subscales. ECERS-E assesses the quality of learning environments for verbal, mathematics and science literacy, as well as for taking care of diversity and individual learning needs (Sylva et al., 2003). Two additional observational measures were used to assess classroom environments: the extended version of the Dortmunder Rating Scale (DO-RESI-E, Fried et al., 2012), and the Early Language and Literacy Classroom Observation (ELLCO). DO-RESI-E assesses pre-academic promotion in mathematics, science and literacy. ELLCO consists of a checklist providing inventory of literacy-related items in the classroom, followed by an observational measure assessing literacy-related classroom practices and a rating scale assessing literacy-related activities (Smith, Dickinson, Sangeorge & Anastasopoulos, 2002). Other observational measures included ECERS-R subscales referring to stimulating activities in various domains (ECERS-R subscales "activities", "language and reasoning" and "interactions") and a self-constructed observational instrument of teacher-managed literacy and math activities. It also entails measures based on staff questionnaires about the frequency of reading activities or developmental stimulating activities in general, as well as self-evaluative quality indicators of Education Scotland (QI- "Improvements in Performance", "Children's Experiences" and "Meeting Learning Needs"; Bradshaw et al., 2014). These quality indicator subscales reflected the provision of learning opportunities and the promotion of children's learning achievements, active involvement and motivation.

Structural quality measures, as originally coded, were amongst the most heterogeneous ones, ranging from various indicators of staff qualification, structural arrangements such as group size, child-staff ratio or m^2 per child, to indicators of group composition such as the proportion of children with migration background or of low-performing children in the group, or group composition by maternal education. We decided to exclude the findings for work experience of staff and for the effects of group composition, as both of these aspects are hard to regulate effectively. Work experience is often measured as the number of years working in ECEC, which is in turn strongly related to the age of the teachers, while using policy regulation for group composition is questionable for ethical and practical reasons. Instead we focused only on those structural aspects which are more clearly subject to legislative regulations (i.e. structural arrangements and staff qualification). These aspects are mainly aspects of the so-called "iron-triangle" dimensions (Hayes, Palmer & Zaslow, 1990). Hence the effect of structural quality is based on measures of child-staff/staff-child-ratios, group size and m^2 per child, as well as different measures of staff qualification, i.e. number of full-time years of education, educational level of teachers, as well as two SCSWIS-scales ("Quality of Staffing" and "Quality of Management and Leadership"; Bradshaw et al., 2014).

Quantity of ECEC experience include findings for variations in the relative amount of ECEC provision received by children, as well as for the absolute effect of ECEC on developmental outcomes. Findings for the absolute ECEC effect are derived from direct comparison between children attending an ECEC setting and a matched group of control children not attending ECEC at all (*Home sample*). Variations in the relative amount of ECEC include findings for measures such as duration, intensity, and entry age. *Duration* referred to the time spent in any type of centre-based care, measured in years or months. *Intensity* of ECEC was measured either in how many hours per week the child spent in centre-based care, or if they were enrolled in the setting on full-time or part-time basis. *Entry age* referred to how old the children were when they first enrolled into centre-based care. The effect of

quantity also entails combinations of these effects when no separate effect was reported, for example a comparison of children with varying ECEC duration with children from a home sample.

When theory implied a reversed relationship for a specific effect, we built the inverse of the reported effects post hoc. For example, in the case of effect sizes for global process quality, the CIS Detachment and Punitive subscales are expected to relate negatively to child outcomes, so we reversed the reported effect sizes. For those structural aspects which according to literature represent challenges to learning situations in institutional contexts, i.e. bigger class sizes, higher child-teacher-ratios (only when measured as ratio of children to teacher and not vice versa) we used the inverse of the effect size. The same was done for the age of entry, as an earlier entry age is linked to a longer ECEC duration.[4]

After that we used Fisher's *z* transformation of effect sizes to normalize their distribution and to stabilize the variances of the effect sizes (see Borenstein, Hedges, Higgins, & Rothstein, 2009). Transformation was done prior to combining the effects and the results were converted back after the analyses.

Aggregation of effects in longitudinal multilevel meta-analysis and assessing heterogeneity

Longitudinal studies, which are the focus of this meta-analysis, report effect sizes for the same relationships at different time-points and various stages of the educational career of children. Due to the complex designs of these studies, sometimes various effect sizes for the same measurement point are given, for example for the relationship to literacy and math outcomes or for two different measures of the same ECEC aspect. We only kept information of the same ECEC aspect if they presented unique information, e.g. relationship of global process quality to child outcome of the same observational instrument, but at different measurement points (e.g. BiKS), or the effects for different subscales of the CIS measure (e.g., EPPNI). Whenever the studies reported effects of the total scale of a specific ECEC aspect, we did not include the effects of the subscales of the same scale (e.g. effect of ECERS-E mathematics was not included when the effect of ECERS-E total score was available). The Utrecht Mixed Preschool Groups study reported findings for two separate samples of different ages, which we included in the analysis as effect sizes for different measurement points of the same study. The findings for different samples of the ECCE study stem from separate European countries and were therefore considered as separate studies in our sample.

Multiple effect sizes within the same study are not independent and share variance or error present in the study. We used longitudinal meta-analyses as outlined by Maas, Hox and Lensvelt-Mulders (2004) to account for the nested structure of our meta-analytic data: effect sizes nested within time-points within studies (see Figure 5).

[4] There is still an unresolved and on-going debate about potential harmful effects of enrolment in the first years after birth (especially under one year of age). Included studies for entry age did not cover very early enrolment ages.

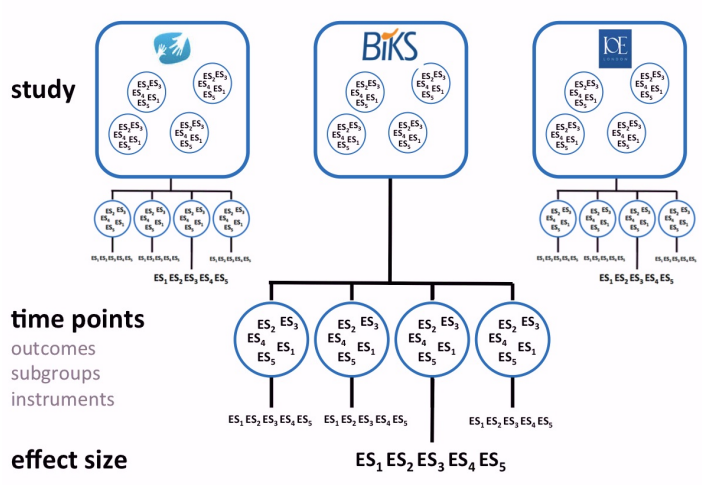

Figure 5. Illustration of the multilevel structure of our meta-analytic data (i.e., effect sizes for different outcomes, subgroups, and instruments nested in measurement points nested in study).

For all analysis we used different packages of the statistical programme R (R Core Team, 2013). For the meta-analytical part of the analysis we used the metafor package (Viechtbauer, 2010). For each ECEC effect separately we specified a three level model. The lowest level included separate effect sizes, the second level specified the measures of associations at the same age of the children (i.e. effects at same measurement points of the same studies), and the third level indicated the various studies. Our models allow the true ECEC effects to differ between and within studies, because we assumed that the effect sizes in our meta-analysis represent random samples of the true ECEC effects. We used a restricted maximum likelihood (REML) estimator in our analysis as recommended by Viechtbauer (2005). We assigned weights to the findings from each study when aggregated into overall effects with more weight given to larger studies. Instead of using sample size for assigning weights we weighted each effect by the inverse of its variance (see Figures 8a-d in the result section for assigned weights for each effect).

For all models we report the number of aggregated effect sizes *(k_{ES})* and studies ($k_{studies}$), as this has important implications for the interpretation of aggregated results. Aggregations based on a smaller number of effect sizes and studies should be interpreted with greater caution with regard to certainty of estimation, and generalizability of findings (Card, 2012; Cooper, 2009). We also provide an estimation of the overall effects with corresponding standard errors (*se*) and confidence intervals ($CI_{95\%}$). We report different measures of the heterogeneity of findings including estimated variance between studies (σ^2_{study}) and within studies (σ^2_{age}) and the Cochran's Q-test (Hedges & Olkin, 1985). Cochran's Q-test of homogeneity with a low *p* value (meaning a large chi-squared statistic relative to its degree of freedom) provides evidence of heterogeneity of ECEC effects. Thus, variations in effect estimates are beyond chance (Higgins & Green, 2008). We also visually display heterogeneity of effects with forest plots for each effect.

Building and Testing Moderators of ECEC effects

Meta-analysis allows for studying potential sources of variations when findings are heterogeneous (Card, 2012; Cooper, 2009). We studied moderators by using mixed-effects model based on the multilevel-models described earlier (Viechtbauer, 2010). Meta-analytic mixed-effects models

resemble mixed-effects models with data of primary studies, but in these models the effect sizes are the criterion variables and the study and/or design characteristics are the predictors (Cooper, 2009). Included studies in our meta-analysis differed in various important characteristics (e.g. study region and its ECEC system, study designs, sample characteristics; see Tables 1 to 4 in the Appendix), and it is likely that some of these differences are linked to strengths of reported effects, i.e. they function as moderators of effects (see Figure 6). So besides the descriptive overview of study characteristics, we investigated if differences in reported effect sizes are associated with differences in characteristics of the studies and applied analyses.

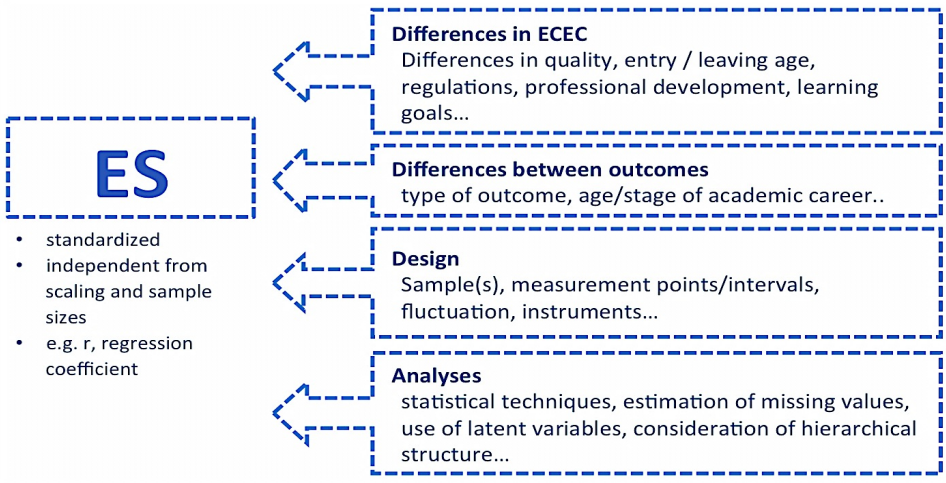

Figure 6. Schematic diagram of possible explanations for differences in effect sizes (*ES*) leading to heterogeneity of overall findings (i.e., potential moderators).

Following our research questions, we studied separately several moderators. Table 1 gives an overview of all moderators and describes how we grouped the effect sizes to build categorical moderators. The following continuous and categorical moderators referred to the studied outcome and were investigated for each of the four overall effects similarly: domain of developmental outcome, and age and phase of academic career at outcome assessment. We built moderators referring to the type of measure for each of the four ECEC effects differently. Additionally, we studied if the effect sizes were derived from a peer-reviewed article or a non-peer reviewed reference to test for publication bias.

The moderators were built as follows:

1.) *Domain of developmental outcome:* As a first moderator, we compared the overall effects in the two developmental domains to study if domain of developmental outcome is a moderator of ECEC effects. Based on the idea of domain-specific promotion and learning processes (Roßbach & Weinert, 2008; Sylva et al., 2015), we tested if the four aspects of ECEC experience related more strongly to literacy than to mathematics or vice versa. To build a moderator we split effect sizes into findings for the relationships to literacy versus mathematical outcomes.

2.) *Age and phase of academic career*: To test the persistency of relationships we introduced age at outcome assessment as a continuous moderator. It is likely that strength of relationships to ECEC experience fade as children grew older (Blok et al., 2005; Burger, 2010), meaning that associations decline gradually with age. Additionally, we composed an additional moderator variable by splitting outcomes into assessments when children were still in ECEC and assessments taking place after children had left ECEC. Exposure to new institutional influences, like the quality and quantity of primary and secondary schools, can lead to a decrease in strength of relationships to ECEC experiences. Every phase representing new institutional influences might be followed by a subsequent decrease in relationship strength. Therefore separating effect sizes according to phase of educational levels might be a more adequate conceptualization of a moderator than age. Based on the limited amount of long-term evidence we split effect sizes into two categories ("in ECEC" and "in school"), even though a more fine-grained categorization would have been preferable. The categorization took into account the specific school enrolment ages of the country and the time-period in which the study was conducted.

3.) *Measures of ECEC aspects*: Even for the same aspect, ECEC measures differed across and within studies, which may partly explain differences in findings. We therefore grouped effects according to types of *measures of ECEC aspects and included them as moderators* in the analysis. For this moderator, grouping was different for each ECEC aspect. For some aspects measures were more comparable than for others.

   a.) For global process quality, for example, most measures were observational and shared similar core facets (e.g. warmth and responsiveness of interaction between child and caregiver). Nonetheless, they differed in some regards, for example in the relevance of the interaction with the material surrounding. The ECERS-R and ITERS, which is widely used in European ECEC research and practice, integrates ratings of the availability of and access to material and of further environmental characteristics into their overall quality ratings. Other measures like the CLASS or CIS do not capture aspects of the material surrounding in their ratings. We compared effect sizes derived from measures which include ratings of the material surrounding in their overall quality ratings, i.e. ECERS-R and its subscales, the ITERS, and the SCSWIS-scales, with effect sizes derived from measures which focus on interactions, i.e. CLASS or CIS and their subscales.

   b.) Measures of pre-academic promotion mostly assess frequency of activities in different academic domains like language, science or mathematics. Not all of the applied measures were observational ones. There is an on-going debate if quality is reliably assessable with staff questionnaires. We compared the effect obtained through questionnaires, such as questionnaires asking educators about the

frequency of stimulating activities, to the effect obtained by observational measures such as the ECERS-E, ELLCO, DO-RESI-E or the subscales of the ECERS-R which refer to the frequency of pre-academic promotion (i.e. “language and reasoning”, “activities”, “interactions”).

c.) Measures of structural quality were the least comparable among the measures of ECEC aspects. Even after excluding aspects of group composition, structural aspects ranged from various indicators of variation in staff qualification to variations in structural arrangements. We compared the effect of aspects of structural quality relating to staff qualification, such as educational level of teachers or number of full-time years of education, to the effect of aspects relating to structural arrangements, like group size, child-staff-ratio, and m^2 per child. We expected qualification to relate more strongly to child outcome, because it should be more directly linked to teacher’s actions, which in turn is relevant to children’s learning (Fukkink & Lont, 2007; Jensen et al., 2015; Kelley & Camilli, 2007); the impact of structural arrangements on children’s learning processes is less direct. Certain structural arrangements like smaller group sizes and smaller staff-to-child-ratio can facilitate and support learning processes by providing better conditions for rich learning situations (Kluczniok, & Roßbach, 2014; Slot, Leseman, Verhagen, & Mulder, 2015b), but existing research evidence suggests that if learning takes place still depends on the pedagogical actions in the setting and not the arrangements of the settings itself.

d.) Quantity effects are combinations of effect measures covering effect sizes for intensity and duration of childcare experience, variations in entry age and comparisons of children with and without childcare experience. We compared quantity effects representing the absolute effect of ECEC, meaning effect sizes for comparison of children with and without ECEC experience, to effect sizes representing relative effects of ECEC, meaning variations in the amount of childcare experience of children within ECEC (i.e. variations in intensity, duration and entry age). Findings for the absolute effect included comparisons of children without ECEC experience to those of varying amount of ECEC experience for studies where no separate effect was reported.

4.) *Non- vs peer-reviewed reference:* We grouped effect sizes as peer-reviewed or non-peer-reviewed and included the categorical moderator in our analysis. It is more likely for stronger effects to get published, which would lead to a stronger overall effect for peer-reviewed effect sizes. This moderation effect is very often discussed in meta-analysis under the term “publication bias” (Card, 2012; Cooper, 2009). A significant positive moderation effect is only one indicator of this bias, and meta-analysis provides further analytic tools to investigate if such a bias is present (see section about publication bias). It is also possible, though less frequently discussed, that peer-reviewed effect sizes prove to be weaker on average. If findings passed a peer-reviewed process, the analysis used should be of a higher quality. It is very unlikely that a finding uncontrolled for certain variables gets published. Overall effects based on findings from a peer-reviewed reference in this case could lead to weaker effects.

Table 1

Overview of all moderator variables and grouping of categorical moderators

Moderator	ECEC effect	coding	Categories	Explanation
Outcome domain	all ECEC effects	0	literacy	relationships to measures of (pre-)reading, (pre-)writing and language skills (e.g. vocabulary, letter recognition, narration skills, sentence comprehension) and grades at a later phase in children's academic careers (i.e. GCSE English)
		1	mathematics	relationships to measures of early (e.g. number identification, counting skills, early number concept) and subsequent mathematical skills (e.g. arithmetic skills) and grades at a later phase in children's academic careers (i.e. GCSE Mathematics)
age at assessment of outcome	all ECEC effects	continuous variable		age in month at assessment of developmental outcome.
phase of academic career at assessment of outcome	all ECEC effects	0	in ECEC	*DE:* BiKS (ages 45 and 68), Early Chances (ages 34, 48, and 61), KIDZ (ages 47, 60, and 73), NEPS (ages 60 and 71), TransKiGs (ages 69 and 75), School-Prepared Child (ages 49 and 64) *FI:* First Steps (ages 74 and 80) *GR:* Attiki (age 57) *NL:* pre-COOL (age 42), PRIMA (age 72), Utrecht Mixed Preschool Groups (ages 60 and 78) *PT:* Contexts and Transition (age 67), Engagement (age 69) *UK – ENG:* EPPSE (age 60), MCS (age 60) *UK – NIR*: EPPNI (age 60) *UK – SCT:* GUS (age 58),
		1	in school	*EU (FI, GR, UK – NIR, IT, PL, ES) and other:* IEA Preprimary Project (age 92) *AT:* ECCE-Austria (age 101) *ES:* ECCE-Spain (age 106) *DE:* BiKS (ages 85, 97, and 110), ECCE-Germany (age 103), School-Prepared Child (ages 78 and 87), TransKiGs (age 87), NEPS (age 85) *FI:* First Steps (ages 86, 92, and 120) *NL:* PRIMA (ages 96 and 120) *PT:* Contexts and Transition (ages 71 and 82) *UK – ENG:* EPPSE (ages 78, 90, 120, 132, 168, and 192) *UK – NIR*: EPPNI (ages 69, 80, and 96)

Table 1 (*continued*)

Moderator	ECEC effect	coding	Categories	Explanation
Measures of ECEC effects	global process quality	0	interaction	relationships to measures of global process quality which focus on interactions (e.g. CLASS, CIS and their subscales)
		1	material	relationships to measures of global process quality which entail ratings of availability and access of material and interaction with material surrounding (i.e. ECERS-R and subscales, ITERS, SCSWIS scales „Quality of care and support“ and „Quality of environment“)
	pre-academic promotion	0	observation	relationships to observational measures of pre-academic promotion, i.e. ECERS-E and subscales, DO-RESI-E, observation of literacy and math activities, ECERS-R subscales referring to frequency of pre-academic promotion (subscales language and reasoning, activities and interactions)
		1	questionnaire	relationships to questionnaire measures indicating frequency of pre-academic promotion, i.e. questionnaires about reading activities, frequency of (developmental stimulating) activities, QI scales “improvements in performance”, “children’s experiences” and “meeting learning needs”.
	structural quality	0	qualification	relationships to staff qualification, i.e. educational level teacher, number of full-time years of education, SCSWIS-scales „Quality of staffing“ and „Quality of management and leadership“
		1	arrangement	relationships to structural arrangements, like group size, child-staff-ratio, and m^2 per child
	quantity	0	relative	relationships to indicators of relative amount of ECEC experience, i.e. variations in intensity, duration and entry-age, of children within ECEC.
		1	absolute	Comparison of developmental outcomes of children with and without ECEC experience (also includes comparisons of children with varying ECEC durations to children without ECEC experience).
peer-reviewed	all ECEC effects	0	not peer-reviewed	Effect sizes extracted from non peer-reviewed references (e.g. technical reports, slides of presentations, brief description of newly computed analysis)
		1	peer-reviewed	Effect sizes extracted from peer-reviewed journal articles

Evaluating possible bias

Although we conducted a thorough review of the literature, the pool of included studies in this meta-analysis may, as in every meta-analysis, not represent all studies conducted in this field of research. Studies investigating longitudinally the relationship of ECEC aspects to developmental outcomes, or particular findings of these studies, could have gone unpublished and maybe not identified for this meta-analysis. Significant results are more likely to get published than non-significant results, potentially leading to an overestimation of overall effects in this meta-analysis (Rosenthal, 1979). Furthermore, it is possible that results in foreign languages were not fairly represented in this meta-analysis. By consulting experts and principle investigators of relevant studies in the field some unpublished results and references in foreign languages could partially be obtained. However, it is important to still estimate the degree of potential bias in various ways.

First of all, we studied if reported findings extracted from peer-reviewed references and other references differed in size (see previous section for details). Then, we compared included effect sizes with the coded information about ECEC aspects measured in each study to find out if relevant evidence entailed in studies was missing in our analysis. This was done by comparing the information about which ECEC aspects the studies measured to the included effect sizes in our meta-analysis. We used the file drawer analysis by Rosenthal (1979) to determine if the mean effect-size calculated was robust against the bias of the missing findings (fail-safe N calculation). The test estimates how many studies with null findings are needed to outreach a significance level. We applied the file drawer analysis only for significant overall effects and used a significance level of p =.05.

The probability of publication bias was also assessed with funnel plots (see Figures 1a-d in the Appendix). A funnel plot is a scatter plot representing the effect sizes from different studies and different measurement points on the horizontal axis against a measure of each study's size or precision. We used the standard error on the vertical axis as recommended by Higgins and Green (2008). With increase of study size, the precision of the estimated effect increases. That is why estimated effects from small studies will be spread more widely at the bottom of the funnel plot and findings from larger studies should be more narrowly aligned around the estimated overall effect. Without bias being present, the graphs should appear like a symmetrical inverted funnel. Apart from visual inspection a further possibility is statistically testing the plot asymmetry. Again following Cochrane's recommendations (Higgins & Green, 2008) we used the rank correlation test for funnel plot asymmetry (Begg 1994; Begg & Mazumdar, 1994) which tests if the effect estimates and the corresponding sampling variances are correlated. A high correlation points to asymmetry of the funnel plot, which may be a result of publication bias.

Results

Descriptive overview about studies and evidence basis for ECEC effects

This subsection gives an overview of the results of the adapted search and selection procedure. All in all, the search resulted in 208 references identified as potentially relevant to investigate impact of ECEC in Europe. Of the collected references 110 included information on literacy and/or mathematics as outcomes and 40 entailed information on the relationships of ECEC aspects to child outcomes.

Often multiple references belonged to the same study, because large-scale longitudinal studies produce separate references for different measurement points, developmental domains, and ECEC aspects, as well as different publication types (e.g. reports, conference slides, or peer-reviewed articles). All in all, the current meta-analysis comprises 226 separate findings.

The final sample of primary studies for the meta-analysis included 22 different studies[5]. The evidence stems from 12 different European countries[6]. Four of the included studies were evaluation studies of ECEC programmes, i.e. the studies KIDZ, TransKiGs, Early Chances, and Utrecht Mixed Preschool Groups. The compendium of studies (Table 5 in the Appendix) shows included and excluded studies as well as additional international studies, which began after 1990 and entail information on aspects of ECEC quantity and quality and developmental outcomes in literacy, mathematics and/or socioemotional outcomes. In sum, the meta-analysis aggregated evidence for ECEC effects on developmental outcomes of 43,527 children from 4,431 institutions/groups.

The overview of the final sample of all included large scale longitudinal studies can be found in Table 1 in the Appendix. Studies differ in various important features. This section and Tables 1 to 4 in the Appendix give an illustrative overview of the differences in design and further study characteristics of included studies. Included studies began from 1992 onwards. The most recent study started in 2012.

Most studies come from Germany (*n* = 6) and the United Kingdom (*n* = 4), while three studies were conducted in Netherlands, two in Portugal, and one in each Finland and Greece. Two studies come from multiple European countries. Most studies conducted first assessments of children around the age of 3, while follow up assessments ranged from several months after the initial assessment, up to at most the age of 16. Generally, the studies were similar in gender distribution. Average proportion of girls in the samples was 49.7 percent across all studies, ranging from 47 to 51 percent. Sample characteristics of each study are summarized in Table 2 in the Appendix.

The studies included in the final sample in general addressed similar questions. However, there was a wide variation between study designs, including differences in social and demographic characteristics of different countries, as well as from adopted measures and instruments.

More than half of the studies reported some sort of measure of disadvantage in the sample (*n* = 12), however, only a subset of studies explored the differential effects of ECEC for disadvantaged children. Most commonly assessed type of disadvantage was economical, with ten studies reporting at least one indicator. The indicators used included parental education (*n* = 8), parental occupation (*n* = 8), and single parent status (*n* = 2). Studies from United Kingdom also included indicators such as receiving social benefits or qualifying for free meals in schools, which, while relevant for the UK setting, are not so relevant or comparable to other countries with different social systems. Seven studies addressed indicators of a migration background. While economic disadvantage indicators were found in studies from all of the included countries, migration background was assessed only in countries with relatively more diverse population: Germany, Netherlands and the UK. The indicators used included language spoken at home or child's own language differing from the official language

[5] The 3 subsamples of the ECCE study in Spain, Germany, Austria were treated as separate studies.
[6] Austria, Germany, Greece, Ireland, Italy, Finland, Netherlands, Poland, Portugal, Spain, UK: Northern Ireland, England, Scotland (Findings for Italy, and Poland are included as findings across the different countries of the IEA Preprimary project, as no country-specific findings were available).

(*n* = 6) and ethnicity other than Caucasian (*n* = 3). These kinds of indicators are not found in studies from other countries, where they might be less relevant due to the more homogenous population, as for example in Finland.

Overall strength of relationships

In a first step, we aggregated results for each of the four ECEC effects to overall effects by using longitudinal multilevel meta-analysis (Maas et al., 2004). As can be seen in Table 2, the aggregation was based on different numbers of studies and effect sizes, which has important implications for the interpretation of aggregated results. Aggregations based on a smaller number of effect sizes and studies should be interpreted with greater caution with regard to certainty of estimation and generalizability of findings (Card, 2012; Cooper, 2009).

Table 2

Parameter estimates for the longitudinal meta-analysis models without predictors for the four ECEC effects

	Sample		Effect size			Heterogeneity			
ECEC aspect	k_{ES}	$k_{Studies}$	*ES*	SE	$CI_{95\%}$	Q_{ES}	df_{ES}	σ^2_{age}	σ^2_{study}
Global quality	73	13	.11**	.04	.04 - .18	2482.15***	72	.02	.00
Promotion	70	11	.10***	.03	.05 - .15	264.55***	69	.01	.00
Structural quality	37	10	.04	.06	-.09 - .16	396.77***	36	.01	.03
Quantity	46	7	.12*	.05	.02 - .23	688.56***	45	.00	.02

*Note. *** p < .001, ** p < .01, * p < .05*

The overall relationships of global process quality (*ES* = .11, *p* <.01, $CI_{95\%}$=.04 –.18), pre-academic promotion (*ES* = .10, *p* < .001, $CI_{95\%}$=.05 –.15) and the quantity of ECEC (*ES* = .12, *p* < .05 $CI_{95\%}$=.02 – .23) to developmental outcomes was positive and significant, but low in overall strength (see Figure 7). No significant relationship was found between the developmental outcomes and the aspects of structural quality (*ES* = .04, *p* = .55 $CI_{95\%}$=-.09 –.16). Transforming our significant correlational measures into the commonly used measure of effect size, Cohen's d (1988), results in effect sizes in the range .20 to .24.

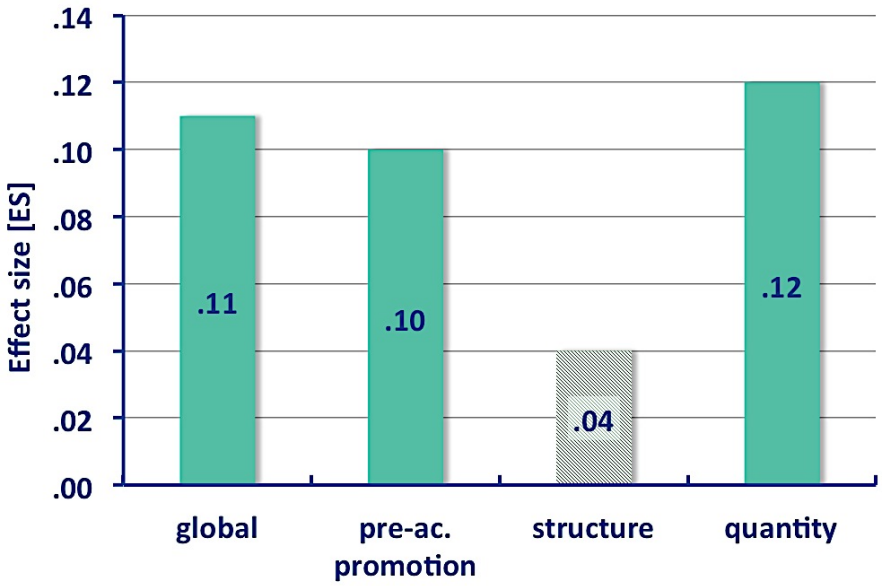

Figure 7. Overall effects for global process quality, pre-academic promotion, structural quality and ECEC quantity indicating their strength of relation to developmental outcomes (non-significant effects are hatched).

Heterogeneity of findings

Table 1 also displays information about the heterogeneity of findings across studies (δ_{study}= .00-.03) and measurement points (δ_{age}= .00-.02). We observed rather small variance components at both levels. We used the weighted least squares extension of Cochran's *Q*-test to tests if the variability in the observed effect sizes was larger than one would expect based on sampling variability alone. The *Q*-test indicated substantial variations in effect sizes for all ECEC effects (global process quality: $Q_{ES}(73) = 2482.15$; pre-academic promotion: $Q_{ES}(69) = 264.55$; structural quality: $Q_{ES}(36) = 396.77$; quantity: $Q_{ES}(45) = 688.56$; p <.001).

Forest plots provide an impression of heterogeneity of the reported effects (see Figures 8 a-d). In a forest plot, each tick mark represents an individual effect size and the line its corresponding confidence interval. The dotted line marks a null effect. Effect sizes at the proximity of the dotted line represent null findings, those on the right side positive and those on the left side negative effects. The diamond shape at the bottom shows the overall effect size. Forest plots for all four ECEC effects show that effect sizes generally varied between and within studies with only partially overlapping confidence intervals.

For global process quality and pre-academic promotion (see Figures 8a and 8b) the majority of effect sizes (and their confidence intervals) lie on the right side of the null effect line (representing positive effects), but some effect sizes lie in the immediate proximity of the null line, and some studies even partly on the left side (representing negative results). For quantity effects (see Figure 8d), the forest plot resembles the previous one, but studies reported fewer negative results. Nonetheless, substantial variation in reported strengths of effects is apparent. The forest plot for structural quality (see Figure 8c) illustrates the enormous amount of heterogeneity of findings with effect sizes scattered on both sides of the null effect line, meaning that studies reported both positive and

negative findings for structural aspects. Whereas for the other ECEC aspects, a tendency towards positive findings was still apparent, this was not the case for structural quality, leading to the previously mentioned non-significant overall effect.

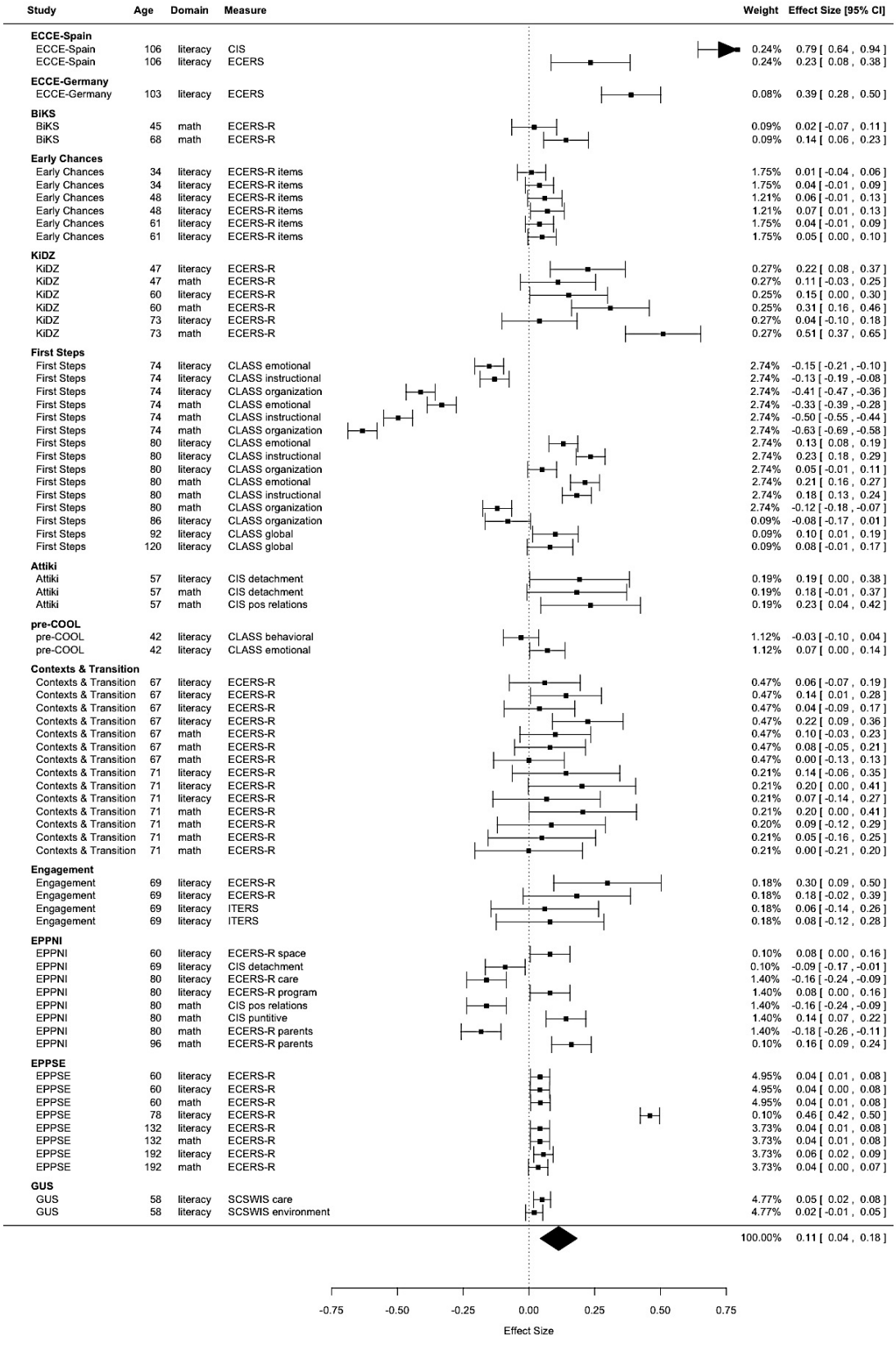

Figure 8a. Forest plot for the relationship of global process quality to developmental outcome. Each tick mark represents an individual effect size and the line its corresponding confidence interval. The diamond shape at the bottom shows the overall effect size (.11**) and the dotted line marks a null effect.

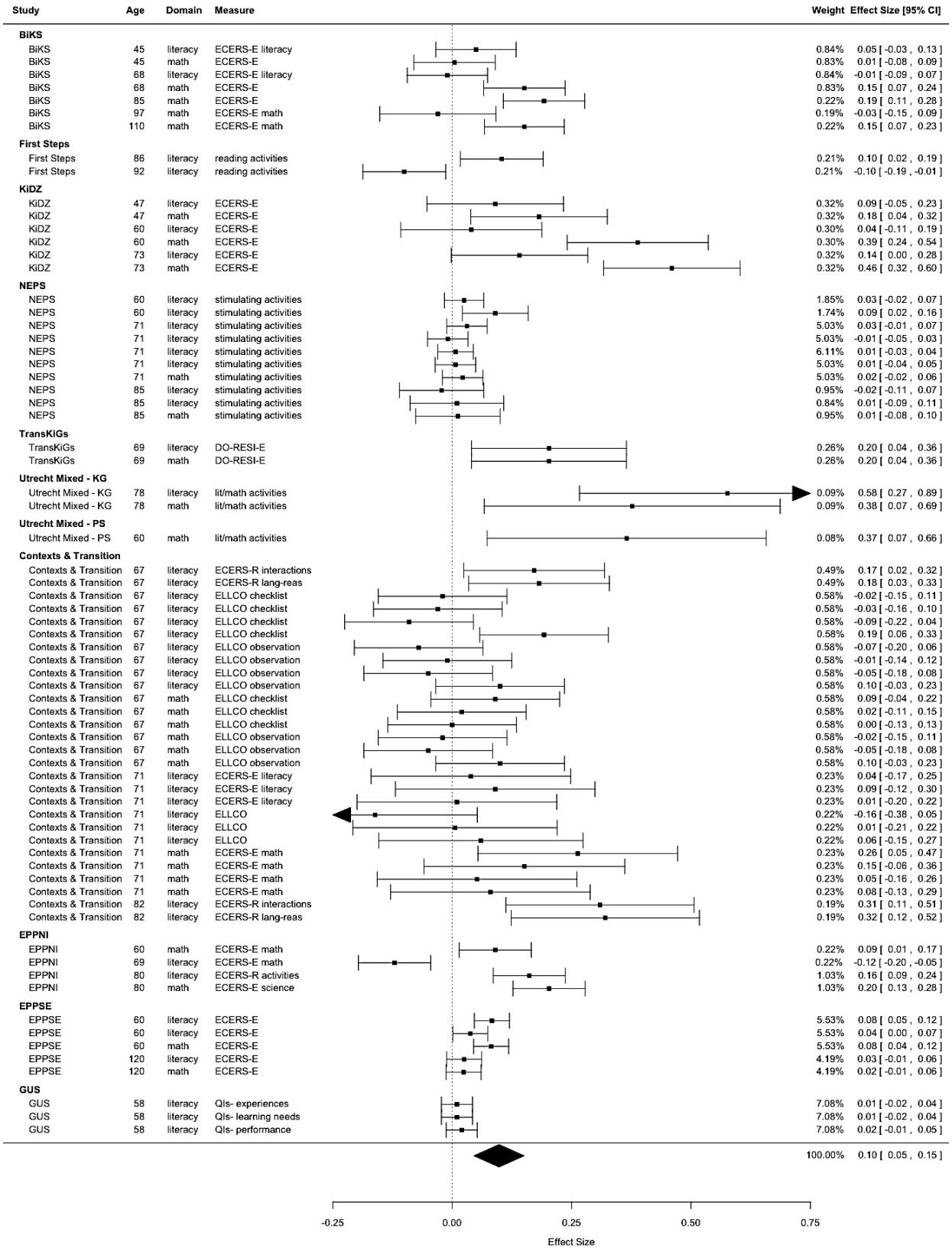

Figure 8b. Forest plot for the relationship of pre-academic promotion to developmental outcome. Each tick mark represents an individual effect size and the line its corresponding confidence interval. The diamond shape at the bottom shows the overall effect size (.10***) and the dotted line marks a null effect.

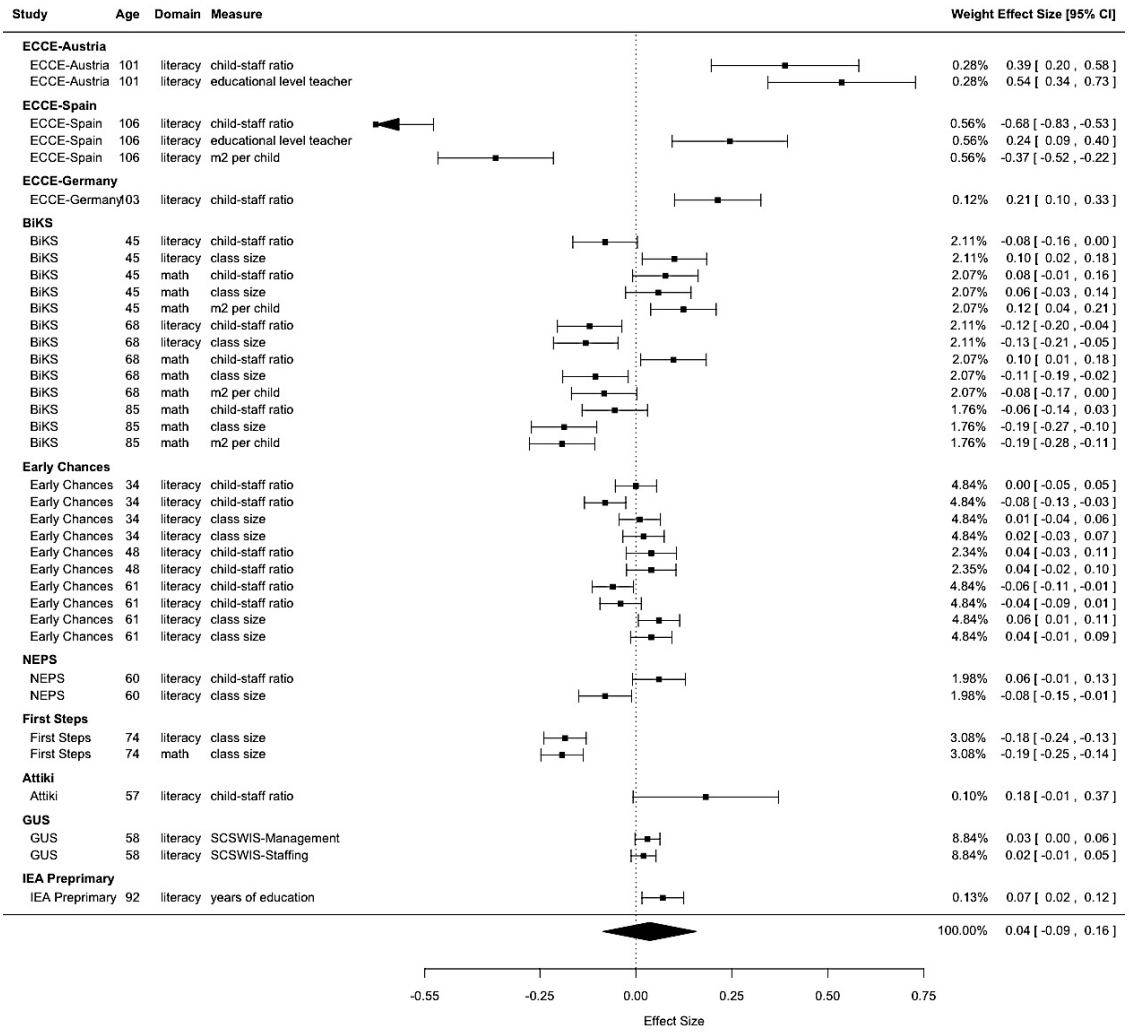

Figure 8c. Forest plot for the relationship of structural quality to developmental outcome. Each tick mark represents an individual effect size and the line its corresponding confidence interval. The diamond shape at the bottom shows the overall effect size (.04) and the dotted line marks a null effect.

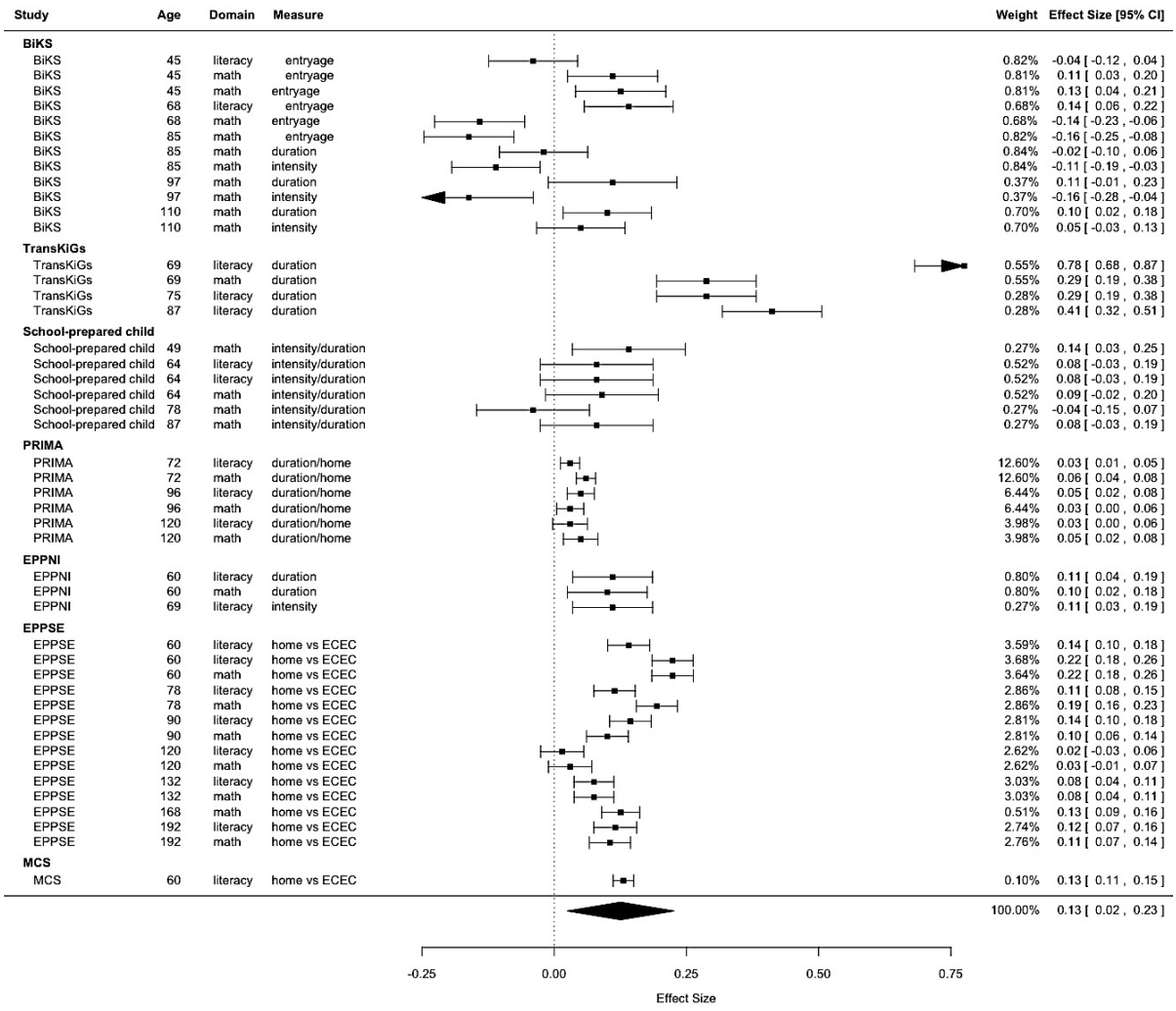

Figure 8d. Forest plot for the relationship of quantity to developmental outcome. Each tick mark represents an individual effect size and the line its corresponding confidence interval. The diamond shape at the bottom shows the overall effect size (.12*) and the dotted line marks a null effect.

Moderators of ECEC effects

As substantial heterogeneity in findings was present for all ECEC effects, we investigated potential sources of variations (Card, 2012; Cooper, 2009). Included studies in our meta-analysis differed in various important characteristics (e.g. study region and its ECEC system, study designs, sample characteristics, see Table 1 to 4 in the Appendix) and it is likely that some of these differences are linked to the strength of the reported effects, i.e. they function as moderators of effects. So besides the descriptive overview of study characteristics, we investigated if differences in reported strength of effects are associated with differences in characteristics of the studies and applied analyses.

We introduced moderators separately, because the samples of effect sizes for each ECEC aspect were very small (k_{ES} = 37 - 73). With few effect sizes introducing moderators simultaneously leads to severe loss of statistical power, as is the case in multiple regressions (Cooper, 2009). If predictors, i.e. moderators in meta-analysis, are highly intercorrelated, introducing them simultaneously may not give valid results (a phenomenon called multicollinearity).

We separately investigated the following continuous and categorical moderators (see Table 1 and method section for details: domain of developmental outcome, age and phase of academic career at outcome assessment, type of measure of ECEC aspect, and if the effect sizes were derived from a peer-reviewed article or a non-peer reviewed reference.

Details of the results of all moderator analyses for categorical moderators are presented in Table 2. For categorical moderators, the table reports the effect sizes separately for groups sharing certain characteristics (e.g. separate mean effects for literacy and mathematics as outcome), confidence intervals for each sub-effect, as well as the numbers of effect sizes and studies included in each group. As recommended by Viechtbauer (2010) we did not compute effect sizes for the subgroups by splitting the effect sizes into two separate data sets, but estimated separate effect sizes for subgroups based on the whole sample of effect sizes. All categorical moderators were dichotomous and the table shows information for both categories of each moderator. Both tables also display the amount of residual variance of effect sizes after including the moderator. The residual variance was significant for all moderator analysis, which means that none of the predictor explained the variation in effect sizes entirely.

Moderators which might explain differential ECEC effects are likely to be correlated with one another, because effect sizes are not independent and the effect sizes from the same studies are generally sharing several characteristics. Tables 6 and 7 in the Appendix show the interrelation of moderators.

Table 4. *Results of meta-regressions*

ECEC sub-effect	Test of moderator	Residual variance		Sample		Effect size of subgroups		
	Q_M[a]	Q_E	df_E	k_{ES}	k_{Study}	*ES*	*SE*	$CI_{95\%}$
Outcome domain								
Global - literacy	36.04***	2346.47***	71	46	12	.13***	.04	.06 - .21
Global - Mathematics				27	7	.07*	.04	.00 - .15
Promotion - Literacy	6.32*	235.83***	68	43	10	.08*	.03	.03 - .14
Promotion - Mathematics				27	9	.11***	.03	.06 - .17
Structural- Literacy	4.70*	377.91***	35	27	10	.03	.06	-.10 - .16
Structural - Mathematics				10	2	.08	.07	-.05 - .21
Quantity - Literacy	0.97	679.04***	44	20	7	.12*	.05	.02 - .22
Quantity - Mathematics				26	6	.13*	.05	.03 - .23
Phase of academic career[b]								
Global - in ECEC	1.78	2332.60***	71	48	11	.09#	.04	.00 - .17
Global - in school				25	6	.17**	.05	.06 - .28
Promotion - in ECEC	1.12	263.75***	68	45	10	.12***	.03	.05 - .18
Promotion - in school				25	6	.07#	.04	.00 - .14
Structural - in ECEC	1.73	395.62***	35	27	6	.07	.07	-.07 - .22
Structural - in school				10	5	-.01	.08	-.17 - .14
Quantity - in ECEC	3.45#	667.47***	44	20	7	.15**	.05	.05 - .25
Quantity - in school				26	6	.10#	.05	.00 - .20
ECEC measures								
Global - interaction-focus	9.35**	2139.65***	71	24	5	.19***	.05	.09 - .29
Global - surrounding				49	10	.09*	.04	.01 - .18
Promotion - observation	4.80*	226.07***	68	55	8	.12***	.03	.07 - .18
Promotion - questionnaire				15	3	.01	.04	-.07 - .10
Structural - arrangements	49.54***	360.33***	35	32	8	-.11	.09	-.28 - .07
Structural - qualification				5	4	.39***	.10	.21 - .61
Quantity - relative	0.23	686.76***	44	25	4	.15*	.07	.00 - .29
Quantity - absolute				21	3	.09	.08	-.07 - .26
Peer vs. not peer reviewed								
Global - not reviewed	2.07	2246.33***	71	45	11	.14**	.04	.06 - .22
Global - reviewed				28	5	.04	.06	-.09 - .16
Promotion - not reviewed	3.34#	254.22***	68	43	8	.13***	.04	.06 - .20
Promotion - reviewed				27	5	.05	.05	-.04 - .14
Structural - not reviewed	1.01	360.84***	35	21	7	.08	.08	-.07 - .23
Structural - reviewed				16	3	-.06	.11	-.28 - .16
Quantity - not reviewed	2.88#	538.17***	44	21	3	.21**	.07	.08 - .35
Quantity - reviewed				25	4	.06	.06	-.06 - .18

Note. ***p < .001. **p < 01. *p < .05. # p < .10. [a]df = 1 for all Q_M.
[b]age as a contionous predictor did not moderate effects.

Do ECEC effects vary by outcome domain?

We studied if the strength of effects differed for the two outcome domains, i.e. if ECEC aspects related more strongly to literacy than to mathematics or vice versa (see Figure 9a). A moderator analysis showed differences in strength of relationships between the two academic domain for all quality aspects ($Q_M(1)$ = 4.70 - 36.04, p < .05). For global process quality, the relationship to literacy was significantly stronger (*ES* = .14, p <.001) than for mathematics (*ES* = .07, p <.05). Pre-academic promotion related more strongly to mathematics (*ES* = .11, p <.001) than to literacy (*ES* = .08, p <.05). Though the strength of effect differed significantly between the two domains, sub-effects of structural quality for both domains were not significant.

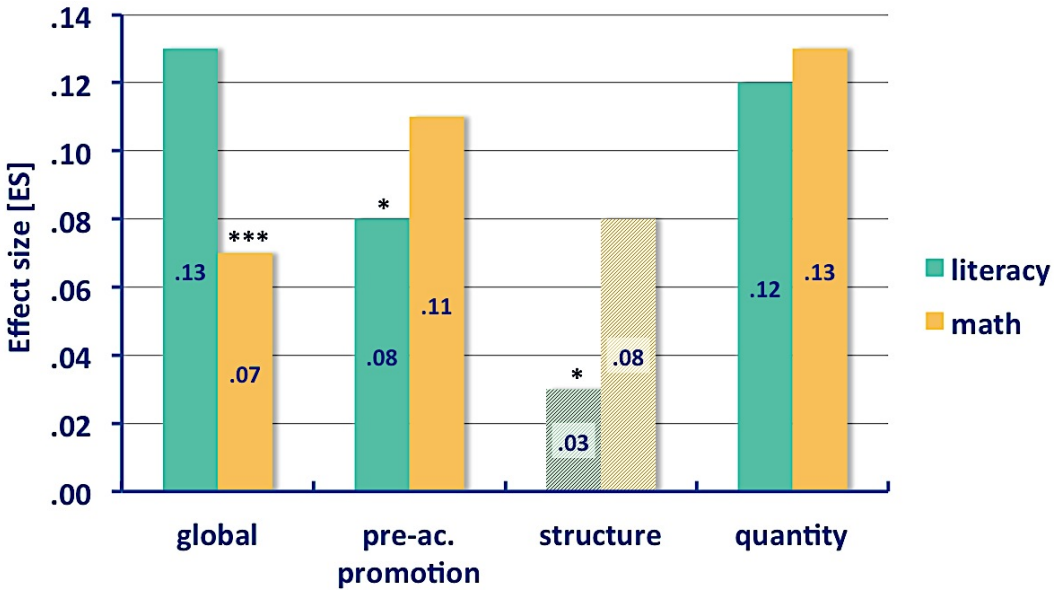

Figure 9a. Results of meta-regression for outcome domain as a moderator (non-significant effects are hatched, significant differences are marked *** p < .001, ** p < .01, * p < .05, # < .10).

Do ECEC effects vary by age and phase of academic career?

To test if the strength of the relationship varied across phases of educational levels of children, for example if relationships to ECEC experience tend to fade as children grew older (Blok et al., 2005; Burger, 2010), we introduced age at outcome assessment as a continuous moderator. Additionally, an extra predictor variable was composed by splitting outcomes into assessπments during and after ECEC phase (see Figure 9b). This was done to investigate if the effects get weaker after children leave ECEC and are exposed to new institutional influences, like primary and secondary schools, rather than gradually declining with age. In general, the analysis found neither age at outcome assessment, nor the phase of academic career to be significant moderators of ECEC effects. Only for quantity effects we observed a marginally significant tendency ($Q_M(1)$ = 3.45, p <.10) for a stronger association for outcomes assessed when children were still in ECEC (*ES* = .15, p<.01), as opposed to assessments of outcomes when children had already left ECEC (*ES* = .10, p<.10).

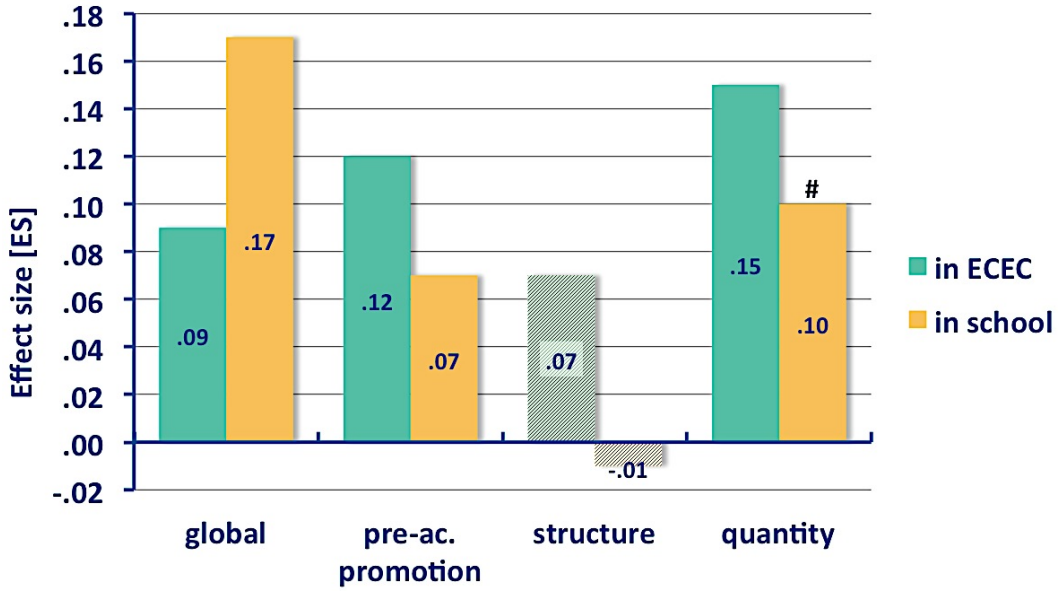

Figure 9b. Results of meta-regression for phase of academic career at outcome assessment as moderator (non-significant effects are hatched, significant differences are marked *** $p < .001$, ** $p < .01$, * $p < .05$, # < .10).

Measures of ECEC aspect as moderators

For the same ECEC aspect measures often differ across and within studies (see method sections for details). Though global process quality measures are generally observational and share similar core facets, they differ in some regards, for example the relevance of the interaction with the material surrounding. We compared effect sizes derived from measures which include ratings of the material surrounding in their overall quality ratings (measure b in Figure. 9c), i.e. ECERS-R and its subscales and the SCSWIS, with effect sizes derived from measures which focus on interactions, i.e. CLASS or CIS and their subscales (measure a in Figure. 9c). We found a difference in size of effects between the two types of measures ($Q_M(1) = 9.35$, $p < .01$). Relationship to developmental outcomes was stronger for interaction-focused measures ($ES = .19$, $p < .001$ versus $ES = .09$, $p < .05$).

Measures of pre-academic promotion mostly captured frequency of activities in academic domains, but not all of the applied measures were observational ones. Four studies used questionnaires (measure b in Figure. 9c) and the strength of the relationship to developmental outcomes in those studies tended to be weaker ($Q_M(1) = 4.80$, $p < .05$) compared to the studies using observational measures (measure a in Figure. 9c). Only the overall relationship assessed by observational measures was significant ($ES = .12$, $p < .001$).

We divided the aspects of structural quality into effect sizes involving the relationship to staff qualification (measure b in Figure. 9c) from those to structural arrangements, like group size, child-staff-ratio, and m^2 per child (measure a in Figure. 9c). Overall developmental outcomes were moderately associated with qualification ($ES = .39$, $p < .001$), but not with aspects of structural arrangement ($ES = -.11$, p = n.s., $Q_M(1) = 49.54$, $p < .001$).

We compared quantity effects representing the absolute effect of ECEC (measure b in Figure. 9c), meaning effect sizes for comparison of children with and without ECEC experience, to effect sizes

representing relative effect of ECEC (measure a in Figure. 9c), meaning variations in intensity, duration and entry age of children within ECEC. No significant moderation effect emerged.

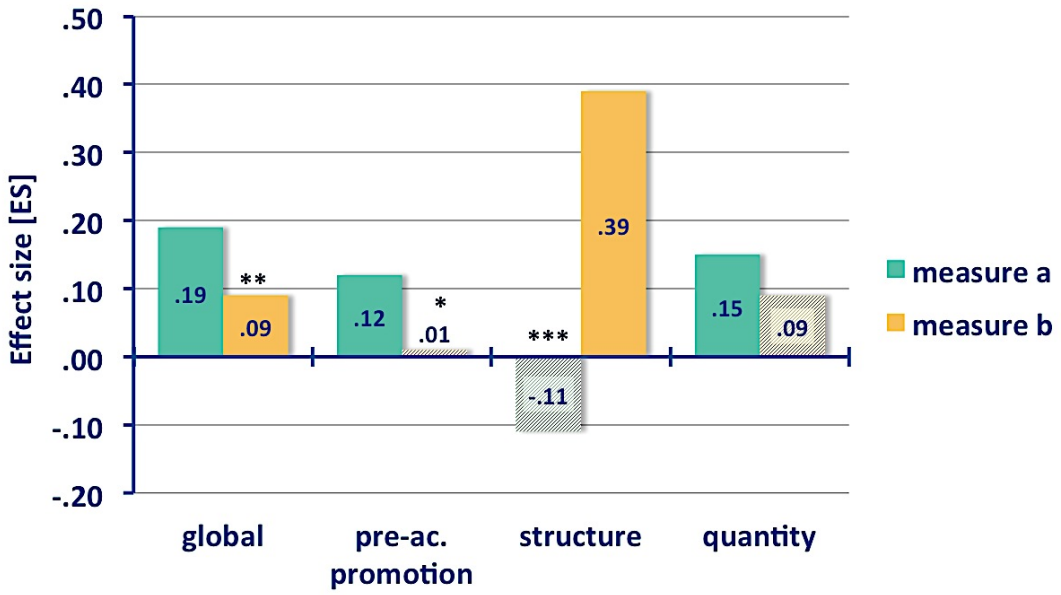

Figure 9c. Results of meta-regression for ECEC measures as moderators (non-significant effects are hatched, significant differences are marked *** $p < .001$, ** $p < .01$, * $p < .05$, # < .10).

Evaluating Possible Bias

Although we conducted a thorough review of the literature, the pool of included studies in this meta-analysis may, as in every meta-analysis, not represent all of the studies conducted in this field of research. Studies investigating longitudinally the relationship of ECEC aspects to developmental outcomes, or particular findings of these studies, could have gone unpublished and maybe not identified for this meta-analysis. Significant results are more likely to get published than non-significant results, potentially leading to an overestimation of overall effects in this meta-analysis (Rosenthal, 1979). Furthermore, it is possible that results in foreign languages were not fairly represented in this meta-analysis. By consulting the experts and the principle investigators of relevant studies in the field, some unpublished results and references in foreign languages could partially be obtained. However, it is important to estimate the degree of bias in various ways (Higgins & Green, 2008; Moher et al., 1999).

First of all, we computed a meta-regression comparing findings from peer-reviewed references to findings from other references (see Figure 10). It is more likely for stronger effects to get published, which would reflect a publication bias. On the other hand, if findings passed a peer-reviewed process, they should be of higher quality, which could lead to more conservative estimations of ECEC effects. Type of reference ermerged as a marginally significant moderator of the effects of pre-academic promotion ($Q_M(1) = 3.34$, $p < .10$) and quantity ($Q_M(1) = 2.88$, $p < .10$) For pre-academic promotion, the overall association to developmental outcomes was only significant for non-peer-reviewed findings (*ES* = .13, $p < .001$ versus *ES* = .05, p = n.s.). The same pattern was observed for quantity effects: reported findings were only significant when extracted from non-peer reviewed references (*ES* = .21, $p < .01$ versus *ES* = .06 = n.s.).

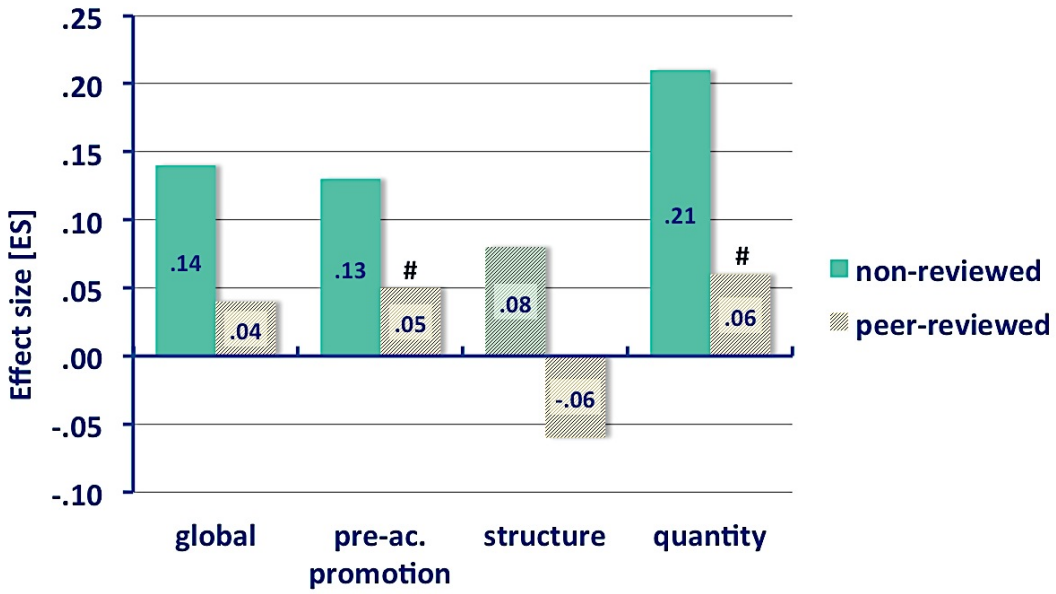

Figure 10. Results of meta-regression for peer-review of reference as moderator (non-significant effects are hatched, significant differences are marked *** $p < .001$, ** $p < .01$, * $p < .05$, # < .10).

We also compared included effect sizes with coded information about ECEC aspects measured in each study to find out if relevant evidence entailed in studies was missing in our analysis. Table 3 in the Appendix shows that, even though most of the studies measured several different aspects of ECEC, the corresponding effect sizes were not always available. This missingness of information was more pronounced for effects of quantity and structural quality. Though most of the studies reported effect sizes for global process quality, we observed a high rate of missing effects for a specific measure, the Caregiver Interaction Scale (CIS; Arnett, 1989). The overall effect for global process quality seems to be biased towards findings from measures like ECERS-R and CLASS. An additional indication of missing information was that we could obviously not include effect sizes for all assessments of child outcomes, especially for very recent assessments (see Table 1 in the Appendix).

Importantly, the numbers of studies varied not only across outcome domains and sample characteristics, but also across assessed ECEC aspects. The majority of studies reported effect sizes for global process quality and promotion in pre-academic domains. Studies reported effect sizes for the other ECEC aspects to a lesser extent, even when assessed (see Table 3 in the Appendix).

We used the file drawer analysis by Rosenthal (1979) to determine if the mean effect-size calculated was robust against the bias of missing findings (fail-safe N calculation). To outreach a significance level of p =.05 over 1,884 null findings are needed for the overall effect of global process quality, over 3,128 null findings for pre-academic promotion, and more than 10,927 for the overall effect of quantity.

The probability of publication bias was also assessed with funnel plots (see Figures 1a-d in the Appendix) plotting the effect sizes from different studies on the horizontal and measurement point against its standard error on the vertical axis as recommended by Higgins and Green (2008; see also Sterne & Egger, 2001). Funnel plots showed slight asymmetry, which indicates bias to be present in our data. Judgment of asymmetry should not be based on visual inspection alone (Sterne et al., 2011), because visual interpretation of funnel plots are known to be misleading (Terrin, Schmid &

Lau, 2005). Results of the rank correlation test for funnel plot asymmetry (Begg, 1994; Begg & Mazumdar, 1994) indicated a significant correlation between effect estimates and sampling variances for pre-academic promotion only (Kendall's τ = .25, p <.01). The test of the other overall effects showed no indication of bias (Kendall's τ = .01 - .13, p =.11 -.91).

Differential effects for disadvantaged children

European evidence from longitudinal studies on differential effects for disadvantage children is sparse and studies address this topic quite differently. Building an overall effect based on few and extremely heterogeneous findings would lead to imprecise estimation of overall effects. A descriptive summary of evidence can provide important insights into differential effects and inspire future research in the field.

Study authors obviously acknowledge the importance of considering indicators of socioeconomic or educational disadvantage as control variable when studying ECEC effects. Slightly different indicators of disadvantage are used across studies and countries, for example foreign language spoken at home (e.g. BiKS study, Anders et al., 2012a, 2012b, language ref), parental occupation (e.g. EPPNI, Melhuish et al., 2002), eligibility to free-school meal (FSM) and multiple risk indicators (composite measure of various indicators (e.g. EPPSE, Sylva, Melhuish, Sammons, Siraj-Blatchford & Taggart, 2008; Sammons et al.,2002). These indicators are shown to be related moderately to strongly to attainments and growth in literacy and math skills (e.g. Anders et al., 2012; Anders et al. 2013; Abreu-Lima, Leal, Cadima & Gamelas, 2013; Kluczniok et al., 2013; Mehuish et al. 2004; Sammons, 2010). Theory and evidence implies that these structural indicators of disadvantage are linked to an impoverished home learning environment where stimulating activities and educational processes in the families are less frequently taking place (Anders et al., 2013; Kleemans, Peeters, Segers, & Verhoeven, 2012; Kluczniok et al., 2013; Niklas & Schneider, 2013; Sammons, 2010).

Though studies frequently control for indicators of disadvantage, only few studies specifically study if ECEC effects differ for disadvantaged children. Summarized the findings imply that disadvantaged children in particular benefit from attending ECEC at a younger age, at a higher intensity or at all, but that they do not benefit more from a higher quality of pedagogical processes than their less disadvantages peers. Most European studies report either that all children benefit equally from higher quality or that disadvantage children do not equally benefit from quality as their peers.

Unfortunately, studies evidence also a phenomenon called social segregation for some ECEC systems meaning that disadvantaged children are more likely to be cared for in groups or centres with higher proportions of disadvantaged children (Becker & Schober, 2015; Lehrl et al., 2015; OECD, 2011). This grouping of disadvantaged children is problematic given the fact that group composition by different indicators of disadvantage was a consistent and comparably strong predictor of child outcomes in various studies (e.g. Anders et al, 2012, 2013; de Haan et al., 2013; Ebert et al., 2013; Montie, Xiang, & Schweinhart, 2006; Pakarinen, Kiuru et al., 2010, 2011; Durda, 2015; Melhuish at al., 2008b). However, some evidence also shows that group composition does not necessarily predetermine process quality (Anders et al., 2015; Resa, 2014). In some countries policy measures tackle this issue by supporting centres with a high percentage of disadvantaged children with financial subsidization or specific structural requirements such as higher qualified staff (Sylva et al., 2015).

Targeted ECEC programmes have been subject to evaluation studies in many countries. These programmes are offered to families and children in need of educational support. Families are motivated to enroll their children timely into ECEC and children and families get (sometimes intense) additional educational support. It may be concluded that some of these ECEC programmes are of high quality, and some of have yielded large and long-lasting beneficial effects on children's development in different areas (see Anders, 2013 for an overview). Multi-facet approaches combining different components, e.g. parent-trainings and home visits, are likely to be more effective than programmes that focus on ECEC as an intervention for disadvantaged children alone (Blok et al., 2005; Nores & Barnett, 2010). Some targeted programmes, like the studies Utrecht Mixed Preschools, Early Chances or Early Education Pilot (Anders et al., 2015; Flöter et al., 2015; de Haan et al., 2013; Smith et al., 2009), try to reach disadvantaged children only by enhanced ECEC experiences: they tackle disadvantage by enhancing pedagogical processes and improving structural arrangements in ECEC centres, or by motivating parents to enroll their children early. Thus, findings of these studies are generalizable to the context of regular provision and provide important hints how to support these children and prevent them from falling behind. Within these programmes children benefit from a greater extent of pre-academic promotion (de Haan et al., 2013) and a higher global process quality (Flöter et al., 2015; Smith et al., 2009). These studies provide important insights into ECEC effects for disadvantaged children, but interpretation is less straight-forward. Although beneficial effects of high quality programmes can be established for groups of disadvantaged children, it remains unclear, if these children benefit more than less disadvantaged peers. One of these studies, the evaluation of the German federal initiative Early Chances, did not find a significant interaction effect between quality and home learning environment, meaning that quality did not compensate for a low home learning environment within this programme.

Following, we give a narrative summary on those European studies that investigated differential effects of ECEC quality drawing on different measures and definitions of educational disadvantage. Findings of Becker (2011) suggest compensatory effects with regard to preschool dose drawing on a UK database. She studied interaction effects with data from the Millennium Cohort Study and found that children from families with low parental education score higher in vocabulary at age 3 when attending preschool compared to children who do not. For children from high educational background preschool attendance did not seem to make a difference. The academic benefit of preschool attendance was still present at age 5 and enrollments at later ages did not show a comparable effect. Though beneficial, early preschool enrollment did not completely compensate for educational disadvantage, but growth between ages 3 and 5 was stronger for disadvantaged children when attending preschool.

Using latent growth curve modeling, the German BiKS study, for example, found besides reduced attainments and growth for non-native language children that only for non-native language children the extent of pre-academic promotion measured by ECERS-E subscale literacy was associated with attainments in vocabulary at age 3, shortly after children entered preschool (Ebert et al., 2013). There was no association to growth from ages 3 to 5 in both samples of native and non-native language children. Other aspects of ECEC experience were not related to vocabulary in the non-native language sample. A smaller group size and a younger age of entry were positively related to initial attainments but only in the sample of native language children. For numeracy skills, Anders et al. (2012) reported significant interaction effects of home learning environment and pre-academic promotion in preschool assessed by ECERS-E total score. However, only children from medium- or

high-quality home learning environment seem to benefit from greater pre-academic promotion in preschool in their growth from age 3 to 5. Benefits from pre-academic promotion in preschool were only maintained at the beginning of primary school when the quality of the home learning environment in middle childhood was high (Lehrl et al., 2015). These effects support a "Matthew effect" rather than a compensatory effect (Walberg & Tsai, 1983)[7].

The Portuguese study Context and Transition (Abreu-Lima et al., 2013) investigated potential moderating effects between mother's educational level in years and preschool quality assessed with ECERS-R and two measure of pre-academic promotion in the literacy domain (ELLCO checklist and observation). The study assessed various outcomes at age 6 in the domains of literacy (concepts about print, vocabulary, and phonological awareness, and letter identification) and math (rote counting, identification of numbers, and basic arithmetic). Only pre-academic promotion was predictive for some outcomes and only for few of the outcomes interaction effects for pre-academic promotion and maternal education emerged (i.e., number identification, rote counting and vocabulary). These interaction effects also suggest a "Matthew effect" (Walberg & Tsai, 1983): children from high educational family backgrounds seem to profit more from a high level of pre-academic promotion in preschool.

Results from hierarchical linear models of the Portuguese Engagement study (Pinto, Pessanha & Aguiar, 2013) also found an interaction between the quality of home learning environment and the quality of global processes in ECEC. Children attending ECEC with high global process quality and high quality of ECEC displayed the strongest growth in basic language skills and literacy attainments from ages 2 to 5. No difference was observed for children of various home learning environments when preschool quality was low.

The Scottish study GUS (Bradshaw et al., 2014) tested the interaction between level of parental education which was the strongest predictor of growth in vocabulary from age 3 to 5 and global process quality measured by the Care Inspectorate subscale care and support. The interaction term was not significant, so that children irrespective of their backgrounds benefit from high quality preschools.

The EPPSE study in England similarly examined the interplay between different forms of disadvantage and preschool attendances and also studied the role of variation in quality of the attended preschool. In early work, the researchers did not find significant interaction effect for ECEC quality and disadvantage, for example families with a low socioeconomic or educational background, for growth in outcomes over the preschool phase (Sammons et al., 2002). At age 10 and 11, however, children from low home learning environments who attended preschool as compared to those not attending preschool achieved better results in English and mathematics, especially when attending a setting where pre-academic promotion was more pronounced (measured by ECERS-E; Sammons et al., 2008; Sylva et al., 2013). Similarly, at age 16, students of low-qualified parents who attended preschools with high pre-academic promotion achieved better grades in GCSE math and English than students without ECEC experience (Sammons et al., 2014).

[7] "Matthew effects" is a concept to describe findings that individuals with advantageous early educational experiences are able to exploit more efficiently new educational experiences.

Summary and Discussion

The results of this meta-analysis confirm that variations in quality and quantity of ECEC have an impact on developmental outcomes of children in two important developmental domains, namely literacy and mathematics. We found overall relationships of global process quality (*ES* = .11), pre-academic promotion (*ES* = .10), and ECEC quantity (*ES* = .12) to developmental outcomes, but we observed no direct overall association of structural quality with academic outcomes. Aggregating evidence on differential effects for disadvantage children was not possible, as European research evidence is still sparse and very heterogeneous. A review of existing evidence suggests that disadvantage children in Europe benefit particularly from an early ECEC enrolment and that they benefit from high quality of ECEC. Some evidence implies that disadvantage children need additional and specific support to catch up with their peers in academic development.

Observed overall effects vary substantially between and within studies, and variations can partly be explained by different moderators: All of the quality effects vary by outcome domain and partly by measure of ECEC aspect. Though all ECEC aspects are associated with outcomes in both domains, global process quality seems to be more strongly related to literacy outcomes, and pre-academic promotion to mathematical outcomes. We did not find indications for a gradual decline of ECEC effects with age in the given age range the included studies focused on, and only for quantity effects our meta-analysis indicated a stronger association with outcomes measured during ECEC period than to outcomes measured at later phase of children's educational career. The available research uses various measures to assess the four ECEC aspects, and the moderator analysis suggests that the choice of measure relates to the strength of observed relationships to child outcomes. Interaction-focused measures tend to be more strongly associated to child outcomes than those including ratings of material surroundings in their overall quality ratings. For pre-academic promotion we found significant overall effects only for observational measures and not for staff questionnaires. Similarly for structural quality, we found only the variations in staff qualification and not variations in structural arrangements to relate to child outcomes. No differences between absolute effects of ECEC versus effects of relative variations in ECEC quantity were apparent across studies. None of the moderators alone explained the variance in effects exhaustively.

Effect sizes published in peer-reviewed journals were on average waker than those found elsewhere for pre-academic promotion and ECEC quantity. Besides differences in reported effects for peer-reviewed as compared to other sources, we found only limited indication of biased results. Comparisons of ECEC aspects measured to reported effect sizes for each study suggested that for some measured aspects effect sizes are missing in our analysis. Visual and statistical testing of possible bias did not suggest extremely biased results and our overall effects appear to be robust against potentially non-included null findings.

Overall effects and comparison to related meta-analytical findings

Results of our meta-analysis support the claim of various others meta-analyses and reviews in the field of ECEC (Anders, 2013; Blok et al., 2005; Burger, 2010): the different experiences children gather within childcare are important and they have developmental impact. We investigated four core aspects of ECEC experience which are generally assumed to have an impact: global process quality, the extent of pre-academic promotion, structural quality, and variations in ECEC quantity. These aspects have been studied in primary studies across and beyond Europe. They have been shown to

be applicable to describe and study the differences of ECEC experiences across countries, various programmes and pedagogical approaches, different types of provisions, and groups of children. Synthesizing knowledge about the effects of these aspects, therefore, allows for drawing comparative conclusions for European ECEC systems. It provides a compressed summary of evidence on the relative impact of different ECEC aspects in Europe. Our overall results imply that children benefit from higher global process quality, more pre-academic promotion, and from a greater amount of ECEC experience.

On first sight, our results appear lower in size than other meta-analytic results for ECEC's impact, but our meta-analysis differs from other meta-analyses on ECEC's impact in several regards, which need to be taken into consideration when interpreting the results:

First of all, meta-analytic results are based on a metric of effect size: though all of the meta-analysis use some type of effect size (*ES*), they are not measured on the same scale. Our overall results ranging from .10 to .12 are based on effect sizes of the correlation family. Transforming our results to Cohen's d values (Cohen, 1988), which is a frequently used measure in other meta-analyses using studies with group contrast designs, results in effect sizes of .20 to .24. Cohen's *d* value for the effects reported elsewhere range from .23 (unweighted; Camilli et al., 2010) to .31 (Nores & Barnett, 2010) for impacts on developmental outcomes in the cognitive domain, including academic outcomes. Hence, sizes of the effects reported here are actually only slightly lower than elsewhere. Our aggregated effects imply that children benefit in their academic development from higher quantity and quality of ECEC in Europe.

Secondly, previous meta-analyses used effects derived from group contrast designs. For sizing the impact of ECEC or early interventions, they compared developmental outcomes of children with some type of ECEC experience to children without ECEC experience or/and compare the developmental outcomes of children in different programmes. Therefore, most previous meta-analyses entailed a large proportion of findings for specific educational programmes sometimes enriching regular ECEC with further early intervention components. These findings additionally reflect partially the absolute effect of ECEC, i.e. the developmental gap between children in ECEC to those without ECEC experience. Our meta-analysis considered mainly variations of children *within regular provision*, and it is even more astonishing, that these variations alone have comparable developmental impact on children.

Thirdly, effects of variations in quantity and quality of ECEC have been studied often indirectly by evaluating programmes in these regards. Unfortunately, these aggregated findings do not allow estimating the developmental gain that is to be expected by increasing the quality and/or quantity in ECEC provision. We, on the other hand, wanted to investigate directly the developmental impact of ECEC quantity and quality. Thereby, we could include the studies on variations in regular provision, without a home sample or other form of group contrast design. This allowed our meta-analysis to follow the shift in primary research from the question of *does ECEC have an impact* to *which aspects of ECEC have developmental impact*.

Last but not least, most previous meta-analyses aggregated findings for cognitive outcomes of extremely differing nature, including IQ-Test or tests of short- or long-term memory. Though, studying ECEC's impact on a variety of outcomes of different domains and nature is important, it is essential to gather specific knowledge about the impact on those outcomes which, according to curricular and guidelines, should explicitly be promoted in ECEC (Burger, 2010; Sylva et al., 2015). This approach is better suited to draw conclusions for the improvement of ECEC practice.

Reported quality effects explain why children within ECEC differ in developmental outcomes, even when they are not exposed to a specific programme. Instead of documenting the effectiveness of specific programmes that are only applied to some children, our meta-analytical approach is a first step to answering the question of how to support the developmental pathways and guarantee a good start for the majority of children attending ECEC in Europe. However, the findings are also applicable for successful programme development. In light of a high mean ECEC coverage rate in Europe finding ways to support children in regular provision and ECEC programmes is important.

At first sight, our results do not seem to confirm Burger's (2010) assumption that high-quality early childhood education and care is more essential to cognitive development than quantitative aspects like age at entry, intensity and duration. However, driven by the conclusion that the nature of ECEC environments is important, many studies have moved their focus on studying quality effects rather than quantity effects. Thus, in the existing studies effects of quality and quantity may not always be completely disentangled. Overall, giving children early access to ECEC, providing sufficient learning opportunities, and improving the quality of global and specific pedagogical processes all seem promising means to improve ECEC's developmental impact in European countries with different ECEC systems, types of provisions, and populations of children.

Moderators of ECEC effect

Though we found significant overall effects across different outcomes, ECEC measures, ages and phases of the educational career we observed substantial variations within and between studies. Studying not only the size, but also the nature of ECEC effects is important. This included investigating the properties of institutional effects like their stability over time, their consistency upon various outcome domains, and differential effects upon children of differing backgrounds (Reynolds et al., 2014). We found domain to be a moderator of all ECEC effects, thus ECEC systems as they stand now seem to have different impact on different domains. Further moderators seem to be related to the strength of some ECEC effects.

ECEC effects are domain-specific

We found ECEC aspects to relate consistently with outcomes in the two academic domains. Nonetheless, all of the quality effects vary significantly by outcome domain. The stronger association of global process quality to literacy outcomes reflects that principles of language learning and promotion are inherent part of various aspects of global quality (Burger, 2014; Grimm & Weinert, 2002; Hoff-Ginsberg, 2000). Measures of global process quality naturally entail ratings of the quality of social and language-based interactions. Whereas global process quality is entwined with beneficial learning opportunities and stimulation of language skills, childcare institutions which score high in global process quality provide rich learning opportunities for language, but not necessarily to an equal extent for early mathematics.

There is a shared understanding that children need to acquire domain-specific knowledge and skills before school enrolment to be best prepared for their later academic development (Roßbach & Weinert, 2008). Early literacy skills, in particular, are seen as essential prerequisites for successful learning in general, because knowledge in all domains is shared orally in social contexts and provided in texts and it also plays a major role for successful integration in society. To acquire early foundations in mathematics, and skills and knowledge in various domains (e.g. early science, and also

pre-reading, pre-writing) specific stimulation and promotion is needed. Domain-specific learning needs to be supported by domain-specific promotion. Children's literacy skills are particularly related to family background and research suggests that parents tend to focus on literacy with regard to stimulating and creating learning opportunities at home (Blevins-Knabe, Berghout, Musun-Miller, Eddy, & Jones, 2002; Kluczniok et al., 2013; LeFevre et al., 2009). This leaves room for institutional influence to make a difference, especially in domains other than literacy. Our results show that mathematical outcomes, as compared to literacy outcomes, relate stronger to the extent of pre-academic promotion.

Despite its potential, research suggests that ECEC systems generally exploit their learning potentials insufficiently. With regard to average scores, many countries do not exceed moderate quality. This is especially true for pre-academic promotion in ECEC, and in particular for domains other than literacy (Kuger & Kluczniok, 2008; Slot et al., 2015a; see also Table 4 in the Appendix). Though early mathematics is implemented as one field of early education in most of the national or federal ECEC curricula and guidelines (Sylva et al., 2015), promotion of early mathematics seldom occurs, especially compared to the promotion of language. Nowadays, there are various ideas for early mathematics education that have been successfully implemented, for example from evaluated programmes, curricular approaches and concepts of teacher trainings (e.g. Dobbs, Doctoroff, & Fisher, 2003; Ginsburg, Lee, & Boyd, 2008; Sarama & Clements, 2004; Sommerlatte, Steinweg, & Gasteiger, 2008; Steinweg, 2007; The Early Math Collaborative - Erikson Institute, 2013), which yield concrete strategies for fostering math development in regular ECEC provision (Burger, 2014).

The consistent differences across quality effects between the two domains reveal a domain-specific accentuation of quality effects. Previous meta-analyses generally subsumed developmental outcomes of extremely heterogeneous nature under one aggregated finding for outcomes in the cognitive domain, e.g. IQ measures, measures of long-term and short-term memory, and other measures of cognitive functioning, as well as outcomes in literacy and mathematics (Blok et al., 2005; Camilli et al., 2010; Nores & Barnett, 2010).

No gradual decline of effects

We did not find a gradual decline of effects with growing age as reported elsewhere (Blok et al. 2005, Burger, 2010; but see Nores & Barnett, 2010) and only for quantity effects strength of association seem to decline after children left ECEC. Generally, this speaks for the persistency of effects across different ages and phases of the academic career. But it must be noted that our meta-analysis covers only European longitudinal studies, and most of them cover – compared to some of the large US-studies – only a restricted age span of children's development. In general, Europe compared to the USA has a comparably shorter history of extensive ECEC research, in particular those involving longitudinal projects. As a consequence our meta-analysis covers a shorter follow-up phase than those meta-analyses, which reported fade-outs of effects.

The strength of ECEC effects depends on the quality dimension and its measure

The strength of the relationship between ECEC aspects and the outcome does not only depend on assessed the outcome, but also on the applied measure to assess the respective ECEC aspect. Interaction-focused measures tend to be more strongly associated to child outcomes than those additionally evaluating the quality of material surroundings. Ratings of the surrounding are important

for indicating that ECEC institutions are safe and well-equipped, providing necessary pre-conditions for children's well-being and rich learning opportunities (Harms et al., 1998). Environmental and material characteristics seem to reflect the structure of learning opportunities (Ditton, 2000; Kluczniok & Roßbach, 2014; Scheerens, & Creemers, 1989), but interactions are the heart of pedagogical processes. Thus, what influences children's pre-academic development the most, is if and how learning opportunities are used. This may explain the stronger association to measures that are interaction-focused.

We found significant overall effects for pre-academic promotion only if assessed by observational measures, and not for findings, which assessed the extent of promotion via staff questionnaires. This finding mirrors the frequent claim of the scientific community that pedagogical processes in ECEC should be measured by observations only (Sylva et al., 2006; Mashburn et al., 2008). Observational approaches as compared to questionnaires are more time- and cost extensive. If assessments of aspects of institutional processes for a huge number of institutions is necessary, for example form part of a country's quality monitoring and assurance systems ECEC (OECD, 2015; Polacek et al., 2011) or as part of a large-scale study like the NEPS or the SOEP studies (Camehl et al., 2012; Durda, 2015) reliable and valid quality questionnaires would be of extreme value for ECEC policy and science. Thus, the development of reliable and valid instruments can be underlined as a relevant desideratum for research and practice. Some recent studies point to aspects of ECEC experience assessed by questionnaires (Anders & Ballaschk, 2014), which may have relevance for the quality of pedagogical interactions (e.g. team development, Anders et al., 2015) and possibly relate to child outcomes (Durda, 2015).

The theoretical framework assumes that aspects of structural quality are prerequisites and predictors of process quality and indirectly linked to children's development (Kluczniok & Roßbach, 2014). Our meta-analysis confirms this view. Only indicators of staff qualifications but not other structural indicators relate on average to child outcomes. This finding confirms other research on the developmental impact of ECEC teacher education (Kelley & Camilli, 2007; Jensen et al., 2015). The qualification of staff is seen as central for teacher practice and the ability to provide enriching teacher-child-interactions. Other structural indicators are associated with process quality, but a well-structured environment does not ensure interactions of high quality (Slot et al., 2015a). Evidencing effects of structural aspects may be difficult in general, because due to regulations we often observe a lack of variance within countries and study regions (European Commission/EACEA/Eurydice/ Eurostat, 2014; OECD, 2006; Polacek et al., 2013). Moreover, the interaction of different structural aspects needs to be taken into account. For example, the impact of staff-child ratios cannot be considered independently of other aspects including staff qualification and group size (Slot et al., 2015a; Munton et al., 2002), as they interact in a complex way. And country-specific combinations of structural aspects may also lead to inconclusive findings. If regulations allow bigger group sizes when teachers are better qualified, neither the positive effect of higher qualification, nor potential negative effects of a larger group size might be observable in this country, if not specifically addressed in analysis.

More effort is needed to extract knowledge of accumulated studies and systematically compare the findings for the relationships between structural aspects and process quality and promotion in pre-academic domains to determine which structural aspects serve to create a supportive environment (Ditton, 2000; Scheerens, & Creemers, 1989) for a high quality of interactions (Slot et al., 2015a, 2015b). Adopting multivariate meta-analysis approaches as proposed by Becker (2000) could actually

investigate direct and indirect influences of structural aspects on child development but the information that would be necessary to adopt these approaches is barely reported.

With regard to the effects of preschool dose and intensity, we observed no differences between absolute effect and effects of relative variations in the amount of ECEC experience, meaning that the overall effect of quantity is not based on the outcome differences of children with ECEC and without ECEC experience alone, but is also a question of how much children are exposed to the institutional environment. This evidence is most relevant for European ECEC when considering that the number of children without any ECEC experience has decreased in most countries.

Limitations and future research

Although we picked up some shortcomings of previous research syntheses, and provided new and additional evidence, some limitations of the presented meta-analysis need to be discussed. A common challenge to the validity of meta-analyses is the possibility of bias. We used various search strategies and different sources (e.g. asking experts for relevant, not yet published and/or translated studies and findings, asking investigators for information which was not available in references) and, thus, reduced bias to a great extent. However, we observed slight indications of bias. Nonetheless, results seem to be robust against potentially not included null-findings, as fail-safe N calculations for each overall effect imply.

The First Steps Study (Pakarinen, Kiuru et al., 2010, 2011, Pakarinen, Lerkkanen et al., 2011, 2015) differs from the other studies in that it assesses quality at a relatively high age (when children are on average 6 and a half years old). Finland has one of the lowest ECEC enrolment rates among OECD countries for children under the age of 6, while at the age of 6 over 90% of children attend a free preschool year (Taguma, Litjens & Makowiecki, 2012). However, at this age in other countries children are often already enrolled in primary education. We conducted a sensitivity analyses (Higgins & Green, 2008) to test the robustness of findings to the exclusion of effect sizes from First Steps, which lead to the exclusion of k_{ES} = 15 for global process quality; k_{ES} = 2 for both pre-academic promotion and structural quality. Obtained overall effect sizes were only marginally higher than the results for the total sample of primary studies: for global process quality *ES* = .13, *p* <.001, $CI_{95\%}$=.07 –.20), for pre-academic promotion *ES* = .11, *p* < .001, $CI_{95\%}$=.05 –.17, and for structural quality *ES* = .06, *p* = *n.s.* $CI_{95\%}$=-.06 –.19).

Included evidence for longer-term effects is restricted to the school phase and seldom goes beyond the primary school stage. This is primarily due to excluding all studies began before 1990. Some syntheses cover older studies and, hence, long-term effects for longer periods of time (Anders, 2013; Burger, 2010, Camilli et al. 2010). The ECEC systems and research have tremendously changed over the past decades. Our time criterion still allowed for an estimation of the persistency of ECEC above the ECEC years, but eliminated very old studies, so that that our overall effects reflect the academic benefit of ECEC nowadays. Only few European studies exist that meet all other criteria except for the time criterion (e.g., two studies in Sweden, Andersson, 1992; 1994; Broberg, Hwang, & Chace, 1993, Wessels, Lamb, & Hwang, 1996). It is important to continue and refine research in this field and to continuously produce updated research syntheses.

There is a scepticism regarding the reliability of combined multivariate effect sizes such as coefficients from multiple regressions or structural equation modelling (Cooper, 2009; Becker & Wu, 2007). Recently, very sophisticated meta-analytical approaches evolved to combine regression

coefficients, but meta-analysts need a lot of specific information, which is almost never provided in references and basically the primary data is needed to adopt them (Becker and Wu, 2007). The effect sizes we obtained may suffer from this limitation to some degree, but using estimation of institutional impact which are controlled for important child and family background characteristics is general standard in educational research and meta-analytic approaches and standards taking account of this are needed in educational science (Card, 2012) including the usage of multivariate coefficients (Bowman, 2012; Kim, 2011).

A further limitation that needs to be addressed is the limitation with regard to the range of the outcome measures. The conclusions of the presented meta-analysis are restricted to the academic domain. There is growing need to synthesize research findings on the impact of ECEC on further aspects of child development, such as socio-emotional development, self-regulation or well-being. Especially for socioemotional outcomes using our meta-analytical approach and established coding procedures can yield important insights into the developmental impact of variations in ECEC experience on this domain, for which previous meta-analyses produced less conclusive overall findings (Blok et al., 2005; Camilli et al., 2010).

Aggregating evidence for differential effects for disadvantage children was not possible, as European evidence is still sparse and very heterogeneous. By providing a descriptive picture of differential effects for disadvantage children, this report sought to inspire future research on this important topic, thereby, allowing for a quantitative synthesis of research evidence in the future.

Conclusion and outlook

This meta-analysis yields important implications for future research syntheses, the planning of primary studies and the funding of studies in the field of early childhood education and care. It highlights the potential of secondary analysis and meta-analysis in the area of educational impact research. The differential findings for different ECEC aspects imply that meta-analysis in this field should finally follow the same direction primary analyses have already taken, moving from the question of *if* ECEC has an impact to study *which* aspects of ECEC have an impact. Research and syntheses have shown that pre-schools, especially if of high quality, have the potential to support children in their development. ECEC is almost a universal experience for children in Europe (European Commission/EACEA/Eurydice/Eurostat, 2014), so that the relevant question is not if we should send children to ECEC institutions, but how we should tailor preschool experience so that it unfolds its beneficial effect. The answer to this question is of great relevance to ECEC policy-makers.

When looking at the public debate on ECEC effects it becomes obvious that this debate is often driven by single studies and findings, often conducted in the U.S. The generalizability of U.S. findings for Europe may be doubted, taking into account the different traditions with regard to governance, provision and pedagogy. Furthermore the dominance of single studies in the debate sometimes masks that we find great heterogeneity in the results of different studies. Thus, meta-analyses on the developmental impact of ECEC are of great importance despite their methodological challenges. This meta-analysis provided new and additional knowledge, which can be taken up by research, practice and policy.

Burger (2010) already argued that research has often not attempted to disentangle potentially distinctive effects of diverse aspects of preschool education. Our meta-analysis confirms this view, it generally supports the framework of the structural-procedural framework of preschool quality, but

the findings clearly show, that different aspects of preschool experience are relevant for different areas of child-development. Practice and policy needs to acknowledge this. At the same time the findings make a general case for investing in ECEC quality, especially in those aspects that are directly related in improving the quality of pedagogical interactions. Initial staff qualification seems to be important, but continuous team development will further improve and enrich the childcare experiences for all children. The general beneficial effects of participating in ECEC and the particular beneficial effects of participating in high quality ECEC could be confirmed.

Teacher beliefs, values and orientations are considered as a factor of staff qualification that may be of increasing importance. This dimension has been subject to a limited number of studies, but should be considered more comprehensively in future research. It is also evident, that the families and the learning environments provided at home are the key for effectively improving the chances of children who grow up in disadvantaged families. Preschool education alone, even if of high process quality often does not provide full potential for these children. It seems to be necessary to find effective means to work with the parents as well as the children. This calls for further accentuation of partnership between preschools and parents as quality dimensions of ECEC.

Children from disadvantaged backgrounds benefit from an early enrolment into ECEC, especially if they experience a different language environment at home. They may also benefit especially from environments and pedagogical approaches specifically adapted to their needs. Research evidence on how to best promote disadvantaged children across Europe is still rare. Some studies even report, that disadvantaged children benefit less from enriched preschool environments than their less disadvantaged peers. In the light of rising numbers of immigrants in Europe, policy and practice is in urgent need of answers to this question. Research needs to address this question with study designs that allow recommendations to directly inform practice and policy.

With regard to measurement, our analyses supported the higher validity of observational measures. It may be concluded that quality monitoring and inspection should rely on observational measures and not on questionnaire-type measures only. At the other hand, the findings underline that the existing questionnaire-type instruments do not seem to capture ECEC experience in a way that it relates significantly to child development. Further developments of instruments may lead to a different situation.

References

*Abreu-Lima, I. M. P., Leal, T. B., Cadima, J., & Gamelas, A. M. (2013). Predicting child outcomes from preschool quality in Portugal. *European Journal of Psychology of Education, 28*, 399-420.

Akgündüz, Y. E., Ünver, Ö., & Plantenga, J., & Nicaise, I (2015). *The Socio-economic dimension of early childhood education and care in Europe*. Retrieved from http://ecec-care.org/fileadmin/careproject/Publications/reports/D5_1_The_Socio-economic_Dimension_of_ECEC_in_Europe.pdf.

Ahnert, L., Pinquart, M., & Lamb, M. E. (2006). Security of children's relationships with nonparental care providers: a meta-analysis. *Child Development*, 77(3), 664-679. doi:10.1111/j.1467-8624.2006.00896.x

Anders, Y. (2013). Stichwort: Auswirkungen frühkindlicher institutioneller Betreuung und Bildung. *Zeitschrift für Erziehungswissenschaft*, 16(2), 237-275.

Anders, Y. (2014). *Literature review on pedagogy, literature review for the OECD*. Paris: OECD.

Anders, Y., & Ballaschk, I. (2014). Studie zur Untersuchung der Reliabilität und Validität des Zertifizierungsverfahrens der Stiftung "Haus der kleinen Forscher". In Stiftung Haus der kleinen Forscher (Hrsg.), *Wissenschaftliche Untersuchungen zur Arbeit der Stiftung "Haus der kleinen Forscher"*, Band 6. Schaffhausen: Schubi Lernmedien AG, 35-116.

*Anders, Y., Ballaschk, I., Dietrichkeit, T., Flöter, M., Groeneveld, I., Lee, H.-J.,...Wieduwilt, N. (2015). Evaluation des Bundesprogramms „Schwerpunkt-Kitas Sprache & Integration" – Abschlussbericht. In H.-G. Roßbach, Y. Anders & W. Tietze (Hrsg.). Evaluation des Bundesprogramms „Schwerpunkt-Kitas Sprache & Integration" – Abschlussbericht. Bamberg und Berlin.

*Anders, Y., Große, C., Roßbach, H.-G., Ebert, S., & Weinert, S. (2013). Preschool and primary school influences on the development of children's early numeracy skills between the ages of 3 and 7 years in Germany. *School Effectiveness and School Improvement*, 24, 195–211.

*Anders, Y., Roßbach, H. G., Weinert, S., Ebert, S., Kuger, S., Lehrl, S., & von Maurice, J. (2012). Home and preschool learning environments and their relations to the development of early numeracy skills. *Early Childhood Research Quarterly, 27*(2), 231-244. doi:DOI 10.1016/j.ecresq.2011.08.003

Anderson, L. M., Shinn, C., Fullilove, M. T., Scrimshaw, S. C., Fielding, J. E., Normand, J., & Carande-Kulis, V. G. (2003). The effectiveness of early childhood development programs. *American Journal of Preventive Medicine*, 24(3), 32-46. doi:10.1016/s0749-3797(02)00655-4

Andersson, B. E. (1992). Effects of day-care on cognitive and socioemotional competence of thirteen-year-old swedish schoolchildren. *Child Development*, 63(1), 20-36.

Andersson, B. E. (1994). Public policies and early childhood education. *European Early Childhood Education Research Journal*, 2(2), 19-32. doi:10.1080/13502939485207581

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

Arnold, D. H., & Doctoroff, G. L. (2003). The early education of socioeconomically disadvantaged children. *Annual Review of Psychology*, 54(1), 517-545.

Bassok, D., French, D., Fuller, B., & Kagan, S. L. (2008). Do child care centers benefit poor children after school entry? *Journal of Early Childhood Research*, 6, 211–231.

Bäumer, T., Aßmann, C., von Maurice, J. & Blossfeld, H.-P. (2013). Möglichkeiten der Analyse von Kontexteffekten im Rahmen des Nationalen Bildungspanels. In R. Becker & A. Schulze (Hrsg.), *Bildungskontexte* (S. 61–83). Wiesbaden: Springer Fachmedien Wiesbaden.

Becker, B. J. (2000). Multivariate meta-analysis*. Handbook of applied multivariate statistics and mathematical modeling*, 499-525.

*Becker, B. (2011). Social disparities in children's vocabulary in early childhood. Does pre-school education help to close the gap? *British Journal of Sociology, 62*(1), 69-88. doi:10.1111/j.1468-4446.2010.01345.x

Becker, B., & Schober, P. S. (2015). Not just any day-care center? Social and ethnic disparities in the choice of high quality early education institutions. *Manuscript in preparation*.

Becker, B. J., & Wu, M. J. (2007). The synthesis of regression slopes in meta-analysis. *Statistical Science*, 22(3), 414-429. doi:Doi 10.1214/07-Sts243

Begg, C. B. (1994). Publication bias. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 399-409). New York: Russell Sage Publications.

Begg, C. B., & Mazumdar, M. (1994). Operating characteristics of a rank correlation test for publication bias. *Biometrics*, 1088-1101.

Belsky, J., Vandell, D. L., Burchinal, M., Clarke-Stewart, K. A., McCartney, K., Owen, M. T., & NICHD ECCRN. (2007). Are there long-term effects of early child care. *Child Development*, 78, 681–701.

Bernal, R., & Keane, M. (2007). *Childcare choices and children's cognitive achievement: The case of single mothers.* Bogota: Universidad de los Andes.

Blevins-Knabe, B., Berghout, A. A., Musun-Miller, L., Eddy, A., & Jones, R. M. (2000). Family home care providers' and parents' beliefs and practices concerning mathematics with young children. *Early Child Development and Care*, 165, 41–58.

Blok, H., Fukkink, R. G., Gebhardt, E. C., & Leseman, P. P. M. (2005). The relevance of delivery mode and other programme characteristics for the effectiveness of early childhood intervention. *International Journal of Behavioral Development*, 29(1), 35-47. doi:Doi 10.1080/01650250444000515

Broberg, A. G., Hwang, C. P., & Chace, S. V. (1993). *Effects of day care on elementary school performance and adjustment.* Paper presented at the 60th Anniversary Meeting of the Society for Research in ChildDevelopment (SRCD), New Orleans, Lousiana, USA.

Broberg, A. G., Hwang, C. P., Lamb, M. E., & Bookstein, F. L. (1990). Factors related to verbal abilities in Swedish preschoolers. *British Journal of Developmental Psychology, 8,* 335–349.

Broberg, A. G., Wessels, H., Lamb, M. E., & Hwang, C. P. (1997). Effects of day care on the development of cognitive abilities in 8-year-olds: A longitudinal study. *Developmental Psychology, 33,* 62–69.

Broekhuizen, M., Leseman, P., Moser, T. & van Trijp, K. (2015). *Stakeholders study: Values, beliefs and concerns of parents, staff and policy representatives regarding ECEC services in nine European countries: First report on parents*. Retrieved from http://ecec-care.org/fileadmin/careproject/Publications/reports/CARE_WP6_D6_2_European_ECEC__Stakeholder_study_FINAL.pdf.

Borenstein, M., Hedges, L. V., Higgins, J., & Rothstein, H. R. (2009). Fixed-effect versus random-effects models*. Introduction to meta-analysis*, 77-86.

Bornstein, M. H., Hahn, C.-S., Gist, N. F., & Haynes, O. M. (2006). Long-term cumulative effects of childcare on children's mental development and socioemotional adjustment in a non-risk sample: The moderating effects of gender. *Early Child Development and Care, 176*, 129–156.

Bos, W., Lankes, E. M., Schwippert, K., Valtin, R., Voss, A., Badel, I., & Plaßmeier, N. (2003). Lesekompetenzen deutscher Grundschülerinnen und Grundschüler am Ende der vierten Jahrgangsstufe im internationalen Vergleich. *Erste Ergebnisse aus IGLU. Schülerleistungen am Ende der vierten Jahrgangsstufe im internationalen Vergleich*, 69-142.

Bowman, N. A. (2012). Effect sizes and statistical methods for meta-analysis in higher education. *Research in Higher Education*, 53(3), 375-382. doi:10.1007/s11162-011-9232-5

*Bradshaw P., Lewis G., & Hughes T. (2014). *Growing Up in Scotland: Characteristics of preschool provision and their associations with child outcomes*. Edinburgh: Scottish Government.

Burger, K. (2010). How does early childhood care and education affect cognitive development? An international review of the effects of early interventions for children from different social backgrounds. *Early Childhood Research Quarter*ly, 25(2), 140-165. doi:10.1016/j.ecresq.2009.11.001

Burger, K. (2014). Effective early childhood care and education: Successful approaches and didactic strategies for fostering child development. *European Early Childhood Education Research Journal*, 1–18. http://doi.org/10.1080/1350293X.2014.882076

*Cadima, J. (2015). [Analyses of the effects of preschool quality on literacy and numeracy outcomes for the contexts and transition study].

Camehl, G.F., Schober, P.S. & Spieß, K. (2012). A SOEP-related study: Early childhood education and care quality in the socio-economic panel (K2ID-SOEP). *SOEP Wave Report 2014*. Berlin: DIW Berlin.

Camilli, G., Vargas, S., Ryan, S., & Barnett, W. S. (2010). Meta-analysis of the effects of early education interventions on cognitive and social eevelopment. *Teachers College Record*, 112(3), 579-620.

Card, N. A. (2012). *Applied meta-Analysis for social science research*. New York: Guilford Publications.

Chambers, B., Cheung, A., Slavin, R. E., Smith, D. and Laurenzano, M. 2010. *Effective early childhood education programmes: A best-evidence synthesis*, Reading, UK: CfBT Education Trust.

Cooper, H. (2009). *Research synthesis and meta-analysis: A step-by-step approach* (Vol. 2). Sage Publications.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale: Lawrence Erlbaum Associates.

Dalli, C., White, E. J., Rockel, J., & Duhn, I. (2011). *Quality early childhood education for under-two-year-olds: what should it look like?: A literature review*. Wellington, New Zealand: Ministry of Education.

*de Haan, A., Elbers, E., Hoofs, H., & Leseman, P. (2013). Targeted versus mixed preschools and kindergartens: effects of class composition and teacher-managed activities on disadvantaged children's emergent academic skills. *School Effectiveness and School Improvement*, 24(2), 177-194. doi:10.1080/09243453.2012.749792

*de Haan, A., Leseman, P. & Elbers, E. (2011). *Pilot gemengde groepen 2007-2010*. Onderzoeksrapportage oktober 2011. Utrecht: Universiteit Utrecht, Orthopedagogiek.

Ditton, H. (2000). Qualitätskontrolle und Qualitätssicherung in Schule und Unterricht. Ein Überblick zum Stand der empirischen Forschung. *Zeitschrift für Pädagogik* (41 Beiheft), 73–92.

Dobbs, J., Doctoroff, G. L., & Fisher, P. H. (2003). The" math is everywhere" Preschool mathematics curriculum. *Teaching Children Mathematics*, 10(1), 20–22.

*Driessen, G. W. J. M. (2004). A large-scale longitudinal study of the utilization and effects of early childhood education and care in the Netherlands. *Early Child Development and Care*, 174(7-8), 667-689. doi:10.1080/0300443042000187158

*Durda, T. (2015). *Eltern als Bildungspartner: Effekte von Elternarbeit in Kindertageseinrichtungen auf den sprachlichen Entwicklungsstand von Kindern (Master).* Bamberg: Otto-Friedrich-Universität.

*Ebert, S., Lockl, K., Weinert, S., Anders, Y., Kluczniok, K., & Roßbach, H. G. (2013). Internal and external influences on vocabulary development in preschool children. *School Effectiveness and School Improvement*, 24(2), 138-154. doi:Doi 10.1080/09243453.2012.749791

European Commission/EACEA/Eurydice/Eurostat. (2014). *Key data on early childhood education and care in Europe. 2014 Edition. Eurydice and Eurostat Report*. Luxembourg: Publications Office of the European Union.

Eurydice. (2009). *Early childhood education and care in Europe: Tackling social and cultural inequalities*. Brussels: Education, Audiovisual and Culture Executive Agency.

*Flöter, M., Weigel, S., & Schmerse, D. (2015, 12.03.2015.). Effekte in der sprachlichen Bildung von Kindern durch den Besuch von Kindertageseinrichtungen. Paper presented at the 3. GEBF-Tagung 2015, Bochum.

*Fried, L., Hoeft, M., Isele, P., Stude J. & Wexeler, W. (2012). *Schlussbericht zur Wissenschaftlichen Flankierung des Verbundprojekts „TransKiGs – Stärkung der Bildungs- und Erziehungsqualität in Kindertageseinrichtungen und Grundschule – Gestaltung des Übergangs"*. Dortmund: TU Dortmund.

Fukkink, R. G., & Lont, A. (2007). Does training matter? A meta-analysis and review of caregiver training studies. *Early Childhood Research Quarterly*, 22(3), 294–311. http://doi.org/10.1016/j.ecresq.2007.04.005

Garces, E., Thomas, D., & Currie, J. (2000*). Longer term effects of Head Start* (No. w8054). National Bureau of Economic Research.

Gorey, K. M. (2001). Early childhood education: A meta-analytic affirmation of the short- and long-term benefits of educational opportunity. *School Psychology Quarterly*, 16(1), 9-30. doi:DOI 10.1521/scpq.16.1.9.19163

Gormley, Jr., W., Gayer, T., Phillips, D., & Dawson, B. (2005). The effects of universal pre-k on cognitive development. *Developmental Psychology*, 41, 872–884.

Gilliam, W. S., & Zigler, E. F. (2001). A critical meta-analysis of all evaluations of state-funded preschool from 1977 to 1998: Implications for policy, service delivery and program evaluation. *Early Childhood Research Quarterly*, 15(4), 441-473.

Ginsburg, H. P., Lee, J. S., & Boyd, J. S. (2008). Mathematics education for young children: What it is and how to promote it. *Social Policy Report*, 22, 3–22.

Gregg, P., Washbrook, E., Propper, C., & Burgess, S. (2005). The effects of a mother's return to work decision on child development in the UK. *The Economic Journal*, 115(501), 48–80.

Grimm, H. & Weinert, S. (2002) Sprachentwicklung. In R. Oerter & L. Montada (Eds.), *Entwiclungspsychologie, Ein Lehrbuch* (5. Eds., pp. 517-550). Weinheim: Psychologie Verlags Union.

Hansen, K., & Hawkes, D. (2009). Early childcare and child development. *Journal of Social Policy*, 38, 211–239.

Harms, T., Clifford, M., & Cryer, D. (1998). *Early Childhood Rating Scale, Revised Edition (ECERS-R)*. Vermont: Teachers College Press

Harms, T., Cryer, D. & Clifford, R. M. (1990). *Infant/Toddler Environment Rating Scale*. New York, NY: Teachers College Press.

Hayes, C. D., Palmer, J. L., & Zaslow, M. J. (1990). Who Cares for America's Children? Child Care Policy for the 1990s. Panel on Child Care Policy, Committee on Child Development Research and Public Policy, Commission on Behavioral and Social Sciences and Education, National Research Council.

Hedges, L., & Olkin, I. (1985). Statistical models for meta-analysis. *New York: Academic Press.*

Hedges, L., & Pigott, TD (2001). *The power of statistical tests in meta-analysis. Psychological Methods,* 6, 203-217.

Heckman, J. J. (2006). Skill formation and the economics of investing in disadvantaged children. *Science*, 312(5782), 1900-1902.

Higgins, J. P. T., & Green, S. (Eds.) (2008). *Cochrane Handbook for Systematic Reviews of Interventions*. Chichester: Wiley.

Hoff-Ginsberg, E . (2000). Soziale Umwelt und Sprachlernen. In H. Grimm (Eds.), *Sprachentwicklung* (pp. 463-494). Enzyklopädie der Psychologie, Bd. 3. Göttingen: Hogrefe.

Jensen, B., Iannone, R. L., Mantovani, S., Bove, C. Karwowska Struczyk, M., & Wyslowska, O (2015). *Comparative review of professional development approaches*. Retrieved from http://ecec-care.org/

fileadmin/careproject/Publications/reports/report_-_Comparative_review_of_professional_development_approaches.pdf

Kane, T. J., & Staiger, D. O. (2012). *Gathering feedback for teaching: Combining high-quality observations with student surveys and achievement gains*. Retrieved from http://www.metproject.org/downloads/MET_Gathering_Feedback_Research_Paper.pdf

Kelchen, R., Magnuson, K., Duncan, G., Schindler, H., Shager, H., & Yoshikawa, H. (2011). *Do the Effects of early childhood programs on academic outcomes vary by gender? A meta-analysis. Retrieved from http://files.eric.ed.gov/fulltext/ED517846.pdf*

Kelley, P., & Camilli, G. (2007). The impact of teacher education on outcomes in center-based early childhood education programs: A meta-analysis. *New Brunswick, NJ: National Institute for Early* Education *Research.* Retrieved from http://nieer.org/resources/research/TeacherEd.pdf.

Keys, T. D., Farkas, G., Burchinal, M. R., Duncan, G. J., Vandell, D. L., Li, W., & Ruzek, E. A. (2012). Preschool center quality and socioemotional readiness for school: Variation by demographic and child characteristics. *Society for Research on Educational Effectiveness*.

Kim, R. S. (2011*). Standardized regression coefficients as indices of effect sizes in meta-analysis*. (Doctor of Philosophy), Florida State University.

Kleemans, T., Peeters, M., Segers, E., & Verhoeven, L. (2012). Child and home predictors of early numeracy skills in kindergarten. *Early Childhood Research Quarterly*, 27(3), 471–477. http://doi.org/10.1016/j.ecresq.2011.12.004

Kluczniok, K., Lehrl, S., Kuger, S., & Roßbach, H.-G. (2013). Quality of the home learning environment during preschool age – Domains and contextual conditions. *European Early Childhood Education Research Journal*, 21(3), 420–438. http://doi.org/10.1080/1350293X.2013.814356

Kluczniok, K., & Roßbach, H.-G. (2014). Conceptions of educational quality for kindergartens. *Zeitschrift Für Erziehungswissenschaft*, 17(S6), 145–158. http://doi.org/10.1007/s11618-014-0578-2

Knudsen, E. I., Heckman, J. J., Cameron, J. L., & Shonkoff, J. P. (2006). Economic, neurobiological, and behavioral perspectives on building America's future workforce. *Proceedings of the National Academy of Sciences*, 103(27), 10155-10162.

Kuger, S., Pflieger, K., & Roßbach, H.-G. (2006a). *Einschätzskalen der Zielkindbeobachtung*. Bamberg: Otto-Friedrich-Universität.

Kuger, S., Pflieger, K., & Roßbach, H.-G. (2006b). *Handbuch zur Zielkindbeobachtung*. Bamberg: Otto-Friedrich-Universität.

Kuger, S., & Kluczniok, K. (2008). Prozessqualität im Kindergarten—Konzept, Umsetzung und Befunde. *Zeitschrift Für Erziehungswissenschaft*, 159–178.

Kuger, S., Sechtig, J. & Anders, Y. (2012). Kompensatorische (Sprach-)Förderung: Was lässt sich aus US-amerikanischen Projekten lernen? *Frühe Bildung*, 1 (4), 181-193.

Lamb-Parker, F., Piotrkowski, C. S., Baker, A. J., Kessler-Sklar, S., Clark, B., & Peay, L. (2001). Understanding barriers to parent involvement in Head Start: A research-community partnership. *Early Childhood Research Quarterly*, 16(1), 35-51

*Leal, T., Abreu-Lima, I. M. P., & Cadima, J. (2013, August 2013). Associations between dimensions of quality and children outcomes in preschool and first grade –a Portuguese study. Paper presented at the 15th Biennial Conference of the European Association for Research in Learning and Instruction (EARLI), Munich.

LeFevre, J. A., Skwarchuk, S. L., Smith-Chant, B. L., Fast, L., Kamawar, D., & Bisanz, J. (2009). Home numeracy experiences and children's math performance in the early school years. *Canadian Journal of Behavioural Science/Revue canadienne des sciences du comportement*, 41(2), 55.

*Lehrl, S., Kluczniok, K., & Rossbach, H.-G. (2016). Longer-term associations of preschool education: The predictive role of preschool quality for the development of mathematical skills through elementary school. *Early Childhood Research Quarterly*, 36, 475–488. http://doi.org/10.1016/j.ecresq.2016.01.013

Loeb, S., Fuller, B., Kagan, S. L., & Carrol, B. (2004). A longitudinal analysis of the effects of child care type, quality, and stability on the social and cognitive development of preschool children of low-income single mothers. *Child Development*, 75, 47–65.

Love, J. M., Harrison, L., Sagi-Schwartz, A., Van IJzendoorn, M. H., Ross, C., Ungerer, J. A., ... & Constantine, J. (2003). Child care quality matters: How conclusions may vary with context. *Child development*, 74(4), 1021-1033.

Maas, C. J., Hox, J. J., & Lensvelt-Mulders, G. J. (2004). Longitudinal meta-analysis. *Quality and Quantity*, 38(4), 381-389.

Mashburn, A. J., Pianta, R. C., Hamre, B. K., Downer, J. T., Barbarin, O. A., Bryant, D., ... & Howes, C. (2008). Measures of classroom quality in prekindergarten and children's development of academic, language, and social skills. *Child development*, 79(3), 732-749.

Melhuish, E. C. (2004). *A literature review of the impact of early years provision upon young children, with emphasis given to children from disadvantaged backgrounds: Report to the Comptroller and Auditor General.* London: National Audit Office. Retrieved from http://www.nao.org.uk/publications/0304/early_years_progress.aspx

*Melhuish, E. (2015). [Analyses of the effects of preschool quality on literacy and numeracy outcomes].

*Melhuish, E., Quinn, L., Sylva, K., Sammons, P., Siraj-Blatchford, I., Taggart, B. & Shields, C. (2002) *The Effective Pre-school Provision in Northern Ireland Project, Technical Paper 5: Pre-school Experience and Cognitive Development at the Start of Primary School.* Belfast, N.I.: Stranmillis University Press.

*Melhuish, E., Quinn, L., Hanna, K., Sylva, K., Siraj-Blatchford, I., Sammons, P. & Taggart, B. (2004). *The Effective Pre-school Provision in Northern Ireland Project, Technical Paper 10: Pre-school Experience and Literacy and Numeracy Development at the End of Year 2 of Primary School.* Belfast, N.I.: Stranmillis University Press.

Melhuish, E.C., Sylva, K., Sammons, P., Siraj-Blatchford, I., Taggart, B., Phan, M. & Malin, A. (2008a). Preschool influences on mathematics achievement. *Science, 321,* 1161-1162.

Melhuish, E.C., Sylva, K., Sammons, P., Siraj-Blatchford, I., Taggart, B., & Phan, M. (2008b) Effects of the Home Learning Environment and preschool center experience upon literacy and numeracy development in early primary school. *Journal of Social Issues, 64,* 95-114**.**

Melhuish, E., Quinn, L., Sylva, K., Sammons, P., Siraj-Blatchford, I., & Taggart, B. (2012). Preschool affects longer term literacy and numeracy: results from a general population longitudinal study in Northern Ireland. *School Effectiveness and School Improvement: An International Journal of Research, Policy and Practice*, *24*, 234-250.

Melhuish, E., Ereky-Stevens, Petrogiannis, K., Ariescu, A., Penderi, E., Rentzou, K., … Leseman, P. (2015). *A review of research on the effects of early childhood education and care (ECEC) on children development*. Retrieved from http://ecec-care.org/fileadmin/careproject/Publications/reports/CARE_WP4_D4__1_review_of_effects_of_ecec.pdf.

Mitchell, L., Wylie, C., & Carr, M. (2008). *Outcomes of early childhood education: Literature review*. New Zealand: Council for Educational Research.

Moher, D., Cook, D. J., Eastwood, S., Olkin, I., Rennie, D., Stroup, D. F., & QUOROM Group. (1999). Improving the quality of reports of meta-analyses of randomised controlled trials: the QUOROM statement. *The Lancet*, 354(9193), 1896-1900.

*Montie, J. E., Xiang, Z. P., & Schweinhart, L. J. (2006). Preschool experience in 10 countries: Cognitive and language performance at age 7. *Early Childhood Research Quarterly*, 21(3), 313-331. doi:DOI 10.1016/j.ecresq.2006.07.007

Moser, T., Melhuish, E., Petrogiannis, K., Pastori, G., Slot, P., & Leseman, P. (2014). *Initial framework for evaluating and monitoring ECEC quality and wellbeing.* Working paper published online for eliciting stakeholder contributions. CARE: Curriculum & Quality Analysis and Impact Review of European Early Childhood Education and Care, Utrecht University. Retireved from http://ecec-care.org/resources/publications/

Munton, T., Mooney, A., Moss, P., Petrie, P., Clark, A., & Woolner, J. (2002). *Research on ratios, group size and staff qualifications and training in early years and childcare settings*. London: Department for Education and Skills, Research report RR320

NICHD ECCRN. (2002a). Early child care and children's development prior to school entry: Results from the NICHD Early Child Care Research Network study of early child care. *American Educational Research Journal*, 39, 133–164.

NICHD ECCRN. (2002b). Child-care structure—process—outcome: Direct and indirect effects of child care quality on young children's development. *Psychological Science*, 13(3), 199–206.

NICHD ECCRN. (2002c). Early child care research network: The interaction of child care and family risk in relation to child development at 24 and 36 months. *Applied Developmental Science, 6,* 144–156.

NICHD ECCRN. (2003). Does amount of time spent in child care predict socioemotional adjustment during the transition to kindergarten? *Child Development, 74,* 976–1005.

NICHD ECCRN. (2005). Early child care and children's development in the primary grades. Follow-up results from the NICHD Study of Early Child Care. *American Educational Research Journal*, 42, 537–570.

Niklas, F., & Schneider, W. (2013). Casting the die before the die is cast: the importance of the home numeracy environment for preschool children. European Journal of Psychology of Education*, 29(3), 327-345. doi:10.1007/s10212-013-0201-6

Nores, M., & Barnett, W. S. (2010). Benefits of early childhood interventions across the world: (Under) Investing in the very young. *Economics of Education Review*, 29(2), 271-282. doi:DOI 10.1016/j.econedurev.2009.09.001

Organisation for Economic Co-operation and Development [OECD]. (2004). *Learning for tomorrow's world. First results from PISA 2003*. Paris: OECD.

Organisation for Economic Co-operation and Development [OECD]. (2006). *Starting strong II: early childhood education and care.* Paris: OECD.

OECD. (2011). *Starting strong III: A quality toolbox for early childhood education and care*. Paris: OECD Publishing.

Organisation for Economic Co-operation and Development [OECD]. (2015). *Starting Strong IV: Monitoring quality in early childhood education and care*. Paris: OECD Publishing. Retrieved from OECD Publishing. http://dx.doi.org/10.1787/9789264123564-5-en

*Pakarinen, E., Kiuru, N., Lerkkanen, M. K., Poikkeus, A. M., Siekkinen, M., & Nurmi, J. E. (2010). Classroom organization and teacher stress predict learning motivation in kindergarten children. *European Journal of Psychology of Education*, 25(3), 281-300. doi:DOI 10.1007/s10212-010-0025-6

*Pakarinen, E., Kiuru, N., Lerkkanen, M. K., Poikkeus, A. M., Ahonen, T., & Nurmi, J. E. (2011). Instructional support predicts children's task avoidance in kindergarten. *Early Childhood Research Quaterly*, 26(3), 376-386. http://doi.org/10.1016/j.ecresq.2010.11.003

*Pakarinen, E., Lerkkanen, M. K., Poikkeus, A. M., Siekkinen, M., & Nurmi, J. E. (2011). Kindergarten teachers adjust their teaching practices in accordance with children's academic pre-skills. *Educational Psychology*, 31(1), 37-53. doi:10.1080/01443410.2010.517906

*Pakarinen, E., Lerkkanen, M. K., Silinskas, G., Salminen, J., Poikkeus, A. M., Siekkinen, M., & Nurmi, J. E. (2015, 25-29 August 2015). *Longitudinal associations between kindergarten classroom Interactions and Finnish children's reading skills*. Paper presented at the 16th Biennial EARLI Conference for Research on Learning and Instruction, Limassol, Cyprus.

Palacios, J., & Lera, M.-J. (1991). *Observation of activities in preschool*. University of Seville: Unpublished Document.

Peisner-Feinberg, E. S., Burchinal, M. R., Clifford, R. M., Culkin, M. L., Howes, C., Kagan, S. L., & Yazejian, N. (2001). The relation of preschool child-care quality to children's cognitive and social developmental trajectories through second grade. *Child Development*, 72, 1534–1553.

Pianta, R. C., & Hamre, B. K. (2009). Conceptualization, measurement, and improvement of classroom processes: Standardized observation can leverage capacity. *Educational Researcher,* 38(2), 109-119.

Pianta, R., Howes, C., Burchinal, M., Bryant, D., Clifford, R., Early, D., & Barbarin, O. (2005). Features of pre-kindergarten programs, classrooms, and teachers: Do they predict observed classroom quality and child-teacher interactions?. *Applied Developmental Science*, 9(3), 144-159.

Pianta, R. C., La Paro, K. M., & Hamre, B. K. (2008). *Classroom Assessment Scoring System (CLASS) Manual: K-3*. Paul H. Brookes Publishing Company.

*Pinto, A. I., Pessanha, M., & Aguiar, C. (2013). Effects of home environment and center-based child care quality on children's language, communication, and literacy outcomes. *Early Childhood Research Quarterly*, 28(1), 94-101. doi:DOI 10.1016/j.ecresq.2012.07.001

Polacek, R., McDaid, D., Fernandez, J., Matosevic, T., Penn, H., Korintus, M. & Verheyde, M. (2011). *Study on social services of general interest*. Final report. Brussels, European Commission, Directorate General for Employment, Social Affairs and Inclusion.

Praetorius, A. K., Pauli, C., Reusser, K., Rakoczy, K., & Klieme, E. (2014). One lesson is all you need? Stability of instructional quality across lessons. *Learning and Instruction*, 31, 2-12.

*Quinn, L., Melhuish, E., Hanna, K., Sylva, K., Siraj-Blatchford, I., Sammons, P., . . . Doyle, A. (2006). Effective Pre-school Provision Northern Ireland (EPPNI) - Technical paper no. 12 - pre-school experience and literacy and numeracy development at the end of key stage 1. Retrieved from http://www.bbk.ac.uk/psychology/our-staff/academic/edward-melhuish/documents/EPPNIP4paper12.pdf

*Quinn, L., Melhuish, E., Sylva, K., Sammons, P., Siraj-Blatchford, I., Taggart, B., . . . Sweeney, G. (2003). Effective Pre-school Provision Northern Ireland (EPPNI) - Pre-school experience and cognitive development at the end of year 1. Retrieved from http://anu.stran.ac.uk/eppni/documents/CognitiveDevelopmentandProgressattheEndofYear1_003.pdf

R Core Team (2013). *R: A language and environment for statistical computing* [Computer software]. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org/.

Ramey, C. T., & Ramey, S. L. (1998). Prevention of intellectual disabilities: early interventions to improve cognitive development. *Preventive Medicine*, 27(2), 224-232.

Resa, E. (2014). Zusammenhänge zwischen dem Anteil von Kindern mit Migrationshintergrund auf Gruppen- und Einrichtungsebene und der sprachlichen Anregungsqualität in Kindertagesstätten. *Unveröffentlichte Masterarbeit*. Freie Universität Berlin.

Reynolds, A. J., Mavrogenes, N. A., Bezruczko, N., & Hagemann, M. (1996). Cognitive and family-support mediators of preschool effectiveness: A confirmatory analysis. *Child Development*, 67(3), 1119-1140.

Reynolds, A. J. R., Ou, S.-R., & Topitzes, J. W. (2004). Paths of effects of early childhood intervention on educational attainment and delinquency: A confirmatory analysis of the Chicago Child-Parent Centers. *Child Development*, 75, 1299–1338.

Reynolds, D., Sammons, P., De Fraine, B., Van Damme, J., Townsend, T., Teddlie, C., & Stringfield, S. (2014). Educational effectiveness research (EER): a state-of-the-art review. *School Effectiveness and School Improvement*, 25(2), 197-230.

Ritchie, S., Howes, C., Kraft-Sayre, M., & Weiser, B. (2001). Emerging academic snapshot. *Unpublished measure,* University of California at Los Angeles.

Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin,* 86(3), 638.

*Roßbach, H. G., Sechtig, J., & Freund, U. (2010). *Empirische Evaluation des Modellversuchs" Kindergarten der Zukunft in Bayern-KiDZ": Ergebnisse der Kindergartenphase* (Vol. 7). University of Bamberg Press.

Roßbach, H.-G. & Weinert, S. (Hrsg.). (2008). *Kindliche Kompetenzen im Elementarbereich: Förderbarkeit, Bedeutung und Messung*. Berlin: Bundesministerium für Bildung und Forschung

Roux, S., & Tietze, W. (2007). Effekte und Sicherung von (Bildungs-) Qualität in Kindertageseinrichtungen. *ZSE: Zeitschrift für Soziologie der Erziehung und Sozialisation*, 27(4), 367-384.

Sammons, P. (2010). Do the benefits of pre-school last? Investigating pupil outcomes to the end of Key Stage 2 (aged 11). In K. Sylva, E. Melhuish, P. Sammons, I. Siraj-Blatchford, & B. Taggart (Eds.), *Early childhood matters: Evidence from the effective pre-school and primary education project* (pp. 114–148). Oxford: Routledge.

* Sammons, P., Sylva, K., Melhuish, E., Siraj-Blatchford, I., Taggart, B., & Elliot, K. (2002). *Measuring the impact of pre-school on children's cognitive development over the pre-school period. Technical paper 8a: Cognitive outcomes.* London: DfES/London Institute of Education.

* Sammons, P., Sylva, K., Melhuish, E., Siraj-Blatchford, I., Taggart, B., Elliot, K., & Marsh, A. (2004a). *The Effective Provision of Pre-School Education (EPPE) Project: Report on Age 6 Assessment*. London: DfES/London Institute of Education.

* Sammons, P., Sylva, K., Melhuish, E., Siraj-Blatchford, I., Taggart, B., Elliot, K., & Marsh, A. (2004b). *The Effective Provision of Pre-School Education (EPPE) Project: The Continuing Effects of Pre-school education at Age 7 Years*. London: DfES/London Institute of Education.

* Sammons, P., Sylva, K., Melhuish, E., Siraj-Blatchford, I., Taggart, B., Grabbe, Y., & Barreau, S. (2007). *Effective Pre-School and Primary Education 3-11 Project (EPPE 3-11): Summary Report, Influences on Children's Attainment and Progress in Key Stage 2: Cognitive Outcomes in Year 5*. London: DfES/London Institute of Education.

* Sammons, P., Sylva, K., Melhuish, E., Siraj-Blatchford, I., Taggart, B., & Hunt, S. (2008). *Effective Pre-school and Primary Education 3-11 Project (EPPE 3-11): Influences on children's attainment and progress in Key Stage 2: Cognitive outcomes in Year 6*. London: DfES/London Institute of Education.

* Sammons, P., Sylva, K., Melhuish, E., Siraj-Blatchford, I., Taggart, B., Toth, K., Draghici, D., & Smees, R. (2011). *Effective Pre-School, Primary and Secondary Education Project (EPPSE 3–14): Influences on students' attainment and progress in Key Stage 3: Academic outcomes in English, maths and science in Year 9.* London: DfES/London Institute of Education.

* Sammons, P., Sylva, K., Melhuish, E., Siraj, I., Taggart, B., Toth, K., & Smees, R. (2014). *Influences on students' GCSE attainment and progress at age 16*. London: DfES/London Institute of Education.

Sarama, J., & Clements, D. H. (2004). Building Blocks for early childhood mathematics. *Early Childhood Research Quarterly*, 19(1), 181–189. http://doi.org/10.1016/j.ecresq.2004.01.014

Scheerens, J. & Creemers, B. (1989). Conceptualizing school effectiveness. *International Journal of Educational Research, 13,* 691-706.

Schweinhart, L. J., Montie, J., Xiang, Z., Barnett, W. S., Belfield, C. R., & Nores, M. (2005). *Lifetime effects: the High/Scope Perry Preschool study through age 40.* Ypsilanti, Michigan: High/Scope Press.

*Silinskas, G., Pakarinen, E., Poikkeus, A. M., Lerkkanen, M. K., & Nurmi, J. E. (2015, 25-29 August 2015). Longitudinal effects of kindergarten quality on the development of literacy skills in Grade 1. Paper presented at the 16th Biennial EARLI Conference for Research on Learning and Instruction, Limassol, Cyprus.

Slot, P.L., Lerkkanen, M.-K., & Leseman, P.P.M. (2015a). *The relations between structural quality and process quality in European early childhood education and care provisions: Secondary analyses of large scale studies in five countries*. Retrieved from http://ecec-care.org/fileadmin/careproject/Publications/reports/CARE_WP2_D2__2_Secondary_data_analyses.pdf

Slot, P.L., Leseman, P.P.M., Verhagen, J. & Mulder, H. (2015b). Associations between structural quality aspects and process quality in Dutch early childhood education and care settings. *Early Childhood Research Quarterly*, *33*, 64–76. doi:10.1016/j.ecresq.2015.06.001

*Slot, P. L., Mulder, H., Verhagen, J., & Leseman, P. P. M. (2014). *Domain-general and domain-specific quality characteristics of early childhood education and care predict growth of two-year-olds' vocabulary and attention skills over one year.* Manuscript submitted for publication.

Smidt, W. (2012). *Zielkindbezogene pädagogische Qualität im Kindergarten. Eine empirisch- quantitative Studie*. Münster: Waxmann.

Smith, M., Dickinson, D., Sangeorge, A., & Anastasopoulos, L. (2002). *Early literacy and language classroom observation scale (ELLCO).* Baltimore, MD: Paul Brookes.

Smith, R., Purdon, S., Schneider, V., La Valle, I., Wollny, I., Owen, R., . . . Lloyd, E. (2009). *Early education pilot for two year old children: Evaluation*. Retrieved from http://dera.ioe.ac.uk/10651/1/DCSF-RR134.pdf

Sommerlatte, A., Steinweg, A. S. & Gasteiger, H. (Hrsg): Lux, M., Meiering, G. & Führlich, S. (2008). *Lerndokumentation Mathematik - Anregungsmaterialien*. Berlin: Senatsverwaltung für Bildung, Wissenschaft und Forschung.

Steinweg, A. S. (2007). Mathematisches Lernen. In Stiftung Bildungspakt Bayern (Hrsg.) *Das KIDZ-Handbuch: Grundlagen, Konzepte und Praxisbeispiele aus dem Modellversuch "KIDZ - Kindergarten der Zukunft in Bayern"* (136 – 203). Köln: Wolters Kluwer.

Sterne J. A., & Egger, M. (2001). Funnel plots for detecting bias in meta-analysis: guidelines on choice of axis. *Journal of Clinical Epidemiology*, *54*, 1046-55.

Sterne, J. A., Sutton, A. J., Ioannidis, J., Terrin, N., Jones, D. R., Lau, J., ... & Tetzlaff, J. (2011). Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. *Bmj*, 343.

Sylva, K., Ereky-Stevens, K., & Ariescu, A.-M. (2015). *Overview of European ECEC curricula and curriculum template*. Retrieved from http://ecec-care.org/fileadmin/careproject/Publications/reports/CARE_WP2_D2_1_European_ECEC_Curricula_and_Curriculum_Template.pdf.

Sylva, K., Melhuish, E. C., Sammons, P., Siraj-Blatchford, I., & Taggart, B. (2004). *The Effective Provision of Pre-school Education (EPPE) project: Final report. A longitudinal study funded by the DfES 1997–2004.* London: DfES/London Institute of Education.

Sylva, K., Melhuish, E., Sammons, P., Siraj-Blatchford, I., & Taggart B. (2008). *The Effective Pre-School and Primary Education 3–11 (EPPE 3–11) Project: Final report from the primary phase: Pre-school, school and family influences on children's development during Key Stage 2 (age 7–11).* London: DfES/London Institute of Education.

Sylva, K., Melhuish, E. C., Sammons, P., Siraj-Blatchford, I., & Taggart, B. (2011a). Pre-school quality and educational outcomes at age 11: Low quality has little benefit. *Journal of Early Childhood Research*, 9, 109-124.

Sylva, K., Sammons, P., Chan, L. L. S., Melhuish, E., Siraj-Blatchford, I., & Taggart, B. (2013). The effects of early experiences at home and pre-school on gains in English and mathematics in primary school: a multilevel study in England*. Zeitschrift Für Erziehungswissenschaft, 16*(2), 277–301. http://doi.org/10.1007/s11618-013-0364-6

Sylva, K., Siraj-Blatchford, I., & Taggart, B. (2003). *Assessing quality in the early years: Early childhood environment rating scale: Extension (ECERS-E), Four curricular subscales.* London: Trentham Books.

*Sylva, K., Siraj-Blatchford, I., Taggart, B., Sammons, P., Melhuish, E., Elliot, K., & Totsika, V. (2006). Capturing quality in early childhood through environmental rating scales. *Early Childhood Research Quarterly*, 21(1), 76-92. doi:10.1016/j.ecresq.2006.01.003

Sylva, K., Stein, A., Leach, P., Barnes, J., Malmberg, L.-E., & the FCCC team (2011b) Effects of early child-care on cognition, language, and task-related behaviours at 18 months: An English study. *British Journal of Developmental Psychology*, 29, 18-45.

Taguma, M., Litjens, I., & Makowiecki, K. (2012). *Quality Matters in Early Childhood Education and Care: Finland*. Paris: OECD Publishing.

Terrin, N., Schmid, C. H., & Lau, J. (2005). In an empirical evaluation of the funnel plot, researchers could not visually identify publication bias. *Journal of clinical epidemiology*, 58(9), 894-901.

The Early Math Collaborative - Erikson Institute. (2013). *Big ideas of early mathematics: What teachers of young vhildren need to know (Practical Resources in ECE)*. US: Pearson Education.

*Tietze, W., Hundertmark-Mayser, J., & Roßbach, H. (1999). European child care and education study. School-age assessment of child development: Long-term impact of pre-school experiences on school success and family-school relationships. *Research and Development RTD Action: Targeted Socio-Economic Research*.

Tietze, W., Meischner, T., Gänsfuß, R., Grenner, K., Schuster, K., Völkel, P., & Roßbach, H.-G. (1998). *Wie gut sind unsere Kindergärten? Eine Untersuchung zur pädagogischen Qualität in deutschen Kindergärten* [How good are our preschools? A study on the educational quality of German preschools]. Weinheim, Germany: Beltz.

*Tsakiri, T. (2012). A sudy of early childhood education and care and child development in Greece. (Doctor of Philosophy Dissertation), University of London, London.

Viechtbauer, W. (2005). Bias and efficiency of meta-analytic variance estimators in the random-effects model. *Journal of Educational and Behavioral Statistics,* 30(3), 261-293.

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, 36(3), 1-48.

von Maurice, J., Artelt, C., Blossfeld, H.-P., Faust, G., Roßbach, H.-G., & Weinert, S. (2007). *Bildungsprozesse, Kompetenzentwicklung und Formation von Selektionsentscheidungen im Vor- und Grundschulalter. Überblick über die Erhebung in den Längsschnitten BiKS-3–8 und BiKS-8–12 in den ersten beiden Projektjahren*. Retrieved from http://psydok.sulb.uni-saarland.de/volltexte/2007/1008.

Votruba-Drzal, E., Coley, R. L., Maldonado-Carreño, C., Li-Grining, C., & Chase-Lansdale, P. L. (2010). Child care and the behavior problems of economically disadvantaged children in middle childhood. *Child Development, 81,* 1460–1475.

Walberg, H.J. & Tsai, S.-L. (1983). Matthew effects in education. *American Educational Research Journal* 20(3), 359-373.

Wessels, H., Lamb, M. E., & Hwang, C. P. (1996). Cause and causality in daycare research: An investigation of group differences in Swedish child care. *European Journal of Psychology of Education, 11(2),* 231-245.

Appendix

Table 1

Overview of included studies

Project	Project period	Age at first assessment[a]	Age at last follow-up[b]
Multiple Countries			
European Child Care and Education Study (ECCE) – Austrian, German, and Spanish samples	1992 - 1998	4 years	8 years
IEA Pre-Primary Project (PPP)	1992 – 1995	4 years	7 years
Finland			
Interaction and Learning within Children-Parent-Teacher Triangle (First Steps)	2006 - 2016	6 years	10 years
Germany			
Bildungsprozesse, Kompetenzentwicklung und Formation von Selektionsentscheidungen im Vor- und Grundschulalter (BiKS)	2005 – ongoing	4 years	9 years
Early Chances	2012 - 2016	2 years	5 years
Kindergarten der Zukunft in Bayern (KiDZ)	2004 – 2013	4 years	6 years
National Educational Panel Study (NEPS) – Kindergarten Cohort	2010 – ongoing	4 years	7 years
Schulreifes Kind (The school-prepared child)	2008 – 2010	4 years	7 years
Stärkung der Bildungs- und Erziehungsqualität in Kindertageseinrichtungen und Grundschule – Gestaltung des Übergangs (TransKiGs)	2005 - 2009	5 years	7 years
Greece			
A Study of Early Childhood Education and Care and Child Development in Greece (Attiki)	2000 – 2002	3 years	4 years
Netherlands			
Cohort study Pre-COOL	2009 – ongoing	2 years	3 years
Cohort study of Primary Education (PRIMA-2) – Kindergarten cohort 1996	1996 - 2000	6 years	10 years
Utrecht Mixed Preschool Groups - preschool and kindergarten cohort	2008 – 2010	3 / 4 years	5 / 6 years
Portugal			
Longitudinal Study of Children's Engagement and Adaptation (Engagement Study)	2001 - 2009	< 1 year	5 years
Contexts and Transition	2005 – 2008	4 and 5 years	6 years

Notes. a.) refers to first assessment, not necessarily entailed in meta-analysis b.) refers to last follow-up included in meta-analysis

Table 1 (*continued*)

Project	Project period	Age at first assessment[a]	Age at last follow-up[b]
United Kingdom			
Millennium Cohort Study (MCS)	2000 - ongoing	< 1 year	5 years
England			
Effective Pre-School, Primary and Secondary Education Project (EPPSE)	1997 - 2013	3 years	16 years
Northern Ireland			
Effective Pre-school Provision in Northern Ireland (EPPNI)	1998 – 2010	3 years	8 years
Scotland			
Growing up in Scotland (GUS) - Birth cohort 1	2005 -ongoing	< 1 year	4 years

Notes. a.) refers to first assessment (not necessarily included in meta-analysis) b.) refers to last follow-up included in meta-analysis

Table 2

Sample characteristics for each included study

Project	Samples (children)	Sample (centre, groups)	Gender (% female)	% Disadvantaged (Type of disadvantage) (and other for now)
ECCE – Austria	T_1: 114, HS = N/R T_L: 107	43 centres	49	28 (Unemployed mothers)
ECCE – Germany	T_1: 396, HS = N/R T_L: 306	103 centres	49	29 (Unemployed mothers)
ECCE – Spain	T_1: 238, HS = N/R T_L: 173	60 centres	49	52 (Unemployed mothers)
PPP[a]	T_1: 2,904 T_L: 2,247	838 centres	51	N/R
First Steps	T_1: 1,268 T_L: 1,239	121 centres, 137 groups	48	6 (Low parental education)
BiKS	T_1: 547 T_L: 320	97 centres, 97 groups	48	10 (Parental language foreign) 24 (Low parental education)
Early Chances	T_1: 1,331 T_L: 1,123	334 centres	N/R	N/R
KiDZ	T_1: 191 T_L: 164	6 centres, 16 groups	50	16 (Home language foreign)
NEPS	T_1: 2458 T_L: 2458	216 centres, 633 groups	50	16 (Home language foreign) 17 (Low parental education) 31 (Unemployed mother) 16 (Low culture capital -<25 books at home)
School-prepared child	T_1: 609 T_L: 340	63 centres	47	N/R
TransKiGs	T_1: 437 T_L: 376	49 centres, 123 groups	49	22 (single parent) 31 (Low culture capital – <50 books at home) 16 (Low parental education) 48 (Low parental occupational status; HISEI)
Attiki	T_1: 150 T_L: 110	20 centres, 21 groups	47	13 (Unemployed, unskilled mothers) 5 (Low parental education) 13 (Low parental occupational status)

Notes. [a] Sample of children comprise 63% of children from European countries (i.e. Finland, Germany, Ireland, Italy, Poland and Spain). T_1 = fist measurement point. T_L = last measurement point. HS = home sample

Table 2 (*continued*)

Project	Samples (children)	Sample (centre, groups)	Gender (% female)	% Disadvantaged (Type of disadvantage) (and other for now)
Pre-COOL	T_1: 1,819, HS = 1,008 T_L: 2,830	263 centres, 295 groups	49	42 (Low to middle parental education) 33 (Migration background) 28 (Home language foreign)
PRIMA	T_1: 11,606, HS = N/R T_L: 3,596	600 centres	N/R	N/R
Utrecht Mixed Preschool	T_1: 48 preschooler, 43 kindergarteners T_L: 48 preschooler, 43 kindergarteners	14 preschool and 12 kindergarten groups	51	80 (ethnicity other than Caucasian)
Engagement Study	T_1: 120 T_L: 95	15 centres, 30 groups	52	N/R
Contexts and Transition	T_1: 215 T_L: 102	60 groups	48	22 (Low maternal education)
MCS	T_1: 10,358, HS = N/R T_L: 4800	N/R	50	5 (ethnicity other than Caucasian) 4 (Home language foreign) 8 (Low maternal education) 16 (Receiving benefits) 7 (Single mother) 14 (Teenage pregnancy)
EPPSE	T_1: 2,857, HS = 315 T_L: 2,484, HS = 262	141 centres	48	21 (ethnicity other than Caucasian) 9 (Home language foreign) 18 (Low maternal education) 16 (Low parental occupation) 23 (Free School Meals) 13 (3 or more siblings)
EPPNI	T_1: 685, HS = 152 T_L: 676, HS = N/R	80 centres, 80 groups	51	2 (Migration background) 23 (Low maternal education) 20 (single parent) 6 (low birth weight) 26 (health problem since birth) 7 (developmental problem since birth) 5 (behavioural problem since birth) 21 (3 or more siblings)
GUS	T_1: 3,658 T_L: N/R	1,296 centres	N/R	N/R

Notes. T_1 = fist measurement point. T_L = last measurement point. HS = home sample

Table 3

Overview of ECEC and outcome measures in each study

Project	Quantity	Structural Quality	Global Process Quality	Pre-academic promotion	Literacy	Mathematics
ECCE	Home sample Duration Entry age	Space per child Child-staff ratio Teacher qualification	ECERS CIS OAP	N/A	PPVT-R	N/A
PPP	N/A	Class size Materials Teacher qualification	CA AB	MOT - preacademic	Items adapted from: Test of early language development Iowa tests of basic skills Level 7 Battelle developmental inventory Test of language development intermediate	N/A
First Steps	N/A	Average age Class size Teaching experience Teacher qualification % male	CLASS	N/A	ARMI - Initial phoneme identification & letter knowledge	Number sequences test Screen number concept test Basic Arithmetic Test
BiKS	Duration Entry age	Space per child Child-staff ratio Class size	ECERS-R	ECERS-E	PPVT-R	K-ABC HRT
Early Chances	N/A	Child-staff ratio	ECERS-R	DO-RESI	PPVT-IV SETK TSVK	N/A
KiDZ	N/A	Child-staff ratio Class size Working hours/week Teacher training	ECERS-R	ECERS-E	PPVT-R TROG-D	K-ABC
NEPS	N/A	Child-staff ratio Class size	N/A	Frequency of stimulating activities	PPVT-R	N/A

Notes. I, II, III, IV or R – Revised versions of the same instrument, E – Extension

Quality measures: AB – Adult Behaviour, CA - Child activity, CIS – Caregiver Interaction Scale, CLASS – Classroom Assessment Scoring System, DO-RESI - Dortmunder Ratingskala zur Erfassung sprachförderrelevante Interaktionen, ECERS - Early Childhood Environment Rating Scale, MOT - Management of time.

Literacy measures: ARMI - A tool for assessing reading and writing skills in Grade 1, BAS – British Ability Scales, PPVT – Peabody Picture Vocabulary Test, SETK – Sprachentwicklungstest für Kinder, TROG-D – Test zur Überprüfung des Grammatikverständnisses, TSVK –Test zum Satzverstehen von Kindern.

Math measures: BAS – British Ability Scales, HRT - Heidelberger Rechentest, K-ABC – Kaufman Assessment Battery for Children.

Table 3 (*continued*)

Project	Quantity	Structural Quality	Global Process Quality	Pre-academic promotion	Literacy	Mathematics
School-prepared child	Intensity	N/A	N/A	N/A	N/A	Test battery for young children (Krajewski, 2005) PIPS DEMAT 1+
TransKiGs	Duration Entry age	Institution size Teaching experience Teacher training	ECERS-R CIS	ECERS-E DO-RESI-E	NEI-KiGs – Narrative competence	OTZ Self-constructed test (knowledge of digits, amounts, basic arithmetic, measurement)
Attiki	Duration Entry age	Child-staff ratio Class size Teacher qualification	ECERS-R CIS	ECERS-E	BAS II - Verbal comprehension & naming vocabulary	BAS II – number concepts
Pre-COOL	Home sample Entry age Intensity	Child-staff ratio Class size Number of staff in institution Change of staff Teaching experience	CLASS	ECERS-E activities questionnaire (self-constructed)	CCC-2-NL PPVT-III-NL CITO – “Taal voor kleuters” VTO	UGT-R CITO - "Ordenen"
PRIMA	Home sample Duration	N/A	N/A	N/A	CITO	CITO
Utrecht Mixed Preschool	Intensity	mixed versus target arrangements (i.e. high % of disadvantage children) Teacher qualification	N/A	Teacher-managed literacy/math activities	PIPS – emergent literacy	PIPS – emergent math

Notes. I, II, III, IV or R – Revised versions of the same instrument, E – Extension

Quality measures: CIS – Caregiver Interaction Scale, CLASS – Classroom Assessment Scoring System, DO-RESI - Dortmunder Ratingskala zur Erfassung sprachförderrelevante Interaktionen, ECERS - Early Childhood Environment Rating Scale.

Literacy measures: CCC - Children’s Communication Checklist, CITO – Central Institute for Test Development , PIPS – Performance Indicators in Primary Schools, OTZ - Osnabrücker Test zur Zahlbegriffsentwicklung , PPVT – Peabody Picture Vocabulary Test, VTO - Voortijdige Onderkenning van Taalstoornissen.

Math measures: BAS – British Ability Scales, CITO – Central Institute for Test Development , DEMAT 1+ - Deutscher Mathematiktest für erste Klassen, HRT - Heidelberger Rechentest, K-ABC – Kaufman Assessment Battery for Children, NFER – National Foundation for Educational Research, PIPS – Performance Indicators in Primary Schools, UGT-R - Utrechtse Getalbegrip Toets-Revised, WPPSI – Wechsler Preschool and Primary Intelligence scale for children.

Table 3 (*continued*)

Project	Quantity	Structural Quality	Global Process Quality	Pre-academic promotion	Literacy	Mathematics
Engagement Study	N/A	N/A	ECERS-R ITERS	N/A	CAP Griffiths Mental Development Scales	N/A
Contexts and Transition	N/A	N/A	ECERS-R	ELLCO ECERS-E	CAP PPVT-R	WPPSI-R-arithmetic subtest Rote counting Number identification
MCS	Home sample	N/A	N/A	N/A	BAS II – Verbal Comprehension & naming vocabulary	BBCS-R
EPPSE	Home sample Duration Entry age Intensity	Child-staff ratio Teacher qualification	ECERS-R CIS	ECERS-E	BAS II – Verbal comprehension & naming vocabulary Pre-reading measure: combined measure of letter recognition (Clay, 1993) and phonological awareness (Bryant & Bradley, 1985) NFER-Nelson Primary Reading (Level 1 and level 2) National Assessment English at KS1 and KS2 Teacher Assessments in English (not further described) GCSE English	BAS II - early number concepts NEFR-Nelson Maths 6 and 10 National Assessment maths at KS1 and KS2 Teacher Assessments in mathematics (not further described) GCSE mathematics
EPPNI	Home sample Duration Intensity	Child-staff ratio Class size Parental involvement Teacher qualification	ECERS-R CIS	ECERS-E	BAS II - verbal comprehension, naming vocabulary, picture similarities, word reading Pre-reading measure: combined measure of rhyme, alliteration and letter recognition (Clay, 1993) NFER-Nelson - Primary Reading Test (Level 1) Literacy measure from school (not further described) K-Stage 2 Assessment English	BAS II - early number concepts NEFR-Nelson Maths 6 Numeracy measure from school (not further described) K-Stage 2 Assessment Math combined with teachers' assessments
GUS	Intensity	Centre size	SCSWIS QI	N/A	BAS II - naming vocabulary, picture similarities	N/A

Notes. I, II, III, IV or R – Revised versions of the same instrument, E – Extension

Quality measures: CIS – Caregiver Interaction Scale, CLASS – Classroom Assessment Scoring System, ECERS - Early Childhood Environment Rating Scale, ELLCO - Early Language and Literacy Classroom Observation, ITERS - Infant/Toddler Environment Rating Scale, QI - Education Scotland quality indicator grade, SCSWIS - Social Care and Social Work Improvement Scotland.

Literacy measures: BAS – British Ability Scales, CAP - Concepts About Print, NFER – National Foundation for Educational Research, PIPS – Performance Indicators in Primary Schools, PPVT – Peabody Picture Vocabulary Test.

Math measures: BAS – British Ability Scales, BBCS – Bracken Basic Concept Scale, NFER – National Foundation for Educational Research, PIPS – Performance Indicators in Primary Schools, WPPSI – Wechsler Preschool and Primary Intelligence scale for children.

Table 4

Means and standard deviations of process quality in studies reporting information for measures with a similar metric ranging from 1 = inadequate quality, 3 = minimal quality, 5 = good quality, and 7 = excellent quality (i.e., ECERS-R, CLASS, ECERS-E and subscales)

Project	**Global Process Quality**			**Pre-academic Promotion**		
	Measure	**Mean**	**SD**	**Measure**	**Mean**	**SD**
ECCE - Austria	ECERS	4.71	0.51	N/A		
ECCE - Germany	ECERS	4.51	0.71	N/A		
ECCE - Spain	ECERS	4.05	0.88	N/A		
First Steps	CLASS – emotional support	5.12	0.72	N/A		
	CLASS – classroom organisation	5.34	0.62			
	CLASS – instructional support	3.96	0.85			
BiKS	ECERS-R	3.73	0.58	ECERS-E	2.98	0.53
KiDZ	ECERS-R in treatment group	4.3	N/R		4.14	0.42
	ECERS-R in control group	4.1	N/R		3.27	0.41
TransKiGs	ECERS-R	3.00	0.62	ECERS-E	1.13	0.84
Attiki	ECERS-R	2.56	0.94	ECERS-E	1.57	0.43
Pre-COOL	CLASS – emotional support	4.94	0.70	ECERS-E	2.24	0.84
	CLASS – behavioral support	5.85	0.48			
	CLASS – engaged support for learning	3.23	0.78			
Engagement Study	ECERS-R	3.32	0.89	N/A		
Contexts and Transition	ECERS-R	3.44	0.72	ECERS-E literacy	3.26	0.89
				ECERS-E mathematics	2.05	0.93
EPPSE	ECERS-R	4.34	1.00	ECERS-E	3.12	1.00
EPPNI	ECERS-R	4.61	0.69	ECERS-E	2.26	0.66
Average		**4.17**	**0.73**		**2.6**	**0.7**

Table 5

Compendium of longitudinal studies on ECEC and developmental outcomes

The compendium of studies includes international and European longitudinal studies that assess ECEC quantity and quality, and child outcomes in the literacy, math, and socio-emotional domain since 1990. The compendium informs about studies that yield information about the developmental impact of ECEC. Findings for the effect of ECEC are not necessarily (yet) available. The compendium shows the potential of studies to address important question regarding the relationship of ECEC quantity and quality in three important developmental domains. Provided information is based on information about the study in reviews (e.g., Anders, 2013; Melhuish et al., 2015; Nores & Barnett, 2010), published results and study description on websites.[8] A short description of the study is provided, if information on the title of the study was not available. If information on project period was missing, we estimated the project period based on the publication date of the study references. Further explanations are listed below.

[8] The compendium and references for the studies will be provided by the authors on request.

Table 6

Association of moderators for global process quality listed below the diagonal and for pre-academic promotion above the diagonal tested with chisquare-tests (for all categorical moderators) and point biserial correlation coefficients (for age)

	Domain	Age at assessment	Phase at assessment	ECEC Measure	Peer-reviewed
Domain		.08	0.00	3.87*^{neg}	2.42
Age at assessment	.07		.63***	.04	-.14
Phase at assessment	0.02	.57***		0.00	9.91**^{neg}
ECEC measure	0.10	-.03	0.14		10.0**^{neg}
Peer-reviewed	0.33	-.13	21.25***^{neg}	1.38	

Note. *** $p < .001$. ** $p < .01$. * $p < .05$. # < .10. All moderators are scored 0 versus 1. ^{pos} significantly higher number of moderator classifications with the same values. ^{neg} significantly higher number of moderator classifications with different values.

Table 7

Association of moderators for structural quality listed below the diagonal and for quantity above the diagonal tested with chisquare-tests (for all categorical moderators) and point biserial correlation coefficients (for age)

	Domain	Age at assessment	Phase at assessment	ECEC Measure	Peer-reviewed
Domain		.10	1.17	0.67	2.00
Age at assessment	.01		.68***	.46**	-.28#
Phase at assessment	0.00	.85***		2.47	0.14
ECEC measure	0.85	.30#	1.55		5.41*^{neg}
Peer-reviewed	14.96***^{pos}	.00	0.00	0.41	

Note. *** $p < .001$. ** $p < .01$. * $p < .05$. # < .10. All moderators are scored 0 versus 1. ^{pos} significantly higher number of moderator classifications with the same values. ^{neg} significantly higher number of moderator classifications with different values.

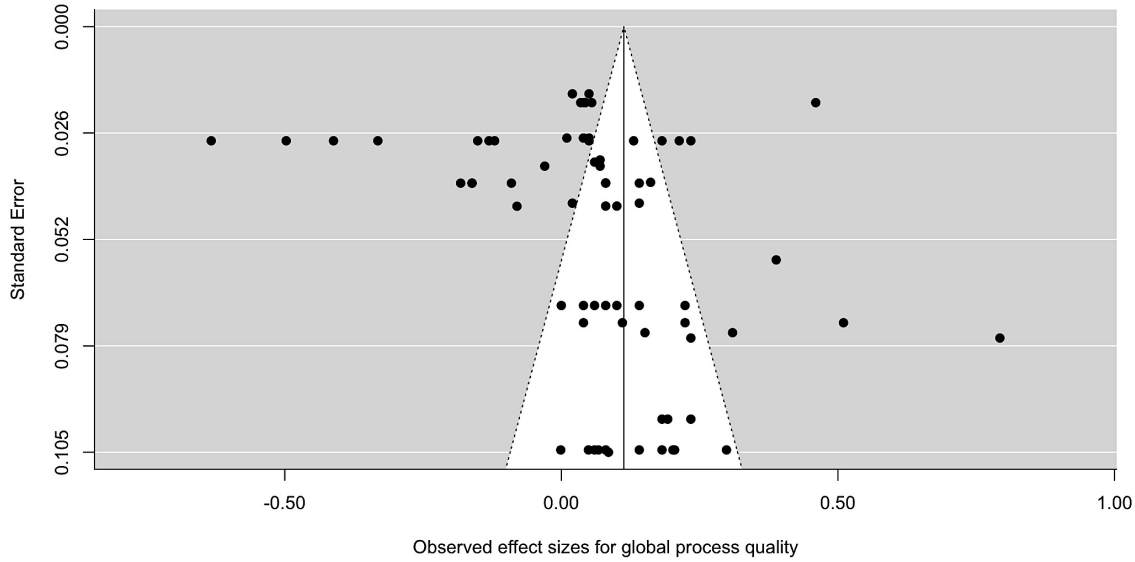

Figure 1a. Funnel plot for the overall effect of global process quality. The plot shows the individual observed effect sizes on the x-axis against the corresponding standard errors on the y-axis. A vertical line indicates the estimate based on the model.

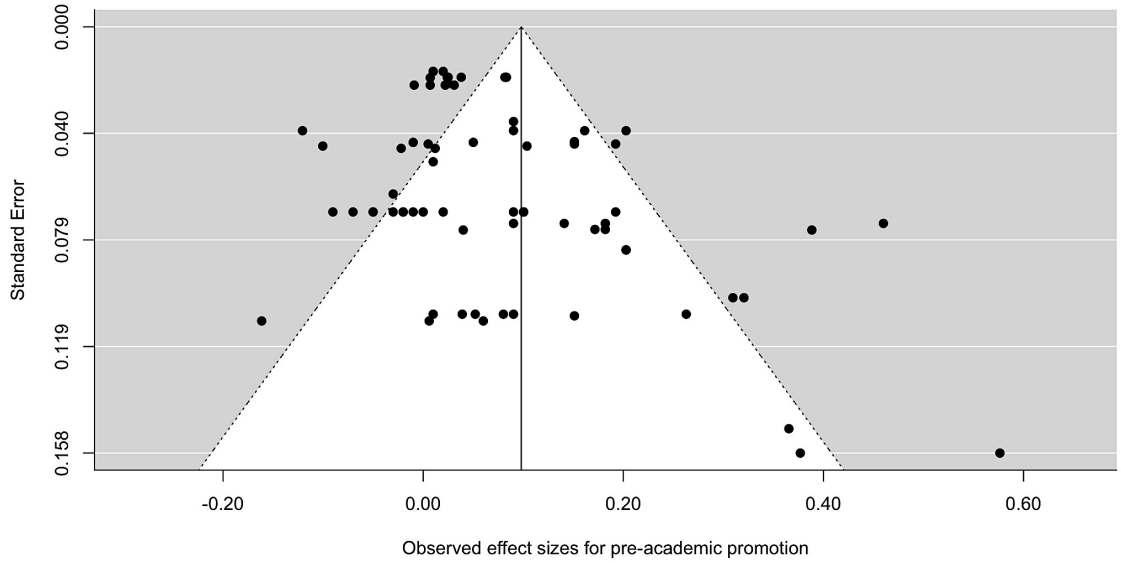

Figure 1b. Funnel plot for the overall effect of pre-academic promotion. The plot shows the individual observed effect sizes on the x-axis against the corresponding standard errors on the y-axis. A vertical line indicates the estimate based on the model.

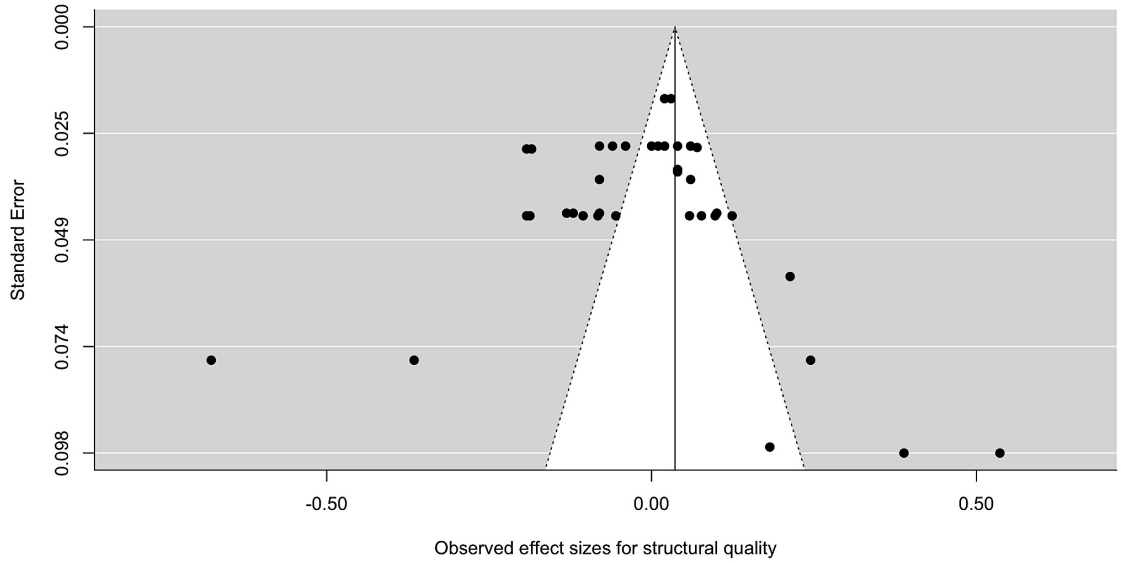

Figure 1c. Funnel plot for the overall effect of structural quality. The plot shows the individual observed effect sizes on the x-axis against the corresponding standard errors on the y-axis. A vertical line indicates the estimate based on the model.

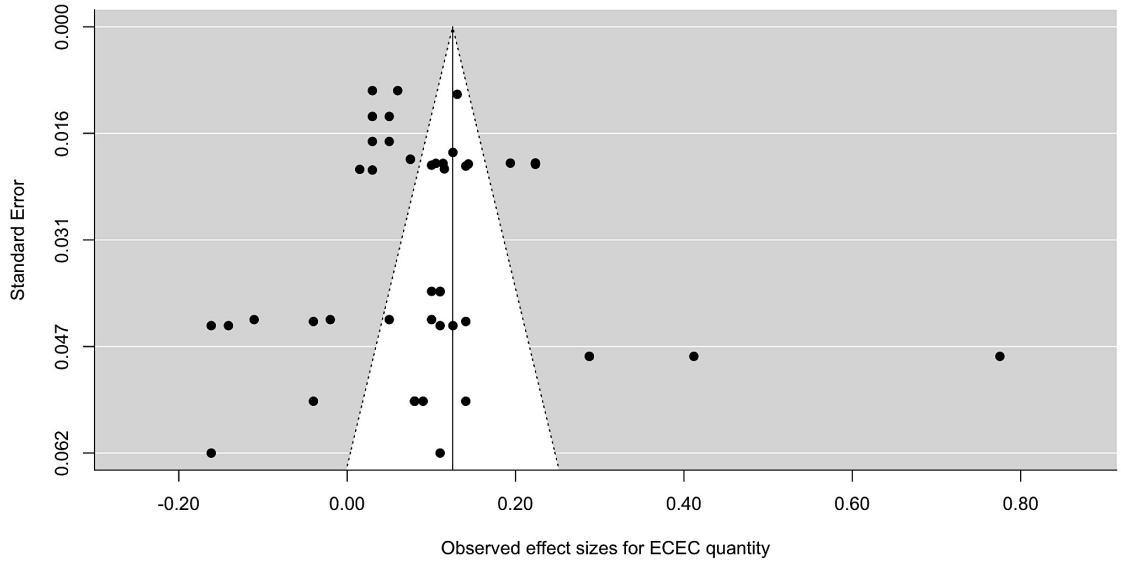

Figure 1d. Funnel plot for the overall effect of ECEC quantity. The plot shows the individual observed effect sizes on the x-axis against the corresponding standard errors on the y-axis. A vertical line indicates the estimate based on the model

Contact list

Finland

Professor Marja-Kristiina Lerkkanen — *E-mail:* marja-kristiina.lerkkanen@jyu.fi

Dr. Gintautas Silinskas — *E-mail:* gintautas.silinskas@jyu.fi

Germany

Professor Yvonne Anders — *E-mail:* yvonne.anders@fu-berlin.de

Ana Susac — *E-mail:* ana.sushac@gmail.com

Hannah Ulferts — *E-mail:* hannah.ulferts@fu-berlin.de

The Netherlands

Professor Paul Leseman — *E-mail:* P.P.M.Leseman@uu.nl

Dr. Martine Broekhuizen — *E-mail:* M.L.Broekhuizen@uu.nl

Dr. Pauline Slot — *E-mail:* P.L.slot@uu.nl

Portugal

Dr. Joana Cadima — *E-mail:* jcadima@fpce.up.pt

UK

Professor Edward Melhuish — *E-mail:* edward.melhuish@education.ox.ac.uk