CampbellCollaboration

Education Coordinating Group

The relationship between teacher qualification and the quality of the early childhood education and care environment

Matthew Manning, Susanne Garvis, Christopher Fleming, Gabriel T.W. Wong

A Campbell Systematic Review
2017:1

Published: January 2017
Search executed: November 2014 – January 2015

CampbellCollaboration

The Campbell Library comprises:

- Systematic reviews (titles, protocols and reviews)
- Policies and Guidelines Series
- Methods Series

Go to the library to download these resources, at:

www.campbellcollaboration.org/library/

Better evidence for a better world

Colophon

Title	The relationship between teacher qualification and the quality of the early childhood education and care environment
Authors	Matthew Manning Susanne Garvis Christopher Fleming Gabriel T.W. Wong
DOI	10.4073/csr.2017.1
No. of pages	82
Citation	Manning M, Garvis S, Fleming C, Wong T. W. G. The relationship between teacher qualification and the quality of the early childhood care and learning environment. Campbell Systematic Reviews 2017:1 DOI: 10.4073/csr.2017.1
ISSN	1891-1803

Copyright	© Manning et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Roles and responsibilities	Matthew Manning, Susanne Garvis, Christopher Fleming and Gabrielle Wong contributed to the writing and revising of this review. Matthew Manning and Susanne Garvis will be responsible for updating this review as additional evidence accumulates and as funding becomes available.
Editors for this review	Editor: Sandra Jo Wilson Managing editor: Carlton J. Fong
Sources of support	Drs Matthew Manning and Susanne Garvis received internal funding from Griffith University to conduct this research.
Declarations of interest	Dr Matthew Manning, Dr Susanne Garvis, Dr Christopher Fleming and Gabriel Wong have no financial interest in any of the existing research being conducted in this area of research. The authors have no vested interest in the treatments that are the subject of this review or in the outcome of the review. The authors also do not have any incentives to represent findings in a biased manner and strive to avoid any potential conflict.

Corresponding author	Associate Professor Matthew Manning ANU Centre for Social Research and Methods The Australian National University Canberra, ACT 4122 Australia E-mail: matthew.manning@anu.edu.au Full list of author information is available at the end of the article.

Campbell Systematic Reviews

Editor-in-Chief	Julia Littell, Bryn Mawr College, USA
Editors	
Crime and Justice	David B. Wilson, George Mason University, USA Charlotte Gill, George Mason University, USA
Education	Sandra Jo Wilson, Vanderbilt University, USA
International Development	Birte Snilstveit, 3ie, UK Hugh Waddington, 3ie, UK
Social Welfare	Brandy Maynard, St Louis University, USA
Knowledge Translation and Implementation	Aron Shlonsky, University of Melbourne, Australia
Methods	Therese Pigott, Loyola University, USA Ryan Williams, AIR, USA
Managing Editor	Chui Hsia Yong, The Campbell Collaboration
Co-Chairs	
Crime and Justice	David B. Wilson, George Mason University, USA Peter Neyroud, Cambridge University, UK
Education	Sarah Miller, Queen's University, UK Gary W. Ritter, University of Arkansas, USA
Social Welfare	Mairead Furlong, National University of Ireland Brandy Maynard, St Louis University, USA
Knowledge Translation and Implementation	Robyn Mildon, CEI, Australia Cindy Cai, AIR, USA
International Development	Peter Tugwell, University of Ottawa, Canada Hugh Waddington, 3ie, UK
Methods	Ariel Aloe, University of Iowa, USA

The Campbell Collaboration was founded on the principle that systematic reviews on the effects of interventions will inform and help improve policy and services. Campbell offers editorial and methodological support to review authors throughout the process of producing a systematic review. A number of Campbell's editors, librarians, methodologists and external peer reviewers contribute.

The Campbell Collaboration
P.O. Box 4404 Nydalen
0403 Oslo, Norway
www.campbellcollaboration.org

Table of contents

LIST OF FIGURES

LIST OF TABLES

Synopsis/plain language summary

Higher teacher qualifications are associated with higher quality early childhood education and care

This review examines the empirical evidence on the relationship between teacher qualifications and the quality of the early childhood learning environment. Higher teacher qualifications are significantly positively correlated with higher quality in early childhood education and care.

What did the review study?

Poor quality early childhood education and care (ECEC) can be detrimental to the development of children as it could lead to poor social, emotional, educational, health, economic and behavioural outcomes.

The lack of consensus as to the strength of the relationship between teacher qualification and the quality of the early childhood learning environment has made it difficult for policy makers and educational practitioners alike to settle on strategies that would enhance the learning outcomes for children in their early stages of education.

What is the aim of this review?
This Campbell systematic review examines the current empirical evidence on the correlation between teacher qualifications and the quality of the early childhood learning environments. The review summarises findings from 48 studies with 82 independent samples. Studies included children from pre-kindergarten and kindergarteners prior to elementary/primary school and centre-based providers.

This review examines the current empirical evidence on the correlation between teacher qualifications and the quality of early childhood learning environments.

What studies are included?

Included studies must have examined the relationship between teacher qualification and quality of the ECEC environment from 1980 to 2014, as well as permit the identification of the education program received by the lead teacher and provide a comparison between two or more groups of teachers with different educational qualifications.

Furthermore, the studies had to have comparative designs and report either an overall quality scale or an environment rating scale.

A total of 48 studies conducted with 82 independent samples were included in the review.

What are the main results in this review?

Overall, the results show that higher teacher qualifications are significantly correlated with higher quality early childhood education and care.

The education level of the teachers or caregivers is positively correlated to overall ECEC qualities measured by the environment rating scale. There is also a positive correlation between teacher qualification and subscale ratings including program structure, language and reasoning.

What do the findings in this review mean?

The review shows a positive statistically significant association between teacher qualification and the quality of early childhood learning environment. This finding is not dependent on culture and context given that the evidence is from several countries.

Mandating qualified teachers, i.e. with tertiary education, may lead to significant improvement for both process and structural quality within centre-based and home-based ECEC settings. However, the evidence is from correlational studies, so evidence is needed from studies with designs which can assess causal effects. Further research should also assess what specific knowledge and skills learnt by teachers with higher qualifications enable them to complete their roles effectively.

How up to date is this review?

The review authors searched for studies published until December 2014. This Campbell systematic review was published in January 2017.

What is the Campbell Collaboration?

The Campbell Collaboration is an international, voluntary, non-profit research network that publishes systematic reviews. We summarise and evaluate the quality of evidence for social and economic policy, programmes and practice. Our aim is to help people make better choices and better policy decisions.

About this summary

This summary was prepared by Ada Chukwudozie and Howard White (Campbell Collaboration) based on the Campbell Systematic Review 2017:1 The relationship between teacher qualification and the quality of the early childhood education and care environment by Matthew Manning, Susanne Garvis, Christopher Fleming and Gabriel T.W. Wong. The summary was designed, edited and produced by Tanya Kristiansen (Campbell Collaboration).

Executive summary/Abstract

BACKGROUND

The notion that a strong early childhood education and care (ECEC) knowledge base, which involves a set of professional competencies, abilities and specific teaching skills, can lead to high-quality ECEC and positive child developmental outcomes is yet to be fully determined (Bowman, Donovan, & Burns, 2001; Vartuli, 1999). This is due, in some instances, to lack of good data, the quality of the method employed to measure the relationship between teacher qualification and the quality of the early childhood learning environment, and the methods used to aggregate the findings of individual empirical studies. The lack of consensus regarding the direction (positive in this case) and strength of the relationship between teacher qualification and the quality of the early childhood learning environment has made it difficult for policy makers and educational practitioners to form strategies that will ultimately enhance the early learning outcomes of children.

OBJECTIVES

The objective of this review is to synthesise the extant empirical evidence on the relationship of teacher qualifications to the quality of the early childhood learning environment. Specifically, we address the question:

> Is there a relationship between the level and type of education of the lead teacher, and the quality of the early childhood learning environment, as measured by the Early Childhood Environment Rating Scale, the Infant Toddler Environment Rating Scale and their revised versions?

SEARCH METHODS

Studies were identified by exploring a large number of relevant academic journals (e.g., Early Childhood Research Quarterly, Early Childhood Research and Practice, Contemporary Issues in Early Childhood, Child Development, Applied Developmental Science, and the Journal of Child Psychology and Psychiatry) and electronic databases (e.g., Academic Search Premier; CBCA-Education; Cochrane Controlled Trial Register; Database of Abstracts of Reviews of Effectiveness (DARE); Dissertation Abstracts; EconLit; Education Full Text; Educational Resources Information Center (ERIC); Journal Storage Archive (JSTOR); Medline; Proquest Digital Dissertations; Proquest Direct; Project Muse; PsychInfo; Scopus; SocINDEX with Full Text; and SSRN eLibrary).

We also searched the reference list of each eligible study, and reviewed the biographies and publication lists of influential authors in the field of early childhood development and education, to determine if there were any relevant studies not retrieved in the original search.

SELECTION CRITERIA

Selection criteria are based on both comparative and correlational studies that examine the relationship between teacher qualification and quality of the ECEC environment (as measured by ECERS/ECERS-R/ITERS/ITERS-R and any subscales) from 1980 (this was when the ECERS was introduced) to 2014. Eligible studies, therefore, report at least one of the following results: (1) the overall ERS ratings (main outcome); (2) ratings of the seven subscales – program structure (i.e. focusing on the schedule, time for free play, group time and provisions for children with disabilities), activities (i.e. focusing on the provision and quality of activities including fine motor, art, music, dramatic play and math/number), language and reasoning (i.e. focusing on the formal and informal use of language, development of reasoning skills and communication), parent and staff needs (i.e. focusing on the provisions for personal and professional needs of staff and parents, and staff interaction and cooperation), space and furnishing (i.e. focusing on the quality of items including indoor space, furniture for routine care, room arrangement and space for privacy), interactions (i.e. focusing on discipline as well as supervision and facilitation of proper interactions between children and staff and among children) and personal care routines (i.e. focusing on teaching and practice of routines including greeting/departing, meals/snacks, toileting/diapering, health and safety); and (3) the two subscales - language and interactions, and provisions for learning.

DATA COLLECTION AND ANALYSIS

The systematic search identified 2,023 unique studies on the relationship of teacher qualifications to the quality of the early childhood learning environment, of which 80 were obtained. A final set of 48 studies was eligible for inclusion in our meta-analysis. Data analysis was conducted using Comprehensive Meta-Analysis 2.0 (CMA), a statistical meta-analysis software package. Both correlation coefficients and mean standardized differences were converted to a common effect size - in this study a correlation coefficient (r).

We examined possible moderators of process quality in ECEC settings including: (1) teacher qualification; (2) baseline characteristics of teacher; (3) country in which the study was conducted; (4) duration of follow-up; (5) outcome measure; and (6) dominant ethnicity of student group. Quality and accessibility of data limited us to exploring only the outcome measure (e.g., ECERS vs. ITERS) and dominant ethnicity of student group.

We employ a random effects model for pooling intervention effects. An assumption is made that there are unexplained sources of heterogeneity across studies. The Q statistic, which was calculated in each fixed effect analysis, was used for the calculation of the τ^2. In addition, we employ the I^2 statistic (Higgins & Thompson, 2002) as an additional, albeit related, method of assessing heterogeneity.

RESULTS

In this review we assess the correlation between teacher qualifications and measures of ECEC quality. There were 82 independent samples available for meta-analysis: 58 assessed the overall quality of ECEC as an outcome and 24 assessed ratings of Environment Rating Scales (ERS) subscales. The relationship between teacher qualifications and overall ECEC quality demonstrate a positive correlation that was statistically significant (mean correlation with robust standard error, assuming ρ = .80 (r=0.198, confidence limits 0.133, 0.263)). When overall quality was disaggregated by measurement method (e.g. ECERS, ECERS-R), studies that measured ECEC quality using different scales produced a non-significant difference.

Below, in descending order of effect size (correlation coefficient r), results (for the 7 factor subscales) show:

- a positive and statistically significant relationship between teacher qualifications and program structure (r= 0.224, 95% confidence limits 0.014, 0.415);
- a positive and statistically significant relationship between teacher qualifications and activities (r=0.204, 95% confidence limits 0.140, 0.);
- a positive and statistically significant relationship between teacher qualifications and language and reasoning (r=0.203, 95% confidence limits 0.122, 0.282);
- a positive and statistically significant relationship between teacher qualifications and parent and staff (r=0.189, 95% confidence limits 0.049, 0.321);
- a positive and non-significant relationship between teacher qualifications and space and furnishings (r=0.122, 95% confidence limits -0.042, 0.280);
- a positive and statistically significant relationship between teacher qualifications and interactions (r=0.122, 95% confidence limits 0.053, 0.189); and
- a positive and non-significant relationship between teacher qualifications and personal care (r=0.095, 95% confidence limits -0.053, 0.239).

In descending order of effect size, the 2 factor subscale outcomes evaluated show:

- a positive and non-significant relationship between teacher qualifications and provisions for learning (mean correlation with robust standard error, assuming ρ = .80 (r=0.173, confidence limits -0.054, 0.399)); and
- a positive and non-significant relationship between teacher qualifications and language and interaction (mean correlation with robust standard error, assuming ρ = .80 (r=0.096, confidence limits -0.172, 0.363)).

AUTHORS' CONCLUSIONS

This review shows the significant association between having lead teachers with higher qualifications and the overall structural and process quality within ECEC settings. In this review, ECEC settings consist of centre-based classroom environments serving children of all ages (birth to prior to elementary/primary school age). The meta-analysis has drawn on a wide range of literature from 1980 onwards to provide statistically significant results on the relationship of

teacher qualification to the quality of the early childhood learning environment. The learning environment consists of program structure, activities, language and reasoning, parent and staff, space and furnishing, interactions and personal care routines. In a two way-factor classification, the meta-analysis also reflects a positive correlation between teacher qualifications and ratings on language and interactions and provision for learning within ECEC settings. This means that higher teacher qualifications are related to improvements in supporting children's development, including supporting language-reasoning experience, supervision and the scheduling of activities, organization and arrangement of the room, providing varied social experiences for children, and creating a warm and friendly environment for interactions.

The results are important for governments and stakeholders wanting to improve early childhood services to enhance children and family outcomes. Quality is closely linked to the level of staff qualification, which may indicate that it is important to have teachers with qualification higher than secondary education working with young children. The professionalization of the early childhood sector through more qualified staff may lead to significant gains for children and their families, contributing towards life-long outcomes that will benefit all of society.

Background

THE PROBLEM, CONDITION OR ISSUE

Not all children are born healthy, provided with adequate health care, have access to good nutrition, or live in acceptable housing conditions (Homel et al., 2006). Further, not all children are born free of disabilities, or are raised by parents who can comfort, nurture, and provide opportunities to develop their language, literacy, social problem-solving and behaviour management skills (Manning, 2008). Evidence demonstrates that an achievement gap (i.e. educational disparities) exists between racial and socio-economic groups (Editorial Projects in Education Research Center, 2011; Lee & Burkam, 2002; Magnuson & Waldfogel, 2005). Economically disadvantaged students and students of racially marginalized groups (e.g., African-American and Hispanic) in the US, for example, are more likely to receive lower grades and scores in standardized tests than Caucasian students (Burchinal et al., 2011; Reardon & Galindo, 2009).

Poor quality early childhood education and care (ECEC) can be detrimental to the development of children from all backgrounds, particularly if they fail to equalise some of the disparities and disadvantages that children face in the early developmental stages of their lives. Disparities, for example, may be present in children's cognitive, physical, and social-emotional development (National Institute of Child Health and Human Development Early Child Care Research Network, 2005; Peisner-Feinberg et al., 1999). Without access to high-quality services (e.g., health care and education) that support the child and his/her family in the early years of life, potential negative pathways may lead to poor social, emotional, educational, health, economic and behavioural outcomes (Manning, Homel, & Smith, 2010; Manning 2008; 2004).

A number of longitudinal research projects support the notion that high-quality (compared with low-quality) ECEC is more likely to support optimal child social, emotional and cognitive development, promote growth experiences (including nurturing and attachment), and facilitate positive interaction among teachers and children (Gordon & Browne, 2014; Wolery, 2004). These positive developmental and social experiences, as well as a supportive and nurturing environment, in the early years (e.g., preschool years) are commonly translated into improved school readiness and subsequent successful academic achievement in areas such as reading and mathematics (Ramey & Ramey, 2004). School readiness, as measured by standardized assessments of cognitive and linguistic performance, is the key to an ongoing positive, successful and enriching educational experience (Manning et al., 2010; Manning 2008; 2004). Further, school engagement (including being prepared for various transitions such as from preschool to elementary school) is positively

correlated with perseverance, school adjustment, and school completion. Being prepared to learn, having positive social experiences in school, and ultimately being successful at school (i.e. academic achievement) minimizes the chances of future antisocial behaviour and engagement in criminal activity (Reynolds, Ou, & Topitzes, 2004; Manning et al, 2010).

High-quality ECEC programs (e.g., the High Scope Perry Preschool Program in Ypsilanti, Michigan) have been shown to produce tangible and intangible societal benefits. These include increased taxes due to higher earnings of program participants, reduced victimisations and their associated personal and criminal justice costs, and improvement in quality of life (Schweinhart et al., 2005). The well-documented relationship between ECEC program quality and the tangible and intangible benefits provided by high-quality programs is a fundamental rationale for addressing and enhancing quality in the ECEC environment (National Institute of Child Health and Human Development Early Child Care Research Network, 2005). This idea is highly congruent with Vitaro, Barker, Brendgen, and Tremblay's (2011) emphasis on the 'school-related pathway' (involving school readiness and engagement) as a child developmental pathway that serves as the theoretical foundation for the development of early childhood education policy.

THE RELATIONSHIP BETWEEN TEACHER QUALIFICATIONS AND THE QUALITY OF EARLY CHILDHOOD EDUCATION

As stated earlier, this review focuses on the relationship between teacher qualifications and classroom quality. The key independent variable is teacher qualifications and the key outcome – classroom quality. This section provides a brief literature review on ECEC quality and their measurements and discusses the relationship between teacher qualifications and the quality of early childhood education. This section also defines the nature of each method of measurements and corresponding issues.

Defining the quality of early childhood education

Katz (1992) suggests that the quality of ECEC can be defined and assessed from four perspectives: top-down; bottom-up; inside; and outside-inside. Assessment based on the top-down perspective examines quality primarily by identifying selected features and characteristics of the program and environment, as recorded or observed by an assessor. The bottom-up perspective gives weight to the children's experience of the program and considers to what extent the young children find themselves being included and respected during the learning process. This approach is especially applicable for ECEC programs that focus on the inclusion and integration of children with developmental or learning delays and disabilities into the environment (Spiker, Hebbeler, & Barton, 2011). The inside perspective defines quality of an ECEC program as perceived by the staff. Three dimensions are considered; collegial relationships, staff-parent relationships, and staff-sponsor relationships. The outside-inside perspective focuses on the parent-teacher relationship. This perspective may reflect whether the parents' expectations or pedagogical goals are compatible with the curriculum. It is also suggested that the learning environment for children is more effective when parents are meaningfully engaged in the program through the teacher or service provider (National Association for the Education of Young Children, 1991).

Defining quality in ECEC has been a continuing challenge in terms of the methodology in research and formulation of policy to enhance quality (La Paro, Thomason, Lower, Kintner-Duffy, & Cassidy, 2012). Among the varying definitions or conceptualizations of quality in reference to ECEC, the predominant approach is the top-down perspective (Katz, 1992). Traditionally, the connotation involves both multiple distal (e.g., broad parameters of program and state policies) and proximal (e.g., curriculum and caregiver-child interactions) features of an ECEC environment that are expected to support children's development in various domains (e.g., cognitive, language and social development) (Dunn, 1993; Kontos, 1991). The relative importance of indicators of either proximal or distal features varies across a multitude of perspectives, including that of the parent, caregiver and child (Layzer & Goodson, 2006). In order to avoid vague and nonspecific operational definitions of quality (e.g., any "all-encompassing" term), early childhood researchers have commonly conceptualized and disaggregated the multidimensional ECEC quality into two measurable interrelated components: (1) structural quality - this refers to structural indicators such as child-to-staff ratios and caregiver characteristics such as teacher formal education; and (2) process quality - including learning opportunities available to the children, and teacher-child and peer-to-peer interactions within the child care environment (Clifford, Reszka, & Rossbach, 2010). Both qualities generally share the same goal (or target), which is the enhancement of child development and learning outcomes (Taguma, Litjens, & Makowiecki, 2012). Previous literature reveals that both structural and process characteristics of quality are not only related, but are also important to children's developmental outcomes (Sammons et al., 2002).

Structural quality

A number of studies have examined the effectiveness of structural features to predict quality outcomes. Three frequently studied features are child-to-staff ratio, class size, and staff educational attainment. Children enrolled in programs with low ratios (i.e., few children per caregivers) tend to be more socially competent, compliant, cooperative and advanced in academic progress compared to children from programs with high ratios (Clarke-Stewart & Gruber, 1984; Guo & Harris, 2000; Howes, 1988; Whitebook, 1989). High child-to-staff ratios are also more frequently associated with negative staff-child interactions (Rosenthal & Vandell, 1996). Smaller group sizes (compared to larger group sizes) are associated with better academic progress, higher levels of cooperation, compliance, consideration and engagement in sophisticated social play. Children in smaller groups also have a lower tendency to answer with antisocial responses during social problem-solving tasks (Clarke-Stewart & Gruber, 1984; Dunn, 1993; Howes, 1988; Whitebook, 1989). When caregivers have higher education levels, children tend to have higher levels of social competence (Clarke-Stewart & Gruber, 1984). Education or training of teachers/caregivers in the ECEC environment is also correlated with higher scores on achievement tests, and higher levels of compliance and cooperation in children (Whitebook, 1989).

It is noted, however, that structural indicators of an ECEC program and caregiver characteristics contribute to the overall quality or child developmental outcomes through relatively indirect ways compared to process features (Vandell & Wolfe, 2000). For example, Blau (2000) proposes a two-stage model, suggesting that structural features contribute to the process quality in ECEC settings during the first stage, which subsequently affects children's developmental outcomes. Similarly, Phillipsen, Burchinal, Howes and Cryer (1997) suggest that the effects of structural variables on

child developmental outcomes are mediated by process quality (e.g., more nurturing and extensive teacher-child interactions).

Process quality

Process quality is found to be more predictive of children's learning and developmental outcomes than structural indicators (Vandell & Wolfe, 2000; Whitebook, 1989). The classroom dynamics, including proximal-level interactions and transactions among teachers, children and materials, reflect the core components within the early childhood learning environment. These components are associated with positive academic and social outcomes for young children (Love, Meckstroth, & Sprachman, 1997). A higher process quality classroom tends to be associated with fewer problem behaviours, better cognitive and social skills, and academic progress - especially math skills (Peisner-Feinberg & Burchinal, 1997; Peisner-Feinberg et al., 2001). The concept that higher process quality supports better educational, social and emotional outcomes is also supported by the developmental prevention literature. This literature highlights that developmental and social experiences vary between groups (e.g., those living in high or low socio-economic status areas) and that these differences in experience may be affected by the environment in which the child grows and learns (Blokland & Nieuwbeerta, 2010; Homel, 2005). As such, ECEC institutions, together with the involvement of parents and caregivers, play a pivotal role in the social, emotional and educational development of the child (Sylva et al., 2006; Warash, Ward, & Rotilie, 2008). Further, these institutions need to move beyond traditional system silos where there may be common goals between institutions, but coordination is poor or non-existent. In this regard, it has been the priority of policy makers, and education and developmental experts to monitor and enhance process quality in the ECEC sector.

Measuring quality

Process quality is assessed primarily by observing the early childhood learning environment (Clifford et al., 2010). In some instances, the measures focus on specific aspects of the teacher-child interaction (e.g., the Caregiver Interaction Scale (CIS; Arnett, 1989), and the Student Teacher Relationship Scale (STRS; Pianta & Nimetz, 1991)). The Classroom Assessment Scoring System (CLASS) is one of the most widely used and reliable measurement tools to assess the quality of classroom interactional processes. Specifically, CLASS focuses solely on several dimensions of teacher-child interactions (Pianta, La Paro, & Hamre, 2007). These dimensions are organised into three broad domains (emotional support, classroom organisation and instructional support) reflecting two aspects (social and instructional) of interactions (Hamre et al., 2012; Mashburn et al., 2008). The social aspects of interactions focus on how sensitive and responsive teachers are with respect to children's needs and cues. The instructional features of interactions focus on teachers' behaviours that promote children's development and performance of skills. Hamre and Pianta (2005), who examined the influence of support from teachers on children's school readiness, found that high levels of instructional and emotional support reduce the probability (for children at risk of early school failure) of developing poor learning outcomes and reduce children's conflict with teachers. Mashburn and colleagues (2008) also discover that instructional interactions are able to predict academic and language skills of children, and teachers' emotional interactions are able to predict teacher-reported social skills.

Environment rating scales

The entire notion of the learning environment, however, is beyond the specific interactional aspects as measured by CIS, STRS and CLASS, and is defined in a broader sense with interactions between program components and among people (e.g., teacher, parent and child) in the ECEC setting (Harms & Clifford, 1983). Hence, several Environment Rating Scales (ERS), as the global measure of ECEC quality, have been developed. They include the Early Childhood Environment Rating Scale (ECERS) (Harms & Clifford, 1980), the Infant/Toddler Environment Rating Scale (ITERS) (Harms, Cryer, & Clifford, 2003) and their revised versions (ECERS-R (Harms, Clifford, & Cryer, 1998) and ITERS-R (Harms et al., 2003)). The ERS focus on multiple processes within the early childhood learning setting and are recognised as a set of standardised measurement tools that are widely used in research assessing ECEC quality (Clifford et al., 2010). Subscales of the ERS include: program structure (i.e. focusing on the schedule, time for free play, group time and provisions for children with disabilities); activities (i.e. focusing on the provision and quality of activities including fine motor, art, music, dramatic play and math/number); language and reasoning (i.e. focusing on the formal and informal use of language, development of reasoning skills and communication); parent and staff (i.e. focusing on the provisions for personal and professional needs of staff and parents, and staff interaction and cooperation); space and furnishing (i.e. focusing on the quality of items including indoor space, furniture for routine care, room arrangement and space for privacy); interactions (i.e. focusing on discipline as well as supervision and facilitation of proper interactions between children and staff and among children); and personal care routines (i.e. focusing on teaching and practice of routines including greeting/departing, meals/snacks, toileting/diapering, health and safety). The ERS, when compared to other commonly used measures (e.g., CLASS), focus not only on items related to teacher-child interactions, but also on available materials, and health and safety issues (Hamre, Goffin, & Kraft-Sayre, 2009).

ECERS and ECERS-R have, since their development, been the primary measurement tool for researchers and professionals in the ECEC field (Tout, Zaslow, Halle, & Forrey, 2009). Although other measurement tools are available that assess the early childhood learning environment (e.g., the Observational Rating of the Care Environment (ORCE) Qualitative Ratings (NICHD Early Child Care Research Network, 1999)), they tend not to be widely applied. This is also the case for the Supports for Early Literacy Assessment (SELA) tool (Committee on Developmental Outcomes and Assessments for Young Children, 2008), which, to date, remains underdeveloped.

Halle, Whittaker and Anderson (2010) produce a compendium of measures, reviewing some 50 instruments, which measure the quality in ECEC settings and compare them in terms of target age, purpose (e.g., accreditation and evaluation), method (e.g., observation, interview and document), child developmental domains (e.g., language development, literacy and general cognition), structure (e.g., family involvement, activities/scheduling, classroom organization and materials), administration (e.g., internal communication and leadership/management), improvement (e.g., professional development and program/staff assessments) and training and administration (e.g., cost).

According to the 'cross-cutting' tables created by the authors, ECERS-R is the most comprehensive measure that assesses most of the quality features within ECEC environments reflecting on child development, administrative structure and staff development. With respect to the quality features that support school readiness and subsequent academic progress, ECERS-R covers eight out of nine domains including language development, literacy, math, science, creative arts, social and emotional development, approaches to learning, and health/physical development. General cognition is not measured by ECERS-R. For quality features related to structure, administration and staff, ECERS-R covers four out of five domains under the structure category, all domains under the administration category, and two out of three domains under the monitoring and improvement category. The two domains that are not measured by ECERS-R are business practices and assessments/monitoring of students.

Both the breadth of multifaceted information measured by the ECERS-R and the general lack of supplemental measurement tools contribute to the predominant use of ECERS-R as the primary scale for assessing quality in ECEC settings (Tout et al., 2009). As a pioneering self-assessment tool in ECEC, the ECERS-R has been used in recent large studies of classroom quality such as the National Centre for Early Development and Learning's (NCEDL) Multi-State Study of Pre-Kindergarten and Study of State-Wide Early Education Programs (SWEEP) (La Paro et al., 2012). Further, ECERS-R has gained additional influence on practice and policy-related decisions in the field, as it has become the measure of quality in the state-wide Tiered Quality Rating and Improvement Systems (TQRISs) in the United States (Tout et al., 2009). In the State of California, for example, the ratings produced by ECERS-R have been used for the selection of mentor teachers in 70 community colleges. California's Compensation and Recognition Encourage Stability (CARES) program also uses ECERS-R to assess teachers continued participation.

The first editions of ECERS and ITERS have been used as comprehensive quality measures in nation-wide studies. For example, the ECERS and its revised version (ECERS-R) are used in the Head Start Family and Child Experiences Survey (FACES) study in the U.S. (Espinosa, 2002). Both ECERS and ITERS were used in the National Child Care Staffing Study (Whitebook, 1989), and the Cost, Quality, and Child Outcomes Study (Helburn et al., 1995). Clifford and colleagues (2010) note that when information is not available for the revised version (i.e., ECERS-R and ITERS-R), the first edition of the scale is suitable. Evidence confirms that data from studies using ECERS are directly comparable to data from studies using the ECERS-R (Sakai, Whitebook, Wishard, & Howes, 2003). Further, the authors of the Environment Rating Scale (ERS) have intended to make ECERS, ITERS and their revised versions as comparable measures of quality (Sakai et al., 2003). Therefore, the scoring and formatting of ITERS and ITERS-R are consistent with that of the ECERS-R. Hence, for the purposes of this systematic review, quality of the ECEC environment will be assessed on studies that have used as their measurement tools, ECERS, ITERS and their revised versions (ECERS-R and ITERS-R).

Evidence of predictive validity of ECERS, ECERS-R, ITERS and ITERS-R

The scientific evidence linking process quality (as measured by the ERS) and child learning and developmental outcomes is compelling, although not unanimous. Some researchers (e.g., Gordon, Fujimoto, Kaestner, Korenman, & Abner, 2013; Gordon, Hofer, Fujimoto, Risk, Kaestner, &

Korenman, 2015) argue that the predictive validity of environment rating scales is not exceptionally consistent. Gordon et al. (2015) used the item response theory (IRT) analyses to reveal that although every indicator (254 ECERS-R indicators as identified by the experts) was generally relevant to child care quality and relevant to at least one of the three meta-domains (i.e., cognitive, socioemotional, and health) and at least one of the eight subdomains (e.g., promote math skills, promote social competence, reduce injuries), there was little concensus with regard to which domains the attribute supported as indicators did not fit together to measure single dimensions. These results were, however, limited as the study only relied on expert ratings of the relevance of indicators for aspects of child development and did not directly look at the indicators' validity in relation to school readiness.

Regarding other researchers who suggest that the practicality of ECERS is sound and supported by evidence, Cassidy and colleagues (2005) argue that any unfavourable results, with respect to the predictive validity of ERS, are due to collapsing the multitude of features of ECEC environments into a unidimensional comprehensive index – omitting the specific mechanisms responsible for development (e.g., language reasoning, and space and furnishings). Scientific examinations of associations between the scores on ERS and children's learning and developmental outcomes, therefore, commonly involve the use of both overall and subscale scores of ERS. For example, Howes and colleagues (2008) decompose the ECERS-R into two distinct factors; language/interactions and physical environment. They discover that language/interactions factors are more predictive of children's development than physical environment factors alone.

Previous research (e.g., Sammons et al., 2003a) suggests that the quality of the ECEC environment is associated with children's learning and developmental outcomes in general. For example, Love and colleagues (2004), who measure child learning outcomes using the Bayley Scales of Infant Development Mental Development Index (BSID-MDI) (Bayley, 1993), discover that higher overall scores on ITERS/ECERS-R can predict higher scores on the intellectual and cognitive development screener. The screener covers several types of abilities including, for example, sensory/perceptual acuities, memory learning and problem solving, vocalization and beginning of verbal communication, mental mapping, complex language and mathematical concept formation. Furthermore, in view of the complexity and multidimensionality inherent in children's learning and developmental outcomes, some researchers have identified several outcome categories (e.g., mathematics and numeracy skills, language and literacy, and social outcomes) and developed measures to investigate the impact of ECEC programs on each outcome category.

ERS have been shown to be predictive of children's performance in standard measures of mathematic achievement and numerical skills. For example, Peisner-Feinberg and colleagues (2001), using a shortened version of the ECERS, discover a positive correlation of ECERS ratings with higher scores on the math achievement applied problems subset of the Woodcock-Johnson Psycho-Educational Battery - Revised (WJ-R) (Woodcock & Johnson, 1990). WJ-R is designed to measure cognitive abilities, scholastic aptitude, and academic achievement in three areas - reading, mathematics, and written language. A positive relationship between the social interaction subscale on the ECERS-R and children's early number concept development has also been revealed (Sammons et al., 2003a). A more recent study (Burchinal et al., 2008), using the ECERS-R as the measurement of quality in pre-kindergarten and kindergarten settings, shows that the Teaching

and Interactions factor is predictive of children's performance on the WJ-R math achievement applied problem subset. Anders and colleagues (2012), using the arithmetic subscale of the Kaufman Assessment Battery for Children (KABC) (Melchers & Preuss, 2003), found a marginally significant relation between the total score on ECERS-R and children's numeracy levels over the preschool period (from the first to the third year of preschool).

The predictive validity of ERS has also covered children's language and literacy performance. Love and colleagues (2004) discover that children in ECEC settings with higher overall ECERS-R or ITERS scores tend to perform better in the Peabody Picture-Vocabulary Test – Third Edition (PPVT-III) (Dunn & Dunn, 1997). PPVT-III is a measure of receptive vocabulary for standard English and a screening test of verbal ability. Peisner-Feinberg and colleagues (2001), using the Peabody Picture Vocabulary Test – Revised (PPVT-R) (Dunn & Dunn, 1981), a former edition of PPVT-III, also discover an association between higher overall scores on a shortened version of the ECERS and children's scores on the PPVT-R. In addition, Bryant and colleagues (2003) note that children from higher quality preschool settings, as assessed by the ECERS, possess greater book knowledge and have higher print awareness scores on Zill and Resnick's (1998) Concepts About Print Assessment. Further, children's expressive language development, as measured by the Oral Expression Scale from the Oral and Written Language Scale (OWLS) (Carrow-Woolfolk, 1995), is shown to be related to the ECERS-R total score (Mashburn et al., 2008), and the Teaching and Interactions subscale (Burchinal et al., 2008). Burchinal and colleagues also discover a positive relationship between receptive language scores on the PPVT-III and scores on the ECERS-R Teaching and Interactions subscale. In summary, Clifford (2010) argues that higher quality environment, as measured by environment rating scales, is associated with children's development in three areas, including receptive language, print awareness, and book knowledge.

With respect to social outcomes, several important elements (e.g., social and behavioural development) can be predicted by using ERS (Sammons, 2010). Sammon and colleagues (2003b), who focus on four measures of social/behavioural development in Goodman's (1997) Strengths and Difficulties Questionnaire (i.e., self-regulation, hyperactivity, pro-social behaviour and anti-social behaviour), discover that the scores on the caring and emotional/relationship aspects of ECERS-R is associated with reduced hyperactivity and increased pro-social behaviour. Additionally, there is a negative relationship between the ECERS-R scores on the space and furnishings subscale, and children's anti-social behaviours (Sammons et al., 2003b). Scores on the language and reasoning subscale are also positively related to children's cooperation and conformity skills (Sammons et al., 2003b). Further, Montes and colleagues (2005) have found considerable effect sizes linking high total score on ECERS-R with the reduction of existing socio-emotional risk factors and prevention of the emergence of new socio-emotional risk factors. A recent study, that used the Teacher-Child Rating Scale (T-CRS 2.1; Hightower et al., 1986) to measure children's social competence scores, has demonstrated a positive relationship between the Teaching and Interactions factor of the ECERS-R and children's assertiveness, frustration tolerance, task orientation, and peer social skills (Burchinal et al., 2008).

The issue of high-quality ECEC provision and teacher policy

High-quality ECEC provision is recognised by international literature, and state and federal governments as a pivotal policy target measure for the foundation of formal education. As stated earlier in this review, such targets are based on the individual (e.g., cognitive, language, and social development) and societal benefits (e.g., human capital development/accumulation) that are generated by high-quality ECEC provision. Hence, governments, especially those in the developed world, with a major responsibility for the funding of ECEC, have been allocating more resources and devoting greater policy attention to the quality of ECEC (International Labour Organization, 2012; New Zealand House of Representatives, 2013; OECD, 2013).

Over the past two decades, governments have been active in designing policies and programs that aim to improve the quality of ECEC and ensure access, irrespective of socio-economic status, to high-quality classroom environments (Dowling & O'Malley, 2009; Pianta et al., 2005). Often, these legislations focus on the manipulation of structural features such as child-to-staff ratio, class size, and staff educational attainment (Phillipsen et al., 1997). Studies have been conducted to identify influential characteristics that promote the quality of ECEC. For example, the relationship between teacher's level and type of education with the quality in ECEC has prompted policy makers and ECEC providers to consider requiring centres to increase the proportion of qualified registered teachers (Dowling & O'Malley, 2009; Elliott, 2006; Mitchell, 2010). The specific rationale behind any ECEC teacher policy (e.g., "bachelor's degree policy") is that lead teachers with a bachelor's degree or higher in ECEC are expected to have higher-quality classrooms (Ackerman, 2005).

Teacher qualification as a predictor variable that influences the quality of ECEC

Teacher qualification has been identified by policy makers as one of the regulatable features and crucial variables that can predict quality in ECEC (Ackerman, 2005). It is, however, noted that the variable "teacher qualification" is not an intervention that was manipulated in the studies comprising the meta-analysis, but a characteristic of the teacher/care provider and a predictor variable in the studies.

Approved early childhood teaching qualifications vary between states and across countries. Based on current literature, teacher qualification can be categorised according to their level and type of education. Level of education can be referred to formal schooling in primary, secondary and tertiary systems. The tertiary system can be categorised as community college, junior college and university. The details of intervention may also include years of education, course credits in any specialized early childhood or child development-related education, and title of program. Further, differentiation of qualification is based on the type of program and the degree earned, such as associate degree (AA), Child Development Associate (CDA) Credential/AA/certificate in ECEC, bachelor's degree (BA), BA in ECEC or primary education, master's degree (MA), MA in ECEC or primary education, PhD, and EdD (professional doctorate). One should note, however, that some of the literature does not separate degrees of ECEC and that of other majors. MA and PhD are sometimes categorised as post-graduate education.

HOW TEACHER QUALIFICATIONS MAY INFLUENCE ECEC QUALITY

The underlying conceptual framework of this review is based on the notion that a strong ECEC knowledge base involves a set of professional competencies, abilities and specific teaching skills, which can lead to high-quality ECEC and positive child developmental outcomes (Bowman, Donovan, & Burns, 2001; Vartuli, 1999). Berk (1985) discovers that tertiary education (e.g., AA and BA) is associated with greater early childhood teaching skills. By comparing early childhood teachers with a high school diploma to those teachers who possess a college education of two years or more, Berk finds that teachers with an AA or BA, regardless of the specific major, were more responsive, encouraging and inspiring when communicating with young children. On the other hand, Snider and Fu (1990) suggest that teacher education with more ECEC content is essential to produce high-quality teaching skills. This idea is supported by other empirical evidence, which demonstrate that teachers' beliefs, knowledge, and actual implementation of developmentally appropriate practice are positively correlated with ECEC education and early childhood coursework (McMullen & Alat, 2002; Snider & Fu, 1990; Vartuli, 1999).

According to File and Gullo (2002), students from child development-related programs also have a stronger and more consistent preference towards developmentally appropriate practices than those in other education programs. Teachers with a BA or tertiary-level specialized ECEC/child development-related qualification often display more sensitive and less harsh and detached behaviours (Howes, Whitebook, & Phillips, 1992). Moreover, the literature suggests that more knowledge in ECEC encourages teachers to adopt developmentally appropriate practices (Vartuli, 1999), which are likely to facilitate supportive and nurturing interactions that are essential to high-quality ECEC (Ackerman, 2005; L. Dunn, 1993).

WHY IT IS IMPORTANT TO DO THE REVIEW

The extant literature

The extant literature, which includes primary studies and systematic reviews, has examined relationships between the common regulation or policy targets (e.g., the classroom features, staff characteristics, and administration) with the process quality of ECEC that is measured by standardised tools (e.g., ERS). Empirical evidence has been used to inform policy deliberation in this area (Vandell & Wolfe, 2000). Specifically, some large-scale studies of childcare quality were conducted, and researchers attempted to demonstrate relationships between program quality (as measured by the ECERS/ECERS-R/ITERS/ITERS-R) and variables such as staff stability and staff background characteristics (Huntsman, 2008; Phillipsen et al., 1997; Sakai et al., 2003).

Among the vast array of variables that claim to impact on the ECEC environment, high staff qualification has been identified as one of the strongest predictors of high ECERS/ITERS ratings (with a score of 5 or higher on a 7-point scale) (Burchinal, Cryer, Clifford, & Howes, 2002; Peisner-Feinberg, Burchinal, Howes, & Cryer, 1997; Phillipsen et al., 1997; Whitebook, 2003b; Whitebook, Sakai, Gerber, & Howes, 2001). By distinguishing bachelor's degrees and specialised child development-related education from other levels of education and training, previous literature

suggests that bachelor's degrees with specialised training in ECEC secure high quality childcare and education outcomes (Ackerman, 2005; Kelley & Camilli, 2007; Whitebook, 2003a).

There are, however, some studies that suggest no significant relationship exists between ECERS scores and the percentage of teaching staff with a bachelor's or advanced training in ECEC in a centre. For example, Phillips, Mekos, Scarr, McCartney, and Abbott–Shim (2000) discover that teacher education did not significantly affect classroom processes for infants and children in preschool rooms. Also, focusing on the school-related learning and social skills over the pre-Kindergarten year, Howes and colleagues (2008) argue that there is a mixed relationship between teacher qualification and classroom quality (as measured by ECERS-R). Early et al. (2007) examined the results of seven major studies regarding the association between teacher qualification and ECEC quality and children's academic outcomes. On the whole they find no association, indicating that relying on policies that focus only on teacher qualifications is not enough to improve and enhance ECEC quality.

Whitebook (2003b) conducts a systematic review regarding the educational level of teachers and how the level of qualification potentially affects the early childhood learning environment. Whitebook's main findings suggest that teacher qualifications matter in terms of high-quality ECEC provision, and that ECERS ratings are positively affected by higher qualifications. To further Whitebook's review, and allow for more meaningful comparisons between individual studies regarding the impact of teachers' qualification on the quality of the ECEC environment, we propose that the target population should be further disaggregated (e.g., sampling at the level of the classroom or at the level of the learning centre) and a common metric (e.g., effect size) for measuring outcomes be calculated.

A more recent meta-analysis was conducted by Kelley and Camilli (2007), who analyse the results of 32 studies (18 treatment-comparison studies and 14 studies with correlations between teacher education and outcomes). The authors examine the relationship between teacher qualifications and the ECEC environment. The authors aggregate four different constructs, including global classroom quality, teacher-child interactions, teacher pedagogical beliefs and knowledge, and classroom instructional activities, into a group of ECEC outcomes. Aggregating these constructs, however, makes it difficult to estimate the additional effect on a specific outcome (e.g., quality of learning environment) as a result of higher educational attainments by teachers (e.g., bachelor's degree or associate degree). In this review, we will not aggregate these constructs. Rather, we focus on quality in the early childhood environment and ratings in the ERS subscales to tease out the relationship between teacher qualifications and process quality in an ECEC setting. This specific information is critical for policy makers, as single measures inform many of their decisions.

Objectives

The objective of this systematic review is to synthesise the extant empirical evidence on the relationship of teacher qualifications to the quality of the early childhood learning environment. Specifically, this review will seek to answer the following question:

> Is there a relationship between the level and type of education of the lead teacher, and the quality of the early childhood learning environment, as measured by the Early Childhood Environment Rating Scale, the Infant Toddler Environment Rating Scale and their revised versions?

Methods

CRITERIA FOR CONSIDERING STUDIES FOR THIS REVIEW

The most common methods used in evaluating the impact of teachers' education level to improve the quality of the ECEC environment are: (1) non-randomized comparative design based on cross sectional survey and secondary data analysis; and (2) correlational designs that report a correlation between levels of education and ratings of classroom quality. Therefore, the following criteria will be used to identify studies for the review.

Types of study designs

A preliminary exploration of published and unpublished literature focusing on authors who have contributed to understanding ECEC quality (e.g., Whitebook, Clifford, and Howes) was conducted to determine the period of time that should be covered in this review. Eligible studies, therefore, are those studies (both comparative and correlational) that examine the relationship between teacher qualification and quality of the ECEC environment (as measured by ECERS/ECERS-R/ITERS/ITERS-R) from 1980 (this was when the ECERS was introduced) to 2014. To be eligible, a study must permit the identification of the education program (e.g., bachelor's degree) received by the lead teacher and provide a comparison between two or more groups of teachers with different educational qualifications. The details of the independent variable include year of education, course credits in ECEC and title of the program.

Since teachers' education qualification cannot be controlled by the researcher, the study designs that will be included in the review are comparative and correlational non-randomized studies. In comparative studies, a group of teachers with high school education is compared to at least one group of teachers with other qualifications (i.e., AA, CDA Certificate, BA, BA in ECEC or primary education, MA, MA in ECEC or primary education, PhD and EdD). In correlational studies, analyses are typically conducted to explore the relationship between level of teacher education and study outcomes, reporting a correlation (e.g., via point biserial or Pearson correlation coefficient) between levels of education and ratings of classroom quality as measured by ERS.

Types of participants

The sample under consideration is ECEC settings where education programs are delivered to children by lead teachers and their assistants. In this review, ECEC programs may be delivered in indoor settings in either a centre-based or home-based classroom environment. We initially

assumed that most studies would focus on centre-based classrooms, thus our primary population of interest is classrooms in ECEC centres that are affiliated with a state licensing agency. Specifically, participants in these ECEC settings are children from pre-kindergarteners and kindergarteners prior to elementary/primary school and centre-based providers.

Types of outcome measures

Eligible studies focus primarily on the quality of the early childhood learning environment as measured by the ECERS/ECERS-R/ITERS/ITERS-R. Seven categories of outcomes are identified according to the subscales and items of those tools, including, for example: Personal Care Routine; Space and Furnishings; Language-reasoning/Listening and Talking; Activities; Program Structure; Interaction; and Parents and Staff. Since some researchers may prefer using the two-factor scale (Activities/Materials and Language/Interactions) for follow-up analysis, these alternative outcomes are also included as applicable and possible.

Duration of follow-up

Multiple independent samples over time are not common for research in this area. Therefore, the literature is expected to be based on evaluation of outcomes in a single event. We did, however, find studies longitudinal in nature. That is, where a cohort is followed-up at subsequent time periods (e.g., one year after intervention, two years after intervention etc.) In such cases, we combined the longitudinal effects (i.e. correlation r) to produce an overall mean effect size (r) for this cohort. To do this, we produced a mean effect size (r) of the time points in those longitudinal studies. We did not treat effect sizes (r) of each time point as separate independent samples.

Types of settings

The review includes independent samples conducted in indoor centre-based settings that serve infants, toddlers, preschool and kindergarten children. Studies conducted in outdoor settings are not included in the review.

DATA COLLECTION AND ANALYSIS

Selection of studies

We began our search for relevant studies by conducting a manual search of key journals for the period 1980 (this was when the ECERS was introduced) to 2014 – examples include, Early Childhood Research Quarterly, Early Childhood Research and Practice, Contemporary Issues in Early Childhood, Child Development, Applied Developmental Science, and the Journal of Child Psychology and Psychiatry. We then scanned relevant review articles, consulted the bibliographies of articles that meet the selection criteria and asked key researchers for assistance in identifying other relevant studies.

We searched a number of electronic databases covering the years 1980 to 2014. Compound terms (e.g., teacher qualification; staff education) were considered as a single term and entered into searches in quotes (i.e., “teacher qualification”). These databases include: Academic Search

Premier; CBCA-Education; Cochrane Controlled Trial Register; Database of Abstracts of Reviews of Effectiveness (DARE); Dissertation Abstracts; EconLit; Education Full Text; Educational Resources Information Center (ERIC); Journal Storage Archive (JSTOR); Medline; Proquest Digital Dissertations; Proquest Direct; Project Muse; PsychInfo; Scopus; SocINDEX with Full Text; and SSRN eLibrary. Within each database, we searched using combinations of keywords from three categories:

1) Outcome: Scores/ratings OR Subscales OR Early Childhood Environment Rating Scale, ECERS/ECERS-R/Infant Toddler Environment Rating Scale, ITERS/ITERS-R

AND

2) Independent variable: Teacher's education OR Teacher's qualification OR Training OR Program OR Levels of qualification OR associate degree (AA) OR Child Development Associate (CDA) OR bachelor's degree (BA) OR BA in ECEC OR master's degree (MA) OR MA in ECEC OR PhD OR EdD

AND

3) Targeted population: Classrooms OR Centre-based (/Center-based) classrooms OR home-based classrooms OR Child care centres (/centers) OR Kindergartens OR Pre-schools OR Pre-K (/PreK) OR Pre-kindergartens (/Prekindergartens)

This strategy ensured that the database searched for the entire term rather than "teacher" AND "qualification", which would clearly produce very different results. In addition, search terms with multiple iterations from a base word stem (e.g., quality) are typed in as word* (e.g., quality*). This approach enabled the researcher to capture relevant literature with fewer searches. Where a database included the function to manipulate the search field option, researchers limited the keyword search to title, abstract, reference list, whole document or a combination of fields. Results from a series of pilots indicated that the search 'anywhere' in the document option produced more hits with a lower inclusion percentage than searches conducted on the abstract only or title, abstract and descriptors. A fundamental objective in this review is to develop a search strategy that can be replicated. Consequently, the focus is to utilize electronic databases/resources that can be generally accessed (i.e., not restricted material through an organization's intranet).

Further, it is also important to locate grey literature or material that is not formally published (e.g., working papers, unpublished dissertations, and reports including government, non-government and technical reports etc.). Websites of any relevant government, research institutes and early childhood associations were searched for published and unpublished studies. Some relevant websites include, hfrp.org, melycaba.com, tnstarquality.org, sped.dpi.wi.gov, ceelo.org, ric.ed.gov, and aplus-education.co.uk.

Efforts were made to track down any unpublished studies highlighted in the search. Authors of published articles were also contacted to ascertain if additional results relating to our key outcomes are available. This strategy aimed to identify other studies that may be overlooked in previous reviews, also to identify relevant data that may have been omitted in the publication of these

studies. In some cases, authors were able to provide more detailed information about their data (e.g., standard deviations) so that effect sizes can be calculated in a similar way.

Researchers recorded search information (date of search, database and search terms used), research information (design, method(s), agency, outcome, population) and reference information in a spreadsheet so that further interrogation of the data could be conducted at a future date.

Criteria for determination of independent findings

We are primarily interested in one outcome measure, process quality as measured by ECERS/ECERS-R/ITERS/ITERS-R. Therefore, the analysis is not affected by any bias that could result from data/outcome dependence. When a study provides data of the subscales, we extracted and coded the effect sizes for each subscale and conducted a separate analysis. Regarding the dependencies on the independent variables, multiple measures of teacher qualifications (e.g. credits of ECEC courses, years of education and level of education) have been used in some of the studies. For studies that report standardised mean differences and compare multiple treatment groups (e.g. teacher with Bachelor degree (as group 1), Associate degree (as group 2)) with a common control group (e.g. teachers with a High school diploma), we code these samples (e.g. Bachelor vs. High school (as sample 1) and Associate vs. High school (as sample 2) as dependent samples that require the use of a robust standard error model (see Section 3.2.4).

Details of study coding

Two of the authors of this review were responsible for reviewing and screening all titles and abstracts found through the search procedures. At the abstract screening stage, studies that were deemed as inappropriate would be those that do not involve the target participants (e.g., ECEC settings that serve all ages of pre-kindergarteners and kindergarteners prior to elementary/primary school) or are descriptive in nature where no relationship between teacher's qualification and ECEC quality is being measured.

The eligibility of relevant articles was screened based on the criteria for eligibility (see Section two "Criteria for Eligibility" in Appendix A). All eligible studies that meet the initial criteria were coded using an instrument developed by the authors to extract the specified information. The coding instrument includes items related to bibliographic information and source descriptors; analysis characteristics; sample characteristics; methodology/research design; and outcome data needed to calculate effect sizes.

All eligible studies were independently coded by two individuals to ensure the reliability of coding procedures and decisions. Coding schemes were refined to resolve any inter-rater discrepancies or differences. Discrepancies that remained unresolved were discussed and resolved by consensus with a third author.

The research team checked the references of eligible studies to determine if there were other studies of interest that were not retrieved in the original search. Any literature of interest was obtained and assessed for eligibility.

Statistical procedures and conventions

The correlation coefficient (r) was selected as the common effect size metric based on the frequency with which the different forms of data are present in the literature. We expected to locate a number of correlational studies, for which the appropriate effect size metric is the r. Where r's are used, the Fisher's r-to-z transformation is applied to all cases (Hedges & Olkin, 1985). We expected that for some studies (i.e., those with comparative designs), the scores on ERS and their subscales would be reported in natural units in the form of mean differences. When means and standard deviations were reported, we transformed these in to r using the method outlined in Lipsey and Wilson (2000). When means and standard deviations were not reported, effect sizes were calculated from other test statistics (e.g., t-tests, F-ratios) using formulas outlined by Lipsey and Wilson (2000). Where there were binary measures of the same outcomes, we used procedures recommended by Sanchez-Meca and colleagues (2003) to convert odds ratios (OR) to r.

Our selection of r as the most appropriate effect size for this study was because: (a) relative to alternative effect size metrics (e.g. Cohen's d and Hedge's g) r has a more direct relationship to other statistical concepts (e.g., statistical power and the general linear model); (b) it can be calculated across a wider range of study designs (correlation coefficients for dichotomous, ordinal, or continuous independent and dependent variables); and (c) it is base rate sensitive. That is, when base rates in a sample reflect those in the population, r provides a more realistic sense for how well one variable correlates to another (McGrath & Meyer, 2006).

We conducted 10 individual meta-analyses that examined the relationship between the level of teacher qualification and the overall ERS scores (main outcome), the seven subscales (e.g., Space and Furnishings, Activities, Interaction, Personal Care Routine, Parent and Staff, Program Structure and Space and Furnishing) and the two subscales (Language and Interactions and Provisions for Learning). Only analyses based on five or more observations were considered. Data synthesis was conducted using a specialized statistical meta-analysis software package (Comprehensive Meta-Analysis (CMA) version 2, 2006). For the calculation of meta-analytic mean effect size, the inverse variance weight (w) is calculated for each study using the appropriate standard error (Lipsey & Wilson, 2001). Regarding the method of variance estimate to compute the random effects mean, the method of moments using weighted least squares, which provides a closed solution for the random effects variance component (τ^2) (Raudenbush, 2009), is adopted by the CMA. The inverse variance was used so that effect sizes with greater precision (i.e., smaller variances) are weighted more heavily in the analysis. We employ a random effects model for pooling effects. An assumption is made that there are unexplained sources of heterogeneity across studies. The Q statistic, which was calculated in each fixed effect analysis, was used for the calculation of the τ^2. In addition, we employ the I^2 statistic (Higgins & Thompson, 2002) as an additional, albeit related, method of assessing heterogeneity for the fixed effect analysis. An I^2 value greater than 50% suggests moderate heterogeneity even if the Q statistic is not statistically significant. All effect sizes are calculated using a 95% confidence interval. For the robust variance estimation approach, unrestricted, intercept-only meta-regression models were run in Stata using a macro provided by Tanner-Smith and Tipton (2013). The estimated mean effect sizes (correlation in this review) with robust standard error take into account dependences in the data due to the use of a same control group for multiple treatment group contrasts.

Forest plots have been used to display the results from the effect sizes. The plot displays the effect size, confidence intervals and significance level with respect to each study.

Moderator analysis

We examined possible moderators of process quality in ECEC settings. We began by including characteristics of the study participants and settings such as: (1) ECEC settings (e.g., centre-based vs. home-based classrooms); (2) baseline characteristics of teacher (e.g., age, gender, race/ethnicity); (3) country in which the study was conducted; (4) duration of follow-up (e.g., 1 year vs. 6 months); (5) outcome measure (e.g., ECERS, ITERS, ECERS-R and ITERS-R); and (6) dominant ethnicity of student group.

Where data was available, we analysed these potential moderating factors using an analog to the Analysis of Variance (ANOVA) and weighted least squares regression when appropriate. The analog to the ANOVA is a method of testing the ability of a single categorical value to explain excess variability in a distribution of effect sizes (Lipsey & Wilson, 2000). CMA was used to conduct moderator analyses. Forest plots are used to graphically display these moderator effects.

Missing data

In the event that a study did not provide suitable data for calculation of effect sizes (e.g., means and standard deviations, valid Ns), we endeavoured to contact the author(s) of the primary studies in order to obtain the missing information.

Sensitivity analysis

A sensitivity analysis was conducted to test the robustness of the conclusions with respect to the quality of the available data and approaches to the analysis. Any differences in terms of the results are examined according to our choice of effect size metric. We inspect funnel plots, and utilise 'trim and fill' methods that estimate correlation by adjusting for the number and outcomes of missing studies. The publication bias of any over - or underrepresentation of particular findings on the basis of their statistical significance were examined by analysing the funnel plot. When we plotted the observed effect sizes against their corresponding sample sizes a more or less symmetric (inverted) funnel is expected in the absence of this form of publication bias. An asymmetrical funnel plot indicates that studies with non-significant findings are suppressed from the published literature. The regression test for funnel plot asymmetry examines whether an association between the effect sizes and the corresponding sample sizes is present in the data. Validity and reliability of the findings would be challenged if such an association is present suggesting that a publication bias does exist (Sterne & Egger, 2005). Regression tests for funnel plot asymmetry were conducted for each of the eight meta-analyses.

Treatment of qualitative research

Qualitative studies was not included in the current study. We do, however, acknowledge seminal pieces of research in our discussion.

Results

DESCRIPTION OF STUDIES

Results of the search

The systematic search identified 2023 unique sources (e.g., published or unpublished documents) on ECEC quality measured by ERS (i.e., ECERS, ECERS-R, ITERS, ITERS-R) and teacher qualifications. We were not able to obtain 12 sources to review for eligibility despite a number of different attempts (through our own efforts as well as through the employment of an information specialist). These sources tended to be university dissertations where the university and/or supervisors could not locate the author, and/or organizational reports. Several of these missing sources had ambiguous citations and may have been unlocatable because they were incorrectly documented in the online database. Of the 2011 sources we obtained, 130 studies were comparative studies and correlational studies, which reported on the relationship between teacher qualifications and ECEC quality, while 1867 of the sources did not report an independent sample on the relationship between teacher qualifications and ERS-measured ECEC quality and 14 of the sources were literature reviews, theoretical articles, or qualitative studies. Of the 130 quantitative studies reporting on a relationship between the two constructs (i.e., ECEC quality and teacher qualifications), 62 were excluded from the review because they had no comparative data and did not report on an outcome of interest. The remaining 68 studies that contained comparative information were further screened for suitability for meta-analysis. Of these, 20 studies did not provide enough data for the calculation of effect size (r). These were excluded. Two additional studies were identified as outliers and subsequently excluded from the analysis. The final set of 46 studies eligible for the meta-analysis contained 58 independent samples. All studies identified were subjected to scrutiny using the coding sheet (see Appendix 1). Table 1 displays the attrition of publications. Of the 46 included studies, 9 were theses (PhD =8; MA = 1). In the analysis of the overall ERS rating, 8 out of 44 were theses. In the analysis of the 7 factor ratings, 3 out of 14 were theses. In the analysis of the 2 factor subscales ratings, 1 out of 10 was a thesis.

Table 1: Attrition of publications

Stage of Review	*N*
Unique sources	2023
Not obtained sources	12
Obtained sources	**2011**
Inventory of interventions	
Not reporting a relationship	1867
Reporting a relationship	**144**
Narrative review	
Not a quantitative sample	14
Quantitative samples	**130**
Meta-analysis	
Not eligible for meta-analysis	62
Studies eligible for meta-analysis	**48**
Independent samples eligible for meta-analysis	82

Included studies

The 48 eligible studies differed according to their classification of teacher qualifications and type of studies, target population and use of measurement tool and outcome. The following section describes these differences, which are summarized in Table 2.

Table 2: Study characteristics

Author	*Sample size (N)*	Type	Scale/comparison (values represent the range of educational level as an ordinal or ratio variable)	Outcomes	No. of independent samples (n)
Antle et al (2008)	47	2	1-4 (1=HS/GED, 4=MA/specialist degree)	Overall ECEC quality	1
Antle et al (2008)	44	2	1-4 (1=HS/GED, 4=MA/specialist degree)	Overall ECEC quality	1
August (2008)	43, 54	1	>HS vs ≤HS	Overall ECEC quality; SF; ACT; INT	4
Bolger & Scarr (1995)	122	2	Years of education	Overall ECEC quality	1
Brooks-Gunn et al (2011)	64	2	Years of education	Overall ECEC quality; SF; ACT; INT; LR; PS; PCR; STR; LI; PL	10
Brown (2005)	15	2	1-4 (1=HS/GED, 4=MA)	Overall ECEC quality; SF; ACT; INT; LR; PS; PCR; STR	8
Bryant et al (2009)	114	2	Years of education	Overall ECEC quality; SF; ACT; INT; LR; PCR; STR; LI; PL	9
Burchinal et al (2000)	51-86	2	Years of education	Overall ECEC quality	1

Author	*Sample size (N)*	Type	Scale/comparison (values represent the range of educational level as an ordinal or ratio variable)	Outcomes	No. of independent samples (n)
Cassidy et al (2003)	1313	2	1-12 (1<HS, 12=graduate degree)	Overall ECEC quality; LI; PL	3
Couse (2001)	12	2	1-4 (1=HS,4=MA)	Overall ECEC quality	1
Cryer et al (1999)	288	2	Not specified	Overall ECEC quality	1
Cryer et al (1999)	82	2	Not specified	Overall ECEC quality	1
Cryer et al (1999)	80	2	Not specified	Overall ECEC quality	1
Cryer et al (1999)	55	2	Not specified	Overall ECEC quality	1
D'Amour (2008)	65	2	1-3(1≤HS,3≥MA)	Overall ECEC quality; SF; ACT; INT; LR; PS; PCR; STR	8
D'Amour (2008)	106	2	1-3(1≤HS,3≥MA)	Overall ECEC quality; SF; ACT; INT; LR; PS; PCR; STR	8
Dennis & O'Connor (2013)	37	2	Years of education	Overall ECEC quality	1
Dove (2003)	8	2	1-4 (1=HS, 4=BA)	SF; ACT;	4

Author	*Sample size (N)*	Type	Scale/comparison (values represent the range of educational level as an ordinal or ratio variable)	Outcomes	No. of independent samples (n)
				INT; LR	
Dunn (1993)	30	2	0-6(0<HS,6=doctoral degree)	Overall ECEC quality	1
Early et al (2001)	238-677	2	Years of education	Overall ECEC quality; SF; ACT; INT; LR; PCR; STR; LI; PL	9
Epstein (1999)	109	2	1-5 (1<HS,5=BA)	Overall ECEC quality	1
Epstein (1999)	110	2	1-5 (1<HS,5=BA)	Overall ECEC quality	1
Epstein (1999)	72	2	1-5 (1<HS,5=BA)	Overall ECEC quality	1
Henry et al (2003)	97	2	1-4 (1<AA, 4≥BA)	Overall ECEC quality	1
Hestenes et al (2015)	101	2	1-4 (1=HS, 4≥BA)	Overall ECEC quality	1
Ho (2011)	231	2	1-10 (1≤8th grade, 10≥MA)	Overall ECEC quality	1
Ho (2011)	97	2	1-10 (1≤8th grade, 10≥MA)	Overall ECEC quality	1
Ho (2011)	141	2	1-10 (1≤8th grade, 10≥MA)	Overall ECEC quality	1
Ho (2011)	201	2	1-10 (1≤8th grade, 10≥MA)	Overall ECEC quality	1

Author	*Sample size (N)*	Type	Scale/comparison (values represent the range of educational level as an ordinal or ratio variable)	Outcomes	No. of independent samples (n)
Holloway et al (2001)	88	2	1-5 (1<HS, 5≥MA)	Overall ECEC quality	1
Honeycutt (2008)	61, 45	1	AA vs <CDA	Overall ECEC quality	1
Honeycutt (2008)	368, 45	1	BA vs <CDA	Overall ECEC quality	1
Honeycutt (2008)	88, 45	1	CDA vs <CDA	Overall ECEC quality	1
Honeycutt (2008)	84, 45	1	MA vs <CDA	Overall ECEC quality	1
Jardine-Ledet (1999)	11	2	1-4 (1=HS, 4=graduate degree)	Overall ECEC quality	1
Jardine-Ledet (1999)	40	2	1-4 (1=HS, 4=graduate degree)	Overall ECEC quality	1
Kitner-Duffy (2011)	41	2	Not specified	LI; PL	2
Kontos & Fiene (1987)	10	2	0-1 (0<BA, 1=BA)	Overall ECEC quality	1
Kwon & Han (2007)	54	2	1-4 (1=HS, 4=graduate school)	Overall ECEC quality	1
Laferney (2006)	269	2	1-8 (1<HS, 8=Post-master's work)	Overall ECEC quality	1
La Paro (2009)	610	2	0-1 (0<BA, 1≥BA & early education credentials)	LI; PL	2
Lee (1994)	34	2	Percentages of teachers with BA or beyond	Overall ECEC quality	1

Author	*Sample size (N)*	Type	Scale/comparison (values represent the range of educational level as an ordinal or ratio variable)	Outcomes	No. of independent samples (n)
Mashburn (2004)	114	2	0-1 (0<BA, 1≥BA)	Overall ECEC quality; SF; ACT; INT; LR; PCR; STR	7
Mashburn et al (2008)	671	2	0-1 (0<BA, 1≥BA)	Overall ECEC quality	1
Melhuish et al (2010)	99, 114	1	BA vs <BA	Overall ECEC quality; SF; ACT; INT; LR; PS; PCR; STR	8
Peisner-Feinberg et al (2008)	151	2	1-3 (1≤1-year degree, 3=BA)	Overall ECEC quality; LI; PL	3
Phillips et al (2001)	98	2	1-9 (1<HS/GED, 9=Ed.D./Ph.D./other post-master's degree)	Overall ECEC quality	1
Phillips et al (2001)	106	2	1-9 (1<HS/GED, 9=Ed.D./Ph.D./other post-master's degree)	Overall ECEC quality	1
Phillips et al (2001)	112	2	1-9 (1<HS/GED, 9=Ed.D./Ph.D./other post-master's degree)	Overall ECEC quality	1
Rous et al (2008)	174	2	1-9 (1≤HS, 9=doctorate)	Overall ECEC quality; SF; ACT; INT; LR; PS; PCR;	8

Author	*Sample size (N)*	Type	Scale/comparison (values represent the range of educational level as an ordinal or ratio variable)	Outcomes	No. of independent samples (n)
				STR	
Sandstrom (2012)	25	2	Not specified	Overall ECEC quality; SF; ACT; INT; LR; PS; PCR; STR	8
Selden & Sowa (2004)	22	2	1-7 (1=Some high school, 7=PhD)	Overall ECEC quality	1
Setodji et al (2012)	307	2	0-1 (0<BA,1≥BA)	Overall ECEC quality	1
Smith (2005)	100	2	Percentages of teachers with AA or beyond	Overall ECEC quality	1
Stein (2010)	343	2	Years of education	Overall ECEC quality	1
Stein-Balock (2007)	110	2	Years of education	Overall ECEC quality; LI	2
Thompson (1992)	66	2	1-10 (1<HS/GED, 10=doctorate degree)	Overall ECEC quality	1
Tout & Sherman (2005)	154	2	Percentages of teachers with CDA or beyond	Overall ECEC quality	1
Vermeer et al (2008)	42	2	Not specified	Overall ECEC quality	1
Vu et al (2008)	LI: 6, 10 PL: 10, 14	1	Head Start and General Childcare (BA or Higher) vs No teacher permit*	LI; PL	2

Author	*Sample size (N)*	Type	Scale/comparison (values represent the range of educational level as an ordinal or ratio variable)	Outcomes	No. of independent samples (n)
Vu et al (2008)	LI: 27, 10 PL: 17, 14	1	Head Start and General Childcare (with Master teacher permit) vs No teacher permit*	LI; PL	2
Vu et al (2008)	LI: 13, 10 PL: 7, 14	1	Head Start and General Childcare (with teacher permit) vs No teacher permit*	LI; PL	2
Vu et al (2008)	LI: 7, 10 PL: 13, 14	1	Private_non-profit (BA or Higher) vs No teacher permit*	LI; PL	2
Vu et al (2008)	LI: 17, 10 PL: 27, 14	1	Private_non-profit (with Master teacher permit) vs No teacher permit*	LI; PL	2
Vu et al (2008)	LI: 10, 10 PL: 6, 14	1	Private_non-profit (with teacher permit) vs No teacher permit*	LI; PL	2
Vu et al (2008)	LI: 34, 24 PL: 25, 24	1	School District and State Preschool (BA or Higher) vs No teacher permit*	LI; PL	2
Vu et al (2008)	LI: 55, 24 PL: 55, 24	1	School District and State Preschool (with Master teaching permit) vs No teacher permit*	LI; PL	2
Vu et al (2008)	LI: 25, 24 PL: 34, 24	1	School District and State Preschool (with teacher permit) vs No teacher permit*	LI; PL	2
Weinraub et al (2005)	35	2	0-1 (0≤HS, 1>HS)	Overall ECEC quality	1
Wheeler (2006)	8	2	1-3 (1=CDA, 3=MA)	Overall ECEC quality; SF; ACT; INT; LR; PS; PCR;	8

Author	*Sample size (N)*	Type	Scale/comparison (values represent the range of educational level as an ordinal or ratio variable)	Outcomes	No. of independent samples (n)
				STR	
Whitebook et al (2004)	42	2	Percentages of teachers with BA or beyond	Overall ECEC quality; ACT; INT; LR; STR	6
Zill et al (2001)	40	2	1-5 (1=HS/GED, 5=graduate degree)	Overall ECEC quality; SF; LR; PCR	4

*Qualifications required for a teacher permit: (a) 24 units CD/ECE coursework, 16 units GE coursework, 6 units in area of specialization, 2 units adult supervision coursework, 350 days teaching experience; (b) BA or higher, 12 units CD/ECE coursework, 3 units supervised field experience; (c) 24 units CD/ECE coursework, 16 units GE coursework, 175 days teaching experience; (d) AA in CD/ECE, 3 units supervised field experience

Note: SF=space and furnishings; PCR=personal care routines; LR=language and reasoning; ACT=activities; INT=interaction; STR=programme structure; PS=parents and staff; LI=language and interaction; PL=provisions for learning; *N* refers to the number of classrooms.

Classification of teacher qualifications and type of studies

The 48 eligible studies were of two distinct types: (Type 1) between-group comparison studies; and (Type 2) correlational studies. Type 1 studies provide ratings of ECEC quality associated with two or more categories of teacher education (i.e., high school, some college, associate's degree and bachelor's degree). Type 2 studies reported correlations between teacher education (typically reported in years) and outcomes, and did not allow for comparisons across different level of qualification. A study may report the subscale ratings separately and contain multiple independent samples (e.g. the same analysis may be conducted in multiple states/countries). In the current review, there were three Type 1 studies (15 independent samples) and 45 Type 2 studies (67 independent samples). It is noted that ratings of each subscale are analyzed separately in the review. In other words, we conduct separate meta-analysis for each subscale ratings and consider each sample to be independent.

Table 2 presents study characteristics of the three Type 1 studies employing between-group comparisons. For each independent sample in studies which compare the classroom quality of

teachers with different educational levels, the "treatment" group was always the group of teachers with a higher education attainment, in most cases, a bachelor's degree. For a study that compared a group of teachers with a bachelor's degree and a group of teachers with an associate degree, the treatment group is teachers with a bachelor's degree and the comparison group is teachers with an associate degree. Since high school diploma is the lowest possible level of education, teachers with this qualification always appear as a comparison group. In some studies, two or more levels of qualification might be combined under a broad category to form a composite qualification. For example, combining an associate degree with some college qualification (e.g., certificate) to form the composite category 'AA/some college', or combining a bachelor's degree with other post-graduate qualifications (e.g. masters or PhD) to form 'BA or above'.

Regarding the 45 Type 2 studies (i.e., correlational analysis) that explore the relationship between teacher qualifications and ECEC quality, the most common correlation reported was between level of education and ERS ratings of classroom quality. Level of education was treated as a continuous variable, but there was variation in how the scale was defined. For these studies, correlations were commonly reported in the form of a Pearson correlation coefficient (r). Table 2 reports the scale of the variable used to measure teacher education. For example, 0-7 indicates that the study measured teacher education using a seven-level classification. Years of education may also be used when teacher qualification was measured as total years of education.

Measurement tool and target population

The studies differed with respect to the measurement tool they used, often according to the age of the target population. Typically, ECERS and its revised version (ECERS-R) were used to measure ECEC quality of pre-school classroom with children aged 2 through 5 years. ITERS and its revised version (ITERS-R) were used to measure ECEC qualities of infant and toddler classroom with children from birth to 2 ½ years of age. In the current review, 61 independent samples used ECERS and ECERS-R, 12 independent samples used ITERS and ITERS-R and nine independent samples used a composite scale. Where possible, we have tried to identify any effects of the measurement tool through moderator analysis. Data reporting styles for each study are included in Table 2.

Outcomes

All of the eligible studies measured ECEC qualities based on one of the ERS. Most of them only reported the rating of the overall quality of their targeted ECEC settings, while some of them used, or reported, ratings of one or more of the ERS subscales. There were also variations with respect to their classification of subscales. Some studies (n=13) reported results based on the authors' original classification of 7 subscales, while other studies (n=9) reported results based on the two-factor classification.

Sensitivity analyses

To examine the impact of our analysis decisions on the results, we conducted a series of sensitivity analyses. We used a funnel plot to inspect possible sources of bias in the data, including publication bias and small-study effects. We used a "trim and fill" method (Duval & Tweedie, 2004) to test for publication bias. We tested for small-study effects using an approach proposed by Egger (1997).

Publication and small-study bias
We attempted to minimize publication bias by including grey literature in our search and document retrieval efforts.

Detecting publication bias
We ran a series of diagnostic tests for publication bias in CMA, based on the effects for the targeted outcomes from each study. The tests were run separately for each outcome.

Overall ECEC quality

Figure 1 presents the funnel plot for the main outcome (overall quality) from the 44 eligible studies. Each point on the figure represents one outcome for one independent sample. Studies with a larger sample size have a smaller standard error, and such points are situated closer to the top of the funnel. Smaller studies are located toward the bottom of the funnel. We expect that large studies with positive and negative effects will be published, so there should be symmetry across both sides of the funnel at the top of the graph. However, in the presence of publication or small-study bias, we expect that the points appearing at the bottom of the funnel will be clustered on the positive effect side of the graph.

Figure 1: Funnel plot of standard error by Fisher's Z: Overall ECEC quality measured by ERS

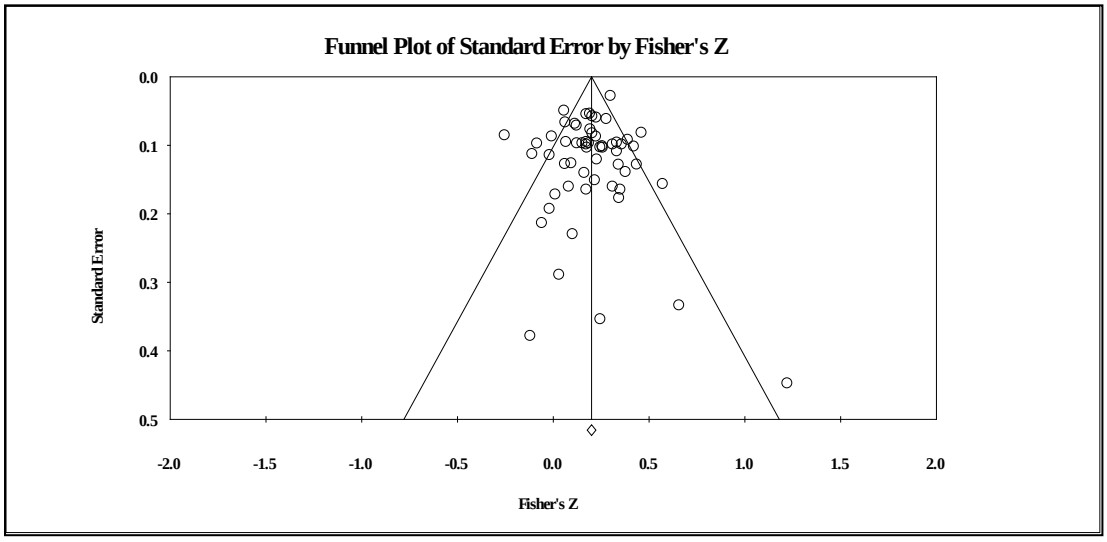


Figure 1 shows a fairly symmetrical distribution with no extreme results. A diagnostic test (Egger's test run in CMA as demonstrated in Rothstein, Sutton, & Borenstein (2005)) was conducted for further examination of potential publication or small-study bias. The above result demonstrates no evidence of publication or small-study bias, t (56) = 0.468, p = 0.641 (2-tailed).

Missing study bias

We attempted to adjust the correlation coefficient (r) for additional publication bias using Duval and Tweedie's (2004) "trim and fill" method. This method uses an iterative procedure to determine where missing studies are likely to fall on the funnel plot, and then adds them to the analysis to determine the overall correlation with the imputed studies included. In our analysis, this method looked for missing studies on the left side of the mean effect using a random effects model. The results suggested that 1 study is potentially missing, and the inclusion of this study would result in an insignificant drop of the effect size - from 0.197 to 0.196. In addition, the results of the Classic Fail-Safe N (run in CMA) indicated that there would need to be an additional 3301 studies to nullify the effect of the meta-analysis. We also conducted Orwin's Fail-Safe N (criterion for 'trivial' correlation = 0.100; mean correlation in missing studies = 0.000 – we selected the criterion as o.100 because a correlation <0.100 would be considered as no to very small). Results indicate the number of missing studies needed to bring the correlation under o.100 would be 58. Further, we also feel confident that our systematic search captured a large number of unpublished and small studies, since a substantial proportion of our eligible studies are dissertations and unpublished reports. Therefore, the effect of publication bias introduced across the studies is likely to be trivial.

Subscale level ECEC quality (7-factor classification)

Figure 2 presents the funnel plot for the seven subscale outcomes from the 14 eligible studies. Each point on the figure represents one outcome for one independent sample.

Figure 2: Funnel plot of standard error by correlation coefficient: Seven-factor subscale ECEC quality measured by ERS

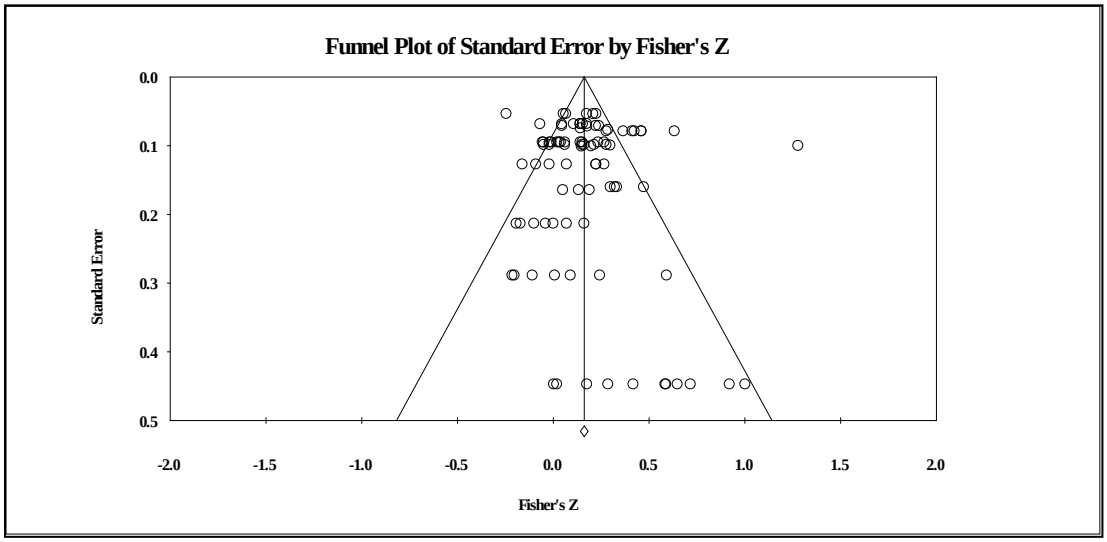


Figure 2 shows a somewhat asymmetrical distribution with a number of extreme positive results on the right. A diagnostic test (Egger's test run in CMA as demonstrated in Rothstein, Sutton, & Borenstein (2005)) was conducted for further examination of potential publication or small-study bias. The result demonstrated that there was no evidence of publication or small-study bias, t(86) = 0.658, p = .513 (2-tailed).

Missing study bias

We attempted to adjust the correlation coefficient (r) for additional publication bias using Duval and Tweedie's (2004) "trim and fill" method. This method uses an iterative procedure to determine where missing studies are likely to fall on the funnel plot, and then adds them to the analysis to determine the overall correlation with the imputed studies included. In our analysis, this method looked for missing studies on the left side of the mean correlation using a random effects model. The results suggested that no studies were potentially missing in the analysis, and therefore the imputed point estimate was the same as that obtained using only observed studies (r = 0.167, CI 0.120, 0.213). In addition, the results of the Classic Fail-Safe N (run in CMA) indicated that there would need to be an additional 4,716 studies to nullify the effect of the meta-analysis. We also conducted Orwin's Fail-Safe N (criterion for 'trivial' correlation = 0.100; mean correlation in missing studies = 0.000). Results indicate that the number of missing studies needed to bring the correlation under o.100 would be 54.

Subscale level ECEC quality (2-factor classification)

Figure 3 presents the funnel plot for the two subscale outcomes from the 10 eligible studies. Each point on the figure represents one outcome for one independent sample.

Figure 3: Funnel plot of standard error by correlation coefficient: Two-factor subscale ECEC quality measured by ERS

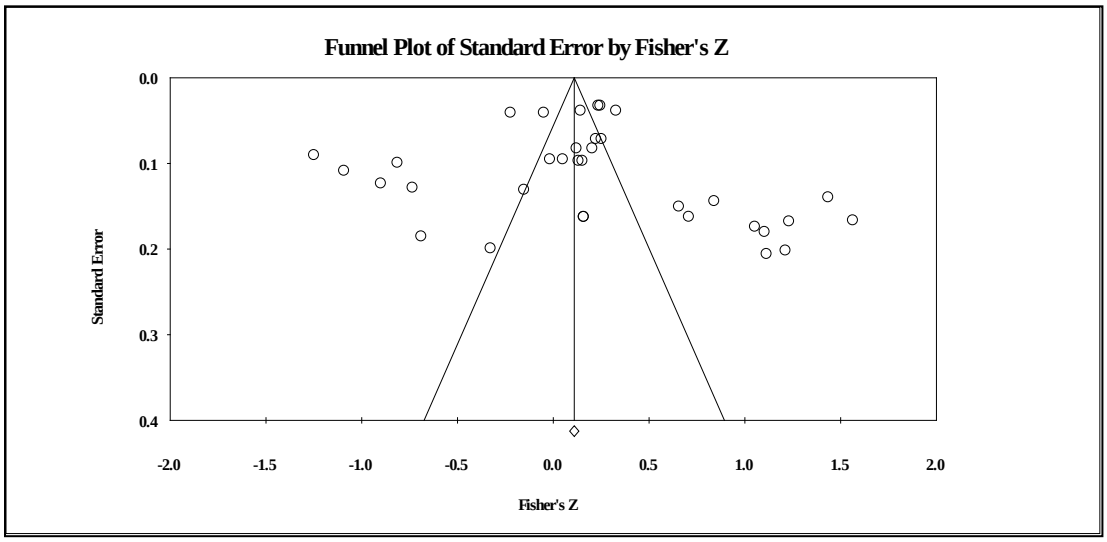


Figure 3 shows a fairly symmetrical distribution with a number of extreme results on the left and the right. A diagnostic test (Egger's test run in CMA as demonstrated in Rothstein, Sutton, & Borenstein (2005)) was conducted for further examination of potential publication or small-study

bias. The result demonstrated that there was no evidence of publication or small-study bias, t(32) = .334, p = .74 (2-tailed).

Missing study bias

We attempted to adjust the correlation coefficient (r) for further publication bias using Duval and Tweedie's (2004) "trim and fill" method. This method uses an iterative procedure to determine where missing studies are likely to fall on the funnel plot, and then adds them to the analysis to determine the overall correlation with the imputed studies included. In our analysis, this method looked for missing studies on the left side of the mean correlation using a random effects model. The results suggested that only one study is potentially missing in the analysis, and its inclusion would lead to the random effects estimate decreasing from 0.189 (CI 0.150, 0.228) to 0.187 (CI 0.147, 0.226). In addition, the results of the Classic Fail-Safe N (run in CMA) indicated that there would need to be an additional 3,301 studies to nullify the effect of the meta-analysis.

SYNTHESIS OF RESULTS

Overall ECEC quality

There were 44 eligible studies that measured overall ECEC quality as an outcome and provided a correlation for teacher or caregiver qualifications. Of the 58 eligible independent samples, 50 had a positive correlation (r greater than zero), indicating that for these independent samples higher teacher or caregiver qualifications were positively correlated with an overall ERS rating. Of the 50 independent samples with r greater than zero, 26 were statistically significant (p<.05).

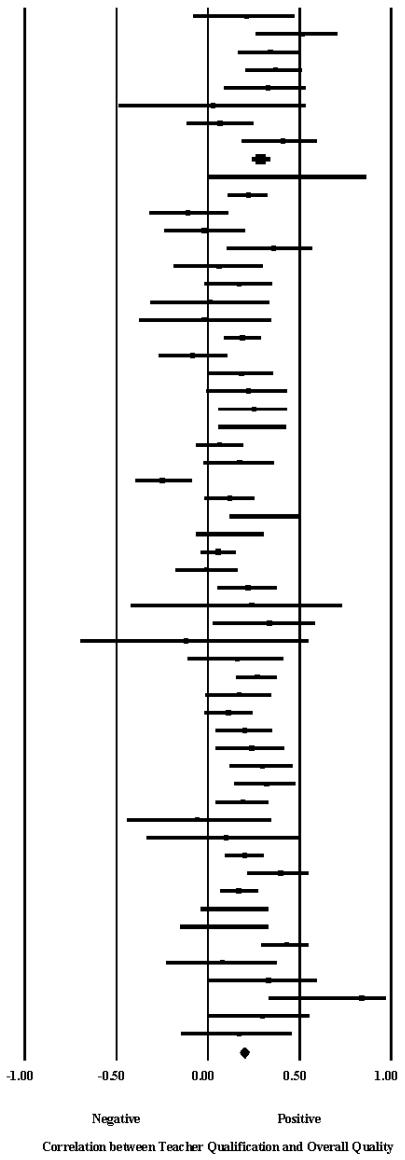
Overall, the education level of the teachers or caregivers is positively correlated (small effect) to overall ECEC qualities measured by ERS. The estimated mean correlation (r) with robust standard error, assuming ρ = .80, was 0.198 (confidence limits 0.133, 0.263) (see Figure 4). The interpretation of correlation produced are: <.10 = trivial; .10 to .30 = small to medium; .30 to .50 = medium to large; and >.50 = large to very large (Cohen, 1988).

According to Hedges and Hedberg (2007), educational researchers have indicated that effect sizes of 0.20 or smaller are of policy interest when they are based on measures of academic achievement or quality of education program. This result also highlights the practical significance of having ECEC program providers with higher education as lower effect sizes are usually attained on assessment of environmental settings based on ERS.

The differences between studies contributed significantly to the variation in effect sizes. The 58 independent samples were significantly heterogeneous according to the Q statistic (Q (57) = 143.85, p < .001). The large I^2 statistic indicated that 60% of the variance in the correlation coefficient may be a result of study factors (I^2 = 60.377, τ^2 = .012, se = .005).

Figure 4: Forest plot of correlation coefficients: Overall ECEC quality measured by ERS

Study name	Subgroup within study	Statistics for each study				
		Correlation	Lower limit	Upper limit	Z-Value	p-Value
Antle et al (2008)	Infant and Toddler	0.213	-0.079	0.471	1.435	0.151
Antle et al (2008)	Preschool	0.516	0.259	0.705	3.655	0.000
August (2008)		0.342	0.162	0.500	3.620	0.000
Bolger & Scarr (1995)		0.370	0.206	0.514	4.237	0.000
Brooks-Gunn et al (2011)		0.328	0.089	0.531	2.660	0.008
Brown (2005)		0.030	-0.490	0.534	0.104	0.917
Bryant et al (2009)		0.066	-0.119	0.247	0.696	0.486
Burchinal et al (2000)		0.410	0.183	0.595	3.406	0.001
Cassidy et al (2003)		0.290	0.240	0.339	10.806	0.000
Couse (2001)		0.576	0.003	0.864	1.969	0.049
Cryer et al (1999)	American	0.220	0.107	0.327	3.776	0.000
Cryer et al (1999)	German	-0.110	-0.319	0.110	-0.982	0.326
Cryer et al (1999)	Portuguese	-0.020	-0.239	0.201	-0.176	0.861
Cryer et al (1999)	Spanish	0.360	0.105	0.571	2.718	0.007
D'Amour (2008)	Infant and Toddler	0.060	-0.187	0.300	0.473	0.636
D'Amour (2008)	Preschool	0.170	-0.021	0.349	1.742	0.081
Dennis & O'Connor (2013)		0.010	-0.315	0.333	0.058	0.954
Dunn (1993)		-0.020	-0.378	0.343	-0.104	0.917
Early et al (2013)	Pre-K	0.188	0.085	0.287	3.560	0.000
Epstein (1999)	Head Start	-0.085	-0.269	0.105	-0.877	0.380
Epstein (1999)	Nonprofit	0.183	-0.004	0.358	1.915	0.056
Epstein (1999)	Public School	0.223	-0.009	0.432	1.884	0.060
Henry et al (2003)		0.252	0.055	0.430	2.497	0.013
Hestenes et al (2015)		0.250	0.057	0.425	2.528	0.011
Ho (2011)		0.062	-0.068	0.190	0.937	0.349
Ho (2011)	English-speaking Hispanic	0.173	-0.027	0.360	1.694	0.090
Ho (2011)	Spanish-speaking Hispanic	-0.249	-0.398	-0.087	-2.988	0.003
Ho (2011)	White	0.121	-0.018	0.255	1.711	0.087
Holloway et al (2001)		0.320	0.118	0.496	3.058	0.002
Honeycutt (2008)	Associate Degree (AA)	0.122	-0.067	0.303	1.269	0.204
Honeycutt (2008)	Bachelor Degree	0.056	-0.040	0.151	1.142	0.254
Honeycutt (2008)	Certificate (CDA)	-0.008	-0.176	0.160	-0.097	0.923
Honeycutt (2008)	Masters Degree	0.219	0.052	0.374	2.557	0.011
Jardine-Ledet (1999)	Infant & Toddler	0.240	-0.420	0.734	0.693	0.489
Jardine-Ledet (1999)	Preschool	0.336	0.027	0.586	2.125	0.034
Kontos & Fiene (1987)		-0.120	-0.697	0.551	-0.319	0.750
Kwon & Han (2007)		0.160	-0.113	0.410	1.153	0.249
Laferney (2006)		0.270	0.155	0.377	4.516	0.000
Mashburn (2004)		0.170	-0.014	0.343	1.809	0.071
Melhuish et al (2010)		0.112	-0.021	0.242	1.651	0.099
Peisner-Feinberg et al (2008)		0.200	0.042	0.349	2.466	0.014
Phillips et al (2001)	Infant	0.240	0.044	0.418	2.386	0.017
Phillips et al (2001)	Preschool	0.300	0.116	0.464	3.141	0.002
Phillips et al (2001)	Toddler	0.320	0.143	0.477	3.462	0.001
Rous et al (2008)		0.190	0.042	0.329	2.515	0.012
Sandstrom (2012)		-0.060	-0.445	0.343	-0.282	0.778
Selden & Sowa (2004)		0.100	-0.336	0.501	0.437	0.662
Setodji et al (2012)		0.200	0.090	0.305	3.535	0.000
Smith (2005)		0.397	0.218	0.550	4.137	0.000
Stein (2010)		0.170	0.065	0.271	3.165	0.002
Stein-Balock (2007)		0.150	-0.038	0.328	1.563	0.118
Thompson (1992)		0.093	-0.152	0.328	0.740	0.459
Tout & Sherman (2005)		0.430	0.292	0.551	5.651	0.000
Vermeer et al (2008)		0.080	-0.230	0.375	0.501	0.617
Weinraub et al (2005)		0.330	-0.004	0.598	1.939	0.052
Wheeler (2006)		0.840	0.332	0.970	2.731	0.006
Whitebook et al (2004)		0.300	-0.004	0.553	1.933	0.053
Zill et al (2001)		0.170	-0.149	0.457	1.044	0.296
Robust standard error model		0.198	0.133	0.263	6.000	0.000



Moderator analyses conducted in CMA showed a significant variation in correlations between studies according to the ethnicity of the dominant students and the different ERS used (see Table 3). The moderator analysis of ethnicity produced indicates that, overall, ECEC quality is positively correlated with teachers' qualification except for the Latino subgroup. The moderator analysis of scale used confirmed that the available standardised ERS measurements used across studies do not contribute to any significant difference about the relationship between teacher qualifications and ECEC quality. Some potential moderators such as socioeconomic status were not included due to the limited specification of such characteristics of the ECEC centres or target population among existing studies.

Table 3: Relationship of study characteristics to effect sizes (correlation coefficient) for the outcome of overall ECEC quality measured by ERS (k=58)

Study Characteristic	k	*r*	Lower limit	Upper limit	Z-value	P-value	Q (df)	P
Ethnicity (dominant group)							11.96 (5)	.035
African American	7	.147	.042	.249	2.736	.006		
Asian	1	.160	-0.165	.454	.965	.335		
Belizean	1	.342	.093	.551	2.654	.008		
Caucasian	22	.175	.117	.231	5.883	.000		
Latino	2	-.061	-.238	.120	-.657	.511		
Not specified*	25	.229	.175	.281	8.161	.000		
Scale used							9.706 (5)	.084
Composite scale	5	.317	.175	.447	4.254	.000		
ECERS	20	.125	.059	.190	3.716	.000		
ECERS-R	27	.204	.150	.256	7.338	.000		
ECERS-R (Modified)	1	.342	.076	.563	2.488	.013		
ITERS	3	.279	.094	.445	2.923	.003		
ITERS-R	2	.129	-.110	.353	1.060	.289		

Subscale outcomes

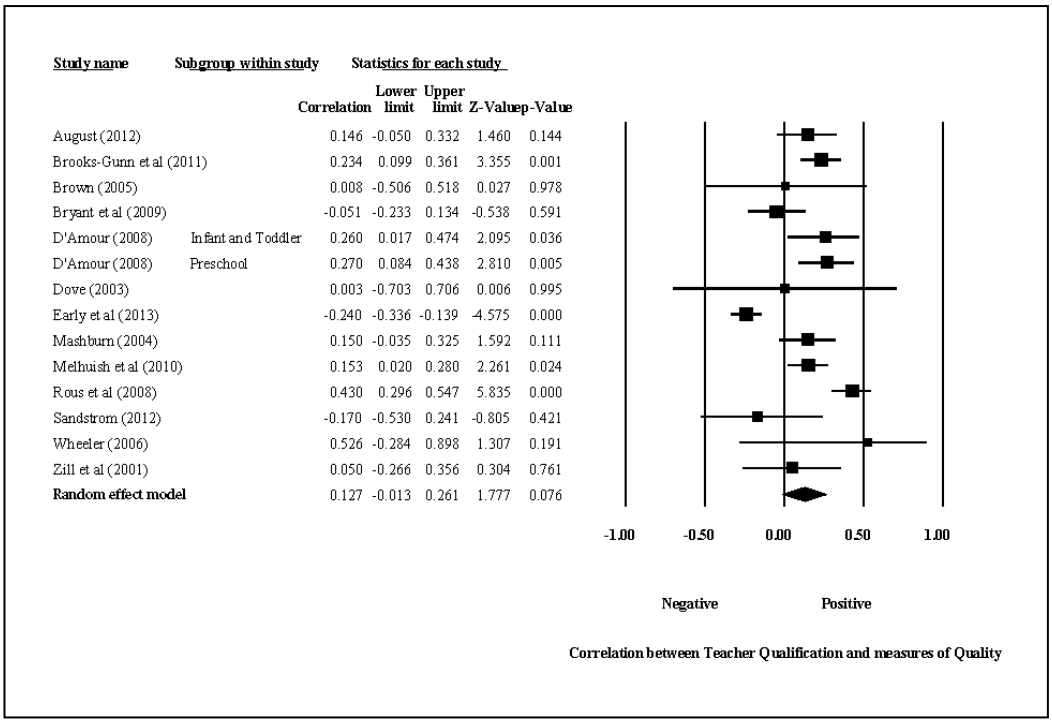
Space and furnishings

Thirteen studies provided outcome data on space and furnishings, giving 14 independent correlations overall. Figure 5 summarizes the 14 independent samples included in the meta-analysis on space and furnishings (note that one study contained two samples). Eleven of the fourteen independent samples had a positive correlation (*r* greater than zero), indicating that for most included studies, higher teacher qualifications was correlated with higher ratings on space and furnishings. Of the 11 independent samples with a correlation greater than zero, five were statistically significant at the .05 level.

Overall, the interventions were associated with a small and non-significant increase in rating on space and furnishing (p=.144). The overall *r* for the 14 independent samples combined was 0.122 using a random effects model, and the 95% confidence interval included zero (lower limit = -.042, upper limit = 0.280).

The 14 independent samples were significantly heterogeneous according to the *Q* statistic (Q (13) = 122.662, p =.000). The I^2 statistic indicated that 89% of the variation in the correlation coefficient could be attributed to study-level factors (I^2 = 89.402, τ^2 = .074, se = .002).

Figure 5: Forest plot of correlation coefficient: Space and furnishings quality measured by ERS



Study name	Subgroup within study	Correlation	Lower limit	Upper limit	Z-Value	p-Value
August (2012)		0.146	-0.050	0.332	1.460	0.144
Brooks-Gunn et al (2011)		0.234	0.099	0.361	3.355	0.001
Brown (2005)		0.008	-0.506	0.518	0.027	0.978
Bryant et al (2009)		-0.051	-0.233	0.134	-0.538	0.591
D'Amour (2008)	Infant and Toddler	0.260	0.017	0.474	2.095	0.036
D'Amour (2008)	Preschool	0.270	0.084	0.438	2.810	0.005
Dove (2003)		0.003	-0.703	0.706	0.006	0.995
Early et al (2013)		-0.240	-0.336	-0.139	-4.575	0.000
Mashburn (2004)		0.150	-0.035	0.325	1.592	0.111
Melhuish et al (2010)		0.153	0.020	0.280	2.261	0.024
Rous et al (2008)		0.430	0.296	0.547	5.835	0.000
Sandstrom (2012)		-0.170	-0.530	0.241	-0.805	0.421
Wheeler (2006)		0.526	-0.284	0.898	1.307	0.191
Zill et al (2001)		0.050	-0.266	0.356	0.304	0.761
Random effect model		0.127	-0.013	0.261	1.777	0.076

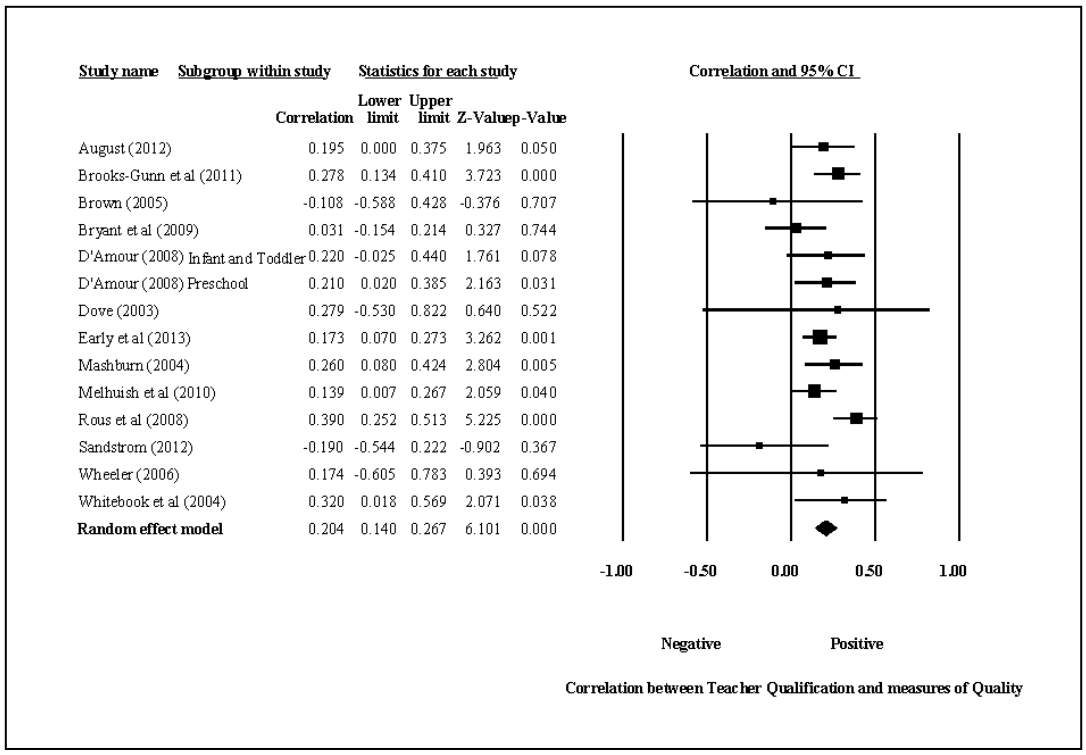
Activities

Thirteen studies measured rating on activities as an outcome, contributing a total of 14 effect sizes (*r*) to the meta-analysis (note that one study contained two independent samples). Of the 14 correlations, 12 had a positive *r*, indicating that for these studies higher teacher education level was correlated with higher ratings on activities. Of the 12 independent samples with a positive correlation coefficient, seven were statistically significant at the .05 level.

Despite some individual studies showing no correlation between teacher qualifications and activities, the meta-analysis showed an overall positive correlation. The weighted correlation coefficient for the 14 independent samples combined was .204 using a random effects model (see Figure 6). The 95% confidence interval for *r* did not include zero (lower limit = .140, upper limit = .267).

The 14 independent samples were not significantly heterogeneous according to the *Q* statistic (Q (13) = 18.317, p = .146). The I^2 statistic indicated that 29% of the variance in *r* may be a result of study factors (I^2 = 29.030, τ^2 = .004, se = .006).

Figure 6: Forest plot of correlation coefficient: Activities quality measured by ERS

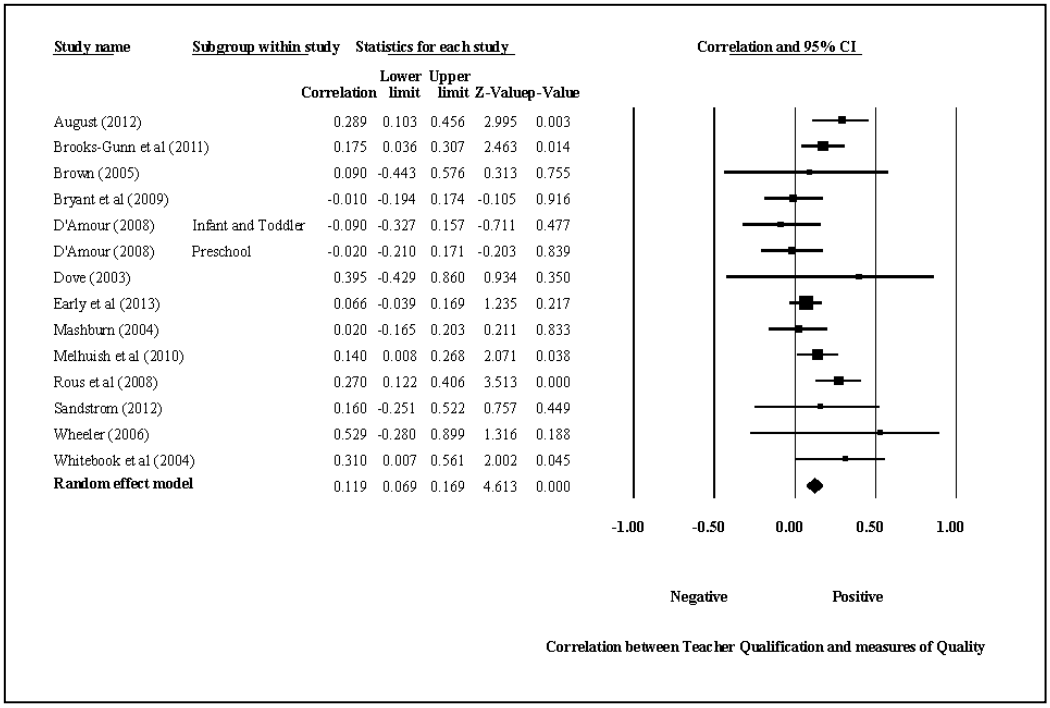


Interaction

Twelve studies measured rating on interaction as an outcome, contributing a total of 14 correlations to the meta-analysis (note that two studies contained two independent samples). Of the 14 effects, 11 had a positive *r*, indicating that for these studies higher teacher education level was correlated with higher ratings on interaction. Of the 11 independent samples with a positive *r*, 5 were statistically significant.

Despite some individual studies showing no correlation between teacher qualifications and interaction, the meta-analysis showed an overall positive correlation. The weighted mean *r* for the 14 independent samples combined was .122 using a random effects model (see Figure 7). The 95% confidence interval for *r* did not include zero (lower limit = .053, upper limit = .189). The 14 independent samples were not significantly heterogeneous according to the *Q* statistic (Q (13) = 19.796, p = .100). The I^2 statistic indicated that 34% of the variance in *r* may be a result of study factors (I^2 = 34.331, τ^2 = .005, se = .006).

Figure 7: Forest plot of correlation coefficient: Interaction quality measured by ERS



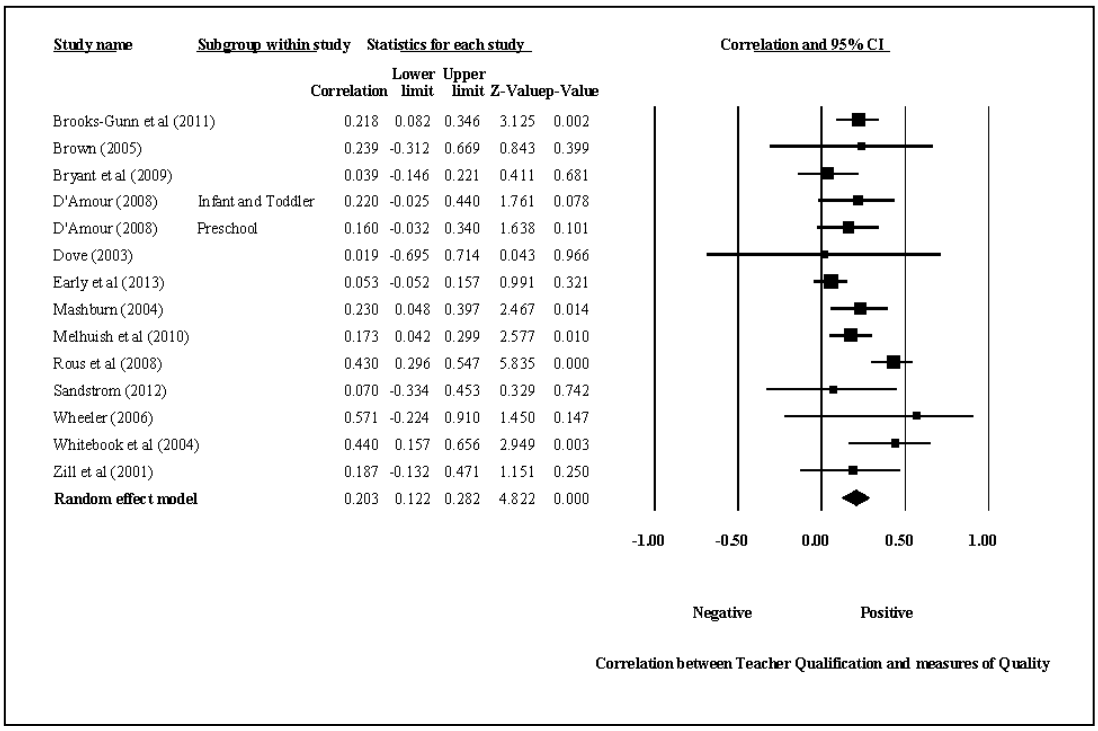
Study name	Subgroup within study	Correlation	Lower limit	Upper limit	Z-Value	p-Value
August (2012)		0.289	0.103	0.456	2.995	0.003
Brooks-Gunn et al (2011)		0.175	0.036	0.307	2.463	0.014
Brown (2005)		0.090	-0.443	0.576	0.313	0.755
Bryant et al (2009)		-0.010	-0.194	0.174	-0.105	0.916
D'Amour (2008)	Infant and Toddler	-0.090	-0.327	0.157	-0.711	0.477
D'Amour (2008)	Preschool	-0.020	-0.210	0.171	-0.203	0.839
Dove (2003)		0.395	-0.429	0.860	0.934	0.350
Early et al (2013)		0.066	-0.039	0.169	1.235	0.217
Mashburn (2004)		0.020	-0.165	0.203	0.211	0.833
Melhuish et al (2010)		0.140	0.008	0.268	2.071	0.038
Rous et al (2008)		0.270	0.122	0.406	3.513	0.000
Sandstrom (2012)		0.160	-0.251	0.522	0.757	0.449
Wheeler (2006)		0.529	-0.280	0.899	1.316	0.188
Whitebook et al (2004)		0.310	0.007	0.561	2.002	0.045
Random effect model		0.119	0.069	0.169	4.613	0.000

Language and reasoning

Twelve studies measured rating on language and reasoning as an outcome, contributing a total of 14 effect sizes to the meta-analysis (note that two studies contained two independent samples). All of the 14 independent samples had a positive *r*, indicating that for these studies higher teacher education level was correlated with higher ratings on language and reasoning. Of the 14 independent samples with a positive r, five were statistically significant.

The meta-analysis showed an overall positive correlation between teacher qualifications and ratings on language and reasoning. The weighted mean *r* for the 14 independent samples combined was .203 using a random effects model (see Figure 8). The 95% confidence interval for r did not include zero (lower limit = .122, upper limit = .282). The 14 independent samples were significantly heterogeneous according to the *Q* statistic (Q (13) = 26.049, p = .017). The I^2 statistic indicated that 50% of the variance in *r* may be a result of study factors (I^2 = 50.094, τ^2 = .011, se = .009).

Figure 8: Forest plot of correlation coefficient: Language and reasoning quality measured by ERS

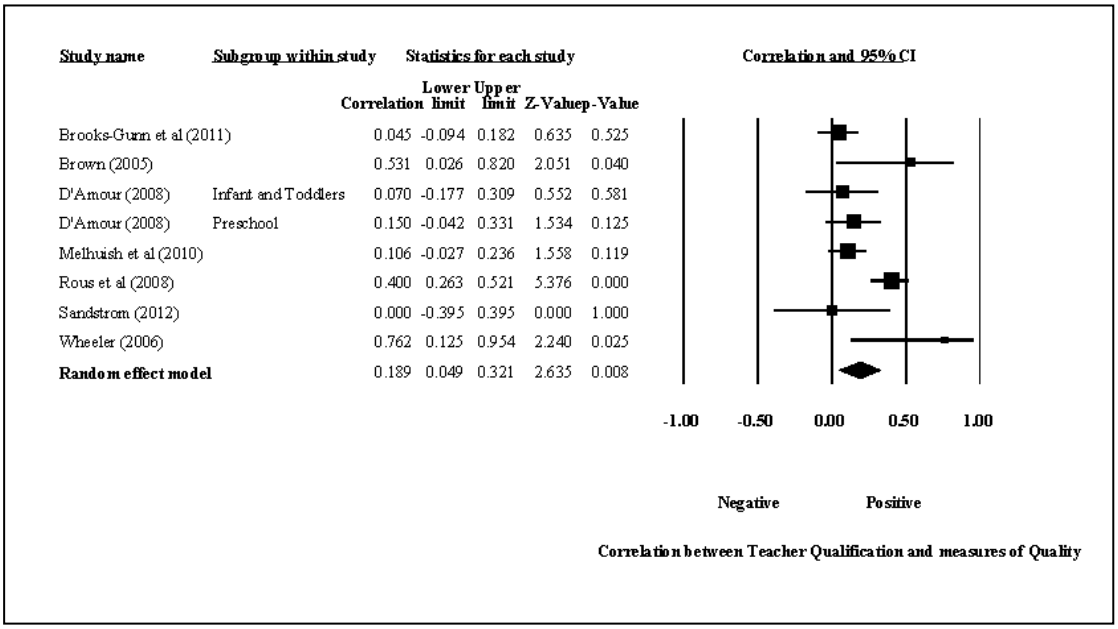


Parents and staff

Seven studies measured rating on parents and staff as an outcome, contributing a total of eight effect sizes to the meta-analysis (note that one study contained two independent samples). Of the eight independent samples, seven of them had an *r* greater than zero, indicating that for these studies higher teacher education level was correlated with higher ratings on parents and staff. Of the seven independent samples with an *r* value greater than zero, three were statistically significant.

The meta-analysis showed an overall positive correlation between teacher qualifications and ratings on parents and staff. The weighted mean *r* for the eight independent samples combined was .189 using a random effects model (see Figure 9). The 95% confidence interval for *r* did not include zero (lower limit = .049, upper limit = .321). The eight independent samples were significantly heterogeneous according to the *Q* statistic (Q (7) = 21.219, p = .003). The I^2 statistic indicated that 67% of the variance in *r* may be a result of study factors (I^2 = 67.011, τ^2 = .023, se = .021).

Figure 9: Forest plot of correlation coefficient: Parents and staff quality measured by ERS



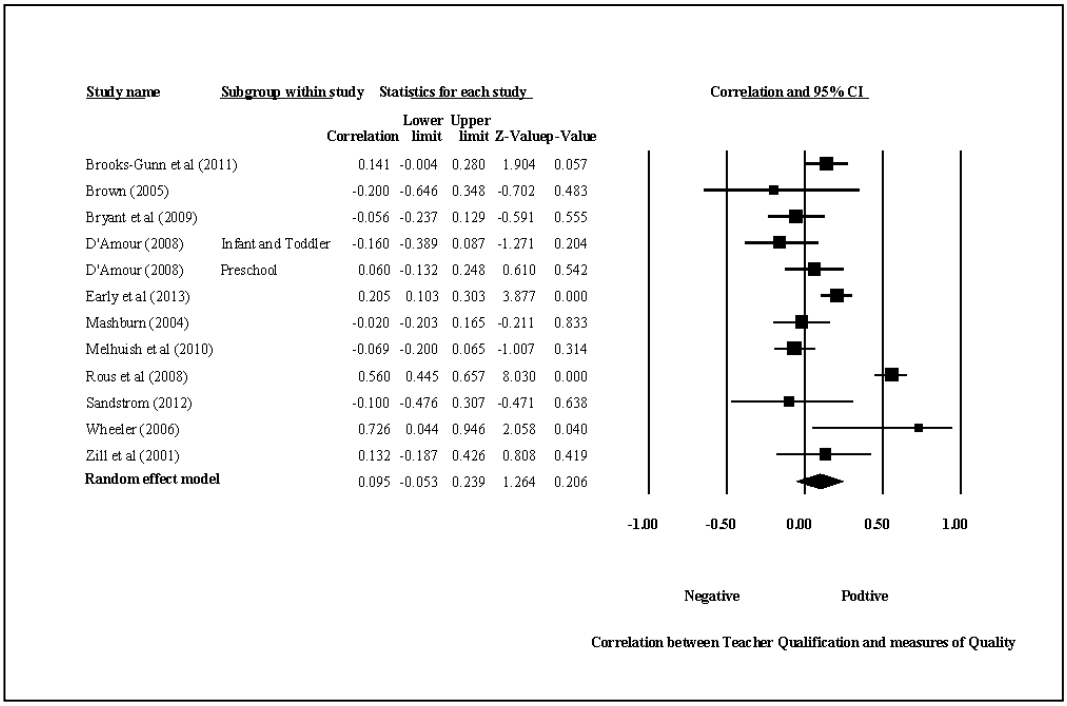
Personal care routines

Eleven studies measured rating on personal care as an outcome, contributing a total of 12 effect sizes to the meta-analysis (note that one study contained two independent samples). Of the 12 independent samples, six had a positive *r*, indicating that for these studies higher teacher education level was correlated with higher ratings on personal care. Of the six independent samples with a positive *r*, three were statistically significant.

The meta-analysis showed an overall positive correlation between teacher qualifications and ratings on personal care routines. The weighted mean *r* for the 12 independent samples combined was .095 using a random effects model (see Figure 10). The 95% confidence interval for *r* included zero (lower limit = -.053, upper limit = .239).

The eight independent samples were significantly heterogeneous according to the *Q* statistic (Q (11) = 69.022, p = .000). The I^2 statistic indicated that 84% of the variance in *r* may be a result of study factors (I^2 = 84.063, τ^2 = .050, se = .030).

Figure 10: Forest plot of correlation coefficient: Personal care quality measured by ERS



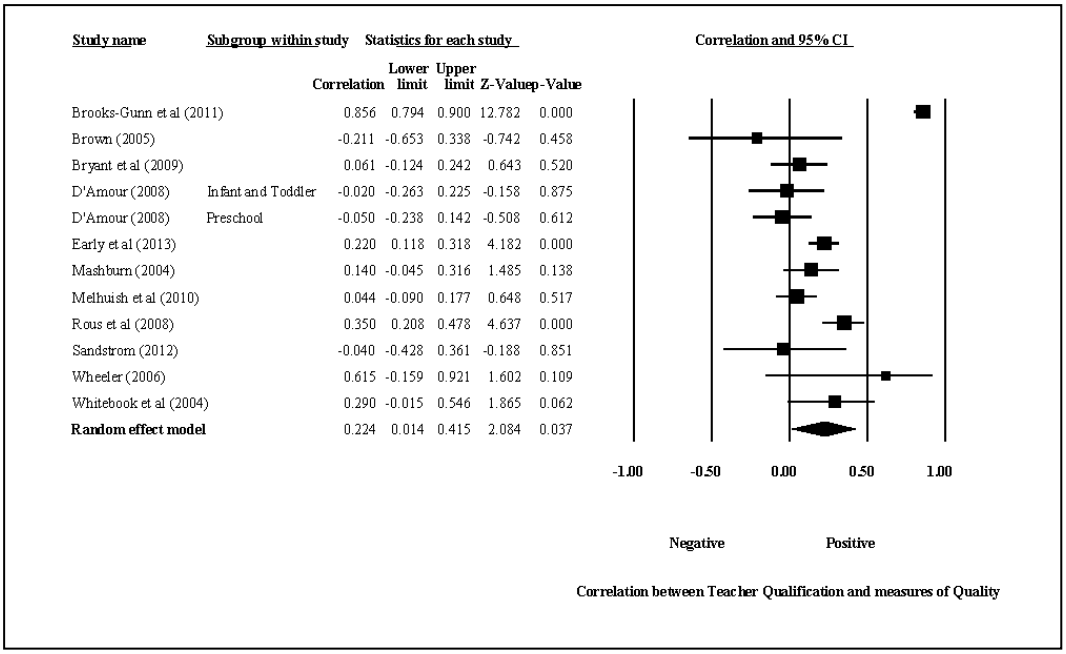
Program structure

Eleven studies measured rating on program structure as an outcome, contributing a total of 12 effect sizes to the meta-analysis (note that one study contained two independent samples). Of the 12 independent samples, eight had a positive *r*, indicating that for these studies higher teacher education level was correlated with higher ratings on program structure. Of the eight independent samples with a positive *r*, three were statistically significant.

The meta-analysis showed an overall positive correlation between teacher qualifications and ratings on program structure. The weighted mean *r* for the 12 independent samples combined was .224 using a random effects model (see Figure 11). The 95% confidence interval for *r* did not include zero (lower limit = .014, upper limit = .415).

The eight independent samples were significantly heterogeneous according to the *Q* statistic (Q (11) = 141.493, p = .000). The I^2 statistic indicated that 92% of the variance in *r* may be a result of study factors (I^2 = 92.226, τ^2 = .119, se = .069).

Figure 11: Forest plot of correlation coefficient: Program structure quality measured by ERS

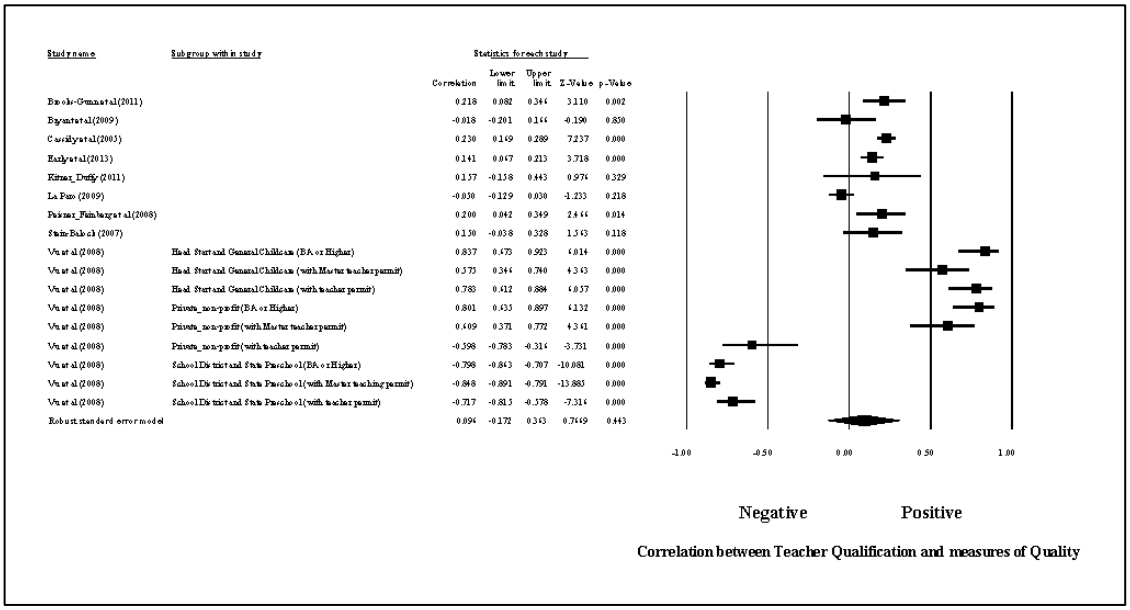


Two-factor classification: Language and interaction

Nine studies measured rating on language and interaction as an outcome, contributing a total of 17 effect sizes to the meta-analysis. Of the 17 independent samples, 11 had a positive *r*, indicating that for these studies higher teacher education level was correlated with higher ratings on language and interaction. Of the 11 independent samples with a positive r, nine were statistically significant. The meta-analysis showed an overall positive correlation between teacher qualifications and ratings on language and interaction. The estimated mean correlation (*r*) with robust standard error, assuming ρ = .80, was 0.096 (confidence limits -0.172, 0.363) (see Figure 12).

The nine studies were significantly heterogeneous according to the *Q* statistic (*Q* (16) = 583.703, *p* =.000). The I^2 statistic indicated that 97% of the variance in *r* may be a result of study factors (I^2 = 97.259, τ^2 = .206, se = .0122). We noted that there were some erratic patterns associated with the Vu et al (2008) study (see Figures 12 and 13). One possible explanation for the variance in the mean difference between the various sub-groups may be differences in sample characteristics. In particular, ERS ratings in school district and state preschool are negatively correlated with teacher qualification. This is particularly unusual, as we would expect the relationship between the two variables to be consistent across samples within the study.

Figure 12: Forest plot of correlation coefficient: Language and interaction quality measured by ERS



Two-factor classification: Provisions for learning

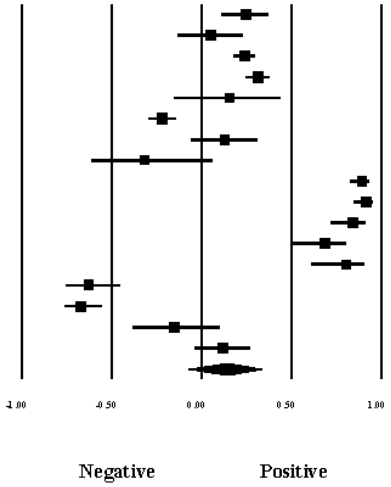
Nine studies measured rating on provisions for learning as an outcome, contributing a total of 17 effect sizes to the meta-analysis. Of the 17 independent samples, 12 had a positive *r*, indicating that for these studies higher teacher education level was correlated with higher ratings on provision for learning. Of the 17 independent samples with a positive *r*, eight were statistically significant.

The meta-analysis showed an overall positive correlation between teacher qualifications and ratings on provision for learning. The estimated mean correlation (*r*) with robust standard error, assuming ρ = .80, was 0.173 (confidence limits
-0.054, 0.399) (see Figure 13).

The nine studies were significantly heterogeneous according to the *Q* statistic (*Q* (16) = 515.764, *p* = .000). The I^2 statistic indicated that 97% of the variance in *r* may be a result of study factors (I^2 = 96.898, τ^2 = .183, se = .110).

Figure 13: Forest plot of correlation coefficient: Provisions for learning quality measured by ERS

Study name	Subgroup within study	Statistics for each study				
		Correlation	Lower limit	Upper limit	Z-Value	p-Value
Brooks-Gunn et al (2011)		0.244	0.111	0.372	3.523	0.000
Bryant et al (2009)		0.049	-0.134	0.231	0.517	0.605
Cassily et al (2005)		0.240	0.179	0.299	7.564	0.000
Early et al (2013)		0.314	0.247	0.382	8.570	0.000
Kitzmar_Duffy (2011)		0.157	-0.138	0.443	0.974	0.329
La Paro (2009)		-0.220	-0.294	-0.143	-5.310	0.000
Stein-Bakoch (2007)		0.130	-0.059	0.310	1.332	0.174
Vu et al (2008)	Head Start and General Childcare (BA or Higher)	-0.317	-0.414	0.042	-1.448	0.099
Vu et al (2008)	Head Start and General Childcare (with Master teacher permit)	0.893	0.821	0.934	10.293	0.000
Vu et al (2008)	Head Start and General Childcare (with teacher permit)	0.914	0.843	0.955	9.404	0.000
Vu et al (2008)	Private_non-profit (BA or Higher)	0.843	0.717	0.915	7.347	0.000
Vu et al (2008)	Private_non-profit (with Master teacher permit)	0.484	0.304	0.808	5.838	0.000
Vu et al (2008)	Private_non-profit (with teacher permit)	0.803	0.411	0.908	3.414	0.000
Vu et al (2008)	School District and State Preschool (BA or Higher)	-0.627	-0.734	-0.450	-5.744	0.000
Vu et al (2008)	School District and State Preschool (with Master teaching permit)	-0.472	-0.745	-0.351	-8.223	0.000
Vu et al (2008)	School District and State Preschool (with teacher permit)	-0.152	-0.388	0.102	-1.174	0.241
Pakner-Feinberg et al (2008)		0.120	-0.041	0.274	1.447	0.142
Robust standard error model		0.173	-0.034	0.399	1.553	0.120



Correlation between Teacher Qualification and measures of Quality

Discussion

SUMMARY OF MAIN RESULTS

We find a positive correlation between teacher qualification and overall ECEC quality with respect to the learning environment (mean correlation with robust standard error, assuming ρ = .80 (r=0.198, confidence limits 0.133, 0.263)). The ERS and subscale ratings are positively correlated to the qualification of the lead teacher. This is true for most of the subscales including language-reasoning, activities, interaction, program structure, and parents and staff. A positive correlation is also found in one of the subscales based on the two-factor classification – namely, provision for learning.

The 7 factor subscales vary with respect to correlation between teacher qualification and measures of ECEC quality. For example:

- The strongest correlation was found for 'program structure' (r= 0.224, 95% confidence limits 0.014, 0.415). Program structure relates to the ability to schedule, provide opportunities for free play, group time and provisions for children with disabilities.
- The second strongest correlation was found for 'activities' (r=0.204, 95% confidence limits 0.140, 0.267). This relates to fine motor, art, music/movement, blocks, sand/water, dramatic play, nature/science, math/number, use of digital technologies and promoting acceptance of diversity.
- The third strongest correlation was found for 'language and reasoning' (r=0.203, 95% confidence limits 0.122, 0.282). Language and reasoning is defined as the formal and informal use of language, development of reasoning skills and communication - encouraging children to communicate, use language to develop reasoning skills and the informal use of language.
- The fourth strongest correlation was found for 'parents and staff' (r=0.189, 95% confidence limits 0.049, 0.321). Teacher qualifications significantly influence the provision for parents, provisions for the personal and professional needs of staff, staff interaction and cooperation and the supervision and evaluation of staff.
- The fifth strongest correlation was found for 'interactions' (r=0.122, 95% confidence limits 0.053, 0.189). This relates to the supervision of gross motor activities, the general supervision of children, discipline, staff-child interactions as well as providing opportunities for children to interact with each other.
- The sixth scale, 'space and furnishings' was not significantly correlated to teacher qualification (r=0.122, 95% confidence limits -0.042, 0.280). This relates to indoor and outdoor space and furnishings.

- The final scale, 'personal care routine' was not significantly correlated to teacher qualification (r=0.095, 95% confidence limits -0.053, 0.239). This relates to hygiene practices, safety, extent to which children's independence/self-help skills are encouraged.

In descending order of effect size, the 2-factor subscale outcomes evaluated are:

- The subscale 'provisions for learning' was not significantly correlated to teacher qualification based on the robust standard error model, assuming ρ = .80 (r=0.173, confidence limits -0.054, 0.399); and
- The subscale 'language and interaction' was not significantly correlated to teacher qualification (mean correlation with robust standard error, assuming ρ = .80 (r=0.096, confidence limits -0.172, 0.363).

Five of the seven factor subscales are positive and statistically significant, regardless of the ERS used or the ethnicity of the dominant student group. The two factor subscales that are not significant are 'space and furnishings' and 'personal care routine'. Neither of the two 2 factor subscale outcomes are statistically significant.

OVERALL COMPLETENESS AND APPLICABILITY OF EVIDENCE

Our meta-analysis explored the relationship between lead teacher qualifications and ECEC quality outcomes. We included studies that measured ECEC classroom quality with ERS and reported comparison data of lead teachers with various levels of education. These evaluations provide two types of evidence; correlations and group mean difference. The evidence also captured a range of definitions or classifications of teacher qualifications by the highest level of formal education attained, years of formal education, earned credits in ECEC-related programs and percentage of teachers with a specific degree at a center-level. Moreover, the studies included in our review had to report at least one outcome measure from an ERS. These outcomes included measures of overall ECEC quality (overall rating), ratings on the original subscale factors (personal care routine; space and furnishings; language-reasoning/listening and talking; activities; program structure; interaction; and parents and staff) and ratings on the 2-factor classification subscales (provisions for learning and language/interactions). Overall, our search of the literature found a relatively small and diverse group of comparative studies that met our review criteria. Moreover, very few studies (n=4) used repeated measures to explore the impact of higher teacher qualifications on ECEC quality.

QUALITY OF THE EVIDENCE

Regarding the quality of the evidence, we focus our assessment on four key types of validity – statistical conclusion validity, internal validity, construct validity and external validity.

Statistical conclusion validity

Statistical conclusion validity (or statistical validity) is concerned with issues or problems that threaten valid inference-making (Kirk, 1995). According to Elvik (1999), results of meta-analyses

cannot be theoretically valid without being statistically valid, at least with respect to some of the criteria of statistical validity. The statistical validity attained in a meta-analysis depends on a list of criteria including choice of techniques of analysis, commensurability of dependent variables, evidence of publication bias, and shape of distribution.

Based on our diagnostic analysis, we adopted a random effects model for some of the reported results due to the identified random heterogeneity of the research subjects. Also, whenever studies with dependent samples were present, we adopted the robust standard error model for the analysis. With regards to the dependent variable of the included studies, the dependent variable is the ratings of measures of ECEC quality using comparable ERSs (i.e. ECERS, ITERS and their revised version). In this respect, our set of included studies with commensurable definitions of dependent variables can be regarded as more valid from a purely statistical point of view than a set of studies with incommensurable definitions of dependent variables. Regarding to potentials of publication bias, several diagnostic tests (e.g. Egger's test, Duval and Tweedie's "trim and fill" method) were performed to demonstrate that there was no evidence of publication or small-study bias. In all, we feel confident that our systematic search captured a large number of unpublished and small studies, since a substantial proportion of our eligible studies are dissertations and unpublished reports. Therefore, the effect of publication bias introduced across the studies is likely to be trivial. Concerning the shape of the distribution of results in our set of studies, the funnel graph diagram for ratings of overall ECEC quality presents a fairly symmetrical distribution with no extreme results, while the diagrams for sub-factor ratings present a somewhat asymmetrical distribution with a number of extreme positive results on the right.

Internal validity

Internal validity focuses on the quality of study design examining the influence of confounding variables. For example, the potential effect or contribution of any uncontrolled/measured variables on the evaluated relationships among concerned variables. A randomized experiment is commonly recognized as the most preferable design that has the highest possible internal validity as it avoids all the potential threats (e.g. preexisting differences between experimental and control conditions) (Farrington, 2003). The UK Centre for Reviews and Dissemination (2001) proposed a hierarchy of research designs with five types of evidence: (1) randomized, controlled, double-blind trials; (2) quasi-experimental studies (experiments without randomization); (3) controlled observational studies (comparison of outcomes between participants who have received an intervention and those who have not); (4) observational studies without a control group; and (5) expert opinion.

In the context of this review, both the randomized experiment trail and quasi-experimental study designs were not adopted as there was no evidence that teachers with different qualification were randomly allocated to an ECEC classroom/center. In addition, a pre-post test experiment was not adopted as most of the studies were cross-sectional controlled observational studies. The class and classroom being evaluated did not have the same chance of being assigned to be managed by a lead teacher with a specific qualification. Further, it was possible that the comparison group differences with regards to ECEC qualities could be attributed to both observed and unobserved characteristics due to systematic cofounding factors (e.g. the classroom characteristics). For example, ECEC centers located in districts with a lower socio-economic level might have a relatively lower overall

rating on ECEC quality when compared to the ratings of centers located in districts with a higher socio-economic level. A lower-income ECEC center may not have the finance to support the management of its classrooms or hire more well-educated/qualified teachers. Some of the nation-wide and multi-state/large-scale studies in this review might have aggregated and compared classroom ratings without controlling for the potential influence of classroom-/center-specific characteristics. However, it is noted that these potential issues do not apply to studies that compared the ECEC quality of multiple classrooms within the same center.

Construct validity

Studies in our review are generally expected to have a reliable construct validity with respect to the accuracy of the operational definition and measurement of the theoretical constructs (ECEC qualities) which guide the evaluations. The use of ERS and its validity has been tested in the ECEC sector.

External validity

Regarding the ability to generalize the results of our current review to other subjects and settings, our moderator analyses suggest that the evidence do not differ across classrooms with different ethnicity or classrooms that were being measured by different ERS.

LIMITATIONS AND POTENTIAL BIASES IN THE REVIEW PROCESS

Limitation with the sole focus on relationship between ECEC quality and lead teacher qualifications

Most of the included studies, except the center-based studies, in the current review are based only on the credentials of the lead teacher in the ECEC classroom. The tendency for solely relying on the lead teacher qualification as a predictor can be attributed to the assumption that the lead teacher, who is most accountable for the management of the classroom, is also responsible for the overall quality of the associated outcomes. Such an assumption may, however, be inaccurate as the potential effects of assistant teachers (i.e., floater teachers, aides, and other staff in the room) are potentially overlooked.

Limitation associated with the dominant presence of correlational studies

The majority of the evaluations included in this review are correlational studies, which treat teacher qualifications as a continuous variable. This type of evaluation collectively allows an examination of the linear relationship between teacher qualifications and ECEC quality. Although an overall positive correlation between the two variables can be depicted, most of these evaluations, except those which used a dichotomous variable for levels of education, are limited in their ability to investigate the marginal change of ECEC quality based on a unit change in teacher qualification. More comparative evaluation studies that measure the marginal effect of a change of teacher levels of education on ECEC quality is desirable. This will also assist the identification of potential saturation of changes with regards to the change in teacher qualifications. For example, there may be a significant difference in ECEC quality when a lead teacher possesses a high school

qualification compared to a teacher who has an associate degree. However, a change across other levels of education (e.g. from AA to BA or BA to post-graduate) may be marginal in terms of effect on the outcome variable. This needs further exploration.

Limitation associated with correlations

Most Campbell reviews neatly fall within the effectiveness/efficacy research domain. This review, however, does not. For this reason, we think it is important to note that causal inference is not the objective of this review (both in the individual studies and the current review), and that a myriad of factors outside of the collected data may affect the direction and magnitude of the observed results.

Limitation due to limited specification of study or sample characteristics in included studies

Many of the analyses (overall ECEC quality and subsequent specific sub-scale outcomes) yielded *Q*-statistics suggesting a significant amount of variability in effect sizes. These results may be attributed to study or sample characteristics, which could be potential moderators of the relationship between teacher qualification and ERS ratings. Future reviews may provide additional sources that may explain this variability (and many unanswered questions). This will be an important point we hope to cover in future updates of this review.

Limitations due to small sample size in sub-scale analyses

The null hypothesis of homogeneity was not rejected in any of the sub-scale analyses. This result may be attributed to the small sample size of the majority of studies and poor statistical power. As such we advise caution when interpreting results of the sub-scale analyses. It may be possible that the relationship between teacher qualification and sub-scale ratings may be inadvertently driven by potential study level variation.

AGREEMENTS AND DISAGREEMENTS WITH OTHER STUDIES OR REVIEWS

Agreement with previous meta-analysis studies

This meta-analysis has provided an extensive review of studies related to teacher qualifications and the early childhood learning environment to provide a current understanding of the relationship. This study agrees with Whitebook's (2003b) systematic review regarding the educational level of teachers and how the level of qualification potentially affects the early childhood learning environment. This study, however, provides a more meaningful comparison between individual studies regarding the impact of teacher's qualification on the quality of the ECEC environment by further disaggregating the target population and providing a common metric for measuring outcomes. The importance of a single measure provides important information for policy makers.

The study also agrees with the meta-analysis from Kelley and Camilli (2007) who identified 32 studies for investigation with a specific focus on aggregating global classroom quality, teacher-child interactions, teacher pedagogical beliefs and knowledge, and classroom instructional activities.

This, however, makes it difficult to estimate the additional effects on the quality of the learning environment. Instead we chose to focus on the quality in the early childhood environment and ratings within the subscale to find relationships between teacher qualifications and the ECEC setting. Again, a focus on specific information is essential for policy makers.

Authors' conclusions

IMPLICATIONS FOR PRACTICE AND POLICY

Importance for practice and policy

The results from this study provide an up-to-date analysis of the correlation between teacher qualifications and the quality of the early childhood learning environment. Based on published research studies since the 1980s, the review provides an in-depth understanding relevant for current practice and policy. The review was able to draw upon information from a number of countries, demonstrating that the relationship between teacher qualification and enhanced quality of the learning environment was not dependent on context or culture.

Results from this study are important for policy makers wanting to enhance policy and practice within ECEC settings to improve children and family outcomes. This meta-analysis provides evidence of a significant and positive correlation between teacher qualification and quality with respect to the learning environment for all young children in ECEC settings, including infants and toddlers. Mandating qualified teachers (with tertiary education) may lead to significant improvement for both process and structural quality within centre-based and home-based ECEC settings.

The economic investment in qualified staff with some tertiary education to enhance the quality of the learning environment may also have lasting affects for children, families and society. Research has shown that children in high quality ECEC settings grow up to be healthier and more productive adults. Government investment in ECEC is internationally recognized as a sound economic driver for the well-being of a nation (Britto, Engle & Super, 2013). This means that improvements in quality in ECEC settings leads to enhanced educational, societal and economic benefits.

IMPLICATIONS FOR RESEARCH

Future research

This meta-analysis provides strong evidence for the relationship between teacher qualification and classroom quality as measured by the ERS. It has been able to provide an up-to-date analysis based on research since the 1980s. While this study has focused on quantitative research outcomes, further research in this area may also want to include an analysis of qualitative data to explore the relationship between teacher qualification and the quality of early childhood learning environment. In addition, future studies may seek to establish a causal relationship between teacher qualification and classroom quality using natural experiments or non-experimental designs. Such analysis

would complement current research and provide a more in-depth understanding of the early childhood learning environment.

Another key consideration for further research would be the consideration of moderators such as socioeconomic status of the ECEC settings or target population among existing studies. Given the limited specification of such characteristics, this review was unable to use this potential moderator.

Further research is also needed into the specific knowledge and skills that are learnt by teachers with higher qualifications that enable them to complete their roles effectively. It would be important for research to identify key enablers during professional training that provide teachers with confidence and competence to complete their role as a teacher effectively. By identifying such enablers, training can become better geared to the learning and development of qualified teachers.

References

REFERENCES TO INCLUDED STUDIES

Antle, B. F., Frey, A., Barbee, A., Frey, S., Grisham-Brown, J., & Cox, M. (2008). Child care subsidy and program quality revisited. *Early Education and Development, 19*(4), 560-573.

August, A. L. (2012). Preschool education in Belize: Research on the current status and implications for the future. *Education, 2*(7), 231-238.

Bolger, K. E., & Scarr, S. (1995). Not so far from home: How family characteristics predict child care quality. *Early Development and Parenting, 4*(3), 103-112.

Britto, P.R., Engle, P.L. & Cuper, C.M. (2013). *Handbook of early childhood development research and its impact on global policy*. London: Oxford University Press.

Brooks-Gunn, J., Garfinkel, I., McLanahan, S. S., & Paxson, C. (2011). Fragile Families and Child Wellbeing Study [Public Use Data] (Publication no. 10.3886/ICPSR31622.v1). (ICPSR31622-v1). Retrieved 17th March 2014, from the Inter-university Consortium for Political and Social Research.

Brown, K. (2005). *The impact of preschool on middle-class children in a public inclusion program*. (Doctor of Philosophy), University of Pennsylvania, Philadelphia, Pennsylvania.

Bryant, D., Wesley, P., Burchinal, P., Hegland, S., Hughes, K., Tout, K., . . . Jeon, H. J. (2009). Quality Interventions for Early Care and Education (QUINCE) - Partners for Inclusion, 2004-2007 [California, Iowa, Minnesota, Nebraska, NorthCarolina] (Publication no. 10.3886/ICPSR28124.v1). (ICPSR28124-v1). Retrieved 15th March 2014, from the Inter-university Consortium for Political and Social Research.

Burchinal, M. R., Roberts, J. E., Riggins Jr, R., Zeisel, S. A., Neebe, E., & Bryant, D. (2000). Relating quality of center-based child care to early cognitive and language development longitudinally. *Child Development, 71*(2), 339-357.

Cassidy, D. J., Hestenes, L. L., Hegde, A., Hestenes, S., & Mims, S. (2005). Measurement of quality in preschool child care classrooms: The Early Childhood Environment Rating Scale-Revised and its' psychometric properties. *Early Childhood Research Quarterly, 20*(3), 345-360.

Couse, L. J. (2001). *Effectiveness of service delivery models in inclusive early childhood programs*. (Doctor of Philosophy), Syracuse University, Syracuse, New York.

Cryer, D., Tietze, W., Burchinal, M., Leal, T., & Palacios, J. (1999). Predicting process quality from structural quality in preschool programs: A cross-country comparison. *Early Childhood Research Quarterly, 14*(3), 339-361.

D'Amour, A. C. (2008). *The relations among childcare provider education, neighborhood poverty, and the quality of childcare classrooms*. College Park, MD: University of Maryland.

Dennis, S. E., & O'Connor, E. (2013). Reexamining quality in early childhood education: Exploring the relationship between the organizational climate and the classroom. *Journal of Research in Childhood Education, 27*(1), 74-92.

Dove, R. A. (2003). *Child engagement: An indicator of quality child care*. (Doctor of Philosophy), Tennessee Technological University, Cookeville, TN.

Dunn, L. (1993). Proximal and distal features of day care quality and children's development. *Early Childhood Research Quarterly, 8*(2), 167-192.

Early, D., Burchinal, M., Barbarin, O., Bryant, D., Chang, F., Clifford, R., . . . Barnett, W. S. (2013). Pre-Kindergarten in Eleven States: NCEDL's Multi-State Study of Pre-Kindergarten and Study of State-Wide Early Education Programs (SWEEP) (Publication no. 10.3886/ICPSR34877.v1). (ICPSR34877-v1). Retrieved 13th March 2014, from Inter-university Consortium forPolitical and Social Research.

Epstein, A. S. (1999). Pathways to quality in Head Start, public school, and private nonprofit early childhood programs. *Journal of Research in Childhood Education, 13*(2), 101-119.
Henry, G. T., Henderson, L. W., Ponder, B. D., Gordon, C. S., Mashburn, A. J., & Rickman, D. K. (2003). Report of the Findings from the Early Childhood Study: 2001-02. Atlanta, Georgia: Georgia State University.
Hestenes, L. L., Kintner-Duffy, V., Wang, Y. C., La Paro, K., Mims, S. U., Crosby, D., . . . Cassidy, D. J. (2015). Comparisons among quality measures in child care settings: Understanding the use of multiple measures in North Carolina's QRIS and their links to social-emotional development in preschool children. *Early Childhood Research Quarterly, 30*, 199-214.
Ho, H. Y. (2011). *Effects, attributes, and predictions of parental involvement during early transition: Does race/ethnicity matter? Evidence from the FACES 1997 Cohort.* (PhD in Psychology in Education), University of Pittsburgh, Pittsburgh, Pensylvania.
Holloway, S. D., Kagan, S. L., Fuller, B., Tsou, L., & Carroll, J. (2001). Assessing child-care quality with a telephone interview. *Early Childhood Research Quarterly, 16*(2), 165-189.
Honeycutt, D. (2008). *The relationship of teacher qualications, quality of care and student achievement outcomes in the Arkansas Better Chance Program.* (Ed.D.), University of Arkansas, Ann Arbor.
Jardine-Ledet, A. S. (1999). *Evaluating the efficacy of inservice training for modifying the knowledge and practices of child care providers in Louisiana.* (Doctor of Philosophy), University of New Orleans, New Orleans, LA.
Kintner-Duffy, V. L. (2011). *"Everybody's different and the same": An inquiry into early childhood teachers' beliefs, knowledge, and practices in relation to children from culturally diverse backgrounds.* (Doctoral of Philosophy), University of North Carolina, Greensboro.
Kontos, S. J. (1991). Child care quality, family background, and children's development. *Early Childhood Research Quarterly, 6*(2), 249-262.
Kwon, J. Y., & Han, Y. M. (2007). A comparison of the quality in public and private child-care center. *The Korean Journal of Community Living Science, 18*(1), 177-187.
La Paro, K. M., Hamre, B. K., Locasale-Crouch, J., Pianta, R. C., Bryant, D., Early, D., . . . Burchinal, M. (2009). Quality in kindergarten classrooms: Observational evidence for the need to increase children's learning opportunities in early education classrooms. *Early Education and Development, 20*(4), 657-692.
Laferney, P. S. (2006). *Early Childhood Professional Development and Classroom Quality in Preschool Classrooms.* (Master of Science), Oklahoma State University, Stillwater, Oklahoma.
Melhuish, E., Belsky, J., MacPherson, K., & Cullis, A. (2010). The quality of group childcare settings used by 3-4 year old children in Sure Start local programme areas and the relationship with child outcomes. London: Department of Education.
Peisner-Feinberg, E., Bryant, D., Fischer, R., & Grayson, L. (2008). Cuyahoga County Child Care Quality Study. Cleveland, OH: Mandel School of Applied Social Sciences, Case Western Reserve University.
Phillips, D., Mekos, D., Scarr, S., McCartney, K., & Abbott–Shim, M. (2001). Within and beyond the classroom door: Assessing quality in child care centers. *Early Childhood Research Quarterly, 15*(4), 475-496.
Rous, B., Grove, J., Cox, M., Townley, K., & Crumpton, G. (2008). The impact of the Kentucky professional development framework on child care, head start and public preschool classroom quality and child outcomes. Lexington, KY: University of Kentucky, Human Development Institute.
Sandstrom, H. (2012). The characteristics and quality of pre-school education in Spain. *International Journal of Early Years Education*, 1-12.
Selden, S. C., & Sowa, J. E. (2004). Testing a multi-dimensional model of organizational performance: Prospects and problems. *Journal of Public Administration Research and Theory, 14*(3), 395-416.
Setodji, C. M., Le, V.-N., & Schaack, D. (2012). Accounting for movement between childcare classrooms: Does it change teacher effects interpretations? *Journal of Applied Developmental Psychology, 33*(1), 1-12.
Smith, W. E. (2005). *Structural characteristics that predict quality in preschool-age classrooms in child care centers.* (Doctor of Philosophy), University of Pittsburgh, Pittsburgh, PA
Stein-Balock, A. (2007). *Predicting the quality of center-based early care and education programs for preschool children: A cumulative asset model.* (Master of Science), Iowa State University, Ames.

Stein, A. (2010). *Family child care provider beliefs and program quality: A longitudinal study investigating the role of consultation.* (Doctor of Philosophy), Iowa State University, Ames, Iowa.
Thompson, D. S. (1992). *The relationship between the amount of training and education of teachers of three to five-year-old children and attitudes and quality of the child care environment.* (Doctor o f Education), University of Illinois at Urbana-Champaign, Champaign, Illinois.
Tout, K., & Sherman, J. (2005). A Snapshot of Quality in Minnesota's Child Care Centers: A Report of the Minnesota Child Care Policy Research Partnership. St. Paul, MN: The Minnesota Child Care Policy Research Partnership.
Vermeer, H. J., Van IJzendoorn, M. H., de Kruif, R. E., Fukkink, R. G., Tavecchio, L. W., Riksen-Walraven, J. M., & van Zeijl, J. (2008). Child care in the Netherlands: Trends in quality over the years 1995-2005. *The Journal of Genetic Psychology, 169*(4), 360-385.
Vu, J. A., Jeon, H.-J., & Howes, C. (2008). Formal education, credential, or both: Early childhood program classroom practices. *Early Education and Development, 19*(3), 479-504.
Weinraub, M., Shlay, A. B., Harmon, M., & Tran, H. (2005). Subsidizing child care: How child care subsidies affect the child care used by low-income African American families. *Early Childhood Research Quarterly, 20*(4), 373-392.
Wheeler, M. (2006). *A comparison of school readiness for preschool children with and without disabilities in inclusive environments.* (Doctor o f Education), Columbia University, New York.
Whitebook, M., Phillips, D., Bellm, D., Crowell, N., Almaraz, M., & Jo, J. Y. (2004). Two Years in Early Care and Education: A Community Portrait of Quality and Workforce Stability. Berkeley, CA: Center for the Study of Child Care Employment.
Zill, N., Resnick, G., Kim, K., McKey, R. H., Clark, C., Pai-Samant, S., . . . D'Elio, M. A. (2001). Head Start FACES: Longitudinal Findings on Program Performance. Third Progress Report. Washington, DC: U.S. Department of Health and Human Services.

REFERENCES TO EXCLUDED STUDIES

Beer, H. (1991). *The Child Development Associate Credential as it relates to preschool day care environmental quality.* (Doctor of Philosophy), Hofstra University, Hempstead, NY.
Blau, D. M. (1997). The production of quality in child care centers. *Journal of Human Resources,* 32(2), 354-387.
Blau, D. M. (2000). The production of quality in child-care centers: Another look. *Applied Developmental Science, 4*(3), 136-148.
Brustkem, M. C. (2002). *CDA, Does It Make a Difference in Preschool Classrooms?* (Doctor o f Education), The University of Dayton, Dayton, OH.
Burchinal, M. R., Cryer, D., Clifford, R. M., & Howes, C. (2002). Caregiver training and classroom quality in child care centers. *Applied Developmental Science, 6*(1), 2-11.
Buysse, V., Wesley, P. W., Bryant, D., & Gardner, D. (1999). Quality of early childhood programs in inclusive and noninclusive settings. *Exceptional Children, 65*(3), 301-314.
Crowell, N. A. (2009). *Language environment and positive caregiving climate in early childhood care and education and their relationship to child language development.* (Doctor of Philosophy in Psychology), Georgetown University, Washington, DC.
Denny, J. H. (2009). *The Relationship between Preschool Teachers' Beliefs about School Readiness and Classroom Practice in Tennessee Child Care Programs.* (Doctor of Philosophy), University of Tennessee, Knoxville, TN.
Denny, J. H., Hallam, R., & Homer, K. (2012). A multi-instrument examination of preschool classroom quality and the relationship between program, classroom, and teacher characteristics. *Early Education & Development, 23*(5), 678-696.
Grisham-Brown, J., Cox, M., Gravil, M., & Missall, K. (2010). Differences in child care quality for children with and without disabilities. *Early Education and Development, 21*(1), 21-37.
Hadeed, J. (2014). Reliability and validity of the Early Childhood Environment Rating Scale, Revised Edition, ECERS-R in Arabic. *Early Child Development and Care, 184*(6), 819-842.
Haustein, S. L. (2012). *The Effects of Literacy Enriched Classroom Environment Partnered With Quality Adult/Child Interaction on the Development of Emergent Literacy Skills in Preschool Children.* (Doctor of Philosophy), Kean University, Union, NJ.

Hemmeter, M. L., Snyder, P., Kinder, K., & Artman, K. (2011). Impact of performance feedback delivered via electronic mail on preschool teachers' use of descriptive praise. *Early Childhood Research Quarterly, 26*(1), 96-109.

Hestenes, L. L., Cassidy, D. J., Hegde, A. V., & Lower, J. K. (2007). Quality in inclusive and noninclusive infant and toddler classrooms. *Journal of Research in Childhood Education, 22*(1), 69-84.

Howes, C. (1995). Reconceptualizing the early childhood work force. In H. S. W (Ed.), *Cost, quality, and child outcomes in child care centers. Technical report* (pp. 159-170). Denver: University of Colorado at Denver, Department of Economics, Center for Research in Economic and Social Policy.

Jackson, R., McCoy, A., Pistorino, C., Wilkinson, A., Burghardt, J., Clark, M., . . . Swank, P. (2007). National Evaluation of Early Reading First. Final Report to Congress. NCEE 2007-4007. Jessup, MD: National Center for Education Evaluation and Regional Assistance.

Lamy, C. E., Frede, E., Seplocha, H., Strasser, J., Jambunathan, S., Juncker, J. A., . . . Wolock, E. (2004). Inch by Inch, Row by Row, Gonna Make This Garden Grow: Classroom Quality and Language Skills in the Abbott Preschool Program. Trenton, NJ: New Jersey Department of Education, Office of Early Childhood Education.

Lawrence, D. L. (2010). *Associations with Teacher Education, Experience, and Program Characteristics In Quality Early Childhood Programming.* (Doctor of Philosophy), Walden University, Minneapolis, MN.

Layzer, J. I., Goodson, B. D., & Moss, M. (1993). Life in preschool: Volume One of an observational study of early childhood programs for disadvantaged four-year-olds. Cambridge, MA: Abt Associates.

Le, V.-N., Schaack, D. D., & Setodji, C. M. (2015). Identifying baseline and ceiling thresholds within the Qualistar Early Learning Quality Rating and Improvement System. *Early Childhood Research Quarterly, 30*(Part B), 215-226.

Lee, S. (1994). *The Relationship of Daycare Program Quality to Children's Play Behavior.* (Doctoral of Philosophy), The Pennsylvania State University, State College, Pennsylvania.

Mashburn, A. J. (2004). *Preschool Quality and Cognitive Development o f Preschoolers in Georgia.* (Doctor of Philosophy), Georgia State University, Atlanta, Georgia.

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.

Peisner-Feinberg, E., Schaaf, J., & LaForett, D. (2013). Children's Growth and Classroom Experiences in Georgia's Pre-K Program: Findings from the 2011-2012 Evaluation Study. Chapel Hill, NC: FPG Child Development Institute.

Pianta, R. C., 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.

Sanders, K., & Downer, J. (2012). Predicting acceptance of diversity in pre-kindergarten classrooms. *Early Childhood Research Quarterly, 27*(3), 503-511.

Setodji, C., Le, V. N., & Schaack, D. (2012). Accounting for movement between childcare classrooms: Does it change teacher effects interpretations? *Journal Of Applied Developmental Psychology, 33*(1), 1-12.

Sitton, L. A. (2007). *The Relationship Among Organizational Practices, Director Level of Education, Experience, and Child Care Center Program Quality in Tennessee.* (Doctor o f Education), Tennessee State University, Nashville, Tennessee.

Spillman, S. D. (2011). *A Comparative Study of Two Professional Development Models' Impact on Preschool Teachers' Classroom Practies.* (Doctorate of Education), University of Houston, Houston, TX.

St Clair-Christman, J., Buell, M., & Gamel-McCormick, M. (2011). Money Matters for Early Education: The Relationships among Child Care Quality, Teacher Characteristics, and Subsidy Status. *Early Childhood Research & Practice, 13*(2), n2.

Thompson, D. S. (1992). *The relationship between the amount of training and education of teachers of three to five-year-old children and attitudes and quality of the child care environment.* (Doctor o f Education), University of Illinois at Urbana-Champaign, Champaign, I.

ADDITIONAL REFERENCES

Ackerman, D. J. (2005). Getting teachers from here to there: Examining issues related to an early care and education teacher policy. *Early Childhood Research & Practice, 7*(1). 1-17.

Anders, Y., Rossbach, H., 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.

Arnett, J. (1989). Caregivers in day-care centers: Does training matter? *Journal of Applied Developmental Psychology, 10*(4), 541-552.

Babbie, E. (2011). *The Basics of Social Research* (5th ed.). Wadsworth: Cengage Learning.

Bayley, N. (1993). *Manual for the Bayley Scales of Infant Development* (2nd ed.). San Antonio. TX: The Psychological Corporation.

Berk, L. (1985). Relationship of educational attainment, child-oriented attitudes, job satisfaction and career commitment to caregiver behavior toward children. *Child Care Quarterly, 14*(2), 103-129.

Blau, D. M. (2000). The production of quality in child-care centers: Another look. *Applied Developmental Science, 4*(3), 136-148.

Blokland, A. A. J., & Nieuwbeerta, P. (2010). Life Course Criminology. In S. G. Shoham, P. Knepper & M. Kett (Eds.), *International Handbook of Criminology* (pp. 51-94). London: CRC Press-Taylor & Francis Froup.

Bowman, B. T., Donovan, M. S., & Burns, M. S. (2001). *Eager to Learn: Educating Our Preschoolers*. Washington, DC: National Academy Press.

Bryant, D., Maxwell, K., Taylor, K., Poe, M., Peisner-Feinberg, E., & Bernier, K. (2003). Smart Start and Preschool Child Care Quality in NC: Change Over Time and Relation to Children's Readiness. Chapel Hill, NC: FPG Child Development Institute.

Burchinal, M. R., Cryer, D., Clifford, R. M., & Howes, C. (2002). Caregiver training and classroom quality in child care centers. *Applied Developmental Science, 6*(1), 2-11.

Burchinal, M. R., Howes, C., Pianta, R., Bryant, D., Early, D., Clifford, R., & Barbarin, O. (2008). Predicting child outcomes at the end of kindergarten from the quality of pre-kindergarten teacher–child interactions and instruction. *Applied Developmental Science, 12*(3), 140-153.

Burchinal, M. R., McCartney, K., Steinberg, L., Crosnoe, R., Friedman, S. L., McLoyd, V., & Pianta, R. (2011). Examining the Black–White achievement gap among low-income children using the NICHD study of early child care and youth development. *Child Development, 82*(5), 1404-1420.

Carrow-Woolfolk, E. (1995). *Oral and Written Language Scales*. Circle Pines, MN: American Guidance Service.

Cassidy, D. J., Hestenes, L. L., Hansen, J. K., Hegde, A., Shim, J., & Hestenes, S. (2005). Revisiting the two faces of child care quality: Structure and process. *Early Education and Development, 16*(4), 505-520.

Centre for Reviews and Dissemination, University of York (2001) *Undertaking systematic reviews of research on effectiveness: CRD's guidance for carrying out or commissioning reviews*. Research Report. CRD Report 4 (2nd Ed.). NHS Centre for Reviews and Dissemination, York, UK.

Clarke-Stewart, K. A., & Gruber, C. (1984). Day care forms and features. In R. C. Ainslie (Ed.), *Quality Variations in Day Care* (pp. 35-62). New York: Praeger.

Clifford, R. M., Reszka, S. S., & Rossbach, H. (2010). *Reliability and Validity of the Early Childhood Environment Rating Scale*. Chapel Hill: University of North Carolina, FPG Child Development Institute.

Cohen, J. (1988). Statistical power analysis for the behavioural sciences (2nd Ed.). Hillsdale, NJ: Erlbaum.

Committee on Developmental Outcomes and Assessments for Young Children. (2008). Measuring Quality in Early Childhood Environments. In C. E. Snow & S. B. V. Hemel (Eds.), *Early Childhood Assessment: Why, What, and How* (pp. 145-178). Washington, D.C.: The National Academies Press.

Comprehensive Meta-Analysis Version 2. (2006). Englewood, NJ: Biostat.

Dowling, A., & O'Malley, K. (2009). Preschool education in Australia. Melbourne: Australian Council for Educational Research.

Dunn, L. (1993). Proximal and distal features of day care quality and children's development. *Early Childhood Research Quarterly, 8*(2), 167-192.
Dunn, L. M., & Dunn, L. M. (1981). Peabody Picture Vocabulary Test - Revised. Circle Pines, MN: American Guidance Service.
Dunn, L. M., & Dunn, L. M. (1997). Peabody Picture Vocabulary Test - Third Edition Manual. Circle Pines, MN: American Guidance Service.
Early, D., Maxwell, K., Burchinal, M., Alva, S., Bender, R., Bryant, D., . . . Griffin, J. (2007). Teachers' education, classroom quality, and young children's academic skills: Results from seven studies of preschool programs. *Child Development, 78*(2), 558-580.
Editorial Projects in Education Research Center. (2011). Achievement Gap. *Education Week*. http://www.edweek.org/ew/issues/achievement-gap/
Elliott, A. (2006). Early childhood education: Pathways to quality and equity for all children. *Australian Education Review*. Camberwell: Australian Council for Educational Research Press.
Espinosa, L. M. (2002). High-Quality Preschool: Why We Need It and What it Looks Like. New Brunswick, NJ: National Institute for Early Education Research.
Elvik, R. (1999). *Assessing the Validity of Evaluation Research by Means of Meta-Analysis: Case Illustrations from Road Safety Research*. PhD, University of Oslo, Oslo.
Farrington, D. P. (2003). Methodological quality standards for evaluation research. *The Annals of the American Academy of Political and Social Science, 587*(1), 49-68.
File, N., & Gullo, D. F. (2002). A comparison of early childhood and elementary education students' beliefs about primary classroom teaching practices. *Early Childhood Research Quarterly, 17*(1), 126-137.
Goodman, R. (1997). The strengths and difficulties questionnaire: A research note. *Journal of Child Psychology and Psychiatry, 38*(5), 581-586.
Gordon, A. M., & Browne, K. W. (2014). Types of Programs. In L. Ganster (Ed.), *Beginnings and Beyond: Foundations in Early Childhood Education* (9th ed., pp. 39-63). Belmont: Wadsworth - Cengage Learning.
Gordon, R. A., Fujimoto, K., Kaestner, R., Korenman, S., & Abner, K. (2013). An assessment of the validity of the ECERS-R with implications for measures of child care quality and relations to child development. *Developmental Psychology, 49*(1), 146-160.
Gordon, Hofer, Fujimoto, Risk, Kaestner, & Korenman. (2015). Identifying high-quality preschool programs: New evidence on the validity of the Early Childhood Environment Rating Scale-Revised (ECERS-R) in relation to school readiness goals. *Early Education and Development, 26*(8), 1086-1110.
Guo, G., & Harris, K. M. (2000). The mechanisms mediating the effects of poverty on children's intellectual development. *Demography, 37*(4), 431-447.
Halle, T., Vick Whittaker, J. E., & Anderson, R. (2010). Quality in Early Childhood Care and Education Settings: A Compendium of Measures (R. A. E. Child Trends for the Office of Planning, Administration for Children and Families, U.S. Department of Health and Human Services, Trans.) (2nd ed.). Washington, DC: Child Trends.
Hamre, B. K., Goffin, S. G., & Kraft-Sayre, M. (2009). Classroom Assessment Scoring System (CLASS) Implementation Guide. http://www.teachstone.org/about-the-class/
Hamre, B. K., & Pianta, R. C. (2005). Can instructional and emotional support in the first-grade classroom make a difference for children at risk of school failure? *Child Development, 76*(5), 949-967.
Hamre, B. K., Pianta, R. C., Burchinal, M., Field, S., LoCasale-Crouch, J., Downer, J. T., . . . Scott-Little, C. (2012). A course on effective teacher-child interactions: Effects on teacher beliefs, knowledge, and observed practice. *American Educational Research Journal, 49*(1), 88-123.
Harms, T., & Clifford, R. M. (1980). *Early Childhood Environment Rating Scale*. New York: Teachers College Press.
Harms, T., & Clifford, R. M. (1983). Assessing preschool environments with the Early Childhood Environment Rating Scale. *Studies in Educational Evaluation, 8*(3), 261-269.
Harms, T., Clifford, R. M., & Cryer, D. (1998). *Early Childhood Environment Rating Scale-Revised Edition*. New York: Teachers College Press.
Harms, T., Cryer, D., & Clifford, R. M. (2003). *Infant/Toddler Environment Rating Scale-Revised Edition*. New York, NY: Teachers College Press.
Hedges, L. V., & Olkin, I. (1985). Statistical methods for meta-analysis. London: Academic Press.

Hedges, L. V., & Hedberg, E. C. (2007). Intraclass Correlation Values for Planning Group-Randomized Trials in Education. *Educational Evaluation and Policy Analysis, 29*(1), 60-87.
Helburn, S. W., Culkin, M. L., Morris, J., Mocan, H. N., Howes, C., Phillipsen, L., . . . Rustici, J. (1995). Cost, Quality, and Child Outcomes in Child Care Centers: Executive Summary. Denver: University of Colorado.
Higgins, J. P., & Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine, 21*(11), 1539-1558.
Hightower, A. D., Work, W. C., Cowen, E. L., Lotyczewski, B. S., Spinell, A. P., & Guare, J. C. (1986). Teacher-Child Rating Scale 2.1. *School Psychology Review, 15*(3), 393-409.
Homel, R. (2005). Developmental crime prevention. In N. Tilley (Ed.), *Handbook of crime prevention and community safety* (pp. 71-106). Cullompton, Devon, U.K: Willan Publishing.
Homel, R., Freiberg, K., Lamb, C., Leech, M., Carr, A., Hampshire, A., . . . Batchelor, S. (2006). *The Pathways to Prevention Project: The first five years 1999-2004*. Brisbane: Griffith University and Mission Australia.
Howes, C. (1988). Relations between early child care and schooling. *Developmental Psychology, 24*(1), 53-57.
Howes, C., Burchinal, M., Pianta, R., Bryant, D., Early, D., Clifford, R., & Barbarin, O. (2008). Ready to learn? Children's pre-academic achievement in pre-Kindergarten programs. *Early Childhood Research Quarterly, 23*(1), 27-50.
Howes, C., Whitebook, M., & Phillips, D. (1992). Teacher characteristics and effective teaching in child care: Findings from the national child care staffing study. *Child and Youth Care Forum, 21*(6), 399-414.
Huntsman, L. (2008). Determinants of quality in child care: A review of the research evidence. Ashfield, NSW: Centre for Parenting and Research, Service System Development, NSW Department of Community Services.
International Labour Organization. (2012). Global Dialogue Forum on Conditions of Personnel in Early Childhood Education. Geneva: International Labour Office, Sectoral Activities Department.
Katz, L. G. (1992). Multiple perspectives on the quality of early childhood programmes. *European Early Childhood Education Research Journal, 1*(2), 5-9.
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, Rutgers University.
Kirk, R. E. (1995). *Experimental design: Procedures for the behavioral sciences* (3rd ed.). Pacific Grove, CA: Brooks/Cole.
Kontos, S. J. (1991). Child care quality, family background, and children's development. *Early Childhood Research Quarterly, 6*(2), 249-262.
La Paro, K. M., Thomason, A. C., Lower, J. K., Kintner-Duffy, V. L., & Cassidy, D. J. (2012). Examining the definition and measurement of quality in early childhood education: A review of studies using the ECERS-R from 2003 to 2010. *Early Childhood Research & Practice, 14*(1), n1.
Layzer, J. I., & Goodson, B. D. (2006). The "quality " of early care and education settings: definitional and measurement issues. *Evaluation Review, 30*(5), 556-576.
Lee, V. E., & Burkam, D. T. (2002). *Inequality at the Starting Gate: Social Background Differences in Achievement as Children Begin School*. Washington, DC: Economic Policy Institute.
Lipsey, M. W., & Wilson, D. B. (2000). *Practical Meta-Analysis* (Vol. 49). Thousand Oaks, CA: SAGE Publications.
Love, J. M., Constantine, J., Paulsell, D., Boller, K., Ross, C., Raikes, H., . . . Brooks-Gunn, J. (2004). The Role of Early Head Start Programs in Addressing the Child Care Needs of Low-Income Families with Infants and Toddlers: Influences on Child Care Use and Quality. Washington, DC: Administration for Children and Families, U.S. Department of Health and Human Services.
Love, J. M., Meckstroth, A., & Sprachman, S. (1997). Measuring the Quality of Program Environments in Head Start and Other Early Childhood Programs: A Review and Recommendations for Future Research. Washington, DC: U.S. Departgment of Education, National Center for Education Statistics.

McGrath, R., & Meyer, G. (2006). When effect sizes disagree: the case of r and d. *Psychological methods, 11*(4), 386-401.
Magnuson, K. A., & Waldfogel, J. (2005). Early childhood care and education: Effects on ethnic and racial gaps in school readiness. *The Future of Children, 15*(1), 169-196.
Manning, M. (2008). *Economic Evaluation of the Effects of Early Childhood Intervention Programs on Adolescent Outcomes.* (PhD), Griffith University, Brisbane.
Manning, M. (2004). Measuring the costs of community-based developmental prevention programs in Australia. Hons., Griffith University, Brisbane.
Manning, M., Homel, R., & Smith, C. (2010). A meta-analysis of the effects of early developmental prevention programs in at-risk populations on non-health outcomes in adolescence. *Children and Youth Services Review, 32*(4), 506-519.
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.
McMullen, M. B., & Alat, K. (2002). Education matters in the nurturing of the beliefs of preschool caregivers and teachers. *Early Childhood Research & Practice, 4*(2), n 1.
Melchers, P., & Preuss, U. (2003). *Kaufman Assessment Battery for Children (KABC)* (6th ed.). Göttingen: Hogrefe.
Mitchell, L. (2010). Constructions of childhood in early childhood education policy debate in New Zealand. *Contemporary Issues in Early Childhood, 11*(4), 328-341.
Montes, G., Hightower, A. D., Brugger, L., & Moustafa, E. (2005). Quality child care and socio-emotional risk factors: No evidence of diminishing returns for urban children. *Early Childhood Research Quarterly, 20*(3), 361-372.
National Association for the Education of Young Children. (1991). Guidelines for Appropriate Curriculum Content and Assessment in Programs Serving Children Ages 3 through 8. Washington, DC: National Association for the Education of Young Children.
National Institute of Child Health and Human Development Early Child Care Research Network. (1999). Child outcomes when child care center classes meet recommended standards for quality. NICHD Early Child Care Research Network. *American Journal of Public Health, 89*(7), 1072-1077.
National Institute of Child Health and Human Development Early Child Care Research Network. (2005). *Child Care and Child Development: Results from the NICHD Study of Early Child Care and Youth Development.* New York: The Guilford Press.
New Zealand House of Representatives. (2013). Inquiry into Improving Child Health Outcomes and Preventing Child Abuse, with a Focus on Preconception until Three Years of Age: New Zealand Kindergartens Inc.
OECD. (2013). Education at a Glance 2013: OECD Indicators: OECD Publishing.
Peisner-Feinberg, E. S., & Burchinal, M. (1997). Concurrent relations between child care quality and child outcomes: The study of cost, quality, and outcomes in child care centers. *Merrill-Palmer Quarterly, 43*(3), 451-477.
Peisner-Feinberg, E. S., Burchinal, M. R., Clifford, R. M., Culkin, M. L., Howes, C., Sharon Lynn, K., & Yazejian, N. (2001). The relation of preschool child-care quality to children's cognitive and social developmental trajectories through second grade. *Child Development, 72*(5), 1534-1553.
Peisner-Feinberg, E. S., Burchinal, M. R., Clifford, R. M., Howes, C., Kagan, S. L., Yazejian, N., . . . Zelazo, J. (1999). The Children of the Cost, Quality, and Outcomes Study Go to School: Technical Report. Chapel Hill: University of North Carolina at Chapel Hill, Frank Porter Graham Child Development Center.
Peisner-Feinberg, E. S., Burchinal, M. R., Howes, C., & Cryer, D. (1997). The prediction of process quality from structural features of child care. *Early Childhood Research Quarterly, 12*(3), 281-303.
Phillips, D., Mekos, D., Scarr, S., McCartney, K., & Abbott–Shim, M. (2000). Within and beyond the classroom door: assessing quality in child care centers. *Early Childhood Research Quarterly, 15*(4), 475-496.
Phillipsen, L., Burchinal, M., Howes, C., & Cryer, D. (1997). The prediction of process quality from structural features of child care. *Early Childhood Research Quarterly, 12*(3), 281-303.
Pianta, R. C., 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. (2007). *Classroom Assessment Scoring System—CLASS*. Baltimore: Brookes.
Pianta, R. C., & Nimetz, S. L. (1991). Relationships between children and teachers: Associations with classroom and home behavior. *Journal of Applied Developmental Psychology, 12*(3), 379-393.
Ramey, C. T., & Ramey, S. L. (2004). Early learning and school readiness: Can early intervention make a difference? *Merrill-Palmer Quarterly, 50*(4), 471-491.
Raudenbush, S. W. (2009). Analyzing effect sizes: Random-effects models. In H. Cooper, L. V. Hedges & J. C. Valentine (Eds.), *The Handbook of Research Synthesis and Meta-Analysis* (2nd ed.). New York: Russell Sage.
Reardon, S. F., & Galindo, C. (2009). The Hispanic-White achievement gap in math and reading in the elementary grades. *American Educational Research Journal, 46*(3), 853-891.
Reynolds, A. J., 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*(5), 1299-1328.
Rosenthal, R., & Vandell, D. L. (1996). Quality of care at school-aged child-care programs: Regulatable features, observed experiences, child perspectives, and parent perspectives. *Child Development, 67*(5), 2434-2445.
Sakai, L., Whitebook, M., Wishard, A., & Howes, C. (2003). Evaluating the Early Childhood Environment Rating Scale (ECERS): Assessing differences between the first and revised edition. *Early Childhood Research Quarterly, 18*(4), 427-445.
Sammons, P. (2010). Does pre-school make a difference? Identifying the impact of pre-school on children's cognitive and social behavioural development at different ages. 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. 92-113). New York: Routledge.
Sammons, P., Sylva, K., Melhuish, E., Siraj-Blatchford, I., Taggart, B., & Elliot, K. (2003a). Measuring the impact of pre-school on children's cognitive progress over the pre-school period. In K. Sylva (Ed.), *The Effective Provision of Pre-School Education (EPPE) Project*. London: Institute of Education.
Sammons, P., Sylva, K., Melhuish, E., Siraj-Blatchford, I., Taggart, B., & Elliot, K. (2003b). Measuring the impact of pre-school on children's social/behavioural development over the pre-school period. In K. Sylva (Ed.), *The Effective Provision of Pre-School Education (EPPE) Project*. London: Institute of Education.
Sammons, P., Sylva, K., Melhuish, E. C., Siraj-Blatchford, I., Taggart, B., & Elliot, K. (2002). The Effective Provision of Pre-school Education Project (EPPE), Technical Paper 8a: Measuring the Impact of Pre-School on Children's Cognitive Progress over the Pre-School Period. London: DfES/Institute of Education, University of London.
Sánchez-Meca, J., Marín-Martínez, F., & Chacón-Moscoso, S. (2003). Effect-size indices for dichotomized outcomes in meta-analysis. *Psychological Methods. 8*(4): 448-467.
Schweinhart, L. J., Montie, J., Xiang, Z., Barnett, W. S., Belfield, C. R., & Nores, M. (2005). *Lifetime Effects: The HighScope Perry Preschool Study through Age 40*. Ypsilanti, Michigan: High/Scope Press.
Snider, M. H., & Fu, V. R. (1990). The effects of specialized education and job experience on early childhood teachers' knowledge of developmentally appropriate practice. *Early Childhood Research Quarterly, 5*(1), 69-78.
Spiker, D., Hebbeler, K. M., & Barton, L. R. (2011). Measuring Quality of ECE Programs for Children with Disabilities. In M. Zaslow, I. Martinez-Beck, K. Tout & T. Halle (Eds.), *Quality Measurement in Early Childhood Settings* (pp. 229-256). Washington, DC: Brookes Publishing.
Sterne, J. A. C., & Egger, M. (2005). Regression methods to detect publication and other bias in meta-analysis. In H. R. Rothstein, A. J. Sutton, & M. Borenstein (Eds.) *Publication bias in meta-analysis: Prevention, assessment, and adjustments* (pp. 99--110). Chichester, England: Wiley.
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.
Taguma, M., Litjens, I., & Makowiecki, K. (2012). Quality Matters in Early Childhood Education and Care: New Zealand 2012. http://www.oecd.org/edu/school/NEW ZEALAND policy profile - published 3-8-2012.pdf

Tanner-Smith EE, Tipton E. Research Synthesis Methods. 2013. Robust variance estimation with dependent effect sizes: Practical considerations including a software tutorial in Stata and SPSS. *Advance Online Publication*.
Tout, K., Zaslow, M., Halle, T., & Forrey, N. (2009). Issues for the Next Decade of Quality Rating and Improvement Systems. Washington, DC: Child Trends.
Vandell, D. L., & Wolfe, B. (2000). Child Care Quality: Does It Matter and Does It Need to be Improved? Madison: Institute for Research on Poverty, University of Wisconsin, Madison.
Vartuli, S. (1999). How early childhood teacher beliefs vary across grade. *Early Childhood Research Quarterly, 14*(4), 489-514.
Vitaro, F., Barker, E. D., Brendgen, M., & Tremblay, R. E. (2011). Pathways explaining the reduction of adult criminal behaviour by a randomized preventive intervention for disruptive kindergarten children. *Journal of Child Psychology and Psychiatry, 53*(7), 748-756.
Warash, B. G., Ward, C., & Rotilie, S. (2008). An exploratory study of the application of early childhood environment rating scale criteria. *Education, 128*(4), 645-659.
Whitebook, M. (1989). Who Cares? Child Care Teachers and the Quality of Care in America. Final Report, National Child Care Staffing Study. Berkeley, CA: Child Care Employee Project.
Whitebook, M. (2003a). Bachelor's Degrees are Best: Higher Qualifications for Pre-Kindergarten Teachers Lead to Better Learning Environments for Children. Washington, DC: Trust for Early Education.
Whitebook, M. (2003b). Early Education Quality: Higher Teacher Qualifications for Better Learning Environments - A Review of the Literature. Berkeley, CA: Center for the Study of Child Care Employment.
Whitebook, M., Sakai, L., Gerber, E., & Howes, C. (2001). Then and Now: Changes in Child Care Staffing, 1994-2000. Washington, DC: Center for the Child Care Workforce.
Wolery, M. (2004). Assessing children's environments. In M. McLean, M. Wolery & D. Bailey (Eds.), *Assessing Infants and Preschoolers with Special Needs* (3rd ed., pp. 204-235). Upper Saddle River, NJ: Pearson Merrill Prentice Hall.
Woodcock, R. W., & Johnson, M. B. (1990). *Woodcock-Johnson Psycho-Educational Battery-Revised*. Allen, TX: DLM Teaching Resources.
Zill, N., & Resnick, G. (1998). *Head Start Family and Child Experiences Survey (FACES)*. Rockville, MD: Westat.

About this review

REVIEW AUTHORS

Lead review author:

Name:	Matthew Manning
Title:	Associate Professor
Affiliation:	Australian National University
Address:	ANU Centre for Social Research and Methods
City, State, Province or County:	Canberra, ACT
Postal Code:	4122
Country:	Australia
Phone:	+61(0) 7 3735 3880
Email:	Matthew.Manning@anu.edu.au

Co-author(s): (There should be at least one co-author)

Name:	Susanne Garvis
Title:	Professor
Affiliation:	University of Gothenburg
Address:	Department of Education, Communication and Learning University of Gothenburg Box 300
City, State, Province or County:	Gothenburg, 41767
Country:	Sweden
Phone:	+46(0) 31 786 50 61
Email:	Susanne.garvis@gu.se

Co-author(s): (There should be at least one co-author)

Name:	Christopher Fleming
Title:	Associate Professor
Affiliation:	Griffith University
Address:	Griffith Business School

City, State, Province or County:	Brisbane, Queensland
Postal Code:	4101
Country:	Australia
Phone:	+61(0) 7 373 57061
Email:	Chris.Fleming@griffith.edu.au

Co-author(s): (There should be at least one co-author)

Name:	Gabriel T.W. Wong
Title:	Mr
Affiliation:	Griffith University
Address:	School of Criminology and Criminal Justice, Mt Gravatt Campus
City, State, Province or County:	Brisbane, Queensland
Postal Code:	4122
Country:	Australia
Phone:	+61(0) 7 3735 1153
Email:	Gabriel.Wong2@griffithuni.edu.au

ROLES AND RESPONSIBILITIES

Content: Dr Matthew Manning has taught and published extensively in the developmental prevention area focusing explicitly on the early years, which includes preschool education. Dr Susanne Garvis has taught and published extensively in the early childhood education area. She has extensive experience working with educators in the early childhood education sector.

Systematic review methods: Dr Matthew Manning has conducted and published several meta-analyses in high-ranking peer-reviewed journals including the Campbell Collaboration.

Statistical analysis: Dr Matthew Manning, Dr Christopher Fleming and Mr Gabriel Wong have extensive training in applied statistics and also those statistical procedures that are applied in meta-analysis.

Information retrieval: Dr Matthew Manning, Dr Christopher Fleming and Mr Gabriel Wong have been involved in developing search strategies using the Campbell and Cochrane technique.

SOURCES OF SUPPORT

Drs Matthew Manning and Susanne Garvis received internal funding from Griffith University to conduct this research.

DECLARATIONS OF INTEREST

None.

PLANS FOR UPDATING THE REVIEW

Dr Matthew Manning and Dr Susanne Garvis will be responsible for updating this review. We anticipate that this review will be updated as new evidence is collected and subsequent empirical papers written.

AUTHOR DECLARATION

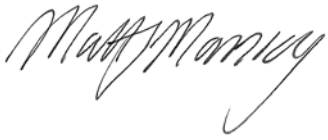
Authors' responsibilities

By completing this form, you accept responsibility for maintaining the review in light of new evidence, comments and criticisms, and other developments, and updating the review at least once every five years, or, if requested, transferring responsibility for maintaining the review to others as agreed with the Coordinating Group. If an update is not submitted according to agreed plans, or if we are unable to contact you for an extended period, the relevant Coordinating Group has the right to propose the update to alternative authors.

Publication in the Campbell Library

The Campbell Collaboration places no restrictions on publication of the findings of a Campbell systematic review in a more abbreviated form as a journal article either before or after the publication of the monograph version in *Campbell Systematic Reviews*. Some journals, however, have restrictions that preclude publication of findings that have been, or will be, reported elsewhere, and authors considering publication in such a journal should be aware of possible conflict with publication of the monograph version in *Campbell Systematic Reviews*. Publication in a journal after publication or in press status in *Campbell Systematic Reviews* should acknowledge the Campbell version and include a citation to it. Note that systematic reviews published in *Campbell Systematic Reviews* and co-registered with the Cochrane Collaboration may have additional requirements or restrictions for co-publication. Review authors accept responsibility for meeting any co-publication requirements.

I understand the commitment required to update a Campbell review, and agree to publish in the Campbell Library. Signed on behalf of the authors:

Form completed by:  **Date:** 4 November 2016

Appendix 1

CODING SHEET

I. Document Information

1. Document ID:
2. Surname of First Author:
3. Title of Study:
4. Publication type
5. Journal:
6. Volume:
7. Issue:
8. Coder's Initials:
9. Date Eligibility Determined:

II. Criteria for Eligibility – A study must meet the following criteria to be eligible for inclusion.

1a. The study is a comparative or correlational study that evaluates the association between teacher qualification and the quality of the early childhood care and education environment (as measured by ECERS/ECERS-R/ITERS/ITERS-R)

(Options include: Yes – go to *1b* / No – go to *1e*)

1b. The study reports on at least one outcome (i.e. total or subscale scores on ECERS/ECERS-R/ITERS/ITERS-R)

(Options include: Yes – go to *1c* / No – go to *1e*)

1c. Can an effect size be calculated from data presented in the paper?

(Options include: Yes – go to *1d* /No – go to *1e*)

1d. Is the language used in the study understandable or translatable to a language that is understandable to the researchers?

(Options include: Yes – go to *2* / No – go to *1e*)

*If the study does not meet the above criteria answer the following question:

1e. The study is a review article that is relevant to this project (e.g. may have references to other studies that are relevant to the project, may have pertinent background information)

2. Eligibility status (tick appropriate box)

☐ Eligible

☐ Not Eligible

☐Relevant to Review

Notes:

III. Coding Protocol for Eligible Papers

Reference Information

1. Document ID:

2. Study Authors:

3. Study Title:

4a. Publication Type: Book (Options include: Book, Book Chapter, Journal Articles, Thesis or Doctoral Dissertation, Government Report, Technical Report, Conference Paper, and Other (specify))

4b. Specify (Other):

5. Publication Date (Year):

6a. Journal Name:

6b. Journal Volume:

6c. Journal Issue:

7. Date Range of Research (When research was conducted):

Start:

Finish:

8. Source of Funding for Study:

9. Country of Publication:

10. Date Coded:

11. Coder's Initials:

IV: Analysis Characteristics

The following questions are regarding the population identified in the study.

1a. Unit of analysis (Individual specify):

1b. Unit of analysis (Classroom specify):

1c. Unit of analysis (Centre specify):

1d. Unit of analysis (Sector specify):

1e. Unit of analysis (Other specify):

V: Sample Characteristics

1a. What was the target population of the study? Classroom (Options include: Classroom, Centre, and Other (specify))

1b. Specify (other):

2. Total population of target population (if known):

3. Gender composition of target population: Mixed (Options include: Mixed, Mostly Male, Mostly Female, and Unknown)

4. Age composition of target population: Baby (Options include: Baby, Infants, Toddlers, and Preschoolers)

5. Socio-economic status of target population: Mostly below poverty line (Options include: Mostly below poverty line, Mostly above poverty line, Unknown/Not mentioned, and Other (specify))

6a. Race/ethnicity of the sample: African_American (Options include: African American, Asian, White/Caucasian, Indigenous, and Other (specify))

6b. Percentage Other (specify):

7. Country of study:

8. What was the initial sample size recruited into the study and what was the final N (sample number related to outcomes examined in the review)?

Initial N:

Final N:

VI: Methodology/Research Design

1a. Type of study: Pre-post Test (no control group) (Options include: Pre-post test (no control group), Non-randomised comparison study, Non-equivalent control group, Correlational study, and Other (specify))

1b. Other (specify):

VII: Outcomes Reported

1. How many outcomes are reported in the study?

2. What is the specific outcome recorded on this coding sheet?

3. Was it the primary outcome of the study? Yes (Options include: Yes, No, and Cannot tell/no priorities outcomes reported)

4a. Was this initially intended as an outcome of the study? Yes (Options include: Yes, No, and Cannot tell)

4b. If no, explain why:

VIII: Dependent Variable

1a. What type of measurement tool was used to measure the outcome in this study? ECERS

1b. Other data (specify):

2. Did the researcher assess the quality of the data collected? Yes (Options include: Yes and No)

3a. Did the researcher(s) express any concerns regarding the quality of the data? Yes (Options include: Yes and No)

3b. If yes, explain:

4a. Does the evaluation data correspond to the initial stated problem? Yes (Options include: Yes and No)

4b. If no, explain the discrepancy:

5. Was the assessor an internal staff member or independent?): Internal assessor (Options include: Internal assessor and Independent assessor)

IX: Effect Size/ Reports of Statistical Significance

1. What is the total sample size in the analysis (based on the unit of analysis for this outcome)? N=

2. How many study groups are identified in the evaluation?

3a. What is the total sample size of the study group 1? N=

3b. What is the total sample size of the study group 2? N=

3c. What is the total sample size of the study group 3? N=

3d. What is the total sample size of the study group 4? N=

4. What is the total sample size of the comparison group? N=

5. Raw difference favours (i.e. shows more success for): Study group 1 (Options include: Study group 1, Study group 2, Study group 3, Study group4, Comparison group, Neither (equal), Cannot tell (statistically insignificant report))

6. Did a test of statistical significance indicate statistically significant differences between either study or comparison groups or the pre and post tested study group? Yes (Options include: Yes, No, Cannot tell, and N/A (no testing completed))

7a. Was a standardised effect size reported? Yes (Options include: Yes and No)

7b. If yes, what was the effect size?

7c. If yes, what page was the effect size found on?

7d. If no, is there data available to calculate an effect size? Yes

7e. Type of data effect size can be calculated from: mean and standard deviations (Options include: Mean and SD, t-value, F-value, Point biserial correlation coefficient, Pearson correlation coefficient, and Other (specify))

7f. Other (specify):

8. Did the evaluation control for validity by using multivariate methods (i.e. regression) to assess the impact of an intervention/program on a given outcome? Yes (Options include: Yes and No)

9. If yes, did the analysis find that the intervention/program reduced the outcome at a statistically significant level (p=.05)? Yes (Options include: Yes and No)

10. What significance level was used? <0.001 (Options include: <0.001, <0.01, <0.05, and <0.1)

X: Data

<u>Means and Standard Deviations</u>

1a. Study group 1 N=

1b. Study group 2 N=

1c. Study group 3 N=

1d. Study group 4 N=

2. Comparison group N=

3a. Study Group 1 mean

3b. Study Group 2 mean

3c. Study Group 3 mean

3d. Study Group 4 mean

4. Comparison group mean

5a. Study group 1 standard deviation

5b. Study group 2 standard deviation

5c. Study group 3 standard deviation

5d. Study group 4 standard deviation

6. Comparison group standard deviation

Significance tests

1. *t*-value

2. *F*-value

3. Chi-square value

4. Calculated effect size

XI: Conclusions by authors

Note: This section provides detail by authors regarding the effectiveness of the intervention with respect to the outcome/problem being addressed on this coding sheet.

1. Conclusion about the impact of the intervention

Study group 1: Authors conclude effect is confirmed (Options include: Authors conclude effect is confirmed, Authors conclude effect is not confirmed, and Unknown/No conclusion provided)

Study group 2: Authors conclude effect is confirmed (Options include: Authors conclude effect is confirmed, Authors conclude effect is not confirmed, Unknown/No conclusion provided, and N/A)

Study group 3: Authors conclude effect is confirmed (Options include: Authors conclude effect is confirmed, Authors conclude effect is not confirmed, Unknown/No conclusion provided, and N/A)

Study group 4: Authors conclude effect is confirmed (Options include: Authors conclude effect is confirmed, Authors conclude effect is not confirmed, Unknown/No conclusion provided, and N/A)

2. Did the authors conclude that the intervention was beneficial? Yes (Options include: Yes, No, and Cannot tell)

3a. Did the authors conclude that a relationship existed between the intervention and the quality of the early childhood learning environment? Yes (Options include: Yes, No, and Cannot tell)

3b. If yes, add notes about conclusions made by authors

CampbellCollaboration

About this review

Poor quality early childhood education and care can be detrimental to the development of children as it could potentially lead to poor social, emotional, educational, health, economic and behavioural outcomes.

The lack of consensus as to the strength of the relationship between teacher qualification and the quality of the early childhood learning environment has made it difficult for policy makers and educational practitioners alike to settle on strategies that would enhance the learning outcomes for children in their early stages of education.

This review examines the current empirical evidence on the correlation between teacher qualifications and the quality of early childhood learning environments.

The Campbell Collaboration
info@campbellcollaboration.org
Phone: (+47) 23 25 50 00

Mailing address:
P.O. Box 4404, Nydalen
N-0403 Oslo, Norway

Visiting address:
Pilestredet Park 7
(Entrance from Stensberggata)

Website:
www.campbellcollaboration.org