Establishing Quality in Preschool Language and Literacy Environments

On the quality of Danish preschools, and pathways

to improving it via professional development

Ph.D. Dissertation

Justin Markussen-Brown

Institute of Language and Communication

University of Southern Denmark

April 2015

Quality is not an act, it is a habit.

-Aristotle

Acknowledgements

First and foremost, I would like to thank my supervisors Professor Dorthe Bleses and Professor Laura Justice. This work would not be possible without their invaluable guidance and encouragement. Thanks for always believing in me. Secondly, I would like to thank my many collaborators, from whom I have also learned so much: Shayne Piasta, Werner Vach, Anders Højen, Jessica Logan, and most certainly Mr. Carsten Juhl. Thank you Laila Kjærbæk Hansen for your coaching. Also thank you to Mette Kjær Andersen for always being available. Thank you, everyone!

I would also like to thank all my friends and colleagues at the University of Southern Denmark who have supported me these past three years. It is an honour to work with such wonderful and interesting people! May we all meet at Carlsens Kvarter every Friday at 7:00 p.m. forever!

In particular, I would like to thank my dear friends and comrades-in-(experimental)-arms Marit Clausen, Burcak Aktürk, and Fabio Trecca. You guys add *quality* to every environment you manifest in. Thank you for your never-ending friendship and awesomeness.

Lastly, I would like to thank my family, (mor, Annetta, Torben and Shannon) for their immeasurable love and support. We may be separated by the North Atlantic (and then some) but thinking of you guys can always make me happy – even in the darkest 11th hour of dissertation-writing. *Elsker jer.*

Justin

Odense, April 2015

Table of Contents

Summary

Childcare holds a prominent position in the Danish welfare system. In fact, 95% of four-year-olds attend state-sponsored preschool (OECD, 2013), and Denmark spends the second highest percentage of its gross national product in the world on childcare (OECD, 2009). Despite this massive investment, little research exists regarding the quality of Danish preschools. In particular, very little is known regarding the quality of the language and literacy environments. This may be problematic because children's language and emergent literacy skills are both influenced by preschool quality (Pianta et al., 2005), and predictive of reading skills in school (Chaney, 1998; Lonigan, Burgess, & Anthony, 2000; National Early Literacy Panel, 2008; Roth, Speece, & Cooper, 2002; Storch & Whitehurst, 2002). Unfortunately, socially disadvantaged children in Danish preschools are more likely to exhibit low language and emergent literacy skills (Bleses, Højen, Jørgensen, Jensen, & Vach, 2010). One potential pathway of improving the quality of preschool learning environments – and thereby children's language and emergent literacy outcomes – is via professional development (PD). However, the degree to which current models of PD succeed in improving preschool quality and child outcomes is unclear in the literature.

This Ph.D. dissertation is an anthology of three research articles that each investigates questions related to language and emergent literacy development in a preschool setting. In its entirety, the dissertation can advise Danish policymakers on both the global quality of Danish preschools, as well as the quality of their language and literacy learning environments. The dissertation also presents research that can advise on the effect of PD interventions for preschool teachers with a language and/or literacy focus, including recommendations regarding which factors improve the likelihood that an intervention will have effect. Individually, the research articles are independent investigations intended for publication in international, peer-reviewed journals.

The first paper is an investigation of the process and structural quality of Danish preschools. Process quality was measured at the teacher level (n = 506) using the Classroom Assessment Scoring System (CLASS; Pianta, La Paro, & Hamre, 2008), which is a measure of the quality of interactions between preschool teachers and children on three subscales of practice. Structural quality was measured at the preschool level (n = 293) using the Classroom Literacy Assessment Profile (CLOP; Crane Center for Early Childhood Research and Policy, 2012). The CLOP measures the quantity of literacy materials such as books and toys within the child's reach, and other physical attributes that support children's learning of language and literacy.

The results of the CLASS investigation revealed medium to high quality on the subscales of *emotional support* and *classroom organization*. In practice, this means that Danish preschool teachers provided children with warm and emotionally sensitive environments, and teachers furthermore demonstrated proficiency in behavioral management skills. However, results for the *instructional support* subscale, which measures the quality of language and cognitive stimulation in interactions, were mostly low. The evidence suggests that preschool teachers were unfamiliar with pedagogical skills and strategies that support children's language and cognitive development. Teachers' education had no effect on their interactional skills. However, significantly higher results were found for teachers who had participated in a four-day PD on language development. Furthermore, socially disadvantaged children were more likely to attend preschools with lower emotional support and classroom organization.

The results from the CLOP investigation indicated mostly low levels of structural supports for children's learning of literacy. Although most preschools provided books to children, the presence of specific kinds of books (such as alphabet books) or other supports (such as a take-home library) was rarer. Overall, the CLOP investigation suggests that Danish

preschools generally lack an intentional, evidence-based approach to the physical learning environment with regards to language and literacy.

The second study in this anthology was a systematic review and meta-analysis of the effects of PD interventions with a language and/or emergent literacy focus. Effect sizes (the standard mean difference, SMD) were estimated for the teacher-level outcomes of process quality, structural quality, and teacher knowledge of language and literacy. At the child level, effect sizes were calculated for receptive vocabulary, phonological awareness, and alphabet knowledge. In addition to the overall effects of the PD interventions, we also investigated the extent to which gains in process quality predicted gains in child outcomes, and we also performed a number of sub-group analyses to investigate which factors potentially explained variation in effects.

The primary research literature was found by searching several databases for key terms regarding PD and children's language and literacy development. To be included in the review, studies had to measure the effects of a PD intervention on process quality, structural quality, and/or teacher knowledge. Furthermore, teachers had to be in-service preschool teachers, children had to be 3-6-years old, and data necessary for calculating effect sizes was required. In total, 22 studies met our inclusion criteria resulting in 27 trials.

Results of the review revealed significant effects for process quality (SMD = 0.52) and structural quality (SMD = 1.07), but not for teacher knowledge (SMD = 0.15). At the child level, a non-significant SMD of 0.21 was estimated for receptive vocabulary, whereas significant effects were found for phonological awareness (SMD = 0.46) and for alphabet knowledge (SMD = 0.18). Effects for process quality did not predict children's gains. Sub-group analyses of the results revealed that interventions that included coaching yielded larger effects for quality than those that did not include coaching, and courses were also effective when combined with another format of PD. We found furthermore that the number of PD

formats predicted effect sizes, as did the combined intensity of coaching and courses, and total intervention duration.

The third study of this dissertation was motivated by cognitive social theory (Bandura, 1977, 1986), and investigated the extent to which teachers demonstrated awareness of their skill performance following three days of PD. Research indicates that if trainees discover that their skill performance is lower than they expected, that they tend to strive harder to learn the new skills and restore feelings of self-efficacy (Bandura & Cervone, 1983).

In our study, preschool teachers were trained to use six scaffolding strategies, and to code their usage of these strategies from video. Three of the strategies that teachers learned provided high support to children having difficulty understanding a learning goal, and three strategies provided lower levels of support to children who required less scaffolding. By coding videos of their own skill performance, teachers were able to calibrate their knowledge of which strategies they used; however, this was only possible if they coded accurately. Therefore, our research goal was to determine the extent to which teachers could code their skill performance of scaffolding accurately. Accurate coding was interpreted as evidence that teachers were well calibrated with regards to their skill performance. Low coding accuracy was interpreted as indicating incomplete learning of the six strategies.

Results of the study indicated that teachers generally coded inaccurately after three days of PD. However, in the cases in which teachers coded accurately, it was usually in recognition that no strategies were used. In particular, we found that teachers generally overrated their usage of the high support strategies, and both under- and overrated their usage of other strategies. We also found evidence that some strategies were more challenging to code than others. Three days of PD is probably too little to create large improvements in teachers' use of scaffolding.

In its entirety, this dissertation finds that there is room for improvement in the quality of Danish preschools. PD initiatives are one pathway of improving quality, and based on the results of the meta-analysis and our study of teachers' awareness of skill performance, we recommend that such interventions utilize several delivery systems of PD, be high intensity, and have a long duration.

Resumé

Børnepasning er en vigtig del af det danske velfærdssystem. I Danmark går 95% af de 4-årige børn i børnehave (OECD, 2013), og Danmark er nr. to i verden i forhold til hvor meget landet bruger på pasning af børn i forhold til bruttonationalproduktet (OECD, 2009). Til trods for denne massive investering eksisterer der meget lidt forskning om kvaliteten af danske børnehaver. For eksempel ved vi meget lidt om kvaliteten af det sproglige og skriftsproglige miljø. Det kan være problematisk fordi børns sprogfærdigheder og før-læsefærdigheder påvirkes af børnehavekvalitet (Pianta et al., 2005), og forudsiger senere læsefærdigheder i skolen (Chaney, 1998; Lonigan, Burgess, & Anthony, 2000; National Early Literacy Panel, 2008; Roth, Speece, & Cooper, 2002; Storch & Whitehurst, 2002). Desværre har socialt udsatte børn i Danmark oftere utilstrækkelige sprog- og før-læsefærdigheder (Bleses, Højen, Jørgensen, Jensen, & Vach, 2010). Man kan forbedre læringen i børnehaver og derved børns sprog- og før-skriftsprogfærdigheder gennem efteruddannelse. Dog er det uklart i forskningslitteraturen om den typiske efteruddannelse virker.

Afhandlingen indeholder tre forskningsartikler, som undersøger udviklingen af sprog- og før-skriftsprogfærdigheder i en børnehavekontekst. Som en helhed giver afhandlingen information til beslutningstagere om den globale kvalitet og det sproglige arbejde i danske børnehaver. Afhandlingen præsenterer også forskning om effekten af pædagogisk efteruddannelse med sprogligt fokus, og den giver anbefalinger af, hvilke faktorer der er vigtige, hvis et tiltag skal lykkes. Forskningsartiklerne er særskilte undersøgelser skrevet til internationale fagfællebedømte tidsskrifter inden for området.

Den første artikel er en undersøgelse af proces-kvaliteten og strukturel-kvaliteten i danske børnehaver. Proces-kvalitet blev målt på pædagogniveauet (n = 506) med anvendelse af *Classroom Assessment Scoring System* (CLASS; Pianta, La Paro, & Hamre, 2008), som er

et måleredskab, der måler kvaliteten af interaktionen mellem pædagogen og barnet på tre subskalaer. Strukturel-kvalitet blev målt på institutionsniveauet (n = 293) med *Classroom Literacy Assessment Profile* (CLOP; Crane Center for Early Childhood Research and Policy, 2012). CLOP måler bl.a. antallet af materialer i børns rækkevide, som fremmer deres udvikling af sprog og før-skriftsprog.

Resultaterne fra undersøgelsen med CLASS viste en mellem til høj kvalitet på subskalaerne *emotional support* og *classroom organization*. I praksis betyder det, at pædagogerne skabte varme og følelsesmæssigt understøttende miljøer for børnene, og at pædagogerne viste stor dygtighed i adfærdspædagogik. Resultaterne var lave på subskalaen *instructional support*, hvilket måler kvaliteten i pædagogernes sprogunderstøttende og kognitivunderstøttende interaktioner med børn. Det indikerer, at pædagogerne ikke var fortrolige med pædagogiske redskaber og strategier som fremmer børns sproglige og kognitive udvikling. Pædagogernes uddannelse havde ingen effekt på resultaterne, men pædagogerne som havde deltaget i *Sprogpakken* scorede signifikante højere på *instructional support*. Socialt udsatte børn var ydermere mere tilbøjelige til at gå i en børnehave med lavere *emotional support* og *classroom organization*.

Resultaterne fra CLOP-undersøgelsen viste lave niveauer af strukturel-kvalitet for det meste. De fleste børnehaver satte bøger frem i børnehøjde, men de havde sjældent bøger med specifikt indhold (fx om alfabetet) eller andet som kunne fremme børns sproglige og før-skriftsproglige udvikling (fx et lånebibliotek). Tilsammen indikerer undersøgelsen, at der mangler en intentionel, evidensbaseret praksis i forhold til det fysiske læringsmiljø.

Den anden artikel i afhandlingen var et systematisk review og en metaaanalyse af den litteratur, der måler effekterne af efteruddannelse med sprogligt fokus. Effektstørrelserne (estimereret som *standard mean difference* [SMD]) blev estimeret på pædagogniveauet for proces-kvalitet, strukturel-kvalitet og pædagogernes viden af sprog og skriftsprog. På

barneniveauet blev effektstørrelserne estimeret for receptivt ordforråd, lydlig opmærksomhed og kendskab til alfabetet. Yderligere undersøgte vi om udbytterne for kvalitet forudså børnenes udbytter, og vi brugte subgruppeanalyse for at undersøge potentielle variabler, som kunne forklare variation i effektstørrelser.

De primære forskningsartikler blev fundet ved at søge i mange databaser med nøgleord om efteruddannelse og børns sprog- og før-skiftsprogsudvikling. For at blive inkluderet i reviewet skulle et studie undersøge effekten af en efteruddannelsesintervention på proces-kvalitet, strukturel-kvalitet og/eller pædagogernes viden. Yderligere skulle pædagogerne være færdiguddannede, børnene skulle være 3-6 år gamle, og de data som bruges for at estimere effektstørrelser skulle angives. I alt blev 22 studier inkluderet med 27 afprøvninger i alt.

Resultaterne for metaanalysen viste signifikante effekter for proces-kvalitet (SMD = 0.52) og strukturel-kvalitet (SMD = 1.07), men ikke for pædagogernes viden (SMD = 0.15). På barnets niveau estimerede vi en ikke-signifikant effekt for receptivt ordforråd (SMD = 0.21), men signifikante effekter for lydlig opmærksomhed (SMD = 0.46) og kendskab til alfabetet (SMD = 0.18). Udbyttet for proces-kvalitet forudså ikke udbyttet for børnene. Subgruppeanalyserne viste, at efteruddannelse som inkluderede coaching var mere effektive end interventioner som ikke gjorde. Kurser var også effektive, hvis de var i kombination med mindst et andet format af efteruddannelse. Vi fandt også frem til, at antallet af efteruddannelsesformater, efteruddannelsesintensitet og interventionens tidsforløb forudsagde effekter.

Den tredje artikel i afhandlingen blev motiveret af kognitivsocialteori (Bandura, 1977, 1986), og undersøgte i hvilket omfang pædagoger efter tre dages efteruddannelse var bevidste om, hvor ofte de brugte en vis strategi (færdighedspræstation). Forskning indikerer, at hvis en person under træning opdager, at hendes præstation er lavere end forventet, så er hun mere

tilbøjelig til at bestræbe sig på at lære den nye færdighed og dermed genoprette sin tiltro på egen evne (Bandura & Cervone, 1983).

I vores undersøgelse blev pædagoger trænet til at bruge seks stilladseringsstrategier, og til at kode deres anvendelse af disse ud fra videooptagelser. Tre af strategierne gav øget hjælp til barnet, og tre af strategierne gav mindre hjælp. Ved at kode deres anvendelse af strategierne fik pædagogerne mulighed for at tilpasse deres viden omkring deres strategianvendelse; imidlertid var det kun muligt, hvis de kunne kode med nøjagtighed. Derfor var forskningsmålet at undersøge i hvilket omfang, pædagogerne kunne kode med nøjagtighed deres anvendelse af de seks stilladseringsstrategier.

Resultaterne indikerede, at pædagogerne kodede unøjagtigt efter tre dages efteruddannelse. I de tilfælde, hvor der blev kodet rigtigt, var det hyppigst, når pædagogen ikke brugte nogen strategi. Nøjere analyse af data viste, at pædagogerne var meget tilbøjelige til at overvurdere deres anvendelse af de meget understøttende strategier, og de både under- og overvurderede deres brug af de mindre understøttende strategier. Vi fandt også evidens for, at nogle strategier var sværere at kode end andre. Tre dages efteruddannelse er tilsyneladende utilstrækkeligt for at skabe markante forbedringer i pædagogernes anvendelse af stilladseringsstrategier.

I sin helhed viser afhandlingen et stort behov for at ændre pædagogernes praksis og uddannelse for at understøtte børns sproglige, før-skriftsproglige og kognitive udvikling. Efteruddannelse er en vej mod bedre kvalitet. På baggrund af resultaterne fra det systematiske review og studiet om pædagogernes kendskab til deres egen brug af strategier og mangel herpå anbefales det at efteruddannelsen får flere formater, højere intensitet og forløber længere.

References

Bandura, A. (1977). *Social learning theory*. Englewood Cliffs, NJ: Prentice-Hall.

Bandura, A. (1986). *Social foundations of thought and action*: Englewood Cliffs, NJ: Prentice Hall.

Bandura, A., & Cervone, D. (1983). Self-evaluative and self-efficacy mechanisms governing the motivational effects of goal systems. *Journal of personality and social psychology, 45*(5), 1017.

Bleses, D., Højen, A., Jørgensen, R., Jensen, K. Ø., & Vach, W. (2010). Sprogvurdering af 3-årige - karakteristika og risikofaktorer. [Langauge screening of 3-year-olds – characteristics and risk factors]. *University of Southern Denmark working papers*, (10). http://www.humaniora.sdu.dk/boernesprog/DK/forsk/doc/E-print_10_2010.pdf

Chaney, C. (1998). Preschool language and metalinguistic skills are links to reading success. *Applied Psycholinguistics, 19*(03), 433-446. doi: doi:10.1017/S0142716400010250

Crane Center for Early Childhood Research and Policy. (2012). *The Classroom Literacy Observation Profile*. Unpublished measure. Columbus, OH.

Lonigan, C. J., Burgess, S. R., & Anthony, J. L. (2000). Development of emergent literacy and early reading skills in preschool children: Evidence from a latent-variable longitudinal study. *Developmental psychology, 36*(5), 596-613. doi: 10.1037/0012-1649.36.5.596

National Early Literacy Panel. (2008). *Developing early literacy: Report of the National Early Literacy Panel: A scientific synthesis of early literacy development and implications for intervention*. Washington, D.C.: National Institute for Literacy.

OECD. (2009). Society at a glance 2009: OECD social indicators: OECD Publishing.

OECD. (2013). Early Childhood Education and Care *Education Today 2013: The OECD Perspective*: OECD Publishing.

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. doi: 10.1207/s1532480xads0903_2

Pianta, R. C., La Paro, K. M., & Hamre, B. K. (2008). *Classroom Assessment Scoring System Manual, Pre-K*. Baltimore, MD: Brookes.

Roth, F. P., Speece, D. L., & Cooper, D. H. (2002). A Longitudinal Analysis of the Connection Between Oral Language and Early Reading. *The Journal of Educational Research, 95*(5), 259-272. doi: 10.1080/00220670209596600

Storch, S. A., & Whitehurst, G. J. (2002). Oral language and code-related precursors to reading: evidence from a longitudinal structural model. *Developmental psychology, 38*(6), 934.

The Structure of This Dissertation

This anthological dissertation is divided into two parts. Part 1 is the so-called "framework" of the dissertation, whose main purpose is to contextualize the research articles that are presented in Part 2. The framework is useful because the typical research article is highly focused on a specific topic, and the genre does not necessarily allow for more general discussions and background information. In the current work, the goal of the framework was to contextualize the research articles, both in how they relate to each other, and how they relate to the larger narrative of language and literacy environments in Danish preschools.

Chapter 1 of the framework is a general introduction, and describes the motivation for the Ph.D. investigation, as well as the general research goals. Chapter 2 surveys the theoretical foundations upon which the research is based. Chapter 3 summarizes the methodologies used in the individual studies, including discussions of the validity and limitations of the approaches. Chapter 4 summarizes the results of each study. Finally, Chapter 5 discusses the results of the individual investigations with respect to the general research goals laid out in Chapter 1. Throughout the framework, the three research articles are referred to as *Paper 1*, *Paper 2*, and *Paper 3*.

All three studies presented in Part 2 are co-authored. The author of the dissertation was the lead author on all three papers, but significant contributions were made by several other researchers. A description of the contribution of co-authors is given in section 1.5. The research articles are formatted in APA style although with a few alterations intended for increasing readability of the manuscripts.

Part 1

The Framework

1. Introduction

1.1. Background

When children start formal education, they are not equal in a range of early outcomes including language and emergent literacy skills (National Institute of Child Health and Human Development Early Child Care Research Network [NICHD ECCRN], 2006). Although there are several theoretical and empirical accounts of why this is, general agreement exists that both children's genetic endowments and language contribute to variation in the development of children's language skills (Chapman, 2000). From an educational perspective, children's preschool language skills – as well as emergent literacy skills – are important because they lay the foundation for learning to read (Lonigan, Burgess, & Anthony, 2000; Whitehurst & Lonigan, 1998). Albeit variation in preschool language and literacy skills need not be considered a problem as long as children have the basic skills that will benefit their acquisition of literacy, challenges arise when a child's preschool skills are so low that he or she is impeded in learning to read. Unfortunately, achievement gaps in reading evident already in the first grade tend to persist throughout the child's ensuing school years (Cunningham & Stanovich, 1997; Juel, 1988). Sowing the seeds of literacy success before children begin school is one way of reducing the likelihood that some children will experience serious challenges in learning to read.

Reading, as a skill, is conceptually difficult to acquire as it requires children both to make associations between letter symbols and abstract phonological and phonetic information, and to link these to their correct meanings (Moats, 1999). Children are aided in this task if they already possess a range prerequisite skills such as behavioral regulation skills (McClelland et al., 2007), and language and emergent literacy skills (van Kleeck, 1998), Essentially, the children who show development in these areas prior to starting school are more likely to succeed in learning to read than children who show little development in these

areas. Unfortunately, social disadvantage due to growing up in poverty is associated with children's lower literacy skills in preschoolers (Smith & Dixon, 1995), which suggests that disadvantaged children grow up in environments that are less supportive of literacy development. NICHD ECCERS (2006) found that variation in a whole range of child outcomes was largely associated with parenting behaviors, which were themselves associated with socioeconomic status (SES). However, for the children in the study who attended child care, the quality of the preschool environments also predicted (although to a lesser extent) child outcomes in language and emergent literacy. One important implication can be drawn from this. For children from disadvantaged families, preschool holds the potential for increasing the odds of a successful school start by giving children at-risk for academic failure the language and literacy experiences they might not receive at home. In this way, preschool attendance can be considered a form of intervention against the effects of growing up in poverty.

The American *Head Start* program, which provides children from low-SES families with free preschool spaces, is a clear example of such an initiative. However, the evidence of the extent to which Head Start improves the life trajectories of Head Start children is mixed (Currie & Thomas, 1995; Garces, Thomas, & Currie, 2000). One often cited possibility for the low effects of preschool attendance against poverty is variation in preschool quality (e.g., Bryant, Burchinal, Lau, & Sparling, 1994; Epstein, 1999; Justice, Mashburn, Hamre, & Pianta, 2008). Indeed, some research has indicated that although high quality preschool can improve the outcomes of at-risk children, low quality preschool can in fact perpetuate achievement gaps (Logan, Piasta, Justice, Schatschneider, & Petrill, 2011).

Thus the question arises: how can we ensure that preschool quality is sufficient to support children's early language and literacy skills? American researchers in particular have approached this question, and within this field of early childhood education research, a sub-

field of research in professional development (PD) of child care providers (henceforth preschool teachers) has emerged. Here, the theory of change is that teachers can be trained to implement curricula and practices that more effectively develop children's early skills including language and literacy (e.g., Dickinson & Caswell, 2007). In the research that has emerged in recent years, several models of PD have been investigated using experimental design (e.g, Landry, Anthony, Swank, & Monseque-Bailey, 2009; Piasta et al., 2012). However, the literature is mixed regarding the degree to which current PD efforts benefit the practice of preschool teachers – not all experimental trials report positive findings (e.g., Lonigan, Farver, Phillips, & Clancy-Menchetti, 2011). Furthermore, some research indicates that large improvements in practice may be necessary to obtain educationally meaningful gains for children (Burchinal, Vandergrift, Pianta, & Mashburn, 2010).

There are many potential reasons for why the effects of PD interventions are unclear (Buysse, Winton, & Rous, 2009). First of all, PD processes can be dynamic and complex. PD in itself is an umbrella term for a plethora of activities, interactions and other experiences that aim to improve teacher knowledge or practice. Secondly, a number of factors can play a role in determining the extent to which a PD intervention will be successful. Teachers, for example, come to PD experiences with varying prerequisite skills, years of experience, beliefs of self-efficacy, and all of these factors may affect their benefit from the PD.

There is also the variation in the interveners. In larger studies in which many teachers participate, a number of individuals may be responsible for teaching courses, conducting coaching sessions, and performing other PD tasks (e.g., Landry et al., 2009). Although it is somewhat common for researchers to document that child interventions were implemented with fidelity (e.g., Hamre et al., 2010), it is less common for researchers to document the fidelity (or quality) with which multiple interveners conducted the same PD intervention.

As such, in the current efforts to improve the school starts of children at-risk for academic failure, two general research issues emerge. First of all, research suggests that some preschools may be of too low quality to benefit children despite the fact that preschool attendance holds the potential for improving the language and emergent literacy outcomes of preschoolers. Secondly, PD efforts to improve the language and literacy practices of preschool teachers do not consistently demonstrate educationally meaningful benefits for children. These two points are central to the research presented in this dissertation.

1.2. The Current Ph.D. Project

This Ph.D. dissertation is an anthology of three research articles motivated by unanswered research questions related to preschool quality and PD, as well as real societal issues facing the nation of Denmark. Each study is a self-contained contribution to the international literature intended for publication in a peer-reviewed journal. However, as a whole, the dissertation is intended as a scientific work that Danish policymakers and other researchers can utilize in their ongoing efforts to improve the language and literacy development of children at-risk for academic failure in Denmark.

The studies presented in the current Ph.D. project were conducted as sub-projects of two larger studies. The first study, Structured Preschool Effort for Language and Literacy (SPELL) was a randomized controlled trial that investigated the effectiveness of a 20 week pre-literacy intervention when implemented at scale (Bleses et al., 2014). SPELL was financed via a grant from the Strategic Research Fund, which is a research fund administered by the Danish Ministry of Higher Education and Science that is dedicated to research that has a potential societal benefit. Over 7000 children and 600 preschools teachers were involved in the SPELL project. For more information about SPELL, see also www.sdu.dk/SPELL.

The second study, Language and Literacy Educational Activities for Preschoolers (LEAP), was equal to SPELL in experimental design in that it too was an effectiveness study

of a pre-literacy intervention conducted at scale, and involved approximately 7000 children and 600 preschool teachers (Bleses et al., 2015) However, whereas the SPELL intervention was based on shared-book reading, the LEAP intervention utilized a multitude of activities as a platform for learning, and it was commissioned by the Danish National Board of Social Services. For more information about the LEAP intervention, see www.sdu.dk/fartpaasproget.

Both large-scale studies were conducted by Center for Child Language, University of Southern Denmark (Prof. Dorthe Bleses as principal investigator) in cooperation with the consulting company Rambøll Management. The three papers presented below draw on data collected for the larger research projects. The author had a lead role in collecting some of the data used in collaboration with the other research members. Data collection of each article is described in more detail in the methods section.

1.3. The Danish Context

In Denmark, approximately 95% of children attend publically funded preschools (OECD, 2013), and arguably, the responsibility of preparing children for school is shared with the preschool teachers who instruct and care for the children. With such a large percentage of children attending preschool prior to school start, the potential for ensuring that all children, regardless of their social background, acquire the prerequisite skills necessary to acquire literacy seems like an obvious advantage. Little is known about the effect of attending preschool in Denmark. A registry study by Gupta and Simonsen (2010) found that the amount of time spent in Danish child care at age three was negatively associated with non-cognitive skills at age seven, but this study did not analyze language or literacy outcomes. In fact, little research has actually investigated the extent to which Danish preschools provide children with environments that foster the development of language and emergent literacy skills.

This is no trivial matter for two reasons. First of all, preschools in Denmark practice a unique holistic form of pedagogy, which emphasizes children's social and emotional development and well-being, but places little focus on developing pre-academic skills (Jensen, 2009). As such, little is known about the effect of such an approach, and whether it produces learning environments that support children's language and literacy development. Secondly, a substantial percentage of Danish youths do not attain functional reading skills by the end of compulsory schooling (Egelund, Nielsen, & Rangvid, 2011), and this despite the fact that most have attended some sort of educational facility since the age of three. Taking into consideration that empirical research in Danish children has also found negative associations for SES and immigrant status on the language skills of preschool children (Bleses, Højen, Jørgensen, Jensen, & Vach, 2010), it is an open question whether Danish preschools could do more to prepare children for reading success in school.

Although the Danish pedagogical approach does not traditionally place much emphasis on the pre-academic preparedness of preschool children, political interest in improving children's language skills has resulted in a series of legislative efforts aimed at improving the language outcomes of children in preschools. These efforts have taken the form of a number of amendments to the legislation surrounding the public child care system commonly referred to as the *Dagtilbudsloven* [The Child Care Law] (Government of Denmark, 2011), and have had the main goal of reducing the effects of social disadvantage on language development.

The first major regulatory change came in 2004 with the addition of § 8 to the Child Care Law requiring all preschools to develop a so-called *læreplan* [learning plan]. The learning plans were intended to serve as a document explaining how teachers intended to stimulate children in the following six areas: all-round personal development, social competence, language development, body and motion, nature and natural phenomena, and

cultural expression and values. Although language development is notably one of the areas of focus, the law does not prescribe specific learning goals. Rather it is up to the individual preschool to decide how each area will be supported in practice.

In 2007, the Danish government again displayed interest in improving children's language outcomes in preschool when they enacted an amendment (§ 11) to the Child Care Law requiring all municipalities to offer the parents of all monolingual three-year-olds a language screening, and language screenings for all multi-language learning children was made mandatory. This passage of the Day Care Law was amended again in 2010 such that preschool teachers only were to screen the children for which they had a suspicion of language delay or impairment, but that screenings were mandatory if deemed necessary. Furthermore, if a child's Danish language skills were found to be far below the norm, the municipality was required to offer free language stimulation services to the family.

Another example of political steps towards improving the language practices of Danish preschool teachers occurred in 2010 as a complement to the regulatory changes described above. The Danish parliament budgeted 34 million kroner to develop and implement a PD course with a language focus. The PD course, referred to as *Sprogpakken* [the Language Package], was conducted either as a four or six day course between 2011 and 2012. The main goal of the course was to give at least one preschool teacher from every preschool across the country a working knowledge of language development, and suggest stimulation methods such as dialogic reading (Whitehurst et al., 1988), and the use of supportive language strategies (e.g., Girolametto, Weitzman, & Greenberg, 2003). Although the government allotted a large sum of public funds to develop and implement the PD, no funds were budgeted for evaluating the effect of the course on teacher and/or child outcomes. As such, we have little empirical evidence that this massive public investment had its intended effect.

According to Jespersen (2006) , there is little practice in Denmark of evaluating the effects of interventions that aim to lessen the influence of social disadvantage. The language and literacy environments of Danish preschools are perhaps a case-in-point regarding this issue. Although the law requires that preschools create pedagogical plans for working with children's language, preschools essentially have full autonomy in how this will be carried out. Furthermore, there are no systems in place to ensure that the language instruction and stimulation that preschools provide children actually have the intended effect.

In sum, Denmark is an interesting case in the international literature on preschool education. Denmark has near universal preschool, and spends the second highest percentage of its gross national product on child care programs in the world (OECD, 2013). Denmark is also a country whose central government has in recent years reformed the public preschool system such that teachers are required to afford children's language development greater focus. Despite this massive public investment and political reform, preschools still retain autonomy in developing and conducting their own teaching plans, which may be highly influenced by the holistic approach of Danish pedagogy, which strays away from pre-academic goal-setting.

1.4. The General Goals of this Dissertation

The three research papers presented in this dissertation each investigate a series of specific research questions, but in its entirety this dissertation has the general purpose of providing empirical research that Danish policymakers can use to improve the language and emergent literacy outcomes of children. The following overall questions were investigated:

i. What is the quality of the language and literacy environments in Danish preschools?

ii. What is the effect of PD interventions aimed at improving the language and literacy environments of preschools, and what factors are associated with successful language and literacy PD interventions?

iii. Following a typical PD course on the use of pedagogical strategies, does as sample of Danish teachers demonstrate awareness of their own skill performance?

The first question is explored in detail in Paper 1, which investigates the global quality of the practice of 506 preschool teachers, and the availability of literacy supports in 293 preschools. The second question is investigated in Paper 2, which is a meta-analysis of the international research on the effects of PD interventions with a language and/or literacy focus. Effects are estimated on teacher outcomes, and children's receptive vocabulary, phonological awareness and alphabet knowledge when possible. Rooted in social cognitive theory, Paper 3 investigated the extent to which teachers demonstrated knowledge of their usage of scaffolding strategies following a three day course. According to the theory, skill performance knowledge is an important aspect of adult learning processes.

The results presented in these research articles can inform policymakers and other pedagogical researchers on a number of issues. First of all, knowledge of the quality of language and literacy environments can be used to evaluate the effect of the current pedagogical practice in Denmark. If quality is found to be too low, then avenues of increasing quality should be studied including providing PD opportunities for in-service preschool teachers, enacting improvements to the current early childhood education degree program, and enacting regulatory changes that can offer preschools assistance in achieving higher quality practice. Secondly, the effects of the meta-analysis can be applied to the Danish context, which for its own part has little tradition for empirically evaluating the effects of preschool interventions on children's outcomes. The meta-analysis offers some insight into

whether typical PD interventions (such as the language package) have the effect they were intended based on a survey of international literature. Finally, the Paper 3, which uses a Danish sample, gives an indication of teachers' skill performance knowledge following a three days of PD. This offers some insight into the learning processes teachers experience when attending PD courses.

1.5. Collaborators

Both SPELL and LEAP were large-scale projects, which together involved many people including researchers, students, consultants, municipal employees, preschool teachers, and of course children. The articles presented in this dissertation are also the culmination of several research collaborations. In the following, the contributions of the co-authors of each article in this dissertation are briefly recounted.

1.5.1. Paper 1

Dorthe Bleses, Jessica Logan, and Laura Justice were co-authors on Paper 1. The study utilized data from both the SPELL and LEAP projects. Dorthe Bleses, as the Ph.D. candidate's main supervisor, and as project leader of SPELL, was involved in the study from its first conceptualization. She had a significant role in the selection and piloting of all measures, and read and commented on drafts of the article. Jessica Logan provided analytic support to the main author, and read and commented on the methods and analysis sections. Laura Justice, as co-supervisor of the Ph.D. candidate and member of the SPELL project's board, was also involved throughout the whole study. She provided supervision to the main author, and read and commented on drafts of the article. The main author wrote all sections of the article, and conducted all analyses, but based on the considerable feedback and supervision he received from the co-authors.

1.5.2. Paper 2

Paper 2 was a comprehensive work and involved several co-authors. Carsten Juhl made significant contributions to study. As a meta-analyst, he guided the main-author through the meta-analytic procedures, and provided invaluable supervision. He also read all drafts of the article, and provided invaluable guidance regarding the methods and results sections. Shayne Piasta – also an experienced meta-analyst – provided general supervision to the main author. She read all drafts of the article, and her comprehensive comments and suggestions influenced the introduction and discussion sections to a large extent. Anders Højen wrote the description of the systematic map in the methods section. He also read drafts of the article, and made comments. Dorthe Bleses, both as main supervisor to the author, and as project leader of the systematic map project, was involved in the conceptualization of the review, and helped frame the research questions. She also read drafts of the paper, and double-coded some of the included studies. Laura Justice, as co-supervisor, was also involved in throughout the entire process, and was also involved in the creation of the systematic map. Although not a co-author of Paper 2, Werner Vach provided valuable advice and guidance for its protocol.

The main author conducted all the analyses. With the exception of Anders Højen's description of the systematic map, he also wrote all sections of Paper 2. However, this was only made possible through the considerable guidance and supervision of the co-authors.

1.5.3. Paper 3

The co-authors of Paper 3 were Dorthe Bleses, Werner Vach, and Laura Justice. Dorthe Bleses, as supervisor and project leader of SPELL, was involved in the conceptualization of the study, provided general supervision throughout the whole process, and read drafts. Werner Vach made essential contributions to the analytic strategy, and made numerous helpful suggestions that aided the main author in analyzing and presenting the findings. Finally, Laura Justice, as co-supervisor, was involved in the conceptualization of the

study from the beginning. She read all drafts of the paper, and provided comments and suggestions that greatly influenced the introduction, and analysis sections of the paper.

The main author wrote all sections of Paper 3, and performed all analyses. As with the other articles, this was only made possible through a close collaboration with the co-authors who offered extensive advice and guidance.

2. Theoretical Approach

2.1. Introduction

This chapter is an overview of pertinent theories related to development, early childhood education, and skill learning. The goal is to thoroughly discuss the theoretic foundations and assumptions on which current empirical research is performed. Where appropriate, the individual papers are related to the theory.

Although the three papers presented in this anthology vary on a number of factors, they are all founded on the overall theoretical understanding that people develop with respect to the environments in which they find themselves (e.g., Bronfenbrenner, 1979). In the case of children, this basic understanding assumes that the quality and quantity of care-givers' speech will contribute to variation in children's language development. Applied to the case of preschool teachers, this understanding assumes that PD interventions can change professional behaviors. It may perhaps seem obvious that environments affect our lives, yet important theoretical questions regarding *how* and *to what extent* environments influence our lives are debated topics, and warrant discussion in relation to the current dissertation.

The following chapter describes the main scientific theories that lay the foundation for the empirical works presented below. Three theoretical perspectives are discussed in particular. The first is the *ecological systems theory* of Urie Bronfenbrenner (1979). Bronfenbrenner's theory posits that humans develop in the context of a series of environmental systems. The theory is foundational to all the studies presented below as it theorizes that individuals are influenced by several layers of environment, but that these environments are also influenced to varying degrees by the individual.

The second foundational theory discussed here deals with children's acquisition of language and emergent literacy, and in particular how these can be supported in the preschool context. The author takes an *interactionist* perspective on children's language development,

which postulates that each child's innate language abilities interact with environmental input to create the individual child's language ability (e.g., Chapman, 2000; Hoff-Ginsberg & Shatz, 1982). Furthermore, the interactionist perspective also has special implications for early childhood pedagogy (e.g., Vygotsky, 1978; Wood, Bruner, & Ross, 1976), which has clear applications for the preschool teacher. The interactionist perspective is highly relevant to all papers in this anthology.

The final theoretical foundation to be surveyed in this chapter is the social cognitive theory of Albert Bandura (1977b, 1986). The theory posits that individuals acquire knowledge and behaviors through observation of others, and that self-efficacy plays a key role in these processes. Social cognitive theory is highly influential in teacher education, and is therefore highly relevant to Papers 2 and 3, which deal with aspects of teacher training.

2.2. The Ecology of Human Development

In 1979, Urie Bronfenbrenner published a work that posited that humans develop at the center of several layers of systems, which both influence and are influenced by the individual at the center of the network. Taking a child as an example, the ecological systems theory postulates that a child's development is heavily influenced by near environmental factors such as parents and attending preschool. On a more distal level, however, other environments, such as a parent's place of work, can influence the development of the child. For example, adults are generally not permitted to look after their children at work, which necessitates the need for child care, and this will inevitably have some kind of effect on the development of the child. Furthermore, the theory views influences as being bidirectional. For example, a child exhibiting aggressive behavior in a preschool can negatively affect the preschool teacher's ability to teach a lesson on the alphabet. Children can even affect more distal systems. For example, an influx of Spanish-speaking children in an English-speaking

community could conceivably change the attitudes of the local community towards immigrants.

2.2.1. The Five Systems

Bronfenbrenner's theory posited five layers of systems that can be "conceived as a set of nested structures, each inside the next, like a set of Russian dolls" (1979, p. 3). At the center of these layers is the individual, who interacts with each layer in accordance with the biological inheritance the individual has. The most immediate and generally most emphasized layer surrounding the individual is the *microsystem*. The microsystem includes other individuals, activities, or physical environments that the individual in the center interacts with on a regular basis. In the case of a preschool child, the microsystem would include parents, peers, siblings, preschool teachers, and environments such as the preschool itself. Routine activities, such as reading books or going to the park, would also be described within the microsystem.

Beyond the microsystem are four more systems that interact with the individual, although to increasingly lesser extents. The layer directly encompassing the microsystem is the *mesosystem*, which houses interactions between one or more units of the microsystem. For example, different actors in a child's microsystem, such as a preschool teacher and a parent, might interact with each other, which results in a consequence for the child (such as the teacher suggesting that the parents read books to the child). The child can also influence the mesosystem. The teacher noticing that the child has a small vocabulary might seek dialogue with the parents about supporting the child's language development with book reading at home.

Following the mesosystem is the *exosystem*, which describes factors that affect the development of the individual even though the factors don't directly involve the individual. For example, a parent's place of work does not directly interact with the child, but can still

influence the child's life as exemplified above. Similarly, a workplace can be influenced by its employees' children by, for example, opening a child care center for the benefit of its employees.

Following the exosystem is the *macrosystem*, which describes the culture and ideologies that characterize the lower systems. The effects of the macrosystem are easily exemplified by language – children in Sweden learn Swedish, whereas children in France learn French. However, Bronfenbrenner (1979) also notes that the macrosystems of people from the same geographical location can vary greatly. For example, disadvantaged families can have views and values that differ from those of wealthy families, all of which interact with the development of the individual in the center of the system.

The final layer in Bronfenbrenner's ecological systems theory is the *chronosytem*. The chronosystem describes the events that occur throughout the course of a life that can interact with human development. For example, a retiring grand-parent may suddenly enter the life of a developing child as a primary care-giver while the parents work. In this way, the events we experience through our lives can have far reaching consequences for those in our peripheries.

2.2.2. Application to Early Childhood Education

The ecological systems theory is useful for framing research in children's development within a preschool context. With regards to former developmental research, Bronfenbrenner criticized what he called "the traditional research model," which measured child outcomes without taking adequate consideration of the environmental factors that could influence child outcomes (1979, p. 164). He also criticized the use of ecologically inappropriate outcome measures such as laboratory measures, which were not developed for use in a preschool context. Essentially, Bronfenbrenner argued that pedagogical research should be ecologically valid, that is to say that it should acknowledge that children develop in the context of the systems described above.

Bronfenbrenner's framing of human development has been influential in pedagogical research, perhaps most so with regards to the microsystem. This is evidenced by the multitude of preschool environment measures that have been developed to measure the environments children interact with when they are in preschool. The Early Childhood Environment Rating Scale Revised (ECERS-R; Harms, Clifford, & Cryer, 1998) is one such measure. The ECERS-R measures the overall quality of a preschool environment mostly in terms of structural characteristics such as the provision of safe equipment and toys, but it also examines the quality of interactions between teachers and children. Such measures operate under the assumption that aspects of the preschool environment will affect child development.

Empirical research has also demonstrated that environmental factors that can be related to the microsystem have effects on a broad range of developmental outcomes. For example, the NICHD ECCRN (2006) followed a cohort of children, and measured a number of factors including the parenting skills of caregivers, and preschool quality. The researchers found that environmental factors explained variation to varying extents on a range of developmental outcomes.

2.2.3. Summary

The ecological systems theory provides an overall framework for the current dissertation. The framework assumes that individuals develop within a complex context, which influences their lives. It is also important to note that Bronfenbrenner's theory is applicable to all individuals regardless of their age. Preschool teachers also exist at the center of an ecological system that contextualizes their development when they, for example, participate in PD. It is also noteworthy that the theory is general with regards to human development, and not language in particular. In the next section, developmental perspectives closer related to cognitive, and specifically language development are discussed.

2.3. The Social Interactionist Perspective

Several theoretical approaches to children's learning of language have been presented in the literature. Often these theories are framed within the nature-versus-nurture debate, which revolves around the question of the extent to which children's language acquisition occurs due to innate mechanisms, or due to environmental factors. Chomsky (1969) postulated that language – and syntactic knowledge in particular – is innate. Chomsky based this stance in part on the logic that no child could learn something as complex as language, when the input from parents is so poor, a problem that he later referred to as the *poverty of the stimulus* (Chomsky, 1980). Rather, the apparatuses needed for learning language are assumed to exist already in the mind, requiring only some basic input throughout the developmental stages at which point the various aspects of universal grammar come online. Chomsky's nativist approach revolutionized the field of language acquisition which had otherwise been greatly influenced by the constructivist theory of Piaget (1959), who postulated that children actively construct their language using general (i.e. non-linguistic) cognitive abilities.

However, many researchers have argued against a strongly nativist approach, and instead placed more emphasis on the child's active engagement in learning language. For example, Tomasello (2005) rejected Chomsky's notion of a universal grammar in favor of a usage-based approached to language acquisition, in which children create linguistic knowledge through their use of it. Tomasello thus argued that children have the ability to recognize linguistic patterns, and essentially reconstruct the adult language in their own minds through experience and practice.

Other child language researchers have also rejected that logic of the poverty of the stimulus on empirical grounds. Work by Snow (1972) demonstrated that mothers' infant-directed speech was in fact simple, which Snow argued demonstrated a natural effort by

mothers to facilitate their children's language acquisition. This empirical finding supports the notion that children's language development is dependent on a special variety of maternal speech that facilitates the language development, and thus reduces the burden on innate language abilities.

In contrast to the theories of Chomsky and Piaget, which exist on different ends of the nature-nurture spectrum, the social interactionist perspective assumes that both innate and acquired aspects of language development exist, and that they interact with each other in a social context (Chapman, 2000). In particular, the pioneering work of Russian psychologist Lev Vygotsky (1978) greatly emphasized the idea that language acquisition occurs within a social context, in which parents, guardians and other individuals within the child's sphere play an active role in mediating the child's development of language. Within this framework, linguistic variation is accounted for not only by the child's genetics, but importantly, by the quality and number of social interactions the child experiences with others. In terms of language acquisition, this assumes that children's language capabilities will reflect the linguistic capabilities and efforts of those by which they are raised

2.3.1. Evidence that Input Accounts for Language Variation

The social interactionist perspective theorizes that language variation is to some extent accounted for by language input. Empirical research supports this stance. Hart and Risley (1995) investigated the linguistic interactions of one and two-year-olds with their care-givers, and found striking differences in the amount and quality of linguistic input children received, which were also reflected in the children's own vocabularies. The variation in input was also found to be in part a function of socio-economic status (SES). SES was positively related to both the amount and quality of parents' input, and children's own vocabularies.

In another study by Hoff (2003), the productive vocabularies of mid-SES and high-SES children also investigated in relation to maternal input, and SES was found to explain

variation in input and child vocabularies completely. At least in terms of children's vocabulary, variation in input appears to be a determining factor. The study also confirmed that there is a positive relation between SES and vocabulary.

The social interactionist perspective thus places great importance on the input of parents and guardians in explaining variation in children's language outcomes. Yet, it also assumes that experiences in formal child care play a role in children's language variation. Evidence of the relative contribution of child care experiences can be found in the longitudinal study by NICHD ECCRN (2006). This study found that parenting behaviors indeed had medium to large effects on a range of child outcomes including language and emergent literacy, but that child care quality also demonstrated small to medium effects for children who spent considerable amounts of time in child care. A number of other studies conducted in the United States have also found small to medium (but statistically significant) associations between the quality of language and literacy environments in preschools, and children's own language and emergent literacy skills (Burchinal et al., 2008; Howes et al., 2008; Justice et al., 2008; Pianta et al., 2005).

2.3.2. The Social Interactionist Perspective on Early Education

The social interactionist perspective not only offers a framework for explaining how the child's surroundings contribute to linguistic variation, but it also offers a framework for developing effective educational practices. Vygotsky (1978) famously proposed that children have a *zone of proximal development* (ZPD) when they learn new skills or tasks. The ZPD can be described as the child's potential for learning when the adult offers appropriate supports. The ZPD therefore places great importance on the ability of the adult to maximize the child's learning, which has clear applications for an early education context. For example, a child's development of alphabet knowledge, which certainly is not innate, is mediated by the degree to which the astute teacher can guide the children from what the child knows (for

example, that his name starts with *J*), to generalizing the knowledge to a higher level (for example, that other names can start with *J*).

The act of supporting children within the ZPD has come to be referred to as *scaffolding* (Wood & Middleton, 1975), a metaphor the describes the way in which a scaffold provides support so that an individual can safely ascend to a higher level. Vygotsky's ZPD and methods of scaffolding have been highly influential in the area of early childhood education. Vygotskyan theory has for example guided the development of curricula (e.g., Bodrova & Leong, 2001; Justice et al., 2010; O'Connor, Notari-Syverson, & Vadasy, 2005), emergent literacy instructional practices (Cabell, Tortorelli, & Gerde, 2013; McGee & Ukrainetz, 2009), and methods of language intervention for children with language impairment (e.g., Schneider & Watkins, 1996).

2.3.3. Summary

This dissertation is theoretically rooted in the social interactionist perspective, which assumes that children learn language within a social context. This assumption is furthermore extended to educational practice, postulating that preschool teachers' usage of pedagogical practices can maximize a child's learning and development. However, doing so clearly requires the teacher to possess these skills, and therefore the theory is particularly important for Paper 1, which investigates quality in preschool environments. In the next section, considerations for how teachers can acquire these practices when they do not possess them are presented.

2.4. Social Cognitive Theory

The final theoretical base on which this dissertation is built is the social cognitive theory of (Bandura, 1977b, 1986). Papers 2 and 3 deal with questions related to how PD can alter pedagogic behavior, and in the same way that Vygotsky has been influential in our

understanding of how children learn language and skills, Bandura has been influential in describing the basis of human behavior – and important for this dissertation – how it changes.

In 1977, Alberta Bandura presented his first unified theory of human behavior and thought, which he referred to as a social learning theory. To emphasize the role of cognition in his model, the theory was referred to as social cognitive theory in his update of the theory (Bandura, 1986). In its essence, the theory posits that human cognitive abilities, and our social nature, have important influences on our behavior. Rather than being a break from purely behavioral approaches, which saw behavior as the results of external conditioning (e.g., Skinner, 1938), Bandura's work is more of an expansion which acknowledges that humans are susceptible to conditioning, but that individuals also retain the abilities to regulate their behavior via cognitive processes. Central to the theory is the understanding that individuals can learn behaviors through vicarious experiences rather than experiencing consequences or rewards themselves, and that personal self-efficacy regulates the degree to which individuals will strive to learn new behaviors.

2.4.1. Observational Learning

As Bandura (1977b) points out, "Learning would be exceedingly laborious, not to mention hazardous, if people had to rely solely on the effects of their own actions to inform them what to do" (p. 22). Rather, people can learn behaviors by watching others. Bandura, Ross, and Ross (1961) for example found that when children viewed someone playing aggressively with a doll, the children tended to do the same, whereas they mimicked non-aggressive behavior if that were the model. On the basis of this and other experiments, Bandura and his colleagues concluded that people readily learn behaviors via observation.

According to social cognitive theory (Bandura, 1977b, 1986), observational learning has four main processes, which govern the extent to which novel behaviors are acquired. The first component is the *attentional process*. The learner must attend to the modelled behavior,

and the degree to which this occurs is dependent on the learners own characteristics such as perceptual skills and arousal level, as well as the stimuli's characteristics such as its distinctiveness and relevance for the learner. Following attention, the next component of successful observational learning is the *retention process*. Learners must be able to remember the details of the behavior being acquired, and this too is dependent on the cognitive abilities of the learner, as well as the relative complexity of the modelled behavior. Following retention are the *motor reproduction processes*. During this process, the learner reproduces the behavior, and receives feedback on the success to which this is done. Learners may actively seek feedback, or notice how others react or do not react. This stage is considered reciprocal, and helpful feedback includes former performance success. The final component of the observational learning model is composed of *motivational processes*. Bandura distinguishes between acquisition of a behavior and reproduction of a behavior. Learners must be motivated to enact the new behavior, and here social cognitive theory states that motivation can be regulated by the response of others who view the learner's performance, or the learner's own performance expectations can also provide motivation.

2.4.2. Self-Efficacy

Another central element of social cognitive theory is *self-efficacy*. First described, in Bandura (1977a), self-efficacy is essentially one's belief that one can succeed in a certain task or situation. Bandura argued that an individual's perceived self-efficacy regulates the extent to which a person strives to accomplish goals, and perseveres in the face of obstacles. Self-efficacy has implications for how people behave when mastering new skills. The theory predicts that those with high self-efficacy will see themselves as being able to master the skill, and therefore exert energy to reach their goals. Others lacking self-efficacy will strive less – even though they might in fact possess the cognitive and physical skills to accomplish the goal.

Bandura (1977a) posits four sources that contribute to people's perceived self-efficacy. First of all, initial *performance accomplishments* provide efficacy information. If learners experience success with the new skill, their belief that they can reach the goal will be strengthened. However, if they experience too many defeats, self-efficacy is reduced, and the learner moves closer to giving up on the task. The second contributor to self-efficacy is *vicarious experience*. If learners view that other individuals who are seemingly similar in skill level are able to carry out the task, then the learner's own self-efficacy will be strengthened. Self-efficacy is also influenced by *verbal persuasion*. Learners can strive more in the face of obstacles when peers and coaches encourage them to persevere. Finally, a learner's emotional arousal contributes to self-efficacy. When a learner is relaxed or in a good mood, self-efficacy can be expected to be higher than if the person were distressed or under pressure.

Of the four contributors of self-efficacy, Bandura (1977a) cited the first, performance accomplishments, as being particularly influential. However, one potential problem with the model occurs when learners mistakenly believe themselves to have mastered a skill when in fact they have not. Interestingly, Bandura and Cervone (1983) found that when trainees discover that their anticipated skill performance was lower than expected, they became motivated to strive more. Bandura and Cervone suggested that high self-efficacy is more desirable for learners than low self-efficacy, and therefore when learners discover that their performance is lower than expected, they redouble their efforts to achieve the goal, and thereby reinstate their beliefs of high self-efficacy.

2.4.3. Summary

Social cognitive theory has clear applications for the PD of preschool teachers. The theory offers a framework for how interveners can develop courses, mentoring programs, or other learning formats that support the general principles of observational learning, and

support teachers' beliefs of self-efficacy. Social cognitive theory is relevant for Paper 2, as the meta-analysis deals with the effects of PD intervention. However, Paper 3 was explicitly motivated by social cognitive theory in that we aimed to support some of the underlying processes of the theory. Using video analysis, we aimed to help teachers observe their true skill performance, which we expected was lower than they believed.

3. Methodology

3.1. Introduction

This chapter accounts for the methodologies used to investigate the overall goals of each paper included in this dissertation. Each paper is described in turn. First, the overall methodology that was used is summarized. Secondly, an account of the decision-making and reasoning behind certain methodological choices is given, such as choice of measure or research design. Furthermore, the validity of the methodologies selected will also be discussed.

3.2. Methodological Considerations of Paper 1

The primary goal of Paper 1 was to investigate the quality of the language and literacy environments in Danish preschools. To do this, we conducted a cross-sectional study in which we observed process and structural quality, and explored associations with children's SES, and a number of teacher background factors. In addition, we also compared process quality scores with an American sample.

3.2.1. Overview of Methodology for Paper 1

A cross-sectional study was an appropriate choice in research design as this investigation aimed to create a snapshot of the language and literacy environments of Danish preschools. Although one limitation of the cross-sectional study is that causality cannot be inferred from it (Babbie, 2015), the design allowed us to investigate a number of empirical questions that have remained unanswered regarding preschools in Denmark. Furthermore, cross-sectional studies have been used by a number of other researchers who also investigated quality in child care programs (e.g., Burchinal, Cryer, Clifford, & Howes, 2002; Pianta et al., 2005).

In Paper 1, process quality was observed at the teacher level, whereas structural quality was observed at the preschool level. In all, 506 preschool teachers were observed with

regards to process quality using the Classroom Assessment Scoring System (CLASS; Pianta, La Paro, & Hamre, 2008). In addition to the CLASS observations, teachers also completed questionnaires that informed us on a number of demographic and professional factors. At the preschool level, 293 preschools were observed using the Classroom Literacy Observation Profile (CLOP; Crane Center for Early Childhood Research and Policy, 2012). Besides the CLOP observations, we also had access to data for five variables related to children's SES (such as belonging to a low-income family). These variables were calculated as the percentage of children displaying the risk factor in each preschool. Thus, the dataset consisted of data from four sources: CLASS observations, CLOP observations, teacher background questionnaires, and SES variables.

3.2.2. Challenges in the Methodology of Paper 1

Our main challenge in Paper 1 was determining how to measure quality. Although our main goal with the cross-sectional study was to provide the first mapping of the quality of language and literacy environments in Danish preschools, we also wanted to be able to compare our results with an American sample. Therefore it was important to choose a measure that was valid across cultures.

In the end, we decided to use two instruments: the CLASS, and a Danish adaptation of the CLOP. The CLASS is an observational tool that measures ten dimensions of process quality on a Likert scale of 1-7. Items are divided into three domains: *emotional support*, *classroom organization*, and *instructional support*. The CLOP is essentially a checklist of structural literacy supports (i.e. the provision of literacy materials), and provides an indication of the literacy materials to which children have access. In the next two sections, the processes that led to our selection of these measures are discussed in detail.

3.2.3. The CLASS

Research indicates that process quality (i.e. teacher-child interactions) predicts children's language and emergent literacy skills (Burchinal et al., 2008; Howes et al., 2008). However, process quality can be evaluated in a number of ways. At a global level, some measures evaluate the quality of interactions that support broad domains such as socio-emotional and cognitive development, whereas more fine-grained tools can evaluate the quality of more detailed interactions such as literacy instruction (Dickinson, 2006).

At the outset of the SPELL project in 2012, it was decided that we would conduct a review to determine which preschool measures might be suitable for evaluating quality in preschools. Early in the review stage, however, an already existing compendium of early childhood quality measures (Halle, Vick Whittaker, & Anderson, 2010) was found, and the review was therefore halted. Of the 51 instruments listed in the compendium from Child Trends, 27 instruments met our age criterion (i.e. 3 to 5-years of age), and were consequently reviewed for suitability for our study. After several meetings, our list of candidate measures was short-listed to the ECERS-R (Harms et al., 1998), Early Childhood Environment Rating Scale Extension (ECERS-E; Sylva & Taggart, 2010) and the CLASS (Pianta et al., 2008).

After comparing the items of all three measures, the CLASS was chosen as the best choice for our purposes. First of all, the CLASS focuses on teacher-child interactions, which according to the literature is a promising predictor of child outcomes (Mashburn et al., 2008; Pianta et al., 2005). Secondly, the CLASS measures interactions at a global level, which we believed was most suitable because we were aiming to make the first large-scale investigation of preschool quality in Denmark. Thirdly, the CLASS was also used in a number of American studies including the efficacy study of *Read It Again!*, the intervention upon which SPELL was based (see Justice et al., 2010). Therefore our usage of CLASS would allow future comparisons with samples from the United States. Furthermore, CLASS had been validated in Finland (see Pakarinen et al., 2010), which gave us confidence that it would also

be a valid measure in Denmark. Finland, like all Nordic countries, has universal preschool, and shares a number of cultural attributes with Denmark. Finally, CLASS was designed to be scored during live observations, or using video. Due to the high cost of live observations, it was a clear advantage that video data could be used taking into consideration the large sample size of the SPELL study.

After selecting CLASS as our main measure of process quality, we conducted two pilot studies, first to determine the relative difficulty of coding CLASS using video data, and secondly to determine whether preschool teachers themselves could be responsible for recording the videos. In the first pilot study, we video-recorded a small number of preschool teachers in interactions with children, and found that video quality was conducive to coding. In the second pilot study, three preschools received a package containing a video-camera, a tripod, memory cards, and explicit instructions on how to video-record instructional scenarios themselves. Evaluation of the returned memory cards indicated that teachers had followed the instructions we provided, and that sound and visual quality was sufficient for coding.

The piloting of CLASS was thus completed in August of 2012. At this point, the author was made responsible for establishing a CLASS coding laboratory. Since the initial teacher sample of SPELL was nearly 700 teachers, we hired a number of undergraduate and master students who were trained and certified to use CLASS. Training was conducted by a certified CLASS trainer, who was made available to us via a research collaboration with Professor Laura Justice of The Ohio State University. Training occurred in the fall of 2012, and in all, six coders were trained and certified (including the author). To reach certification, coders had to code five videos online, and achieve an average inter-rarer reliability of at least 80%.

Although at this point all coders were technically certified to code with CLASS, we took further steps to ensure that coders would apply the CLASS reliably with Danish data.

Three addition training rounds were therefore conducted. This resulted in an average inter-rater reliability of 94.8%, at which point we commenced coding of the data that is presented in Paper 1.

It took approximately one year to code the CLASS data that is used in Paper 1. Monthly maintenance meetings were held to prevent coding drift, and preserve reliability. At these meetings, one or two videos were coded, and any discrepancies were discussed and resolved.

Paper 1 thus utilizes a CLASS dataset that was the culmination of an extensive process that included an elaborate selection procedure, thorough piloting, the establishment of a training laboratory, certification with the developers of CLASS, extra training with Danish data, and monthly meetings that maintained coding skills. With turnover, ten different individuals coded data at some point. Although the CLASS manual considers 80% inter-rater reliability to be acceptable, our final dataset was coded with an inter-rater reliability of 91.2%.

One potential methodological limitation of coding CLASS evident in Paper 1 involves response rate from teachers. As described in the article, 638 teachers were originally selected for participation in the study (and SPELL as a whole). Our final sample used for analysis contained 506 participants. Half of the attrition was explained by early drop-out from the entire SPELL project, which was an anticipated side-effect of implementing an effect study at scale. However, it is unknown why the remaining 66 teacher, who did not record their practice, failed to do so.

3.2.4. The CLOP

We complemented our investigation of the process quality of preschools with an investigation of the structural quality of the literacy environment. This was operationalized using the CLOP. The CLOP is essentially a checklist of the number of various literacy

materials, which either are available at child level in the cases of books, toys and implements, or visible to children in the case of displays and posters. Some research has indicated that the availability of literacy materials such as books can contribute to the number and quality of literacy interactions children have in preschool (Neuman, 1999; Neuman & Roskos, 1993). Furthermore, high process and structural qualities in the preschool literacy environment appear to reinforce each other (Guo, Justice, Kaderavek, & McGinty, 2012).

As no research to our knowledge has investigated the availability of literacy materials in Danish preschools, the use of the CLOP seemed highly relevant with regards to evaluating the general quality of the literacy environments. However, adapting the CLOP for use in Danish preschools presented unique methodological challenges. First of all, the CLOP was designed and validated in the United States (Dynia, 2013), where classrooms generally function as independent units in a preschool. Although in Denmark preschools typically consist of three classrooms, classrooms in Danish preschools function more as an entirety. For example, one physical classroom might contain a library section or an art area that other classrooms may use at certain times during the day. Children may also have the freedom to move between classrooms during free play, which can last for extended periods in Danish preschools. Furthermore, other preschools in Denmark do not follow the typical three-classroom design. Some modern preschools are built with an open-concept design, and essentially do not use classrooms. Yet another subtype of preschools in Denmark are the *skovbørnehaver* [forest preschools], which often have just one smaller building where children are dropped off and picked up. Otherwise, the children spend most of their time playing and receiving instruction in the forest. Using the CLOP at the classroom level was thus problematic, since it is a level of analysis that does not truly exist in Danish preschools. Therefore we chose to fill out one CLOP form for each preschool as a whole. We evaluated

this to be an acceptable compromise since our intention with the CLOP was to obtain a broad, first-time view of the structural literacy environment in Danish preschools.

Other methodological challenges occurred in adapting the CLOP to the Danish context. While most of the items were easily transferable to a Danish preschool, others were more difficult. For example, the original CLOP has items related to the materials available in the classroom's so-called writing center. Danish preschools do not contain writing centers because a focus on learning to write is traditionally viewed as being developmentally inappropriate for preschool children in the Danish pedagogical view. To deal with these and other possible misalignments between American and Danish preschools, we translated the CLOP to Danish, and asked Danish preschool teachers to give feedback on which items may not apply to a Danish preschool. The teachers suggested that we for example change "writing center" to "drawing center," because there is a special focus on drawing in Danish pedagogics, and teachers might teach children some writing skills during these activities. Otherwise, teachers reported back that the CLOP items would apply to varying degrees in preschools.

We added some items to our Danish adaption of the CLOP as well. In particular, we added checklists regarding the presence of materials that could support multi-language learners, and materials about language and literacy directed towards parents. For example, our Danish CLOP investigated the presence of a home-loan library, the number of books in it, and how many of the books were available in non-Danish languages (i.e. the heritage languages of children with immigrant background). Furthermore, we investigated whether preschools had pamphlets or other kinds of information about language and/or literacy development available to parents in the entrance areas. Another adaption was made regarding the use of computers. The CLOP contains an item for the number of turned-on computers, and the literacy programs available to children on them. At the time of the SPELL and LEAP

studies, Danish municipalities had begun purchasing programs to replace computers with iPads. Therefore we added an item that investigated the number of available iPads, and the number of language and/or literacy applications on them.

Another methodological challenge with CLOP occurred with regards to inter-rater reliability. Conducting the CLOP investigations was very costly. In all, 293 institutions were observed all over the country. Observations were mostly carried out by the author for the SPELL study, and by Laila Kjærbæk Hansen (a post-doc researcher) for the LEAP study. Our observations occurred simultaneously, and therefore it was not possible to double-code each other's observations. We sought therefore to minimalize this issue in a number of ways. First of all, we collaborated closely on translating the CLOP to Danish, and in this way we were both very familiar with the checklist. Secondly, in the first days of conducting CLOP observations, we held telephone meetings to discuss how we coded the CLOP, challenges we experienced, and how we resolved them.

After CLOP observations were conducted, the author compared the total mean scores. The average score for preschools in the SPELL project was 19.62 (SD = 4.85; min = 5; max = 32), whereas the average score in LEAP preschools was 18.07 (SD = 4.31; min = 7; max 31). An independent t-test revealed that the difference in mean scores was significantly different: $t(289) = -2.87$, $p = .005$. However, we do not know if this is due to lack of reliability, variation in the structural quality, or a combination of both. Nevertheless, it should be noted that the difference in mean scores is very small in practical terms, and the similar minimum and maximum scores give us confidence that our coding was reliable.

3.3. Methodological Considerations of Paper 2

The primary goal of Paper 2 was to investigate the effects of PD interventions with a language and/or literacy focus on teacher and child-level outcomes. In addition to this primary goal, we investigated the extent to which improvements in quality outcomes

predicted child outcomes. A systematic review of the literature and meta-analysis was selected as the method of approaching these goals. Teacher level outcomes of interest were process quality, structural quality, and teacher knowledge. Child level outcomes of interest were receptive vocabulary, phonological awareness, and alphabet knowledge. Although Paper 2 contains a detailed description of the specific methodology used to conduct the literature search and meta-analysis, the overall methodology is summarized briefly below followed by a discussion of why they were appropriate methods of exploring our research goals.

3.4. Summary of Methodology for Paper 2

In the initial stages of the systematic review on which Paper 2 is based, we published a protocol stipulating all aspects of the review process for Paper 2 on the Prospero database (Markussen-Brown, Juhl, Piasta, Bleses, & Højen, 2014). As stated in the protocol, the review utilized database searches that were conducted in creation of a *systematic map* that was part of the LEAP project. The protocol for the systematic map was designed by Dorthe Bleses, Anders Højen, Philip Dale, Werner Vach, Shayne Piasta and Laura Justice. The author also contributed by suggesting search terms for retrieving studies having to do with PD. The data collection and the initial screening of papers was supervised by Dorthe Bleses, Anders Højen and Philip Dale. The database for the systematic map ended up containing 65, 037 studies before screening began. Paper 2 utilized this initial pool of studies to find studies that should be included in the meta-analysis.

The review process followed the following steps. First databases were searched to identify potentially relevant studies. Then studies were screened according to our inclusion and exclusion criteria. The following numbers of trials were included: n = 25 (process quality); n = 16 (structural quality); n= 10 (teacher knowledge); n = 5 (receptive vocabulary); n = 5 (phonological awareness); n = 6 (alphabet knowledge). Included studies were then

coded for a number of variables, and data was extracted for calculating effect sizes. Each study was permitted to contribute an average effect size for each outcome of interest, but as a minimum, included studies had to contribute with at least one teacher-level outcome due to our wish to explore the mediating effect of preschool quality on child outcomes. Effect sizes were calculated as Cohens *d* and estimated as the standard mean difference (SMD), and were corrected to Hedge's *g* due to a slight tendency to overestimate effect sizes, especially in small studies. Once the effect size was estimated for each applicable outcome for each included study, a random-effects meta-analysis was conducted for each outcome. Meta-regression and sub-group analysis were used to explore variables that potentially accounted for variation in effect sizes.

3.4.1. Systematic Review and Meta-Analysis: Introduction

When searching for research on any given social-scientific topic, it is not uncommon to find that the literature contains seemingly similar investigations that have yielded diverging results. The field of preschool teachers' PD is no exception. For example, Lonigan et al. (2011) investigated the effects of a pre-literacy curriculum plus PD course, and then investigated the added-value of PD coaching in addition to the course teachers received in connection with the new curriculum. The researchers found that the curriculum plus course yielded moderate effects on children's outcomes compared to a control group, but that the additional coaching contributed little additional effects. In contrast, Neuman and Cunningham (2009) also investigated the isolated effects of coaching in addition to a language and literacy oriented course, and found that teachers who received coaching demonstrated significant improvements in literacy practices (child outcomes were not measured). Thus, two seemingly similar studies produced different results. Cooper and Hedges (2009) described this situation as the raison d'être for the systematic review: "If

results that are expected to be very similar show variability, the scientific instinct should be to account for the variability by further systematic work" (p. 4).

Systematic review is a method by which we can attempt to investigate the aggregated effect and variability in a niche of research results, and is well used in educational research (see Ahn, Ames, & Myers, 2012 for a review). Systematic review can help us approach this issue by moving beyond the question of whether an intervention works or not, and towards an understanding of the *conditions* under which a treatment works. This knowledge then lays the foundation for future empirical endeavors that can refine this understanding with specific research designs that answer more narrowed questions.

All systematic reviews are similar in that they aim to synthesize an overview of the results of the existing literature on a certain topic, but this can be achieved using different methods. In a configurative or narrative systematic review, the research synthesis compares the results of the primary literature in a descriptive fashion, often using a series of tables to categorize and group studies (Gough, Oliver, & Thomas, 2012). This method is very accessible for the layman, and allows the reader to obtain an efficient overview of the existing literature. However, another common method of analyzing the results of the primary research involves aggregating the empirical data using the method of meta-analysis.

Meta-analysis is a statistical method that is especially useful with empirical research that publishes effect sizes, or sufficient data needed to estimate effect sizes. Treating individual studies as units of analysis, the meta-analytic approach can combine the effects of multiple studies achieving a mean effect size (Card, 2011). Although an overall effect size does not tell us under which conditions a treatment works, the meta-analytic approach includes a number of tools that can reveal patterns of variation. Two of these tools are *sub-group analysis* and *meta-regression*. Using these methods, the researcher can investigate whether certain variables such as study design, treatment intensity, or sample size are

systematically associated with variation in outcomes. When systematic variation is revealed, the researcher can examine the characteristics of the responsible variable, which potentially can elucidate why variation was found. As such, meta-analysis has both the potential to inform researchers on the overall efficacy of a treatment or intervention, and to provide evidence of factors that potentially moderate or mediate effects.

3.4.2. Systematic Review and Meta-Analysis: Sub-group analysis

Sub-group analysis essentially involves grouping studies according to a categorical variable, and comparing the aggregated effect sizes for each group. For example, in Paper 2, we compared PD interventions that included coaching against studies that did not. If we found that interventions that included the coaching format yielded significantly stronger effect sizes then this would provide evidence that coaching may be a format of PD that increases quality.

Sub-group analysis can also investigate the effects of study quality or bias in effect sizes. For example, other meta-analyses have found that small sample sizes are generally more likely to produce larger effect sizes (e.g., Fukkink & Lont, 2007). This may due to smaller studies being easier to control, or publication due to a variety of reasons. In Paper 2, we used sub-group analysis to compare the effects of studies with 50 or fewer participants with studies with over 50 participants to see if sample size was also a design-related variable that could explain variation.

3.4.3. Systematic Review and Meta-Analysis: Meta-regression

Meta-regression is another analytic tool that can be used to probe deeper into the main results of a meta-analysis. Meta-regression is basically a regression model that estimates effect of a predictor variable on study effect sizes. Like with regular regression, if the slope is statistically significant, then the researcher gains confidence in the association between the moderator and the effect size. In our investigation, we used meta-regression to investigate the

extent to which teacher-level effect sizes predicted child outcome effect sizes. This is an important relationship to investigate, because despite the fact that preschool quality is the immediate target of PD interventions, the most important target is of course children's language and literacy outcomes. The threshold of preschool quality required to make meaningful improvements in children's outcomes is also a current issue in the literature (see Zaslow et al., [2010] for a review), and therefore the meta-regression analysis had the potential to contribute to this ongoing discussion.

3.4.4. The Challenge of Statistical Dependence

One of the main challenges of meta-analysis involves dealing with statistical dependence in effect sizes. Dependence occurs when a single control group is compared to multiple experiment groups, or when multiple measures are used to measure the same construct (Scammacca, Roberts, & Stuebing, 2013). In these situations, some participants are used in estimations more than once which can result in inflated variance, and studies with multiple effect sizes come to outweigh studies that used only a single outcome measure.

In Paper 2, we were also faced with challenges related to statistical dependence, as many of the included studies either contained trials comparing multiple experiment groups against a single control group, or utilized more than one instrument to measure a single construct – or both. To alleviate the effects of dependence in our study, we treated multiple experiment groups as independent trials, and divided the control group by the number of experiment groups, as recommend by Higgins and Green (2008). Thus a study such as Neuman and Wright (2010), which compared two different forms of PD against a single control group, was treated as two separate trials. Another recommend solution could have been to combine the experiment groups, but doing so would have eliminated the differential effects that the Neuman and Wright were investigating. Given that we were also investigating the effects of different PD formats, combining the experiment groups was not a desirable

option. Instead, we chose the method that retained the contribution of the original research design even though it reduced the size of the control group.

To deal with statistical dependence due to multiple outcome measures, we followed guidelines by Cooper (1998), which involved grouping outcomes for each trial by their construct, and then estimated an average effect size. This method is referred to as the shifting-unit-of-analysis, and using it ensured that included trials in Paper 2 contributed only one effect size for any six of the teacher or child-level outcomes of interest.

3.5. Methodological Considerations for Paper 3

The SPELL study included an experimental condition in which teachers received an enhanced PD module that sought to increase teachers' usage of scaffolding strategies. We aimed to do this by improving teachers' self-awareness of their scaffolding performance. As described in Chapter 2, cognitive learning theory predicts that awareness of one's own performance motivates behavioral change (Bandura & Cervone, 1983). Applied to preschool teachers, this theory predicts that if teachers discover that their usage of scaffolding is lower than expected, they will probably strive harder to increase usage. Being able to discover one's skill performance, however, requires the ability to recognize with accuracy the skill in question. In Paper 3, we investigated the extent to which teachers could accurately recognize their usage of six scaffolding strategies following three days of PD. Measuring teachers' self-awareness of skill performance posed some methodological challenges, however. In the following section, the way in which we operationalized our measure of self-awareness is described followed by a discussion of the limitations that the method had.

3.5.1. Summary of Methodology for Paper 3

In Paper 3, we used a cross-sectional study to evaluate the extent to which a sample of 73 teachers could accurately code their usage of six scaffolding strategies that supported children's learning of language and early literacy. Teachers were all participants in the

SPELL project, and in particular, they participated in an extended-PD experiment group in which they received more PD than other teachers in the project.

To evaluate how well teachers coded their performance of the six scaffolding strategies, teachers first attended a two-day course in which they were taught to use the strategies (this was the SPELL basis PD). Shortly following the basis course, teachers received one more daylong course in which they refreshed the six scaffolding strategies, and were furthermore trained to code videos for usage of the six strategies (this was part of the SPELL extended PD). Following training, teachers completed a so-called *self-coding task*.

The self-coding task required teachers to video-tape themselves performing the SPELL book-reading intervention, and then to code the video afterwards for their own use of the six strategies. Teachers submitted their coding sheets to the research team, and accuracy scores were calculated for each teacher by master-coding the videos. Some teachers also coded the videos of a colleague in addition to their own video (n = 53), which allowed us to investigate whether there were systematic differences in how teachers coded themselves versus a colleague. Thus our data set consisted of three codings per video: one by the teacher in the video, one by the teacher's colleague, and one by the master coder.

3.5.2. Validity of Measuring Self-Awareness Using a Self-Coding Task

The interpretation of the results of the self-coding task relied on certain assumptions. First of all, we assumed that accurate coding was an indication that teachers were able to calibrate (i.e. become aware of) their skill performance level. Secondly, we assumed that teachers who coded inaccurately still lacked the ability to calibrate their skill performance level. However, it is questionable whether coding accuracy is a valid measure of a preschool teacher's self-awareness of skill performance. Although it is logical to assume that teachers who code their scaffolding usage accurately are aware of their skill performance, it is more difficult to conclude that teachers who code poorly are unaware of their skill performance.

For example, some teachers who code poorly may simply be poor at coding. Although some research has found that teachers can be taught to code accurately in a relatively short period of time (Prusak, Dye, Graham, & Graser, 2010), coding may still be a skill that requires considerable practice. This could be especially true for scaffolding strategies, which teachers may be unfamiliar with (McGee & Ukrainetz, 2009).

Therefore in Paper 3, the self-coding task can only tell us with certainty if teachers learned to code accurately following three days of PD. This is not to say that teachers' accuracy scores do not reflect their self-awareness, but interpretations should be treated cautiously.

One way to add validation to the self-coding task could be to conduct a longitudinal study, in which teachers' gains in coding accuracy overtime could be compared to overall usage of scaffolding. If it were found that gains in coding accuracy predicted increased usage of scaffolding strategies, then this could provide empirical evidence that self-awareness has the motivational effect it is described to have in social cognitive theory. Future research endeavors could explore this.

4. Summary of Findings

4.1. Summary of Findings for Paper 1

In Paper 1, we investigated the process and structural quality of Danish preschools. We asked the following research questions:

i. What is the process of Danish preschools and how does it compare to an American sample?

ii. What is the structural quality of Danish preschools?

iii. To what extent does children's SES predict structural and process quality?

iv. To what extent do teacher background variables such as education and years of experience predict process quality?

v. To what extent does process quality vary as a function of instructional setting?

With regards to question 1, we found evidence of medium to high quality in the domains of emotional support and classroom organization, but low quality in instructional support. The average score for emotional support was 5.82 (out of 7), which is considered high-medium quality on the CLASS scale. The scores for the *positive* and *negative climate* dimensions were in the high end, but a medium score for *regard for student perspectives* brought the overall score down somewhat. Classroom organization was also in the high-medium range (5.67). Here we observed that Danish preschool teachers provided excellent *behavior management*, and displayed high *productivity*, but their own involvement in children's activities and learning resulted in a medium score for *instructional learning formats*. Finally, we found evidence of mostly low quality in instructional support (2.41). The three dimensions of instructional support were of particular interest to us because they are associated with children's language and emergent literacy skills (Pianta et al., 2005).

We compared our results with an American sample taken from the MyTeacherPartner (MTP) study, which is reported in the CLASS manual (Pianta et al., 2008). The comparison

revealed that emotional support and classroom organization was higher in the Danish sample, whereas instructional support was lower. It should be noted though that the MTP study was not nationally representative of American preschools.

With regards to the second research question, we found some evidence of the provision of materials that promote preliteracy skills in children, but mostly children lacked access to a broad range of materials and supports, which indicated a lack of intentional practice with regards to structural quality. For example, most children had access to books (although a handful of preschools had no books at all), but most preschools only had one or two alphabet books, or books about numbers or forms. Few institutions provided games or puzzles that dealt with words or the alphabet, and there was virtually no evidence that children produced writing. There was also only limited evidence of structural support for dual-language learners, such as books in other languages.

Regarding question 3, we found some associations between quality outcomes and SES. For structural quality, children with lower SES were more likely to go to preschools with less access to books. However, children with lower SES also went to preschools with more outreach materials to parents (such as loan-home libraries), and which had more supports directed towards dual-language learners. This indicates that some preschools recognized that they had a higher proportion of children with potential risk factors, and thus made structural changes. In terms of process quality, we found that children with lower SES were more likely to go to preschools with lower emotional support and classroom organization.

Regarding question 4, we found that teacher education had no association with CLASS scores. For the most part, teacher experience also had no relation to quality with the exception of the teachers with the least amount of experience. Their scores on the instructional support domain were somewhat higher. Furthermore, we found that teachers

who had attended a four-day PD intervention called *Sprogpakken* [The Language Package] also demonstrated somewhat higher scores in instructional support. Finally, male teachers scored lower on all CLASS domains by approximately 0.3 points.

The results of question five indicated that teachers exhibited less emotional support and higher classroom organization during structured activities such as shared-book reading and language activities. Slightly higher instructional support was also found for language activities, but the difference had little practical importance: instructional support was low regardless of whether teachers were eating with children, or conducting a language activity.

Altogether, these findings indicate that the quality Danish preschools are low in key areas that support children's language and emergent literacy development. However, we found evidence of high quality in teachers' supports for children's social-emotional development, and behavioral regulation.

4.2. Summary of Findings for Paper 2

Paper 2 was a systematic review and meta-analysis of literature testing the effects of a PD intervention for preschool teachers with a language and/or emergent literacy focus. Through extensive coding and data extraction of the included studies, we sought to answer the following research questions:

i. What is the estimated effect of PD on process quality, structural quality, and teacher knowledge?
ii. What is the estimated effect of PD on children's language and emergent literacy outcomes?
iii. To what extent do the proximal effects of PD (i.e., teacher outcomes) mediate effects on child outcomes?
iv. Which formats of PD are most effective?

v. What additional factors explain variation in the results according to a sensitivity analysis?

For question one, we found significant effects for process quality (SMD = 0.52) and structural quality (SMD = 1.07). We found a small effect for teacher knowledge (SMD = 0.15), but it was not significant.

Question two was investigated in relation to children's receptive vocabulary, phonological awareness, and alphabet knowledge. A non-significant SMD of 0.21 was estimated for vocabulary. Significant effects were found for phonological awareness (SMD = 0.46) and for alphabet knowledge (SMD = 0.18).

Due to the limited number of studies that contained extractable data for both teacher and child outcomes, our investigation of question 3 was limited to studies that contained outcomes for process quality, and at least one of the child outcomes of interest (vocabulary [n=5], phonological awareness [n = 5], and alphabet knowledge [n = 6]). No significant relations were found between effect sizes for process quality, and effect sizes for the selected child outcomes.

In answer to question four, we found that PD interventions that consisted only of courses had no significant effect on process and structural quality. However, when combined with at least one other format, significant effects were found. PD interventions that included coaching produced larger effect sizes for process and structural quality than did interventions that did not include it. Interestingly, PD interventions that consisted only of coaching were just as effective for process and structural quality as ones that combined coaching with another format of PD (however, this sample size was small).

The intensity of courses had no relation to effects, but for coaching, intervention intensity predicted process quality. Combined together, the intensity of coaching and courses predicted both process and structural quality. We also found that the number of PD formats in

an intervention significantly predicted effect sizes for quality. The overall length of the intervention period predicted structural quality, but not process quality.

Finally, our investigation of question five revealed a number of factors that had influences on effect sizes. We found that studies with small sample sizes produced larger effects for structural quality, and that studies that included a majority of children at-risk for academic failure yielded larger effects on both process and structural quality. Furthermore, studies that randomized at the institution level rather than the classroom or teacher levels produced larger effects for structural quality. Finally, studies that reported accurate procedural fidelity yielded smaller effects for structural quality, and studies that utilized appropriate blinding procedures yielded larger effects for process quality.

The results of the investigation indicated that PD interventions have significant effects on process and structural quality, as well as phonological awareness and alphabet knowledge. However, gains in quality did not predict gains in child outcomes. Furthermore, we found that both courses and coaching are beneficial formats of PD, but that coaching appears to contribute more, whereas courses ought to be combined with at least one other format of PD. However, this finding is potentially confounded by course or coaching intensity. Coaching interventions tended to have higher intensity.

4.3. Summary of Findings for Paper 3

In Paper 3, we investigated whether teachers could be taught to code with accuracy their performance of six scaffolding strategies following three days of PD. We assumed that accurate coding would indicate that teachers were able to calibrate their skill performance, whereas inaccurate coding indicated an incomplete learning of the strategies. In particular, we asked the following research questions:

i. To what extent do teachers code their usage of six scaffolding strategies accurately following three days of PD?

ii. To what extent do teachers demonstrate the same coding accuracy for all strategies?

iii. Does teacher coding accuracy change as a function of whether they are coding themselves or a colleague?

With regards to question one, we found that teachers largely scored their performance of the six scaffolding strategies inaccurately, but with two different patterns emerging from the data. We found that teachers generally overrated their usage of the high support strategies, which are those used to offer children the most assistance. In contrast, teachers both under- and overrated their usage of the low support strategies. This indicated that some teachers thought they were using the strategies when they were not, and others did not recognize that they had used the strategies. Interestingly, when teachers did code their performance accurately, it was usually in recognition of not having used any strategies.

With regards to our second question, we found only some evidence that a teacher's coding accuracy for one strategy was shared with the other strategies. Teachers demonstrated a moderate tendency to code some of the high support strategies similarly, but for the low support strategies, there were only weak or non-significant correlations. Altogether, the results suggested that the strategies functioned as discrete units, and were not equal in their learnability.

Our investigation of question three revealed little evidence that coding accuracy increased or decreased as a function of whether teachers coded themselves or a colleague. Rather, first and second-coders tended to code similarly.

Overall, the study demonstrated that teachers may have difficulty perceiving their performance of new skills despite three days of PD. Furthermore, individual scaffolding strategies may function as discrete units with regards to teachers' learning of them. Some

strategies appeared to be harder to code accurately for some teachers than for others. Finally, teachers did not display any bias when coding the videos of others.

5. Discussion

The results of the three studies will be discussed in two parts. First, they will discussed individually with regards to how they relate to previous research, how they contribute to the literature, what limitations they might have, and what perspectives for future research there may be. Then they will be discussed in their entirety as a contribution to our knowledge of early childhood education in Denmark.

5.1. Discussion of Paper 1

Paper 1 was unique in that it was the first large-scale investigation of the quality of Danish preschools. The investigation of process quality revealed both examples of high and low quality. True to the roots of the Danish pedagogical tradition, which tends to focus on children's socialization (Jensen, 2009), we found that Danish teachers provided high levels of emotional support. However, we also found that Danish preschool teachers appeared to lack interactional skills that support children's language and general learning, which is a finding similar to other research from the United States (e.g., Burchinal et al., 2008). This is a concerning finding considering other research that indicates that preschool only facilitates gains in children if quality is high (Logan et al., 2011; NICHD Early Child Care Research Network, 2006). It was also concerning to find that children from socially disadvantaged families were more likely to attend preschools with lower emotional support and classroom support. However, in contrast to American research that has found that socially disadvantaged children are more likely to attend preschools with lower instructional support (e.g., Justice et al., 2008; Pianta et al., 2005), we did not find strong evidence of an association between SES and instructional support. This domain was equally low for all groups of children.

Teachers' education did not relate to their scores on the CLASS scale, which mirrors a similar finding by Justice et al. (2008). However, whereas Justice and colleagues also found

no association with years of work experience, we found a small advantage for teachers with least experience. Although this might seem unintuitive, it could be that teachers with least experience are also new college graduates, and thereby possess a more up-to-date knowledge of language and emergent literacy practices. The fact that teachers demonstrated significantly higher scores on the instructional support domain if they had attended a four day PD on language and preliteracy development supports the notion that in-service teachers may benefit from up-to-date evidence based PD opportunities.

With regards to structural quality, we found little evidence of an intentional practice in providing children with materials that support their interactions with literacy. Furthermore, we found that socially disadvantaged children had less access to books than other children. This is also an unfortunate finding considering other research that found that socially disadvantaged children have fewer interactions with literacy materials in their home communities (Neuman & Celano, 2001). However, the finding that preschools with higher percentages of socially disadvantaged children provided more outreach to parents, and more supports for dual-language learners indicates awareness and responsiveness to the challenges that face children growing up in poverty. Nevertheless, our findings indicate that Danish preschools require considerable improvement in their support of children's early interactions with literacy.

Finally, we found that process quality did not change dramatically as a function of instructional setting. We found that during structured activities (shared-book reading and language activities) that emotional support was slightly lower, and classroom organization was slightly higher. Furthermore, we found that instructional support was slightly higher during language activities. However, the differences in practical terms were not large. This is an interesting finding for a few reasons. First of all, that language activities did not yield much higher scores in instructional support suggests that teachers might lack the tools and

strategies to maximize the benefit of these activities for children. Alternatively, the findings might also be interpreted as demonstrating that all instructional situations have the potential for learning opportunities.

This study is limited in that it is a cross-sectional study, and the results cannot be used to draw causal conclusions. Our knowledge of the quality of Danish preschools could potentially be increased through the usage of longitudinal observational studies, or through experimentation. For example, we found that children with socially disadvantage were more likely to attend preschools with lower behavior management. However, it is possible that these children also contribute to the lower behavior management score on the CLASS by making the teacher's work more challenging. The research design in Paper 1 does not elucidate more on this topic. Other questions about directionality of effects could be studied using another type of research design.

5.2. Discussion of Paper 2

Paper 2 found that PD interventions had a medium effect on process quality, a large effect on structural quality, and no effect on teacher knowledge. Furthermore, we found significant effects for phonological awareness and alphabet knowledge, but not for vocabulary. These findings are only partially in line with a previous meta-analysis of the effects of training on child-care providers by Fukkink and Lont (2007). They found that training had positive effects on teacher processes as well as knowledge. Furthermore, they did not find significant effects for child outcomes. However, Paper 2 is limited in its comparability with the meta-analysis by Fukkink and Lont. Our investigation dealt exclusively with PD interventions that focused on language and/or literacy, and our outcome for teacher knowledge was also limited to teachers' knowledge of language and literacy development. Research has previously found that preschool teachers may have low levels of language and literacy knowledge (Cunningham, Zibulsky, & Callahan, 2009).

Although we found that PD had some positive effects on child outcomes, we did not find that process quality effect sizes mediated children's gains. This finding calls into question the role of process quality in children's development of language and emergent literacy skills. Previous research has found that process quality is associated with child outcomes (e.g., Mashburn et al., 2008; Pianta et al., 2005), but our meta-analysis indicates that improvements in process quality might not result in equal effects in children's outcomes. Burchinal et al. (2010) found that associations between process quality and child outcomes were stronger when quality was higher. It is therefore possible that the process quality in our included studies was still too low (despite statistically significant improvements) to demonstrate strong relations to children's outcomes. It should be noted, however, that the sample of studies that included child outcomes was small (5-6 studies per outcome). It is entirely possible that a larger sample would reveal significant associations.

Our analysis of PD formats revealed that courses were beneficial when they were combined with at least one other format of PD, but coaching was effective in isolation or combined with more formats. This finding could be due to coaching being a more effective form of PD, but it should be noted that the intensity of coaching interventions was higher than the intensity of courses, which introduces a possible confound. However, Neuman and Wright (2010) compared a course-based PD intervention against a coaching-based intervention with the same intensity. They found no significant effects on process quality, but coaching did benefit structural quality more than courses.

We also found evidence that the number of PD formats was related to quality outcomes. This *more-is-more* finding suggests that teachers benefit from multiple learning platforms when they receive PD, and we find support for this notion from studies such as Landry et al. (2009), who also found that comprehensive PD models were more effective than models consisting of fewer formats. In addition to the cumulative effect of multiple formats

of PD, we also found that PD duration was a significant predictor of quality. Grace et al. (2008) was the included study with the longest duration (three years), and the effects of their study were some of the highest. This indicates that PD interventions are most effective when they have a long-term framework.

Paper 2 was limited in a number of ways. Several experimental studies that otherwise met inclusion criteria had to be excluded because they failed to present the data needed for estimating effect sizes. The study was also limited by the small subsets of studies that included outcome data for receptive vocabulary, phonological awareness, and alphabet knowledge. Another cautionary point is that the review consisted mostly of studies conducted in North America. This is not to say that we should assume that the learning processes of preschool teachers in other countries are fundamentally different, but the need for experimental research from countries other than Canada and the United States is needed to better understand how PD interventions function in other linguistic and cultural contexts.

Future research should work towards identifying the underlying factors and processes that make PD interventions successful. Sheridan, Edwards, Marvin, and Knoche (2009) recommended a paradigm shift in which researchers moved beyond evaluating the overall effect of PD interventions, and investigated the conditions under which PD works, and for whom. We find support for this recommendation in our own meta-analysis. Although we were able to identify evidence of factors that potentially explain variation in the effects of PD, future research should more systematically investigate the effects of factors such as intervention intensity and duration. Furthermore, researchers might also consider studying more the effects of processes that underlie PD experiences such as quality of PD delivery (i.e. how well courses are taught), or even take a more theoretical approach, and investigate the psychological processes teachers experience while receiving PD interventions. Social cognitive theory (Bandura, 1977b, 1986) cites a number of factors that are important when

learning new skills, such as attention to the learning target, and ability to retain information. A greater understanding of how or if these factors support or facilitate learning could lead to interventions that are more effective.

5.3. Discussion of Paper 3

Paper 3 is unique in comparison to the other two studies presented in this dissertation in that the investigation was primarily theoretically motivated. The study was founded in a social cognitive theory approach to PD that posits that teachers' awareness of low skill-performance can motivate behavioral change. We investigated teachers' ability to capitalize on this motivational effect by investigating their awareness of own skill performance. This was operationalized by teaching teachers to use six scaffolding strategies, and then by giving them a coding-task to evaluate if they were able to code their usage of the strategies. If their coding was accurate, we assumed that they had calibrated knowledge of their skill level, and could use this information to improve their practice further. If coding was inaccurate, it indicated for us an incomplete learning process of the strategies.

The results of the coding task indicated that teachers mostly coded inaccurately, but with two distinct patterns emerging. Teachers overrated their usage of the high support scaffolding strategies when in fact they rarely used them. Furthermore, they both under- and overrated their usage of the low support strategies. This finding mirrors the results of a smaller study by Pentimonti and Justice (2010), who found that teachers rarely used high support strategies, favored low support strategies, and generally displayed lack of calibration with regards to their skill performance. Whereas the study by Pentimonti and Justice was based on a small sample of teaches (n = 5), Paper 3 utilized a much larger sample (n = 73).

We also found that coding accuracy tended to be more strategy-specific than it was general for all strategies (especially for the low support strategies), meaning that a teacher's ability to code one strategy accurately did not necessarily transfer to the other strategies. This

finding has potential implications for future PD intervention models. In our PD course, we assumed an equal learnability of all six scaffolding strategies, and taught teachers the strategies in very short succession. However, the results of Paper 3 indicate that an incremental method of teaching might potentially be more effective, since teachers appear to learn each strategy as independent units. Allowing teachers to master one strategy before moving on to the next is a method of teaching that could be explored in future research.

Finally, we found little evidence that teachers coded their own skill-usage any more or less accurately than the skill performance of teacher colleagues. This finding suggests that teachers coded without bias. Being able to code without bias is an important factor of using self-coding as a tool for PD.

The results of the study have two other implications, the first pertaining to the effectiveness of PD interventions, and the second pertaining to the usefulness of self-coding tasks. That teachers demonstrated poor calibration of their skill performance following the PD indicates that teachers may require increased learning supports if they are to acquire deep knowledge of novel practices over a short period of time. In the case of Paper 3, a relatively short workshop did not appear to be sufficient for instilling a deep understanding of the six scaffolding strategies in teachers.

Coaching is designed to help teachers transfer knowledge from the classroom to their practices (Gupta and Janese, 2012), and could be a useful supplement to courses and workshops. However, repeat usage of self-coding tasks may also hold the potential for helping teachers hone skills in a classroom-setting. Although teachers in our study generally coded inaccurately, we do not consider the self-coding task to be ineffective. In fact, there were a number of occurrences of teachers coding some strategies accurately. Interestingly, this was most often in recognition that no strategies were used. We consider this an encouraging finding because the intention of the self-coding task was to help teachers

"discover" that their true skill performance could be lower than expected, and thereby motivate them to redouble their efforts. Despite the fact that the three days of PD courses were not sufficient to reach coding accuracy of all strategies, some teachers did appear to calibrate their skill performance knowledge. Perhaps with more practice, teachers' scoring accuracy would increase.

This study is limited by its cross-sectional design. The study does not investigate whether self-coding increases awareness of skill performance, but rather if teachers could code accuracy following three days of PD. A longitudinal study could follow teachers' coding skills over time, and furthermore investigate the extent to which gains in calibration are linked to increased usage of the strategy in question. Such a finding could be used to validate the theoretical approach of social cognitive theory, and contribute to our knowledge of how to design effective PD models.

5.4. Discussion of the Dissertation in its Entirety

In its entirety, this dissertation aimed to provide Danish policymakers with new research that could be used to improve the language and emergent literacy outcomes of children attending Danish preschools via improved quality in preschool learning environments. To do so, we sought to describe the process quality of Danish preschools, and their provision of literacy supporting materials. Then we investigated the pooled effects of previous PD interventions with a language and/or literacy focus. Finally, we investigated whether Danish teachers demonstrated awareness of their skill performance of strategies that they recently had learned during a PD intervention. Altogether, this dissertation describes the quality of Danish preschools, what effect can be expected of PD interventions that seek to improve their quality, and an indication of how much teachers learn following a typical PD course.

Although we found evidence of high quality in the socio-emotional environment, the quality of the language and literacy environments was consistently low, and our general finding is that there is considerable need for improvement. These results are not necessarily surprising given that the traditional Danish holistic approach to pedagogy focuses on children's social development and general well-being, but strays away from systematic instruction of children's pre-academic skills (Jensen, 2009). It is also important to note, that our evaluation of process quality may have even overrated the true quality. This is because we scored the CLASS based on videos in which teachers interacted with no more than five children, which is a teacher-child ratio that is preferable to the ratio they normally work with.

Our investigation of the structural environments also revealed that children have limited access to a wide range of materials that support early interactions with literacy. For example, it was very typical for children to have access to books, but the books did not appear to be intentionally selected or positioned for the children, and very few of the books were alphabet books, or books about numeracy or shapes. Overall, we surmise that Danish preschools do not have a clear conceptualization or plan of how they can support children's early development of literacy. The variability in findings points towards a general lack of intentionality.

It is our recommendation that steps be taken to support preschools in adopting evidence-based knowledge and practices that can support children's language and literacy development. Doing so will require extensive PD initiatives and other forms of teacher training. Such steps have already occurred in a limited fashion. For example, the *Sprogpakken* course (which occurred in 2011-2012) was a large-scale initiative that offered one teacher from every preschool in Denmark a place in the course. We found in Paper 1 evidence that *Sprogpakken* still had observable effects on instructional practices one year after the four day course took place. However, *Sprogpakken* was not far-reaching enough.

Birman, Desimone, Porter, and Garet (2000) advised that all teachers from a preschool should be involved in PD initiatives, because it can be hard for a single individual who has attended PD to spread change to the practice of all the other preschool personal. Comprehensive, more long-term initiatives are likely needed if we are to increase the language and literacy environments of Danish preschools.

Designing and implementing PD initiatives that can increase the quality in Danish preschools is thus an important new goal, and Papers 2 and 3 offer some insight here. Our meta-analysis revealed that PD interventions do indeed have effects on quality, although structural quality was more malleable than process quality. We also saw that some models of PD were more effective than others. Factors that were found to increase the effectiveness of interventions were intensity, duration, and numbers of formats of PD. Thus, future PD initiatives in Denmark ought to be comprehensive and provide teachers with the possibility to learn over a longer period of time with a high number of PD supports.

We find support for our long-term view of PD from Paper 3 that found that teachers displayed difficulty in recognizing their skill usage of six scaffolding strategies after three days of typical PD workshops. Although the self-coding task that we employed was not a true test of teachers' learning, it did suggest that most teachers had not mastered the skills we attempted to teach them. Rather, they appeared to overrate their skill usage, and to a large extent in the case of the high support strategies. We find this to be demonstrative of the error in relying on short-term PD initiatives to change teacher practice. If we expect teachers to make comprehensive changes to their practice, they should be afforded the time and supports needed to transform PD opportunities to permanent changes in practice.

As a final recommendation, we suggest that the Danish pedagogical college degree program be reexamined. Denmark is unique in comparison to countries such as Canada and the United States in that most teachers have the same 3½ year pedagogical degree. Because

low quality in the language and literacy environments of the sampled preschools was widespread with few exceptions to the rule, it is logical to investigate the need for infusing more education about language and literacy development into the basic degree program. Future research should investigate the extent to which this could be beneficial.

6. References

Ahn, S., Ames, A. J., & Myers, N. D. (2012). A review of meta-analyses in education: Methodological strengths and weaknesses. *Review of Educational Research, 82,* 436-476.

Babbie, E. (2015). *The practice of social research*. United States: Wadsworth Cengage Learning.

Bandura, A. (1977a). Self-efficacy: toward a unifying theory of behavioral change. *Psychological review, 84*(2), 191.

Bandura, A. (1977b). *Social learning theory*. Englewood Cliffs, NJ.: Prentice-Hall.

Bandura, A. (1986). *Social foundations of thought and action*. Englewood Cliffs, NJ: Prentice Hall.

Bandura, A., & Cervone, D. (1983). Self-evaluative and self-efficacy mechanisms governing the motivational effects of goal systems. *Journal of personality and social psychology, 45*(5), 1017.

Bandura, A., Ross, D., & Ross, S. A. (1961). Transmission of aggression through imitation of aggressive models. *The Journal of Abnormal and Social Psychology, 63*(3), 575-582. doi: 10.1037/h0045925

Birman, B. F., Desimone, L., Porter, A. C., & Garet, M. S. (2000). Designing professional development that works. *Educational leadership, 57*(8), 28-33.

Bleses, D., Højen, A., Dybdal, L., Dale, P., Justice, L., Piasta, S., . . . Aktürk Ari, B. (2014). A randomized controlled trial of a large scale systematic and explicit language and literacy intervention in Danish daycares: The SPELL study. *British Journal of Education, Society & Behavioural science, 4*(11), 504-1524.

Bleses, D., Højen, A., Justice, L., Dale, P., Dybdal, L., Piasta, S., . . . Andersen, M. K. (2015). *Unpacking early childhood language and literacy curricula: Effectiveness of*

individual components on children's outcomes. Fourthcoming. University of Southern Denmark.

Bleses, D., Højen, A., Jørgensen, R., Jensen, K. Ø., & Vach, W. (2010). Sprogvurdering af 3-årige - karakteristika og risikofaktorer. [Language screening of 3-year-olds – Characteristics and risk factors]. *University of Southern Denmark working papers,* (10). http://www.humaniora.sdu.dk/boernesprog/DK/forsk/doc/E-print_10_2010.pdf

Bodrova, E., & Leong, D. J. (2001). Tools of the Mind: A Case Study of Implementing the Vygotskian Approach in American Early Childhood and Primary Classrooms. *Innodata Monographs 7*.

Bronfenbrenner, U. (1979). *The ecology of human development: Experiments by nature and design*: Harvard university press.

Bryant, D. M., Burchinal, M., Lau, L. B., & Sparling, J. J. (1994). Family and classroom correlates of head start children's developmental outcomes. *Early Childhood Research Quarterly, 9*(3–4), 289-309. doi: http://dx.doi.org/10.1016/0885-2006(94)90011-6

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. doi: 10.1207/S1532480XADS0601_01

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 Development Science, 12*(3), 140-153.

Burchinal, M. R., Vandergrift, N., Pianta, R., & Mashburn, A. (2010). Threshold analysis of association between child care quality and child outcomes for low-income children in pre-kindergarten programs. *Early Childhood Research Quarterly, 25*(2), 166-176. doi: http://dx.doi.org/10.1016/j.ecresq.2009.10.004

Buysse, V., Winton, P. J., & Rous, B. (2009). Reaching consensus on a definition of professional development for the early childhood field. *Topics in early childhood special education, 28*(4), 235-243.

Cabell, S. Q., Tortorelli, L. S., & Gerde, H. K. (2013). How Do I Write…? Scaffolding Preschoolers' Early Writing Skills. *The Reading Teacher, 66*(8), 650-659. doi: 10.1002/trtr.1173

Card, N. A. (2011). *Applied meta-analysis for social science research*. New York: Guilford Press.

Chapman, R. S. (2000). Children's Language Learning: An Interactionist Perspective. *Journal of Child Psychology and Psychiatry, 41*(1), 33-54. doi: 10.1111/1469-7610.00548

Chomsky, N. (1969). *Aspects of the Theory of Syntax* (Vol. 11). MIT press.

Chomsky, N. (1980). Rules and representations. *Behavioral and brain sciences, 3*(01), 1-15.

Cooper, H. (1998). *Synthesizing research: A guide for literature reviews* (Vol. 2): Sage.

Cooper, H., & Hedges, L. V. (2009). Research synthesis as a scientific process. *The handbook of research synthesis and meta-analysis, 2*, 3-17.

Crane Center for Early Childhood Research and Policy. (2012). *The Classroom Literacy Observation Profile*. Unpublished measure. Columbus, OH.

Cunningham, A. E., & Stanovich, K. E. (1997). Early reading acquisition and its relation to reading experience and ability 10 years later. *Developmental psychology, 33*(6), 934-945. doi: 10.1037/0012-1649.33.6.934

Cunningham, A. E., Zibulsky, J., & Callahan, M. D. (2009). Starting small: Building preschool teacher knowledge that supports early literacy development. *Reading and Writing, 22*(4), 487-510.

Currie, J., & Thomas, D. (1995). Does Head Start Make a Difference? *The American Economic Review*, 341-364.

Dickinson, D. K. (2006). Toward a Toolkit Approach to Describing Classroom Quality. *Early Education and Development, 17*(1), 177-202. doi: 10.1207/s15566935eed1701_8

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

Dynia, J. M. (2013). *The Literacy Environment of Early Childhood Special Education Classrooms: Predictors of Print Knowledge.* (Dissertation). The Ohio State University.

Egelund, N., Nielsen, C. P., & Rangvid, B. S. (2011). PISA Etnisk 2009: Etniske og danske unges resultater i PISA 2009. Copenhagen: AKF.

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. doi: 10.1080/02568549909594732

Fukkink, R. G., & Lont, A. (2007). Does training matter? A meta-analysis and review of caregiver training studies. *Early Childhood Research Quarterly, 22*(3), 294-311.

Garces, E., Thomas, D., & Currie, J. (2000). *Longer term effects of Head Start*. California Center for Population Research.

Girolametto, L., Weitzman, E., & Greenberg, J. (2003). Training day care staff to facilitate children's language. *American Journal of Speech-Language Pathology, 12*(3), 299-311.

Gough, D., Oliver, S., & Thomas, J. (2012). *An introduction to systematic reviews*: Sage.

Government of Denmark. (2011). Dagtilbudsloven. [Day care legislation]. LBK numberr 668.

Grace, C., Bordelon, D., Cooper, P., Kazelskis, R., Reeves, C., & Thames, D. G. (2008). Impact of Professional Development on the Literacy Environments of Preschool Classrooms. *Journal of Research in Childhood Education, 23*(1), 52-81. doi: 10.1080/02568540809594646

Guo, Y., Justice, L. M., Kaderavek, J. N., & McGinty, A. (2012). The literacy environment of preschool classrooms: contributions to children's emergent literacy growth. *Journal of Research in Reading, 35*(3), 308-327. doi: 10.1111/j.1467-9817.2010.01467.x

Gupta, N. D., & Simonsen, M. (2010). Non-cognitive child outcomes and universal high quality child care. *Journal of Public Economics, 94*(1–2), 30-43. doi: http://dx.doi.org/10.1016/j.jpubeco.2009.10.001

Gupta, S. S., & Daniels, J. (2012). Coaching and professional development in early childhood classrooms: Current practices and recommendations for the future. *NHSA Dialog, 15*(2), 206-220.

Halle, T., Vick Whittaker, J., & Anderson, R. (2010). Quality in early childhood care and education settings: A compendium of measures. *Washington, DC: Child Trends. Prepared by Child Trends for the Office of Planning, Research and Evaluation, Administration for Children and Families, US Department of Health and Human Services.*

Hamre, B. K., Justice, L. M., Pianta, R. C., Kilday, C., Sweeney, B., Downer, J. T., & Leach, A. (2010). Implementation fidelity of MyTeachingPartner literacy and language activities: Association with preschoolers' language and literacy growth. *Early Childhood Research Quarterly, 25*(3), 329-347.

Harms, T., Clifford, R. M., & Cryer, D. (1998). *Early childhood environment rating scale*: Teachers College Press.

Hart, B., & Risley, T. R. (1995). *Meaningful differences in the everyday experience of young American children*. Baltimore, MD, US: Paul H Brookes Publishing.

Higgins, J. P., & Green, S. (2008). *Cochrane handbook for systematic reviews of interventions* (Vol. 5). Wiley Online Library.

Hoff-Ginsberg, E., & Shatz, M. (1982). Linguistic input and the child's acquisition of language. *Psychological bulletin, 92*(1), 3-26. doi: 10.1037/0033-2909.92.1.3

Hoff, E. (2003). The Specificity of Environmental Influence: Socioeconomic Status Affects Early Vocabulary Development Via Maternal Speech. *Child Development, 74*(5), 1368-1378. doi: 10.1111/1467-8624.00612

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. doi: http://dx.doi.org/10.1016/j.ecresq.2007.05.002

Jensen, B. (2009). A Nordic approach to Early Childhood Education (ECE) and socially endangered children. *European Early Childhood Education Research Journal, 17*(1), 7-21.

Jespersen, C. (2006). *Socialt udsatte børn i dagtilbud* (Vol. 6): Socialforskningsinstituttet.

Juel, C. (1988). Learning to read and write: A longitudinal study of 54 children from first through fourth grades. *Journal of Educational Psychology, 80*(4), 437-447. doi: 10.1037/0022-0663.80.4.437

Justice, L. M., Mashburn, A. J., Hamre, B. K., & Pianta, R. C. (2008). Quality of language and literacy instruction in preschool classrooms serving at-risk pupils. *Early Childhood Research Quarterly, 23*(1), 51-68. doi: http://dx.doi.org/10.1016/j.ecresq.2007.09.004

Justice, L. M., McGinty, A. S., Cabell, S. Q., Kilday, C. R., Knighton, K., & Huffman, G. (2010). Language and Literacy Curriculum Supplement for Preschoolers Who Are Academically At Risk: A Feasibility Study. *Language, Speech, and Hearing Services in Schools, 41*(2), 161-178. doi: 10.1044/0161-1461(2009/08-0058)

Landry, S. H., Anthony, J. L., Swank, P. R., & Monseque-Bailey, P. (2009). Effectiveness of comprehensive professional development for teachers of at-risk preschoolers. *Journal of Educational Psychology, 101*(2), 448-465. doi: 10.1037/a0013842

Logan, J. R., Piasta, S., Justice, L., Schatschneider, C., & Petrill, S. (2011). Children's Attendance Rates and Quality of Teacher-Child Interactions in At-Risk Preschool Classrooms: Contribution to Children's Expressive Language Growth. *Child & Youth Care Forum, 40*(6), 457-477. doi: 10.1007/s10566-011-9142-x

Lonigan, C. J., Burgess, S. R., & Anthony, J. L. (2000). Development of emergent literacy and early reading skills in preschool children: Evidence from a latent-variable longitudinal study. *Developmental psychology, 36*(5), 596-613. doi: 10.1037/0012-1649.36.5.596

Lonigan, C. J., Farver, J. M., Phillips, B. M., & Clancy-Menchetti, J. (2011). Promoting the development of preschool children's emergent literacy skills: A randomized evaluation of a literacy-focused curriculum and two professional development models. *Reading and Writing, 24*(3), 305-337.

Markussen-Brown, J., Juhl, C. B., Piasta, S., Bleses, D., & Højen, A. (2014). The effects of professional development (PD) on early childhood educators measured in terms of knowledge, process quality, and structural quality, and children's emergent literacy and language skills at 3-6 years of age: A meta-analysis. Prospero databse. http://www.crd.york.ac.uk/PROSPERO/display_record.asp?ID=CRD42014009361.

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. doi: 10.1111/j.1467-8624.2008.01154.x

McClelland, M. M., Cameron, C. E., Connor, C. M., Farris, C. L., Jewkes, A. M., & Morrison, F. J. (2007). Links between behavioral regulation and preschoolers' literacy, vocabulary, and math skills. *Developmental psychology, 43*(4), 947-959. doi: 10.1037/0012-1649.43.4.947

McGee, L. M., & Ukrainetz, T. A. (2009). Using Scaffolding to Teach Phonemic Awareness in Preschool and Kindergarten. *The Reading Teacher, 62*(7), 599-603. doi: 10.1598/RT.62.7.6

Moats, L. C. (1999). *Teaching reading is rocket science*. Washington D.C.: American Federation of teachers

Neuman, S. B. (1999). Books Make a Difference: A Study of Access to Literacy. *Reading Research Quarterly, 34*(3), 286-311. doi: 10.1598/RRQ.34.3.3

Neuman, S. B., & Celano, D. (2001). Access to Print in Low-Income and Middle-Income Communities: An Ecological Study of Four Neighborhoods. *Reading Research Quarterly, 36*(1), 8-26. doi: 10.1598/RRQ.36.1.1

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

Neuman, S. B., & Roskos, K. (1993). Access to Print for Children of Poverty: Differential Effects of Adult Mediation and Literacy-Enriched Play Settings on Environmental and Functional Print Tasks. *American Educational Research Journal, 30*(1), 95-122. doi: 10.2307/1163191

Neuman, S. B., & Wright, T. S. (2010). Promoting language and literacy development for early childhood educators: A mixed-methods study of coursework and coaching. *The Elementary School Journal, 111*(1), 63-86.

NICHD Early Child Care Research Network. (2006). Child-care effect sizes for the NICHD Study of Early Child Care and Youth Development. *American Psychologist, 61*(2), 99-116. doi: 10.1037/0003-066X.61.2.99

O'Connor, R., Notari-Syverson, A., & Vadasy, P. F. (2005). *Ladders to Literacy: A kindergarten activity book* (2nd ed.). Baltimore, MD: Brookes.

OECD. (2013). Early Childhood Education and Care *Education Today 2013: The OECD Perspective*: OECD Publishing.

Pakarinen, E., Lerkkanen, M.-K., Poikkeus, A.-M., Kiuru, N., Siekkinen, M., Rasku-Puttonen, H., & Nurmi, J.-E. (2010). A validation of the Classroom Assessment Scoring System in Finnish kindergartens. *Early Education and Development, 21*(1), 95-124.

Pentimonti, J. M., & Justice, L. M. (2010). Teachers' use of scaffolding strategies during read alouds in the preschool classroom. *Early Childhood Education Journal, 37*(4), 241-248.

Piaget, J. (1959). *The language and thought of the child*. Psychology Press.

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. doi: 10.1207/s1532480xads0903_2

Pianta, R. C., La Paro, K. M., & Hamre, B. K. (2008). *Classroom Assessment Scoring System Manual, Pre-K*. Baltimore, MD: Brookes.

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

Prusak, K., Dye, B., Graham, C., & Graser, S. (2010). Reliability of pre-service physical education teachers' coding of teaching videos using studiocode analysis software. *Journal of Technology and Teacher Education, 18*(1), 131-159.

Scammacca, N., Roberts, G., & Stuebing, K. K. (2013). Meta-Analysis With Complex Research Designs Dealing With Dependence From Multiple Measures and Multiple Group Comparisons. Review of Educational Research. doi: 10.3102/0034654313500826

Schneider, P., & Watkins, R. V. (1996). Applying Vygotskian Developmental Theory to Language Intervention. *Language, Speech, and Hearing Services in Schools, 27*(2), 157-170. doi: 10.1044/0161-1461.2702.157

Sheridan, S. M., Edwards, C. P., Marvin, C. A., & Knoche, L. L. (2009). Professional Development in Early Childhood Programs: Process Issues and Research Needs. *Early Education and Development, 20*(3), 377-401. doi: 10.1080/10409280802582795

Skinner, B. F. (1938). The behavior of organisms: An experimental analysis. Oxford: Appleton-Century.

Smith, S. S., & Dixon, R. G. (1995). Literacy Concepts of Low- and Middle-Class Four-Year-Olds Entering Preschool. *The Journal of Educational Research, 88*(4), 243-253. doi: 10.1080/00220671.1995.9941305

Snow, C. E. (1972). Mothers' speech to children learning language. *Child Development*, 549-565.

Sylva, K., & Taggart, B. (2010). *ECERS-E: The Four Curricular Subscales Extension to the Early Childhood Environment Rating Scale (ECERS-R)*: Teachers College Press.

Tomasello, M. (2005). *Constructing a language: A usage-based theory of language acquisition*: Harvard University Press.

van Kleeck, A. (1998). Preliteracy Domains and Stages: Laying the Foundations for Beginning Reading. *Communication Disorders Quarterly, 20*(1), 33-51. doi: 10.1177/152574019802000105

Vygotsky, L. (1978). *Mind in society: The development of higher mental process*: Cambridge, MA: Harvard University Press.

Whitehurst, G. J., Falco, F. L., Lonigan, C. J., Fischel, J. E., DeBaryshe, B. D., Valdez-Menchaca, M. C., & Caulfield, M. (1988). Accelerating language development through picture book reading. *Developmental psychology, 24*(4), 552-559. doi: 10.1037/0012-1649.24.4.552

Whitehurst, G. J., & Lonigan, C. J. (1998). Child Development and Emergent Literacy. *Child Development, 69*(3), 848-872. doi: 10.1111/j.1467-8624.1998.tb06247.x

Wood, D., Bruner, J. S., & Ross, G. (1976). the role of tutoring in problem solving. *Journal of Child Psychology and Psychiatry, 17*(2), 89-100. doi: 10.1111/j.1469-7610.1976.tb00381.x

Wood, D., & Middleton, D. (1975). A STUDY OF ASSISTED PROBLEM-SOLVING. *British Journal of Psychology, 66*(2), 181.

Zaslow, M., Anderson, R., Redd, Z., Wessel, J., Tarullo, L., & Burchinal, M. (2010). Quality Dosage, Thresholds, and Features in Early Childhood Settings: A Review of the Literature, OPRE 2011-5. *Washington, DC: Office of Planning, Research and Evaluation, Administration for Children and Families, US Department of Health and Human Services*.

Part 2

The Three Studies

Paper 1

The Quality of the Language and Emergent Literacy Environments in Danish Preschools: Evidence from a Scandinavian Model

Justin Markussen-Brown and Dorthe Bleses

The University of Southern Denmark

Jessica A. R. Logan and Laura M. Justice

The Ohio State University

Author Note

Justin Markussen-Brown, Institute for Language and Communication, University of Southern Denmark; Dorthe Bleses, Institute for Language and Communication, University of Southern Denmark; Jessica A. R. Logan, Crane Center for Early Childhood Research and Policy, The Ohio State University; Laura M. Justice, Crane Center for Early Childhood Research and Policy, The Ohio State University.

Correspondence to concerning this article should be addressed to Justin Markussen-Brown, Institute for Language and Communication, University of Southern Denmark, Campusvej 55, Odense 5230, Denmark. E-mail: justinmb@sdu.dk.

Abstract

Little research has been conducted in Denmark regarding the quality of the language and emergent literacy environments in preschools. This is despite the fact that approximately 95% of four-year-olds attend preschool (OECD, 2013). In this study, we evaluated the structural quality of the literacy environments in 293 Danish preschools using the Classroom Literacy Assessment Profile (Crane Center for Early Childhood Research and Policy, 2012), and furthermore investigated the process quality of 506 preschool teachers using the Classroom Assessment Scoring System (Pianta, La Paro, & Hamre, 2008). The results for structural quality were generally low with children having limited access to materials that support literacy development. In terms of process quality, emotional support was generally high, classroom organization was medium-high, and instructional support was low. Negative associations were found between children's socioeconomic status and process quality, but the relation was weaker with instructional support, possibly because it was low for all children. Teacher background variables such as age and education were not associated with CLASS scores, but male teachers scored lower on all CLASS domains. Overall, Danish preschools provide warm and emotionally supportive environments to children, but the level of support for language, literacy and cognitive development was consistently in the low levels.

Keywords: process quality, structural quality, teacher-child interactions, emergent literacy, Denmark

The Quality of the Language and Emergent Literacy Environments in Danish Preschools: Evidence from a Scandinavian Model

Research demonstrates that emergent literacy and language skills in preschool predict literacy skills in elementary school (Chaney, 1998; Lonigan, Burgess, & Anthony, 2000; National Early Literacy Panel, 2008; Roth, Speece, & Cooper, 2002; Storch & Whitehurst, 2002). However, despite general awareness of this relationship, very little research has explored the extent to which preschools in Denmark, which is a country with universal preschool access, support the development of these early skills. In light of other research that has found that attending preschool – although it can predict a whole range of developmental outcomes – does not necessarily benefit child outcomes more than a full-time rearing by a mother (NICHD ECCRN, 2006), it is important to investigate the extent to which preschools provide children with learning environments that add value to their lives. In the current study, we aimed to address this gap in the literature by evaluating Danish preschools in terms of process and structural quality indicators associated with supporting children's language and emergent literacy skills.

Preschool Attendance as Intervention for Disadvantaged Children

There is some evidence that attendance in high quality preschool programs may benefit the school start of children growing up in poverty. For example, Burchinal et al. (2008) found that the quality with which teachers supported children's language, pre-academic, and social skills in preschool predicted child outcomes in kindergarten. Other research has found that such benefits are even detectable in the early grades of elementary school (Peisner-Feinberg et al., 2001). Multi-year research has also found that a range of life outcomes of highly impoverished children can be improved if the children attend high quality child care (Campbell, Pungello, Miller-Johnson, Burchinal, & Ramey, 2001). The age at

which disadvantaged children enter formal child care also appear to moderate the effects of attendance. Yazejian, Bryant, Freel, and Burchinal (2015) found a positive relation between the receptive language skills of disadvantaged Latino children growing up in the United States, and how young they were when they began attending child care (from birth to 5 years). Furthermore, children in the study maintained their Spanish proficiency.

However, it is also important to interpret the relative benefit of high quality preschool with caution. A recent meta-analysis has reaffirmed that quality is associated with pre-academic skills such as language and literacy, but that effects are mainly small (Keys et al., 2013). Furthermore, there is evidence that attending low quality preschool may perpetuate language learning gaps of preschoolers growing up in poverty (Logan, Piasta, Justice, Schatschneider, & Petrill, 2011). Unfortunately, research also suggests that disadvantaged children are also more likely to attend low quality preschool (Justice, Mashburn, Hamre, & Pianta, 2008; LoCasale-Crouch et al., 2007; NICHD ECCRN, 2006; Pianta et al., 2005). Altogether, the current literature indicates that preschools can be a tool in efforts against poverty's effects on children's language and literacy development, but that preschools can have a sustaining effect on learning gaps if quality is too low.

Defining High Quality Preschool

Preschool quality has been conceptualized in a number of different ways, but in the broadest sense it can be described as a multidimensional construct in which pertinent indicators are those aspects of preschool environments that have a positive association with children's outcomes (Pianta et al., 2005). In general, researchers discern two main domains of preschool quality: *structural* and *process quality* (Dowsett, Huston, Imes, & Gennetian, 2008; Pianta et al., 2005; Scarr, Eisenberg, & Deater-Deckard, 1994). Structural quality deals either with regulatory factors such as teacher-child ratios, teacher training, and teacher salaries (Phillipsen, Burchinal, Howes, & Cryer, 1997), or with the provision of materials and

equipment in the preschool (e.g., Koh & Neuman, 2009; Neuman & Cunningham, 2009). In contrast, process quality describes the direct experiences of children in terms of interactions with teacher, peers and materials (Pianta et al., 2005).

In much research, regulatory aspects of structural quality have been regarded as distal indicators of process quality in that they facilitate the ability of teachers to engage in high-quality proximal processes. For example, research has found that lower teacher-child ratios are related to more sensitive interactions with children (Russell, 1990), and fewer problem behaviors in children (NICHD, 2001). Likewise, Phillipsen et al. (1997) found that teacher education predicted the basic aspects of global quality, such as the provision of developmentally appropriate furnishing and play areas as measured by the Early Childhood Environment Rating Scale (ECERS; Harms & Clifford, 1980).

However, other research indicates that some indicators of regulatory structural quality may not be predictive of teacher-child interactions. Pianta et al. (2005) found that teacher-child ratios did not predict process quality as measured by the Classroom Assessment Scoring System (CLASS; Pianta et al., 2008), which measures the quality of teacher-child interactions. Similarly, Howes et al. (2008) found that teacher-child ratios did not predict children's pre-academic skills. In other research looking more closely at emergent literacy, Justice et al. (2008) found that teachers' with advanced degrees provided lower quality language instruction than less educated teachers, while teacher education and work experience did not predict literacy instruction at all. Together, these research findings indicate that although structural indicators may predict general aspects of preschool quality, they may not be the best predictors of the classroom processes that facilitate children's learning of language and literacy skills.

Rather than using regulatory structural indicators as a proxy of teacher-child interaction quality, researchers have increasingly turned to observing teacher-child

interactions directly as a means of predicting children's outcomes. In doing so, two subtypes of teacher-child interactions – the quality of instruction, and socio-emotional support – have emerged as being predictive of children's outcomes. For example, Mashburn et al. (2008) found that teacher's socio-emotional interactions, such as teacher warmth and sensitivity, predicted children's social competence, while instructional quality, such as the use of non-contextual questioning, predicted language and pre-academic skills. Other studies have confirmed these findings, but have additionally found that socio-emotional interactions also supported language and pre-academic skills (Burchinal et al., 2008; Howes et al., 2008). This growing base of research suggests that of the many aspects of process quality, instructional and socio-emotional teacher-child interactions may be important factors that foster children's language and literacy development.

The Role of Structural Supports in the Classroom

Although the evidence is mixed regarding the extent to which regulatory indicators of structural quality predict children's language and literacy outcomes, some research indicates that structural quality supports for literacy (i.e. the availability of literacy toys and materials) may have an effect on children's literacy behaviours and development, especially when teachers participate actively in interactions (Christie & Enz, 1992; Neuman & Roskos, 1993). For example, Guo, Justice, Kaderavek, and McGinty (2012) found that although the provision of literacy materials on its own did not predict children's growth in alphabet knowledge and name writing, a significant interaction with instructional support was found indicating that process and structural features complement each other when quality is high.

Interestingly, even the availability of reading and writing materials in dramatic play areas without direct adult involvement can significantly increase literacy behaviours in children over a sustained period of time (Morrow, 1990). Books in particular are a literacy material that can benefit children. For example, Neuman (1999) "flooded" preschool

classrooms with books, and instructed teachers on how to use them in interactions with children. Results of the study indicated that children spent more time reading and playing with books than a control group, and demonstrated significant gains on several measures of emergent literacy. Besides the quantity of books, the quality of books also appears to be important. For example, research shows that Children who are read to with alphabet books make greater gains in alphabet knowledge than children who only read story books (Murray, Stahl, & Ivey, 1996).

These studies indicate that the provision of literacy materials can increase the number and quality of children's early literacy experiences in preschool classrooms especially when combined with intentional teacher involvement. Furthermore, children at risk for academic failure may benefit from increased access to quality literacy materials, since research suggests that children from families with low socioeconomic status (SES) have fewer literacy interactions than middle-class children (Neuman & Celano, 2001).

Preschool in Denmark

In Denmark, 95% of four-year-olds attend state-sponsored preschools (OECD, 2013). When universal childcare was established in the 1960's, the main goal was to increase Danish mothers' access to the job market at a time when unemployment was historically low. During this period, a massive childcare expansion project essentially doubled the number of preschool and nursery places within a short number of years (Borchorst, 2000). However, the role of preschool in Denmark has not remained stagnant. In the decades that have ensued, preschools have increasingly taken on the role of preparing children for life in Danish society, and a unique Danish pedagogical tradition has emerged. Whereas preschools in the United States provide children with early education, in Denmark, preschool teachers are trained to take a holistic approach to pedagogy, which is reflected in the 3.5 year pedagogical bachelor degree required of most preschool teachers. In the Danish model, encouraging

democratic values and social skills are cornerstones of practice, whereas early academic goal-setting is generally treated warily (Jensen, 2009). Nevertheless, in 2004 the Danish government enacted a new law requiring all preschools to create *learning plans*, which would describe how each preschool intended to support children's development in six general areas including language development (Ministry of Children, 2014) . Beyond this, however, there are no specific learning goals mandated by government in Danish preschools.

Academic success in school has become a major theme in current educational debates in Denmark in light of recent PISA studies revealing that Danish pupils score in the middle range of Western nations, and that approximately 40% of youths with immigrant status lack functional reading skills at the end of primary schooling (Egelund, Nielsen, & Rangvid, 2011). These findings – and in particular the findings for youths with immigrant background – have raised questions regarding the extent to which Danish preschools prepare children for succeeding in school. Although there is currently little evidence of the quality of language and literacy environments in Danish preschools, a language norming study by Bleses, Højen, Jørgensen, Jensen, and Vach (2010) found that preschool-aged children with immigrant background and/or social disadvantage scored significantly lower on a battery of language tests compared to other children. Taking into consideration that the majority of youths who lack functional reading skills at the end of primary school attended preschool (due to universal preschool), it is possible that Danish preschools could do more to prepare at-risk children for school entry.

The Contribution of the Current Study

The main goal of the current study was to investigate the structural quality of Danish preschools in terms of children's access to literacy materials, and process quality in terms of teacher-child interactions. Danish preschools are often assumed to be of higher quality than American ones, ostensibly due to regulatory structural factors such as universal enrollment,

extensive expenditure of public funds, and low teacher-child ratios (e.g., Esping-Andersen et al., 2012). For example, Denmark spends 1.4% of its gross national product (GDP) on childcare programs, which is the second highest in the world after Iceland, and considerably higher than other Western nations without universal preschool access such as the United States, which spends only 0.4% of its GDP (OECD, 2009). Furthermore, the teacher-child ratio for Danish preschools is relatively low at one teacher to 6.7 children, but it should be noted that the ratio was slightly higher than previous studies (Dalsgaard, Nøhr, & Tenney Jordan, 2014).

A small number of recent studies have raised questions regarding the overall quality of care children receive in preschool. Kragh-Muller and Ringsmose (2015) compared the environments of newer large institutions versus traditionally smaller preschools and concluded that teachers working in small preschools had better opportunities for supporting language development than teachers in larger preschools. Another recent study found that children in a small number of Danish nurseries (i.e. 0- to 3-year-olds) received low amounts of contact and stimulation from teachers (Hansen, 2013), but it is unknown if these results are generalizable to preschools. Finally, Bauchmüller, Gørtz, and Rasmussen (2011) investigated the longitudinal effects of several preschool regulatory structural quality factors on Danish test scores at the end of the 9th grade. They found that teacher-child ratios, the percentage of staff with a pedagogical degree, and percentage of teachers with a non-Danish background positively predicted test scores. However, the effects, although statistically significant, were rather small, and may not have been educationally meaningful.

Taking into consideration the evidence that preschool might only be beneficial for children's language and literacy skills when quality is high (e.g., Logan et al., 2011; NICHD 2006), it is problematic that we know so little about the extent to which Danish preschools provide children with high quality language and literacy experiences. In particular, the degree

to which Danish teachers support domains of process quality linked to literacy development (i.e. socio-emotional and cognitive/language development) is currently unknown, as is the extent to which children have access to structural literacy supports.

Research Questions

In the current study, we aimed to fill this gap in the literature by conducting a large scale investigation of process and structural quality within Danish preschools. The investigation had three main goals. Our first goal was to investigate the structural and process quality of Danish preschools in terms of literacy materials and teacher-child interactions, and furthermore to contextualize results by comparing process quality in Denmark to an American sample.

Our second goal was to investigate factors that predicted and explained variation in quality outcomes. As research in the United States demonstrates that children at risk for academic failure due to poverty are more likely to attend lower quality preschools (e.g., Justice et al., 2008; LoCasale-Crouch et al., 2007), we investigated whether there was evidence of a similar pattern in Denmark. In particular, we investigated the extent to which preschool SES factors predicted structural and process quality at the institution level. Furthermore, we investigated whether teacher background variables such as age, work experience, education, and participation in professional development predicted teacher-child interactions.

As a final objective, we investigated how quality varied as a function of daily routines. Research by Gest, Holland-Coviello, Welsh, Eicher-Catt, and Gill (2006) found that the type of language support teachers provided children depended on what activity the children were partaking in. Therefore, we investigated whether teacher-child interactions changed dependent on whether children were engaging in mealtimes, shared-book reading, or language stimulation activities.

Method

Participants

Two-hundred and ninety-three preschools participated in the study of structural quality supports for language and literacy. Preschools were participating in one of two larger randomized controlled trials: Structured Preschool Effort for Language and Literacy (SPELL; Bleses et al., 2014) or Language and Literacy Educational Activities for Preschoolers (LLEAP). Both studies were effectiveness trials testing the effects of bringing language and emergent literacy interventions to scale. Between the two projects, a total of 13 (out of 98) municipalities were recruited for participation, and from these, preschools were randomly selected for participation in the effectiveness studies, and thereby the current study of structural and process quality.

Process quality was evaluated using teacher-submitted video recordings from the SPELL study. In all, 506 teachers participated in the observation study out of an original sample of 638 eligible teachers. Half of the attrition was due to entire preschools dropping out of the project in the early stages of the study (66 teachers in all), which is a more common occurrence in effect studies versus efficacy studies (Gartlehner, Hansen, Nissman, Lohr, & Carey, 2006). There was an even number of dropouts in all experimental groups. It is unknown why the 66 remaining missing teachers failed to send in their videos. Background characteristics for participating teachers are given in Table 1.

Procedure and Measures

Structural quality was observed live by trained research staff during pre-scheduled visits. Observations took approximately 25 minutes to complete for each preschool. Process quality was evaluated from video. Participating teachers were instructed to film three daily routines (a mealtime, a shared-book reading, and a language activity of the teacher's choice),

and submit them to the project team. To increase the comparability of video recordings, we instructed teachers to include approximately five children in each video.

Table 1

Teacher Characteristics

	n	%
Age of teachers		
Under 25	2	0.4
25-35	101	20.0
36-45	136	26.9
46-55	123	24.3
Over 55	63	12.5
Missing	81	16.0
Years of work experience		
under 5	76	15.0
6-10	109	21.5
11-15	78	15.4
16-20	45	8.9
over 20	122	24.1
Missing	76	15.0
Education		
ECE degree	375	74.1
No ECE degree	102	20.2
Missing	29	5.7
Gender		
Femal	390	77.1
Male	35	6.9
Missing	81	16.0
Language PD		
4 day course	129	25.5
No 4 day course	301	59.5
Missing	76	15.0
6 day course	9	1.8
No 6 day course	421	83.2
Missing	76	15.0

Structural quality. To measure the accessibility of literacy materials in preschools, we used a Danish adaptation of the Classroom Literacy Observation Profile (CLOP; The Crane Center for Early Childhood Research and Policy, 2012). Psychometric attributes of the

CLOP are described in Dynia (2013). The CLOP is a checklist of physical literacy objects, such as books and displays that are accessible and/or visible to children. Our adaptation of the CLOP differed in two regards. First of all, our level of measurement was the preschool and not the classroom. Danish preschools normally consist of *stuer* ("rooms"), which are similar to the American concept of preschool classrooms, but teachers and children do not necessarily spend the whole day in one classroom. Because some preschools are also open-concept, and do not use classrooms at all, we evaluated the availability of literacy materials at the institution level. Secondly, we added items that we considered pertinent to the Danish context such as the availability of home-loan children's libraries, and consideration for Danish-as-a-second-language learners (such as books in other languages than Danish). Items in our adaption were collected into seven categories. *Books* is a checklist of the number and types of books accessible in the preschool. *Technology* counts the number of iPads and computers with language and literacy programs. *Areas* is a checklist of classroom areas that support reading or writing, such as a reading corner with accessible books. *Materials* counts the number of types of literacy implements available to children such as letter blocks, crayons and pencils. *Visibility of Print* is a checklist of literacy signs, posters, and other displays visible to children such as displays of the alphabet. *Parents* is a checklist of parent-directed literacy supports such as the availability of home-loan children books, and pamphlets about children's acquisition of language and literacy. *Diversity* counts the presence of objects that supported literacy in foreign languages. These could include displays of foreign writing systems or word labels in foreign languages. CLOP checklists were completed by the first author and two other researchers trained by the first author. Regular meetings were held to discuss data collection procedures, and maintain reliability.

Process quality. We used the Classroom Assessment Scoring System (CLASS; Pianta et al., 2008) to evaluate process quality from teachers' videos. The CLASS measures

ten dimensions of teacher-child interactions which are given in Table 2. A factor analysis of CLASS's items determined that a three-factor solution was the most suitable (Pianta et al., 2008). As such, CLASS dimensions are grouped into the domains of *emotional support* (positive climate, negative climate, teacher sensitivity, and regard for student perspectives), *classroom organization* (behaviour management, productivity, and instructional learning formats) and *instructional support* (concept development, quality of feedback, and language modeling). CLASS dimensions are scored on a Likert scale from 1-7, in which scores of 1-2 indicate low quality, 3-5 indicate medium quality, and 6-7 indicate high quality interactions.

Table 2

The CLASS dimensions

Emotional Support	
Positive Climate	Measures positivity in the classroom's environment. Evidence can include warm interactions, respectful language, smiling and shared activities.
Negative Climate	Measures negativity in the classroom's environment. Evidence can include irritability or anger, the use of threats, harsh punishments, teasing, and bullying.
Teacher Sensitivity	Measures the extent to which the teacher is in tune with the children's needs. Evidence can include teacher awareness of problems, comforting behaviour, resolution of problems, and student comfort.
Regard for Student Perspective	Measures the extent teachers demonstrate openness of children's interests and points of view rather than more rigid teacher. Evidence can include showing flexibility in lesson plans, supporting autonomy, and allowing wiggling.
Classroom Organization	
Behaviour Management	Measures the extent to which the teacher provides clear behavioural expectations and uses effective methods for preventing and addressing misbehaviour. Evidence can include teacher monitoring and consistency in rules.
Productivity	Measures the teacher's balance of routine and instructional time. Evidence can include little wandering and brief transitions between activities.
Instructional Learning Formats	Measures how the teacher maximizes and maintains children's interest and engagement. Evidence can include teacher involvement in activities, children's active participation, and clarity of learning objectives.

Instructional Support	
Concept Development	Measures the degree to which the teacher promotes higher-order thinking via discussion and instruction. Evidence can include asking children to analyze and reason, brainstorming, and the coupling of learning targets with real-world applications
Quality of Feedback	Measures the quality of teachers' feedback to children's utterances with regards to expanding understanding and learning. Evidence can include the use of scaffolding, providing information, and encouraging children's persistence.
Language Modeling	Measures the extent to which the teacher uses supportive language strategies. Evidence can include conversations, open-ended questions, and the use and explanation of advanced language.

Certified CLASS trainers trained coders, and coders had to pass a certification test with a minimum 80% reliability score before they could work on the project. Furthermore, we made extensive use of additional videos in the Danish language to ensure that all coders reliably applied CLASS in the Danish context prior to coding CLASS for the project. Ten percent of project videos were randomly selected and double coded to determined inter-rater reliability. Reliability was 91.2%. Although CLASS is a measure developed in the United States, there is precedent for using it in European countries such as Portugal (Cadima, Leal, & Burchinal, 2010), and the Netherlands (Slot, 2014). Furthermore, the CLASS has been validated and used in Finland (Pakarinen et al., 2010), which has a preschool system similar to that of Denmark.

Children's socioeconomic status. Five indicators of children's SES were obtained from Statistics Denmark using the Danish Central Person Registry. These indicators were combined at the institutional level to create a SES indicators for each preschool. The SES indicators for each institution were the percentage of children living with both parents, the percentage of families with an income below the 10th percentile, percentage of families receiving welfare payments, percentage of families where one parent had a higher education, and percentage of families with immigrant status.

Teacher characteristics. Teachers filled out an online questionnaire with demographic and professional questions. Included items were age, gender, education, years of experience, and participation in professional development.

Analytic Approach

Our data analyses were conducted either at the preschool or teacher levels, as SES data and CLOP scores were measured by institution, whereas CLASS scores and teacher background factors were measured at the teacher level. To evaluate the effects of children's SES on CLASS scores, we calculated the average CLASS score for each institution. The first goal of this study was to describe the structural and process quality of Danish preschools, and to compare CLASS scores from the Danish sample with an American sample. This was conducted using descriptive statistics, and independent *t*-tests compared the Danish sample to the American one. Our second goal of this study was to investigate factors that predicted and accounted for variation in quality outcomes. To do so, a series of multiple regressions were conducted, first to examine the role of SES variables on CLOP and mean CLASS scores at the institution level, and then to investigate the role of teacher background variables at the teacher level. Our final goal was to investigate whether CLASS scores changed as a function of instructional setting. This was investigated by conducting a multiple regression for each CLASS domain with instructional setting as the predictor variable.

Results

The Structural and Process Quality of Danish Preschools

The first goal of this study was to describe the structural and process quality of Danish preschools. Furthermore, we contextualized CLASS scores by comparing them to an American sample.

Structural quality. Descriptive statistics of the CLOP scores are given in Table 3. In general, most preschools scored low on all subscales indicating low availability of literacy

materials. In terms of books, we found that most institutions had a large assortment of books available at child level (however, 4.4% of preschools did not provide books at all), but the books were mostly narrative in nature. 78.5% of preschools did not provide children with books about the alphabet, and 59% did not provide books about numeracy and shapes. Despite the apparent lack of variation in the books provided at child level, we did find that 80.2% of preschools provided children with a reading corners indicating that storybook reading was a common occurrence.

The use of technology to support interactions with literacy was not widespread. For example, only 84 of the 293 preschools had at least on tablet computer at its disposal. Of these preschools, only 43 (14.7% of all preschool) had a large assortment of language and literacy apps.

Table 3

CLOP sub-scores

	n	M	SD	Range	
Books	293	9.17	2.42	0	16
Technology	293	0.97	1.50	0	6
Areas	293	1.20	0.67	0	3
Materials	292	1.17	0.94	0	5
Print visibility	292	5.07	2.11	1	12
Parents	291	1.05	1.81	0	7
Diveristy	293	0.18	0.63	0	4
Total	291	18.83	4.64	5	32

The availability of literacy materials was mixed. Most preschools (85.7%) provided children with things they could write or draw with, but only some preschools provided children with paper they could write or draw on at child level (37.5%). The availability of letter blocks was also low (37.2%), and 91.4% of preschools did not provide children with literacy props or materials in the dramatic play areas.

Visibility of print in preschools was mixed. Alphabet displays were observed in 85.7% of preschools, and 22.5% had several visible displays indicating a focus on learning the alphabet. However, there were few displays of children's own writing (20%) suggesting that writing activities were uncommon.

In terms of family supports, there were sporadic examples of preschools interacting with parents regarding language and literacy. Some preschools (15.1%) had a parent information board where daily language activities such as story-book reading were described. In 29.6% of preschools, information was available on the importance of language development, and 12.7% of preschools had home-loan libraries with books parents could borrow to read to their children.

There were few examples of literacy materials that acknowledged linguistic diversity. Only 6.5% of preschools had displays of words in other languages, and 7.0% of preschools had visible displays of non-Roman-alphabet writing systems. Of the preschools that had home-loan libraries, only 5.4% had books in foreign languages. However, two preschools had high quantities of foreign language books to loan indicating a concerted effort by those particular preschools to encourage parents with immigrant status to read to their children.

Process quality. Descriptive statistics are given for CLASS scores in Table 4 together with CLASS scores from the MyTeacherPartner (MTP) study in the United States (MTP results taken from Pianta et al., 2008), and results from independent t-tests. The MTP scores are not intended to be nationally representative of American preschools. They are presented here only to give an indication of how Danish preschools compared to a sample of American preschools.

Table 4

CLASS Results of the Danish Sample Compared Against an American Sample

	Denmark *N*=506		MTP (USA) *N*=164			
	M	*SD*	*M*	*SD*	*t*	*p*
Emotional Support	**5.82**	**0.45**				
Positive Climate	6.21	0.56	5.21	0.9	40.11	.001
Negative Climate*	1.08	0.20	1.63	0.69	61.46	.001
Teacher Sensitivity	5.72	0.67	4.34	0.94	46.69	.001
Regard for Student Perspectives	4.44	0.83	4.36	0.97	2.30	.022
Classroom Organization	**5.67**	**0.48**				
Behavior Management	6.39	0.51	4.94	0.88	64.26	.001
Productivity	6.23	0.55	5.41	0.82	33.52	.001
Instructional Learning Formats	4.38	0.81	4.57	0.78	-5.24	.001
Instructional Support	**2.41**	**0.57**				
Concept Development	1.77	0.60	2.69	0.68	-34.18	.001
Quality of Feedback	2.38	0.68	2.87	0.85	-16.20	.001
Language Modeling	3.09	0.69	2.85	0.73	7.84	.001

Note: MTP = MyTeachingPartner (see technical appendix; Pianta, 2008); M = Mean; SD = Standard deviation; *Negative Climate is scored negatively and therefore inverted when calculating the mean for Emotional Support

Results for emotional support revealed that Danish preschool teachers provided exceptionally high socio-emotional support with moderately high teacher sensitivity. However, they demonstrated only moderate regard for children's perspectives indicating that there is a noticeably degree of teacher-directedness in their practices. In terms of classroom organization, Danish teachers provided excellent behavioural management, and provided activities to the children. The score for instructional learning formats was moderate, and indicates that teachers only to some degree actively facilitated children's learning during interactions. Finally, the results for instructional support were very low to moderately low. Teachers did not promote children's higher order thinking skills beyond a low, rote level. The quality of feedback teachers gave to children was also low. There were low levels of scaffolding, prompting of thought processes, or extended feedback loops. Language modeling

was in the low end of medium quality. This suggests some usage of strategies to promote children's acquisition and usage of language, but in mostly unintentional occurrences.

Independent t-tests revealed that the Danish scores were significantly different from the MTP scores on all dimensions. Notably, the Danish scores were significantly higher on dimensions that were more socio-emotionally oriented, while instructional and learning oriented dimensions were significantly lower than the American sample (with the exception of language modeling).

Moderators of CLOP and CLASS Scores

The second goal of this study was to investigate the significance of preschool SES variables on structural and process quality, and teacher background characteristics on process quality. We calculated CLASS average scores for each preschool to control for SES at the institution level.

Effects of SES at the Institution Level. Multiple regression analyses revealed that SES was a significant predictor for three CLOP sub-scores (Table 5), but with varying directions in the results. SES was a negative predictor of the provision of books, but in all it only explained 5% of variation. Low SES was, however, a positive predictor of the Parents and Diversity subscales. As such, parents with low SES backgrounds had better access to support materials (accounting for 13% of variation). This pattern was less pronounced but still significant for the diversity subscale with SES variables explaining 5% of variation in favour of low SES children.

In terms of CLASS scores, multiple regressions revealed that SES accounted for 11% of variation in emotional support, 19% in classroom organization, and just 8% in instructional support (Table 6). This indicates that social background was associated with quality, but most so for classroom organization.

Table 5

Multiple regressions of the effects of SES on CLOP subscores

	Books					Parents					Diversity				
Variable	B	SE	ß	*T*	*p*	B	SE	ß	*t*	*p*	B	SE	ß	*t*	*p*
constant	7.08	3.00		2.36	.019	5.51	2.31		2.38	.018	0.52	0.81		0.64	.524
SES 1	4.34	2.47	0.14	1.76	.080	1.96	1.90	0.08	1.03	.304	0.35	0.67	0.04	0.52	.602
SES 2	2.08	2.59	0.12	0.80	.424	4.19	2.00	0.31	2.09	.037	0.50	0.70	0.11	0.71	.479
SES 3	-10.13	3.73	-0.28	-2.72	.007	0.83	2.87	0.03	0.29	.772	-1.77	1.01	-0.18	-1.76	.080
SES 4	-1.90	2.63	-0.09	-0.72	.472	-6.52	2.03	-0.39	-3.22	.001	-0.79	0.71	-0.14	-1.12	.265
SES 5	-0.74	1.42	-0.06	-0.52	.602	-5.14	1.09	-0.57	-4.70	.000	0.42	0.38	0.14	1.11	.268
R^2			.06					.13					.05		

Note: SES 1: percentage of children living with both parents; SES 2: percentage of families with an income below the 10th percentile; SES 3: percentage of families receiving welfare payments; SES 4: percentage of families where one parent has a higher education; SES 5: percentage of families with immigrant status.

Table 6

Multiple regressions of the effects of SES on CLASS domains

	Emotional Support					Classroom Organization					Instructional Support				
Variable	B	SE	ß	*t*	*p*	B	SE	ß	*t*	*p*	B	SE	ß	*t*	*p*
constant	6.02	0.55		10.86	.001	6.47	0.58		11.19	.001	3.12	0.77		4.07	.001
SES 1	0.58	0.46	0.14	1.26	.212	0.10	0.48	0.02	0.21	.832	0.92	0.63	0.16	1.45	.149
SES 2	-0.48	0.47	-0.26	-1.03	.305	-1.49	0.49	-0.73	-3.04	.003	-1.34	0.65	-0.53	-2.07	.041
SES 3	0.99	0.56	0.28	1.76	.082	1.36	0.59	0.35	2.30	.023	1.32	0.78	0.28	1.69	.094
SES 4	-0.66	0.51	-0.28	-1.30	.195	-0.78	0.53	-0.30	-1.47	.144	-1.52	0.70	-0.48	-2.17	.032
SES 5	-0.57	0.29	-0.43	-1.97	.052	-0.31	0.30	-0.21	-1.02	.311	-0.31	0.40	-0.17	-0.77	.441
R^2			.11					.19					.08		

Note: SES 1: percentage of children living with both parents; SES 2: percentage of families with an income below the 10th percentile; SES 3: percentage of families receiving welfare payments; SES 4: percentage of families where one parent has a higher education; SES 5: percentage of families with immigrant status.

Effects of teacher background factors on process quality. Multiple regressions of the effects of teacher background variables on CLASS scores by domain are given in Table 7. Teacher age was not a significant predictor of any CLASS domains, but we found that teachers with fewer than five years of work experience provided significantly more instructional support than teachers with more work experience. Level of education was not a significant predictor of CLASS scores. Male teachers scored significantly lower on all CLASS domains. Finally, we found that teachers who had participated in *Sprogpakken* [The Language Package], a four-day professional development program focusing on language and emergent literacy development, scored higher on all CLASS domains, reaching statistical significance for classroom organization and instructional support. However, the six-day version of the same program did not appear to have an effect, although the sample size was considerably smaller.

The Quality of Interactions in Routine Situations

Our third goal was to investigate variation in CLASS scores as a function of the preschool instructional setting´. These included mealtimes, shared-book reading sessions, and a language activity of the teacher's choice. Results of the multiple regression analyses are given in Table 8. Setting mealtimes as the constant, we found that emotional support significantly decreased during shared-book readings and language activities. In contrast, teachers provided more classroom organization when conducting book-readings and language activities. Finally, we found that teachers provided equal amounts of instructional support during mealtimes and shared-book readings, and slightly more during language activities (although the level was still very low).

Table 7

Multiple regression analysis of teacher background variables on CLASS domain scores

	Emotional Support					Classroom Organization					Instructional Support				
	B	SE	ß	*t*	*p*	B	SE	ß	*t*	*p*	B	SE	ß	*t*	*p*
Constant	5.74	0.06		99.70	.001	5.63	0.06		89.10	.001	2.30	0.07		30.98	.001
AGE															
Under 25	-0.56	0.33	-0.09	-1.71	.089	0.28	0.36	0.04	0.77	.439	-0.68	0.42	-0.08	-1.62	.105
25-35	0.05	0.10	0.05	0.47	.640	0.13	0.11	0.12	1.20	.231	0.02	0.13	0.02	0.15	.878
36-45	0.14	0.09	0.15	1.60	.111	0.14	0.09	0.14	1.50	.135	0.08	0.11	0.07	0.75	.455
46-55	0.09	0.07	0.10	1.33	.183	0.07	0.08	0.07	0.93	.352	0.10	0.09	0.08	1.09	.275
Over 55 yrs (comparison variable)															
WORK EXPERIENCE															
Under 5 yrs	0.10	0.09	0.09	1.08	.281	-0.06	0.10	-0.05	-0.61	.542	0.25	0.12	0.17	2.09	.037
6-10 yrs	0.00	0.08	0.00	-0.06	.956	-0.10	0.09	-0.09	-1.15	.250	0.03	0.10	0.02	0.26	.795
11-15 yrs	-0.03	0.08	-0.03	-0.37	.712	-0.08	0.08	-0.07	-0.95	.341	-0.04	0.10	-0.03	-0.39	.693
16-20 yrs	0.12	0.08	0.09	1.54	.124	0.08	0.09	0.05	0.86	.388	0.08	0.10	0.05	0.81	.417
over 20 yrs (comparison variable)															
EDUCATION															
Pedagogical degree	-0.04	0.07	-0.03	-0.59	.553	0.03	0.07	0.02	0.38	.701	-0.05	0.09	-0.03	-0.59	.552
No higher education	-0.10	0.11	-0.04	-0.86	.388	-0.09	0.12	-0.04	-0.74	.457	-0.06	0.14	-0.02	-0.44	.662
GENDER															
Male	-0.30	0.08	-0.19	-3.84	.001	-0.32	0.09	-0.19	-3.80	.001	-0.29	0.10	-0.14	-2.84	.005
LANGUAGE PD															
4 day course	0.08	0.05	0.09	1.77	.077	0.14	0.05	0.13	2.73	.007	0.17	0.06	0.14	2.92	.004
6 day course	0.07	0.07	0.05	0.96	.337	0.01	0.08	0.01	0.13	.896	0.02	0.09	0.01	0.24	.809
Note: R^2			0.10					0.07					0.08		

Table 8

Multiple regressions of the effect of preschool routine situation on CLASS domain scores

	Emotional Support					Classroom Organization					Instructional Support				
	B	SE	ß	*t*	*p*	B	SE	ß	*t*	*p*	B	SE	ß	*t*	*p*
Constant (eating situation)	5.99	0.03		218.21	.001	5.43	0.03		179.12	.001	2.39	0.04		67.68	.001
Reading situation	-0.32	0.04	-0.25	-8.16	.001	0.33	0.04	0.23	7.65	.001	0.00	0.05	0.00	0.01	.995
Language Activity	-0.15	0.04	-0.11	-3.71	.001	0.42	0.04	0.28	9.57	.001	0.13	0.05	0.08	2.64	.008
Note: R^2			.05					.07					.01		

Discussion

The current study is unique in that it is the first large-scale study of the quality of Danish preschools, operationalized in terms of structural supports for literacy, and teacher-child processes. The results can be used to describe the aspects of Danish pedagogy that appear to contribute to higher quality language and literacy environments, and to pinpoint other areas where quality can be improved.

First, our investigation of structural quality revealed a lack of access to literacy supports in most institutions. Most striking was perhaps the lack of variety in books available to children. Although most institutions provided children with free access to narrative story books, the prevalence of alphabet books and concepts books was very low. Furthermore, most preschools did not provide story-books in foreign languages to families with immigrant status with the exception of a handful of preschools that in fact provided large numbers of books, indicating that some preschools did have a special interest in supporting families with immigrant background.

In terms of process quality, our investigation revealed evidence of both high and low quality. We found that socio-emotional support and behavior management was high, which was not surprisingly given the focus on social and behavioral skills in Danish pedagogy. We also found evidence of high productivity, but these scores may have overestimated the true productivity, since teachers video-recorded activities themselves, and possibly did not record time spent doing managerial tasks. Scores for the instructional support domain, however, were mostly low, suggesting that Danish preschool teachers may lack interactional skills and strategies for stimulating language and cognitive growth in children, which mirrors a similar finding from the United States (Burchinal et al., 2008). Taking into consideration that teacher-child ratios in the teacher-submitted videos (i.e. one teacher to five children) were more favorable than the ratios teachers normally work with, it is possible that the low

instructional support scores even overestimated the true values. Although it is positive that socio-emotional support was so high – since this domain also predicts language and literacy outcomes (Burchinal et al., 2008; Howes et al., 2008) – it is concerning that the scores for the concept development, quality of feedback and language modelling dimensions were so low, since research indicates that these preschool interactions support children's learning of language and literacy most (Mashburn et al., 2008; Pianta et al., 2005).

We also found evidence of the influence of SES on quality outcomes. For example, we found that structural supports for families and diversity were higher in preschools serving larger proportions of children with risk status, and SES explained variation in all CLASS domains indicating that disadvantaged children were more likely to attend preschools with lower process quality. However, this relation was most evident for the classroom organization domain, where SES accounted for nearly 20% of variation. Only 8% of variation in instructional support was explained by SES. This indicates that it may not be as important a factor, possibly because quality was so consistently low in this domain. These results reflect to some extent American research which also found that children from low SES families were more likely to attend lower quality preschools, but in all CLASS domains (Justice et al., 2008; LoCasale-Crouch et al., 2007).

Our examination of the contribution of teacher characteristics to CLASS outcomes revealed that teacher age and level of education had no effect on the quality of teacher-child interactions. The null effect of teacher education on quality is consistent with other large-scale findings (Early et al., 2006; Early et al., 2007). However, significant effects were found for work experience, participation in a language PD course, and gender. More specifically, teachers with the least amount of experience provided better instructional support than their more experienced colleagues. This finding could be due to changes in the preschool teacher education that have increased the focus on cognitive and language development, such that

new teachers may have more evidence based knowledge and skills. We also found that teachers who attended a recent nationwide four-day course on language and literacy development scored higher on all CLASS domains, with the largest effect occurring in the instructional support domain. This finding seems to support similar finding by Justice et al. (2008), who found that PD had a small, but significant positive effect on language instruction. Finally, we found a clear gender effect favouring female teachers on all CLASS scores. On average, men scored nearly a third of a CLASS point lower on all three domains. This finding is at odds with NICHD ECCRN (2006), which found that the proportion of male staff members in a preschool was positively associated with child outcomes. It should be noted, however, the number of males in our study was relatively small.

The final goal of our study examined whether CLASS scores changed as a function of preschool routine. Here we found that emotional support was significantly lower during shared-book reading and language activities, which may be due to teachers taking on a more directive role during these more structured activities. We also found that language activities had a slight advantage over mealtimes and shared-book readings in terms of instructional support, and this is likely explained by the teachers taking on a more didactic role in these situations. It is also noteworthy that the magnitude of the advantage was very small. In practical terms, the amount of language and cognitive stimulation children receive was equally low in all three scenarios.

Implications for Practice

This study contributes to the literature with a more nuanced view of Denmark's universal preschool system. The high socio-emotional support we found seems to confirm that the holistic approach of Danish pedagogy has succeeded in creating warm and caring environments where children's individual social skills can develop. In contrast, the lack of literacy materials and low instructional support indicate that Danish preschools do not

necessarily provide high quality language environments to children. This is a concerning finding taking into account that preschool children with immigrant status score significantly lower on a range of language outcomes (Bleses et al., 2010), and could potentially derive benefit from higher quality learning environments. More empirical research is needed, however, to investigate the extent to which preschool quality meditates children's language and literacy outcomes in Danish preschools, and whether there is an association with functional reading skills later in life.

The findings of the present study suggest that improving the quality of language and literacy environments needs to come into focus in Denmark. Despite the existence of universal preschool, the quality of cognitive and language support is low – for the most part lower than the sample of American preschools we compared it with. As such, pedagogical approaches opposed to supporting children's pre-academic skills in preschool (e.g., Klitmøller & Sommer, 2015) may not be well aligned with international research that finds relations between preschool quality and children's language and academic skills in school (e.g., Peisner-Feinberg et al., 2001). Together with recent PISA studies (Egelund et al., 2011), which indicate that vulnerable population groups in Denmark fail to gain functional reading skills despite attending state institutions from the age of three, these findings suggest that an improvement in the instructional quality in Danish preschools could potentially mediate a better school start for children, and thereby future academic success.

Why Danish teachers exhibit low use of language and cognitively stimulating techniques, even when engaging in a language stimulation activity of their choice with a favorable teacher-child ratio, is potentially due to the nature of their formal education. The acquisition of discrete interactional techniques that support language and literacy is not a large part of the Danish pedagogical education. Pathways to improving instructional practices may include professional development and reform of the current bachelor education. For

example, research by Hamre et al. (2012) found that a skill-based approach to professional development trained teachers to recognize and implement new practices. However, research is needed in Denmark to determine exactly how new training initiatives can improve the instructional skillsets of preschool teachers, so that children attending one of the most expensive preschool systems in the world receive maximum benefit.

References

Bauchmüller, R., Gørtz, M., & Rasmussen, A. W. (2011). Long-run benefits from universal high-quality pre-schooling. Copenhagen: AKF.

Bleses, D., Højen, A., Dybdal, L., Dale, P., Justice, L., Piasta, S., . . . Aktürk Ari, B. (2014). A randomized controlled trial of a large scale systematic and explicit language and literacy intervention in Danish daycares: The SPELL study. *British Journal of Education, Society & Behavioural science, 4*(11), 504-1524.

Bleses, D., Højen, A., Jørgensen, R., Jensen, K. Ø., & Vach, W. (2010). Sprogvurdering af 3-årige - karakteristika og risikofaktorer. [Language screening of 3-year-olds – Characteristics and risk factors]. *University of Southern Denmark working papers,* (10). http://www.humaniora.sdu.dk/boernesprog/DK/forsk/doc/E-print_10_2010.pdf

Borchorst, A. (2000). Den danske børnepasningsmodel - kontinuitet og forandring. [The Danish child care model – continuity and change. *Arbejderhistorie, 4*, 55-69.

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 Development Science, 12*(3), 140-153.

Cadima, J., Leal, T., & Burchinal, M. (2010). The quality of teacher–student interactions: Associations with first graders' academic and behavioral outcomes. *Journal of School Psychology, 48*(6), 457-482. doi: http://dx.doi.org/10.1016/j.jsp.2010.09.001

Campbell, F. A., Pungello, E. P., Miller-Johnson, S., Burchinal, M., & Ramey, C. T. (2001). The development of cognitive and academic abilities: Growth curves from an early childhood educational experiment. *Developmental psychology, 37*(2), 231-242. doi: 10.1037/0012-1649.37.2.231

Chaney, C. (1998). Preschool language and metalinguistic skills are links to reading success. *Applied Psycholinguistics, 19*(03), 433-446. doi: doi:10.1017/S0142716400010250

Christie, J. F., & Enz, B. (1992). The Effects of Literacy Play Interventions on Preschoolers' Play Patterns and Literacy Development. *Early Education and Development, 3*(3), 205-220. doi: 10.1207/s15566935eed0303_1

Crane Center for Early Childhood Research and Policy. (2012). *The Classroom Literacy Observation Profile*. Unpublished measure. Columbus, OH.

Dalsgaard, C., Nøhr, K., & Tenney Jordan, A. L. (2014). *Personale og børn i kommunernes dagtilbud – En undersøgelse af perioden 2007-2012*. [Personale and children in municipal preschools – An investigation from the period of 2007-2012]. Copenhagen: KORA.

Dowsett, C. J., Huston, A. C., Imes, A. E., & Gennetian, L. (2008). Structural and process features in three types of child care for children from high and low income families. *Early Childhood Research Quarterly, 23*(1), 69-93. doi: http://dx.doi.org/10.1016/j.ecresq.2007.06.003

Dynia, J. M. (2013). *The Literacy Environment of Early Childhood Special Education Classrooms: Predictors of Print Knowledge.* (Dissertation). The Ohio State University.

Early, D. M., Bryant, D. M., Pianta, R. C., Clifford, R. M., Burchinal, M. R., Ritchie, S., . . . Barbarin, O. (2006). Are teachers' education, major, and credentials related to classroom quality and children's academic gains in pre-kindergarten? *Early Childhood*

Research Quarterly, 21(2), 174-195. doi: http://dx.doi.org/10.1016/j.ecresq.2006.04.004

Early, D. M., Maxwell, K. L., Burchinal, M., Alva, S., Bender, R. H., Bryant, D., . . . Griffin, J. A. (2007). Teachers' education, classroom quality, and young children's academic skills: Results from seven studies of preschool programs. *Child Development, 78*(2), 558-580.

Egelund, N., Nielsen, C. P., & Rangvid, B. S. (2011*). PISA Etnisk 2009: Etniske og danske unges resultater i PISA 2009*. [PISA Ethnic 2009: Ethnic and Danish youths' results in PISA 2009]. Copenhagen: AKF.

Esping-Andersen, G., Garfinkel, I., Han, W.-J., Magnuson, K., Wagner, S., & Waldfogel, J. (2012). Child care and school performance in Denmark and the United States. *Children and youth services review, 34*(3), 576-589.

Gartlehner, G., Hansen, R., Nissman, D., Lohr, K., & Carey, T. (2006). Criteria for Distinguishing Effectiveness From Efficacy Trials in Systematic Reviews. *Technical Reviews*, (12). http://www.ncbi.nlm.nih.gov/books/NBK44029/

Gest, S. D., Holland-Coviello, R., Welsh, J. A., Eicher-Catt, D. L., & Gill, S. (2006). Language development subcontexts in Head Start classrooms: Distinctive patterns of teacher talk during free play, mealtime, and book reading. *Early Education and Development, 17*(2), 293-315.

Guo, Y., Justice, L. M., Kaderavek, J. N., & McGinty, A. (2012). The literacy environment of preschool classrooms: contributions to children's emergent literacy growth. *Journal of Research in Reading, 35*(3), 308-327. doi: 10.1111/j.1467-9817.2010.01467.x

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.

Hansen, O. H. (2013). *Stemmer i Fællesskabet.* [Voices in the community]. (Dissertation). Aarhus Universitet, Institut for Uddannelse og Pædagogik.

Harms, T., & Clifford, R. M. (1980). *Early childhood environment rating scale*. New York: Teachers College Press.

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. doi: http://dx.doi.org/10.1016/j.ecresq.2007.05.002

Jensen, B. (2009). A Nordic approach to Early Childhood Education (ECE) and socially endangered children. *European Early Childhood Education Research Journal, 17*(1), 7-21.

Justice, L. M., Mashburn, A. J., Hamre, B. K., & Pianta, R. C. (2008). Quality of language and literacy instruction in preschool classrooms serving at-risk pupils. *Early Childhood Research Quarterly, 23*(1), 51-68. doi: http://dx.doi.org/10.1016/j.ecresq.2007.09.004

Keys, T. D., Farkas, G., Burchinal, M. R., Duncan, G. J., Vandell, D. L., Li, W., . . . Howes, C. (2013). Preschool Center Quality and School Readiness: Quality Effects and Variation by Demographic and Child Characteristics. *Child Development, 84*(4), 1171-1190. doi: 10.1111/cdev.12048

Klitmøller, J., & Sommer, D. (2015). *Læring, dannelse og udvikling: Kvalificering til fremtiden i daginstitution og skole*. [Learning, cultural upbringing and development: Enhancing for the future of day care and school]: Hans Reitzel.

Koh, S., & Neuman, S. B. (2009). The Impact of Professional Development in Family Child Care: A Practice-Based Approach. *Early Education and Development, 20*(3), 537-562. doi: 10.1080/10409280902908841

Kragh-Muller, G., & Ringsmose, C. (2015). *Pædagogisk kvalitet i store og små daginstitutioner.* [Pedagogical quality in large and small day cares]. Danmarks Pædagogiske Universitetsskole.

LoCasale-Crouch, J., Konold, T., Pianta, R., Howes, C., Burchinal, M., Bryant, D., . . . Barbarin, O. (2007). Observed classroom quality profiles in state-funded pre-kindergarten programs and associations with teacher, program, and classroom characteristics. *Early Childhood Research Quarterly, 22*(1), 3-17.

Logan, J. R., Piasta, S., Justice, L., Schatschneider, C., & Petrill, S. (2011). Children's Attendance Rates and Quality of Teacher-Child Interactions in At-Risk Preschool Classrooms: Contribution to Children's Expressive Language Growth. *Child & Youth Care Forum, 40*(6), 457-477. doi: 10.1007/s10566-011-9142-x

Lonigan, C. J., Burgess, S. R., & Anthony, J. L. (2000). Development of emergent literacy and early reading skills in preschool children: Evidence from a latent-variable longitudinal study. *Developmental psychology, 36*(5), 596-613. doi: 10.1037/0012-1649.36.5.596

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. doi: 10.1111/j.1467-8624.2008.01154.x

Ministry of Children, Gender Equality, Integration and Social Affairs. (2014). Pædagogiske læreplaner. [Pedagogic learning plans]. Retrieved April 1, 2015 from http://sm.dk/arbejdsomrader/dagtilbud/paedagogiske-laereplaner.

Morrow, L. M. (1990). Preparing the classroom environment to promote literacy during play. *Early Childhood Research Quarterly, 5*(4), 537-554. doi: http://dx.doi.org/10.1016/0885-2006(90)90018-V

Murray, B., Stahl, S., & Ivey, M. G. (1996). Developing phoneme awareness through alphabet books. *Reading and Writing, 8*(4), 307-322. doi: 10.1007/BF00395111

National Early Literacy Panel. (2008). *Developing early literacy: Report of the National Early Literacy Panel: A scientific synthesis of early literacy development and implications for intervention*. Washington, D.C.: National Institute for Literacy.

Neuman, S. B. (1999). Books Make a Difference: A Study of Access to Literacy. *Reading Research Quarterly, 34*(3), 286-311. doi: 10.1598/RRQ.34.3.3

Neuman, S. B., & Celano, D. (2001). Access to Print in Low-Income and Middle-Income Communities: An Ecological Study of Four Neighborhoods. *Reading Research Quarterly, 36*(1), 8-26. doi: 10.1598/RRQ.36.1.1

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

Neuman, S. B., & Roskos, K. (1993). Access to Print for Children of Poverty: Differential Effects of Adult Mediation and Literacy-Enriched Play Settings on Environmental and Functional Print Tasks. *American Educational Research Journal, 30*(1), 95-122. doi: 10.2307/1163191

NICHD. (2001). Nonmaternal care and family factors in early development: An overview of the NICHD Study of Early Child Care. *Journal of Applied Developmental Psychology, 22*(5), 457-492. doi: http://dx.doi.org/10.1016/S0193-3973(01)00092-2

NICHD Early Child Care Research Network. (2006). Child-care effect sizes for the NICHD Study of Early Child Care and Youth Development. *American Psychologist, 61*(2), 99-116. doi: 10.1037/0003-066X.61.2.99

OECD. (2009). Society at a glance 2009: OECD social indicators: OECD Publishing.

OECD. (2013). Early Childhood Education and Care *Education Today 2013: The OECD Perspective*: OECD Publishing.

Pakarinen, E., Lerkkanen, M.-K., Poikkeus, A.-M., Kiuru, N., Siekkinen, M., Rasku-Puttonen, H., & Nurmi, J.-E. (2010). A validation of the Classroom Assessment Scoring System in Finnish kindergartens. *Early Education and Development, 21*(1), 95-124.

Peisner-Feinberg, E. S., Burchinal, M. R., Clifford, R. M., Culkin, M. L., Howes, C., Kagan, S. L., & Yazejian, N. (2001). The Relation of Preschool Child-Care Quality to Children's Cognitive and Social Developmental Trajectories through Second Grade. *Child Development, 72*(5), 1534-1553. doi: 10.1111/1467-8624.00364

Phillipsen, L. C., 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. doi: http://dx.doi.org/10.1016/S0885-2006(97)90004-1

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. doi: 10.1207/s1532480xads0903_2

Pianta, R. C., La Paro, K. M., & Hamre, B. K. (2008). *Classroom Assessment Scoring System Manual, Pre-K*. Baltimore, MD: Brookes.

Roth, F. P., Speece, D. L., & Cooper, D. H. (2002). A Longitudinal Analysis of the Connection Between Oral Language and Early Reading. *The Journal of Educational Research, 95*(5), 259-272. doi: 10.1080/00220670209596600

Russell, A. (1990). The Effects of Child-Staff Ratio on Staff and Child Behavior in Preschools: An Experimental Study. *Journal of Research in Childhood Education, 4*(2), 77-90. doi: 10.1080/02568549009594789

Scarr, S., Eisenberg, M., & Deater-Deckard, K. (1994). Measurement of quality in child care centers. *Early Childhood Research Quarterly, 9*(2), 131-151. doi: http://dx.doi.org/10.1016/0885-2006(94)90002-7

Slot, P. L. (2014). *Early childhood education and care in the Netherlands: Quality, curriculum, and relations with child development.* (Dissertation). Utrecht University.

Storch, S. A., & Whitehurst, G. J. (2002). Oral language and code-related precursors to reading: evidence from a longitudinal structural model. *Developmental psychology, 38*(6), 934.

Yazejian, N., Bryant, D., Freel, K., & Burchinal, M. (2015). High-quality early education: Age of entry and time in care differences in student outcomes for English-only and dual language learners. *Early Childhood Research Quarterly, 32*(0), 23-39. doi: http://dx.doi.org/10.1016/j.ecresq.2015.02.002

Paper 2

The Effects of Language and Emergent Literacy Professional Development for Teachers and Children: A meta-analysis

Justin Markussen-Brown & Carsten B. Juhl,

University of Southern Denmark

Shayne B. Piasta

The Ohio State University

Dorthe Bleses & Anders Højen

University of Southern Denmark

Laura M. Justice

The Ohio State University

Author Note

Justin Markussen-Brown, Institute for Language and Communication, University of Southern Denmark; Carsten B. Juhl, Institute for Sports Science and Clinical Biomechanics, University of Southern Denmark; Shayne B. Piasta, Department of Teaching & Learning, The Ohio State University; Dorthe Bleses, Institute for Language and Communication, University of Southern Denmark; Anders Højen, Institute for Language and Communication, University of Southern Denmark; Laura M. Justice, Department of Teaching & Learning, The Ohio State University.

Correspondence to concerning this article should be addressed to Justin Markussen-Brown, Institute for Language and Communication, University of Southern Denmark, Campusvej 55, Odense 5230, Denmark. E-mail: justinmb@sdu.dk.

Abstract

Professional development interventions are increasingly used to improve the language and emergent literacy environments that preschool teachers provide children. However, the literature does not clearly indicate the extent to which such efforts reach their goals, or whether improvements in teacher outcomes translate to learning gains in children. In the current study, meta-analysis was used to evaluate the effects of PD interventions on the teacher level outcomes of process quality, structural quality, and teacher knowledge. Furthermore, effects were also estimated for children's receptive vocabulary, phonological awareness, and alphabet knowledge. A medium effect was found for process quality, a large effect for structural quality, and no effect was found for teacher knowledge. Interventions yielded a medium effect for phonological awareness, and a small effect for alphabet knowledge, but these were not predicted by gains in teacher outcomes. Interventions that included coaching tended to yield larger effects than interventions that did not. The number of PD formats used by an intervention was also positively associated with effect sizes. In general, PD interventions appear to be a viable method of improving language and literacy processes and structures in preschools, but gains may need to be substantial if they are to improve child outcomes.

Keywords: professional development, preschool quality, teacher-child interactions, process quality, emergent literacy

The Effects of Language and Emergent Literacy Professional Development for Teachers and Children: A Meta-Analysis

Stakeholders and researchers have increasingly turned to professional development (PD) as a means of increasing preschool quality in order to improve the language and literacy outcomes of children (Buysse, Winton, & Rous, 2009). The logic of using PD to improve children's outcomes is supported by two findings in empirical research. First, research indicates that children's early language and emergent literacy skills are related to their literacy skills in early elementary school (National Early Literacy Panel, 2008; Whitehurst & Lonigan, 1998). Secondly, preschool quality is correlated with children's language and emergent literacy skills (Burchinal et al., 2008; Howes et al., 2008; Keys et al., 2013). Thus, it seems plausible that if preschool quality can be increased via PD interventions, children's outcomes may also improve.

PD refers to experiences aimed at enhancing teachers' knowledge base, skill set, or practices (Sheridan, Edwards, Marvin, & Knoche, 2009). The basic premise behind every language and/or emergent literacy PD intervention assumes that PD opportunities (such as coaching or courses) can lead to changes in preschool quality, which in turn benefit children's outcomes (Buysse et al., 2009). Although the logic of this model is in line with interactionist theories of child development (e.g., Bronfenbrenner, 1979; Vygotsky, 1978), the literature is mixed with regards to (a) the extent to which PD interventions are effective, and (b) what constitutes effective PD. For example, some research has found that PD interventions that utilize coaching (e.g., Pianta, Mashburn, Downer, Hamre, & Justice, 2008) or courses (e.g., Hamre et al., 2012) have resulted in improvements in preschool quality. However, other research has not found significant improvements with the inclusion of coaching (e.g., Lonigan, Farver, Phillips, & Clancy-Menchetti, 2011) or courses (e.g., Neuman & Cunningham, 2009). Furthermore, some research indicates that improvements on teacher-level outcomes, such as overall quality, do not necessary have a broad effect on children's

language and literacy outcomes (e.g., Buysse, Castro, & Peisner-Feinberg, 2010). In contrast, other experimental research has found that improvements in global preschool quality predict child gains in vocabulary and phonological awareness (Wasik & Hindman, 2011).

Thus, the extent to which current PD efforts improve the quality of language and environments in preschools, and how these relate to children's outcomes is unclear. Furthermore, we still lack insight regarding what makes PD effective. In the current work, we conducted a systematic review of the literature as to the effects of PD interventions with a language and/or literacy focus. By conducting a series of meta-analyses, we estimated the pooled effects of PD interventions on constructs of preschool quality and teacher knowledge, as well as key child outcomes of language and literacy. We also investigated which variables might explain variation in study results.

Using PD to Enhance Process Quality, Structural Quality and Teacher Knowledge

Typical theories of change in preschool PD research revolve around improving teachers' skills or knowledge with the expectation that teachers will carry these changes over to their classroom practices thereby increasing the quality of the preschool environment (e.g., Hamre et al., 2012). Measuring preschool quality is therefore one way of evaluating the effect of a PD intervention with a language and/or literacy focus.

In general, preschool quality refers to variables in the language and literacy learning environment that are empirically associated with the child precursor skills that facilitate learning to read (Pianta et al., 2005). Historically, indicators of preschool quality have been divided into the broad domains of *structural* and *process quality* (Dowsett, Huston, Imes, & Gennetian, 2008; Pianta et al., 2005; Scarr, Eisenberg, & Deater-Deckard, 1994). Here, structural quality refers either to regulatory factors such as teacher qualifications or teacher-child ratios, or features of the physical environment. In contrast, process quality refers mainly to interactions that occur in the preschool

environment, such as teacher-child interactions, or child interactions with equipment and materials (Mashburn et al., 2008; Pianta et al., 2005). Regulatory structural quality has traditionally been regarded as a distal measure of process quality under the assumption that, for example, well-educated teachers provide better interactions to children (Dickinson, 2006). However, some research has demonstrated that indicators such as teacher credentials neither predict process quality nor child outcomes (Early et al., 2006; Howes et al., 2008). Direct observations of classroom processes, on the other hand, have been found to predict children's outcomes (Burchinal et al., 2008; Howes et al., 2008), and thus process quality has garnered increased attention as an important outcome in PD intervention research. A number of studies have demonstrated that process quality can be improved via PD interventions (e.g., Hamre et al., 2012; Landry, Anthony, Swank, & Monseque-Bailey, 2009; Powell, Diamond, Burchinal, & Koehler, 2010)

In addition to process quality, the structural quality of physical classroom literacy environments such as the provision of literacy materials may be an important outcome of PD interventions as well. Research indicates that structural quality can positively influence children's early literacy experiences, especially in combination with quality interactions (Guo, Justice, Kaderavek, & McGinty, 2012; Neuman, 1999; Neuman & Roskos, 1992). Research has also demonstrated that PD can help teachers to provide higher quality structural literacy environments (Grace et al., 2008; Neuman & Wright, 2010). Furthermore, high structural quality in the literacy environment may have special importance for socially disadvantaged children, because research suggests that they have less access to literacy materials in their home communities (Neuman & Roskos, 1993).

Due to its importance in the theory of change of PD, teacher knowledge of language and literacy is another potentially significant outcome of PD intervention research. In elementary education, school teachers' knowledge of literacy instruction is considered essential to children's

reading success (Moats, 1999). It is therefore reasonable to assume that teachers who lack content knowledge about literacy will have difficulty teaching children literacy skills. Some evidence suggests that little focus is placed on ensuring that preschool teachers have adequate knowledge of language and emergent literacy. Cunningham and colleagues (2009) found that preschool teachers scored low on actual knowledge of emergent literacy, tended to overestimate their knowledge level, and mostly failed to acquire new knowledge from a PD intervention. Other research, however, has demonstrated that when PD succeeds in improving teacher knowledge, changes in classroom practices manifest as well (Hindman & Wasik, 2011). As such, measuring teacher knowledge is also a potential method of evaluating the effect of PD interventions with a language and literacy focus.

Although teacher-level outcomes such as preschool quality and teacher knowledge are proximal objectives of PD interventions, the true test of the effect of a PD intervention lies in its effect on children's key outcomes. According to a meta-analysis conducted by the National Early Literacy Panel (2008), a number of language and emergent literacy skills have medium to strong associations with conventional reading skills. These include alphabet knowledge, phonological awareness, oral language skills, as well as others. However, it is unclear how well PD interventions improve these outcomes. Some researchers have found that PD interventions have positive effects on children's language and literacy outcomes (e.g., Landry et al., 2009), but others found that PD had little or no effect on children's outcomes (e.g., Cabell et al., 2011). Other researchers have reported gains for teachers, but did not report child outcomes (e.g., Dickinson & Caswell, 2007).

Challenges in Interpreting the Results of PD Intervention Research

As described above, the main goals of PD interventions are to improve proximal outcomes at the teacher level in order to benefit child outcomes at the distal level. In recent years, a growing number of researchers have investigated the effects of such PD interventions experimentally. Yet,

the literature is mixed with regards to the extent to which PD interventions are effective, and what constitutes effective PD. Some studies have demonstrated positive results across most measures. For example, Wasik and Hindman (2011) implemented the Exceptional Coaching for Early Language and Literacy (ExCELL) PD model, an intensive PD model, in which 19 teachers in an experiment group received a PD intervention consisting of a summer course, monthly training cycles including coaching, materials, and lessons plans, and were compared against a control group (n = 11). After one year of implementation, teachers demonstrated increases in process quality as measured by the Classroom Assessment Scoring System (CLASS; Pianta, La Paro, & Hamre, 2008), and structural quality as measured by a subset of the Early Language & Literacy Classroom Observation (ELLCO; Smith & Dickinson, 2002). Furthermore, children demonstrated significant improvements in receptive vocabulary and phonological awareness.

However, the results of other PD interventions have demonstrated fewer effects. For example, Buysse et al. (2010) implemented the Nuestros Niños PD program, a language and literacy PD for teachers working in classrooms with a majority of Latino children in the United States (experiment group n = 26, control group n = 29). The PD intervention consisted of several components aimed at supporting teachers' learning including the use of courses, coaching, and a community of practice. After a period of one school year, the researchers observed significant gains on half of the teacher-level outcomes of preschool quality as measured by the ELLCO and the ELLCO Addendum for English language learners (Castro, 2005). However, only one out of five child outcomes (phonological awareness) demonstrated significant gains.

Interpreting why two seemingly similar studies diverge in their results can be challenging. One possible explanation of the differences in findings may be due to intervention dosage (the study by Wasik and Hindman [2011] employed nearly twice as much courses and coaching), but it might also be that the coaching and courses in the study by Wasik and Hindman were more effective. In

both studies, it is furthermore difficult to ascertain which PD formats accounted for which effects. For example, in the study by Buysse and colleagues, we do not know whether it was the courses, coaching or community of practice (or some combination of them) that mediated the improvement in children's phonological awareness.

Some researchers have attempted to isolate the effects of individual PD formats. Neuman and Wright (2010) conducted an experiment in which they compared the effects of courses (n = 58) against coaching (n = 58) when both experiment groups had received the same dosage, and compared them to a control group (n = 32). Teachers in the course group received 30 hours of coursework at a community college, whereas the coaching group received 30 hours of in-class coaching. Effects were measured on process and structural quality using the ELLCO, as well as teacher knowledge of language and literacy. Results indicated that neither experiment group made gains in process quality or knowledge, but that the coaching group made significant gains in structural quality. Thus the results of the study indicated that coaching – when dosage is equal – may be a preferable form of PD than courses.

However, in another study, Hamre et al. (2012) designed a college course with a language and literacy focus for preschool teachers. Teachers who received the course (n = 160) received 42 hours of instruction, and were compared to a control group (n = 174). Teachers' process quality was evaluated at post-test using the CLASS. Teachers in the experiment group displayed significant gains in the emotional support (effect size = .41) and instructional support (effect size = .66) domains. Thus, the findings of Hamre et al. (2012) demonstrated that a course format of PD could improve process quality on its own, which contrasts with the findings by Neuman and Wright (2010), who isolated the effects of both courses and coaching, and found no effect on process quality. Notably, however, neither study investigated the effects of their interventions on children's outcomes.

The studies reviewed above exemplify some of the challenges in interpreting the results of experimental PD research. When results conflict, it can be difficult to ascertain which variables accounted for which effects. Some researchers have argued that the difficulty in untangling the effect of PD research lies in the current research paradigm, which tests over-arching formats of PD without taking much consideration of the myriad internal processes and other factors that potentially moderate and mediate outcomes (Sheridan et al., 2009). This critique is perhaps supported by Schachter (2015), who reviewed the research designs of PD interventions in the early childhood education field, and identified 35 unique models of delivering intervention. Schachter called for future researchers to develop better means of assessing PD interventions in order to facilitate our understanding and interpretations of the results that manifest.

Using Synthesis to Explore Variation in the Literature

In consideration of the mixed results in the literature of PD interventions for preschool teachers, Sheridan and colleagues (2009) called for a paradigm shift regarding how scientists research the impacts of PD. Schachter (2015) also made a number of recommendations including better evaluation of intervention effects. Although it can be difficult to understand PD impacts on a study-to-study basis, it can also be helpful to conduct systematic review and synthesis of the existing literature. In situations in which seemingly similar research investigations yield dissimilar results, systematic reviews can reveal patterns of results that could indicate the overall effect, or reveal study variables that potentially account for variability in results (Cooper & Hedges, 2009).

Notably, some researchers have already used systematic review to draw new information from the existing literature on the PD of preschool teachers and other professional child care providers. Fukkink and Lont (2007) used systematic review and meta-analysis to investigate the effects of PD on child care providers, and found medium effects on caregivers' knowledge, attitudes, and skills. The researchers also estimated an effect of PD at the child level, and found a

positive but non-significant result. However, the review had a broad scope, and did not report findings specific to language and emergent literacy.

Zaslow, Tout, Halle, Whittaker, and Lavelle (2010) conducted an exhaustive configurative systematic review of the literature on PD for child care providers with the goal of identifying the features of effective PD interventions. They surveyed experimental and descriptive studies alike. Their review was extensive and included literature focusing on language and literacy, mathematics, and social skills. The review of literature with a language and literacy focus found that studies generally reported positive effects of PD on teacher knowledge, and teacher practice in some cases. Furthermore, most studies reported positive findings for children's outcomes. However, Zaslow and colleagues noted that the complexity of most PD interventions made it difficult to disentangle the individual contributions of study features, but they did find that PD interventions that aided teachers in goal-setting tended to report positive findings.

Aikens and Akers (2011) conducted a systematic review on the effects of one format of PD, specifically use of a coaching format. They defined coaching as the use of a coach or other knowledgeable person that provided ongoing support to teachers with the goal of helping teachers learn and implement skills or curricula. Like the review by Zaslow et al. (2010), the review was also configurative in nature, and in particular categorized studies according to whether their outcomes related to classroom instruction, curriculum implementation, classroom environmental indicators, teacher-child interactions and child academic and socio-emotional outcomes. The researchers found that most studies reported positive findings for all categories. However, in most cases, the effects of coaching were not isolated in the research design due to most researchers' use of multi-format PD models. Thus, Aikens and Akers also found it challenging to determine coaching's unique contribution to study effects. Furthermore, the review included non-experimental studies, and participants came from a wide range of teaching backgrounds including preschool, kindergarten and

early elementary school. Although informative, causality cannot be inferred from non-experimental studies, and results partially based on the inclusion of school teachers may not be applicable to early childhood education.

Collectively, these reviews suggest that PD interventions can have positive effects on a number of adult outcomes, and possibly on child outcomes as well. They also describe in detail the diversity in PD approaches, especially in terms of design. However, there still lacks clarity regarding the variation in results that we currently observe in the literature. Furthermore, a greater understanding of the extent to which teacher-level outcomes mediate meaningful gains for children's language and literacy outcomes is also still needed.

The Contribution of the Current Study

In the current work, we conducted a systematic review of the literature on PD interventions with a language and/or literacy focus. Using meta-analysis, we synthesized intervention effects for process quality, structural quality, and teacher knowledge. Through extensive coding, results were further broken down into sub-categories that allowed us to examine the roles of frequently used formats of PD, as well as a range of other pertinent variables such as intervention intensity, or whether child populations were predominantly at risk for academic failure or not. Effects for the child outcomes of receptive vocabulary, phonological awareness, and alphabet knowledge were also estimated when data was provided.

The systematic review presented here contributes to the literature in a number of ways. First, it is to our knowledge the first meta-analysis to have a scope narrowed to PD interventions with a language and/or literacy focus within a preschool setting. Second, it includes several primary research articles that were published since the reviews named above were published. Finally, by estimating effects at both teacher and child levels, we were able to investigate the mediating effects

of quality on children's early language and literacy skills. Specifically, the current review aimed to answer the following five research questions:

i. What is the estimated effect of PD on process quality, structural quality, and teacher knowledge?
ii. What is the estimated effect of PD on children's language and emergent literacy outcomes?
iii. To what extent do the proximal effects of PD (i.e., teacher outcomes) mediate effects on child outcomes?
iv. Which formats of PD are most effective?
v. What additional factors explain variation in the results according to a sensitivity analysis?

Method

Search Strategies

A published protocol specifying search strategy, selection criteria, and analytic approach is published on the PROSPERO database under protocol number CRD42014009361 (see Markussen-Brown, Juhl, Piasta, Bleses, & Højen, 2014). As stated in the protocol, studies included in the present work were retrieved from databases as part of a larger project to generate a web-based systematic map at the University of Southern Denmark. The goal of the systematic map was to create an interactive interface containing research regarding the proximal and distal factors related to children's language and emergent literacy development. The search strings designed to find studies for the systematic map also contained the necessary terms for conducting the current review of PD. The systematic map consists of more than 2,300 studies.

The following databases were searched in the creation of the systematic map: Cumulative Index to Nursing and Allied Health Literature (CINAHL), Modern Language Association (MLA), PsycINFO, Linguistics and Language Behavior Abstracts, Australian Education Index (AEI), British Education Index (BEI), Sociological Abstracts, Canadian Business & Current Affairs

Education (CBCA Education), Web of Science, and Scopus. Each database was searched on a single day, with actual searches being conducted between October 11th 2013 and March 13th 2014.

Through extensive piloting, four categories of search terms were created and combined in each database using the boolean operators OR within each category, and the operators AND between the four categories. The search categories were: *language and communication*, *development period of interest*, e*ffect factor*, and e*nvironmental factors*. The full list of search terms is given in Appendix A.

The database searches returned 65,037 studies after duplicates were removed. A team of 26 bachelor and master students conducted the first screening of studies, in which studies were flagged for possible suitability for the systematic map, and for a number of systematic reviews including the current PD review. Coders were initially trained to 90% inter-rater reliability before they began screening papers for inclusion in the systematic map. If a coder was in doubt about the suitability of an article, it could be tagged for a second opinion from the research staff. As an extra precaution for ensuring that potentially relevant PD articles proceeded to full text screening, the first author rescreened all imported studies in the systematic map containing the term "professional development."

Studies passing this initial screening were screened again by the first author using the inclusion criteria defined below. Finally, the reference lists of included studies were checked for additional studies that would also meet inclusion criteria.

Inclusion Criteria

As stated in the protocol, studies included in this review had to be peer-reviewed and test a PD intervention with the aim of improving children's language and/or emergent literacy development. Participants had to be in-service preschool teachers providing care to 3-6 year-old children. Home-based child-care providers were also included if there was evidence of the provision

of educational instruction (however, in practice, this only resulted in one additional study). Furthermore, studies had to compare an experiment group against a comparison group, and there had to be at least one non-self-reported teacher-level outcome of interest (i.e. process quality, structural quality, or teacher knowledge). Finally, studies had to report the data necessary for calculating effect sizes (i.e., the standardized mean difference).

Study Coding

Once a study was included in the review, it was coded by the first author for a number of variables including main outcomes and covariates. A full list of study codes is given in Table 1. To ensure reliability of the coding scheme, 20% of included studies were double-coded by one of the co-authors, and the average kappa was determined to be .86, which we considered to be acceptable.

The presence of child outcomes was not an inclusion criterion, but they were extracted for meta-analysis if available. Our original intention was to extract preschool outcomes that the National Early Literacy Panel (2008) identified as predictors of conventional literacy. However, through pilot testing of our coding scheme, it became apparent that the size of the subsets of studies including both teacher and child outcomes was small. Therefore, it only made practical sense to investigate receptive vocabulary, phonological awareness, and alphabet knowledge.

A number of variables were used to evaluate the quality of included studies including a number of suggestions adapted from Gersten et al. (2005). These included evidence of experimental and control groups being comparable, evidence of whether measures were commercially available, the degree to which measures aligned with the PD-intervention, and evidence of post-test occurring within a few weeks of intervention stop. Four variables of the Cochrane Collaborations risk of bias tool (Higgins & Green, 2008) were also coded including sequence generation, blinding of personnel, incomplete outcome data, and selective reporting. The coding of these variables was used to conduct a risk of bias analysis.

Table 1

Codes used for each included study

Study code	Description
Outcomes	
Classroom/teacher outcomes	
Process quality	Measures of teacher-child interactions at either the global or a more domain-specific level.
Materials quality	Measures of the structural features of process quality such as materials. E.g., the Literacy Environment Checklist from the ELLCO.
Teacher knowledge	Measures of teachers' content knowledge of language and literacy.
Child outcomes	
Receptive vocabulary	Any measure of children's receptive vocabulary such as the PPVT.
Phonological awareness	Any measure of children's manipulation of phonemes such as rhyming or segmenting.
Alphabet knowledge	Knowledge of letter names and sounds.
PD characteristics	
Format	Format of PD employed. Options included courses and coaching (included mentoring and consultation), as well as the additional use of an experimental curriculum, assessment data software, and other.

Courses	Courses were also subdivided as either college courses or a workshop (summer institutes included).
Coaching	Coaching was also subdivided into coaching with video versus coaching without video.
Duration	Length of the study at posttest.
Intensity	Total intensity of the PD format (courses and coaching only). Calculated as the number of hours of PD x the number of sessions.
Teacher characteristics	
Work experience	Teachers' mean years of experience.
Quality	
Trial size	Teacher population is less than or equal to 50 participants versus over 50.
At-risk children	Preschools serve primarily at-risk children.
Level of randomization	Randomized at individual, classroom or preschool level.
Procedural fidelity	Evidence regarding the extent to teachers actually received the PD.
Comparability of groups	Evidence regarding the extent to which experiment and control groups are comparable at the beginning of the study.
Appropriate timing of data collection	Evidence that posttest data was collected shortly after the end of the intervention.
Measures	
Commercially available	

instrument	Measure is commercially available.
Alignment of measures	The extent to which the measure is aligned with the PD content.
Cochrane's Risk of Bias	
Sequence generation	Confidence that groups were randomized.
Blinding	Evidence regarding the blinding of assessors.
Incomplete outcome data addressed	The extent to which researchers address attrition of participants (teachers only).
Free of selective reporting	Evidence that all pre-specified outcomes are reported.

Note. ELLCO = Early Language and Literacy Classroom Observation (Smith & Dickinson, 2002); PPVT = Peabody Picture Vocabulary Test (Dunn & Dunn, 1997).

Statistical Methods

Intervention effects were calculated as the standard mean difference (SMD), which was estimated as the difference between post-test scores for experimental and control groups divided by the pooled standard deviation (see da Costa et al., 2013 regarding the use of post-test scores versus change-scores). The magnitude of effect sizes was interpreted using Cohen's (1988) guidelines. All effect sizes were calculated using means scores, standard deviations, and number of participants in experimental and control groups. Because the SMD slightly overestimates small samples, SMDs were corrected by converting effect sizes to Hedges' *g* using the formula described in Card (2011). Finally, the scores of measures that used falling scales rather than rising scales were inverted (e.g., the Negative Climate dimension from the Classroom Assessment Scoring System [CLASS]; Pianta, La Paro, et al., 2008).

Included studies were typical to educational science in that they varied in the number of experiment groups being tested, and the number of outcomes being measured. These experimental

designs introduce the problem of statistical dependence between effect sizes because a single control group is compared multiple times to experiment groups, or because participants are measured more than once on different outcomes (Scammacca, Roberts, & Stuebing, 2013). For example, Neuman and Wright (2010) tested the effects of coaching in one experiment group, coursework in another, and compared both against a single control group. Furthermore, teachers in this study were measured on multiple outcomes including classroom quality and teacher knowledge.

In the current work, we dealt with the dependence of experiment groups by treating each experiment group as an independent trial and split the shared control group in two as suggested by the Cochrane Handbook (Higgins & Green, 2008). Other recommended options included combing the experiment groups into a single group, or selecting the most important group. Although both of these methods would preserve the size of the control group, they would also essentially eliminate the relevance of Neuman & Wright's research goal, which was to compare the impacts of two different formats of PD. Given that answering this question was also a goal of the current work, the split control group method was favored.

To deal with dependence caused by multiple outcomes, we employed the shifting-unit-of-analysis approach in which outcomes are grouped according to construct, and then averaged within studies (Cooper, 1998). This method essentially entails conducting one meta-analysis for each construct of interest (i.e. process quality, structural quality, and teacher knowledge), with studies being allowed to contribute an effect size to each construct. The method is well used in educational research (see Ahn, Ames, & Myers, 2012), as it allows the meta-analyst to avoid the situation in which different constructs become combined into a single one. The standard error of each averaged SMD that a trial contributed with was calculated from the mean variance of the sub-scores making up the averaged score.

There were also cases in which a measure contained a subscale that was process in nature, and another other that was structural (the ELLCO, for example, has process and structural subscales). In these situations, the subscales were treated as separate measures and allowed to contribute with an effect size to the meta-analysis that best fit their construct.

A random-effects meta-analysis (DerSimonian & Laird, 1986) was conducted for each outcome using the STATA 13 statistical package. A random-effects model was an appropriate choice as heterogeneity was assumed due to the fact that studies varied on a great number of factors, and because outcomes were measured using many different and sometimes multiple instruments. Heterogeneity was calculated as I^2, which is the percentage of variation between studies that cannot be explained by chance. The contribution of each trial to the pooled SMD for each outcome was weighted according to a variation of the inverse variance method called the DerSimonian & Laird method by which a measure of the between study variance (often referred to as tau-squared [τ^2, or Tau^2]) is incorporated in the variance.

Moderator Analysis

In addition to the six main meta-analyses that were conducted at the teacher and child levels, the moderating effects of covariates were also evaluated. When variables were categorical (such as randomization level), we used sub-group analysis, and then tested changes in SMD for statistical significance. If the statistical difference between subgroups were significant, then the covariate was considered to be important with regards to the PD outcome. The moderating effects of continuous variables, such as study duration, were modelled using meta-regression.

Results

Description of Studies

The flowchart in Figure 1 describes the process by which 22 studies (27 trials altogether) were included in the systematic review. Using the data in these trials, six meta-analyses were carried: process quality (n = 25), structural quality (n = 16), teacher knowledge (n = 10), receptive vocabulary (n = 5), phonological awareness (n = 5), and alphabet knowledge (n = 6). Additional study characteristics pertaining to number of participants, PD formats, intensity, duration and outcomes are listed in Appendix B.

The Effects of PD on Process Quality, Structural Quality, and Teacher Knowledge

Our first question explored the extent to which PD interventions affected process quality, structural quality, and teacher knowledge. As displayed in Figure 2, the overall pooled SMD for process quality was 0.52 (95% CI 0.34, 0.70) (n = 25) with moderate inconsistency in the results (I^2=50.5%). PD therefore had a medium effect on process quality. The overall pooled SMD for structural quality was 1.07 (95% CI 0.69, 1.45) (n = 16) (Figure 3), which is considered a large effect, but with substantial variation between effect sizes being detected (I^2 = 83.1%). This suggests that other factors not related to chance or the intervention explain the variation between individual trial results. In terms of teacher knowledge, a statistically non-significant pooled SMD of 0.15 (95% CI -0.02, 0.31) (n = 10) was detected (Figure 4**)** with no statistical heterogeneity (I^2 = 0%). Since all trials testing teacher knowledge yielded non-significant results, and no variation between individual effect sizes was detected, this outcome is not investigated further in the current work.

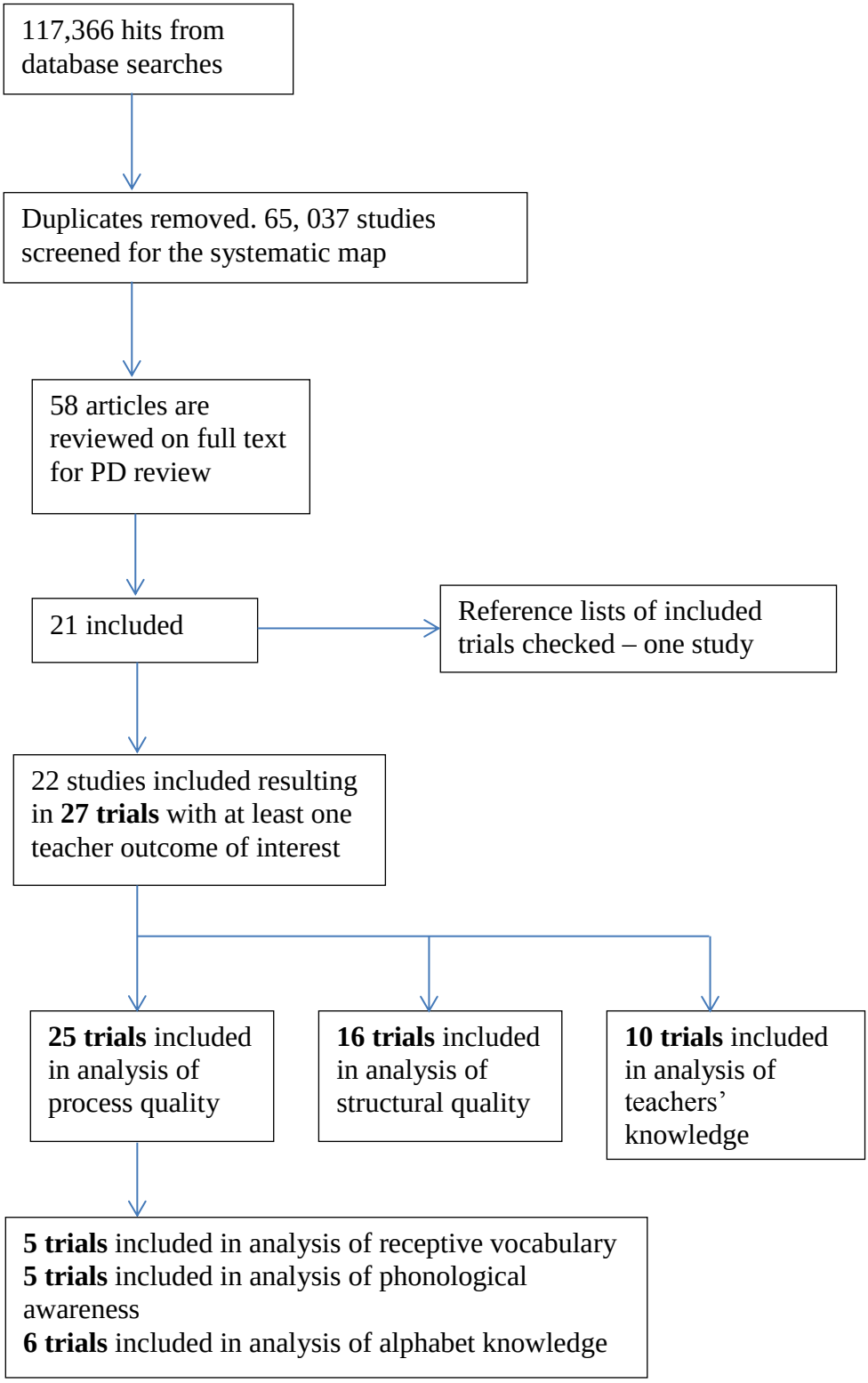

Figure 1. A flowchart of the process for identifying trials included in the review.

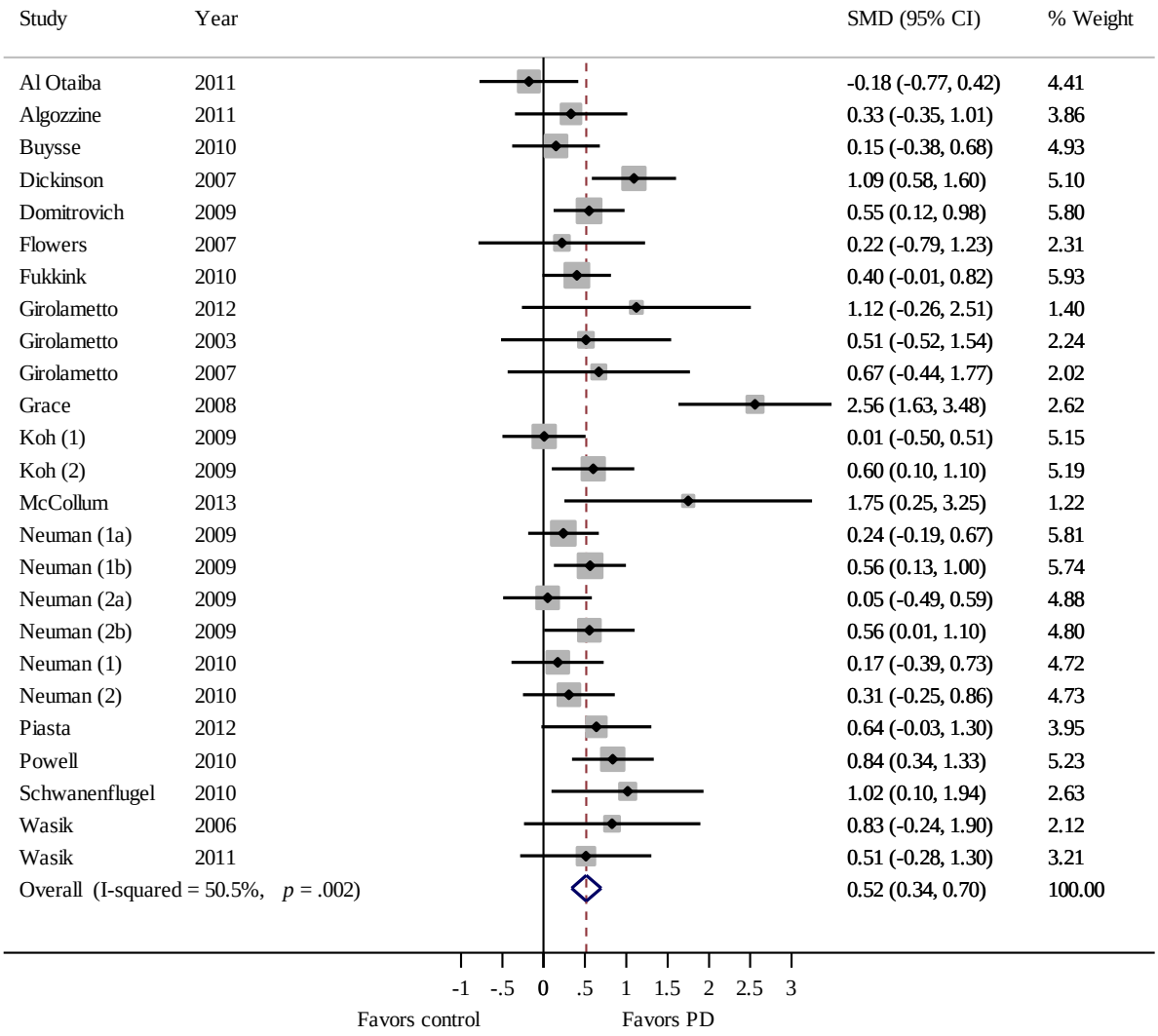

Figure 2. The effects of PD on process quality.

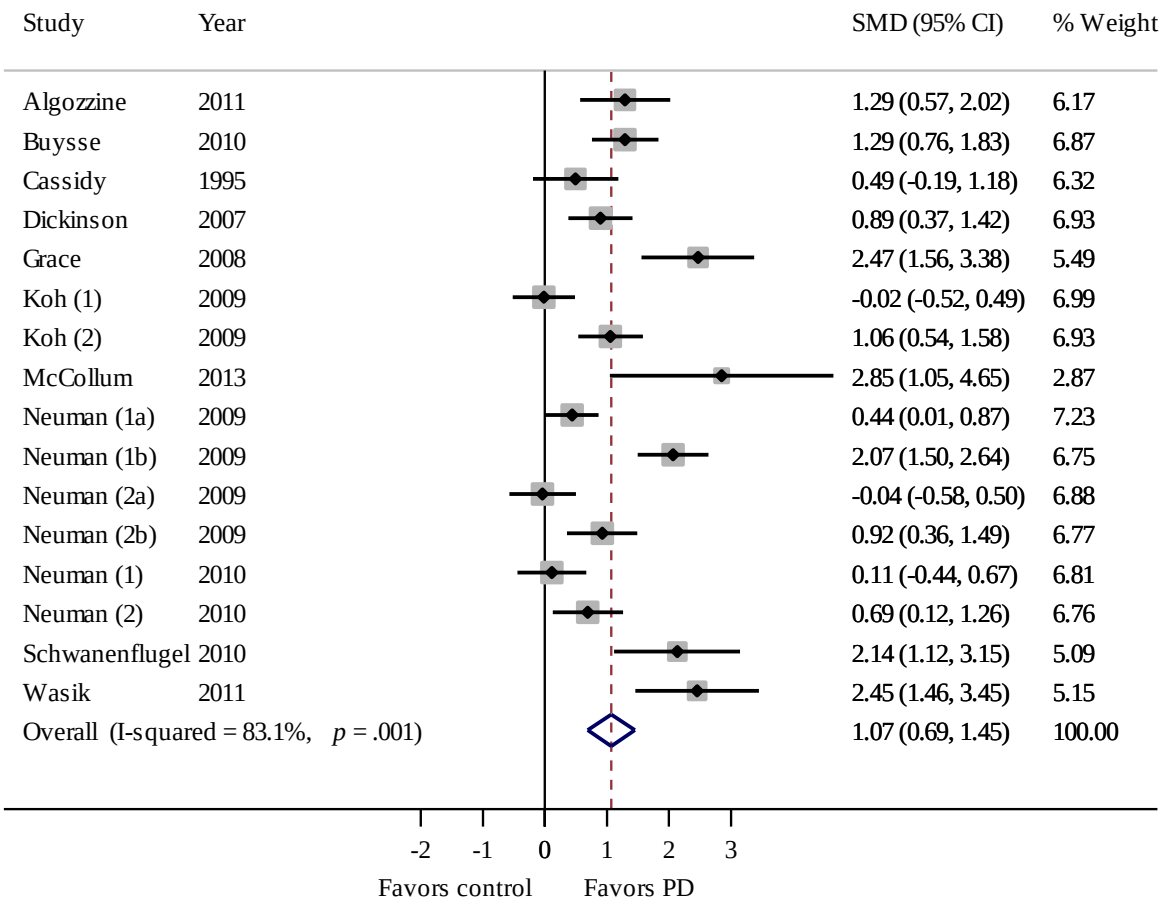

Figure 3. The effects of PD on structural quality.

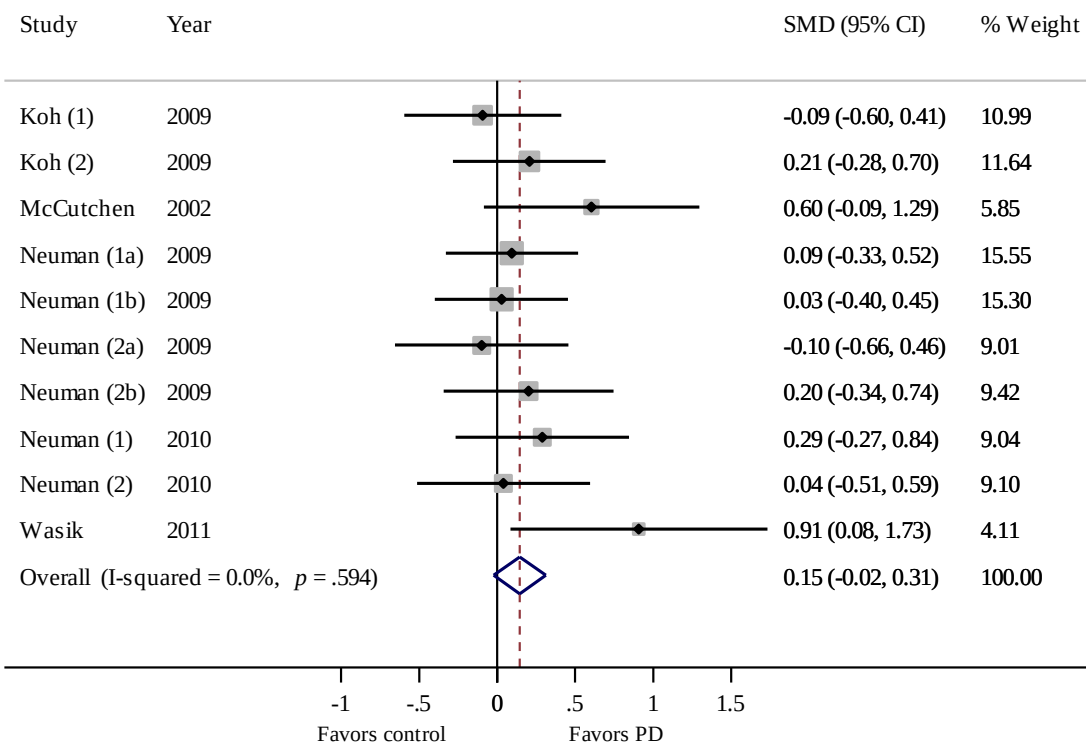

Figure 4. Effects of PD on teacher knowledge.

The Effects of PD on Receptive Vocabulary, Phonological Awareness, and Alphabet Knowledge

Our second research question investigated the effect of PD on the child outcomes of language and emergent literacy in the included trials. A subgroup of five trials measured the effects of the PD on children's receptive vocabulary (Figure 5). Here, a small SMD of 0.21 (95% CI -0.01, 0.43) (n = 5) was estimated that was approaching significance. For phonological awareness (Figure 6), a subgroup of five trials contributed to a pooled SMD of 0.46 (95% CI 0.2, 0.71) (n = 5) indicating a medium effect. Finally, a small pooled SMD of 0.18 (95% CI 0.1, 0.27) (n = 6) was estimated for alphabet knowledge (Figure 6). Although considerable heterogeneity was detected for receptive vocabulary and phonological awareness (I^2 = 79.6% and I^2 = 84.6% respectively), no heterogeneity was detected for alphabet knowledge (I^2 = 0%).

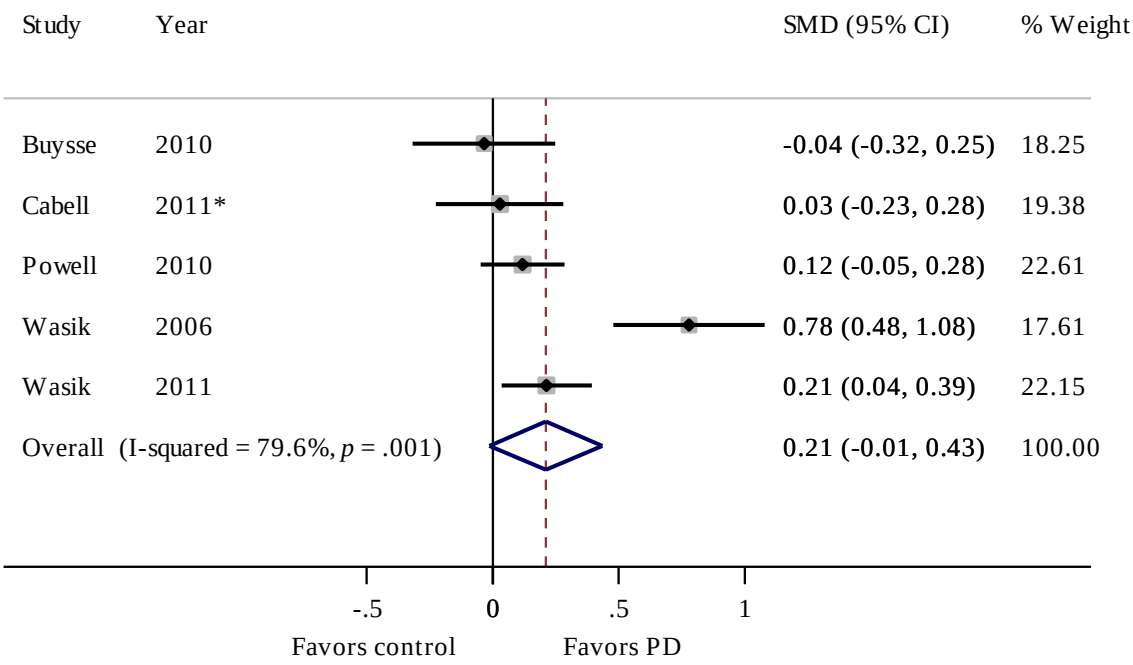

Figure 5. Effects of PD on children's receptive vocabulary. *Teacher outcomes for this trial are reported in Piasta et al. (2012).

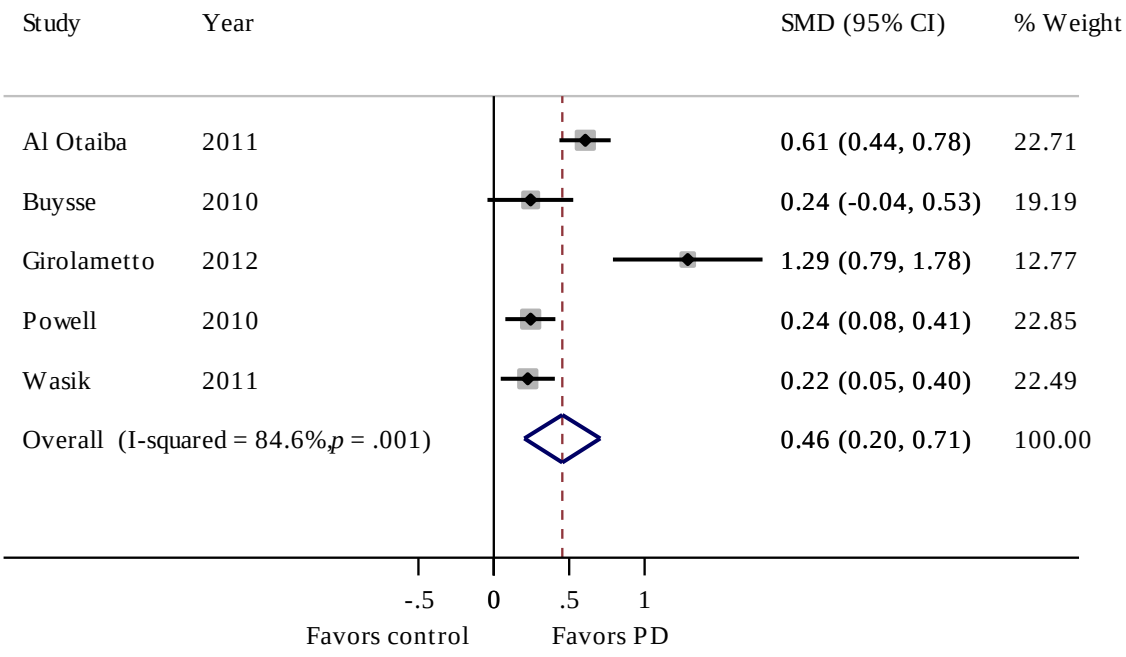

Figure 6. Effects of PD on children's phonological awareness.

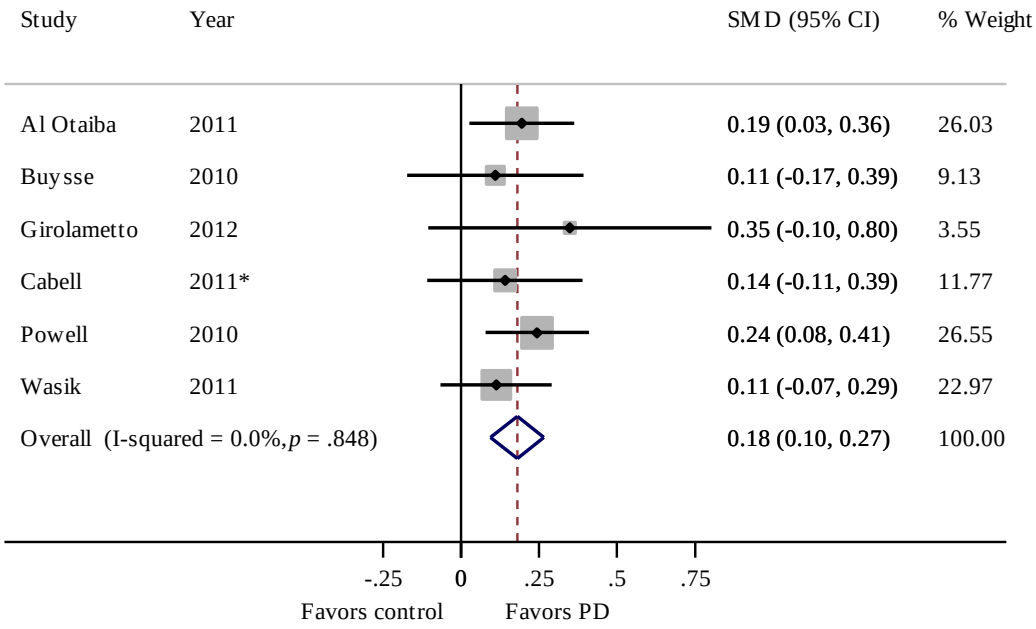

Figure 7. Effects of PD on children's alphabet knowledge. *Teacher outcomes for this trial are reported in (Piasta et al. [2012]).

The Extent to Which Process Quality Mediates Child Outcomes

We used meta-regression to address our third question regarding the extent to which teacher outcomes of quality mediated child outcomes of language and emergent literacy. There was an insufficient number of trials measuring both structural quality and child outcomes; however, we were able to investigate the extent to which SMDs for process quality predicted effects in children's receptive vocabulary (n = 5), phonological awareness (n = 5), and alphabet knowledge (n = 6).

For receptive vocabulary, process quality predicted a slope of 0.62 (95% CI -1.11, 2.35), but the wide confidence interval indicated an unstable relation between the two outcomes, and the result was not significant. Similarly, teachers' improvements in process quality did not significantly predict improvement to children's phonological awareness (ß = 0.30 [95% CI -1.14, 1.74]). Finally,

changes in children's alphabet knowledge were also unrelated to changes in teachers' improvements in process quality (ß=0.05 [95% CI -0.24, 0.34]).

The Effects of PD Formats

Our fourth question investigated the effectiveness of common formats of PD. This question was investigated by coding each trial for the types of PD format utilized. For nearly all included studies, PD interventions primarily consisted of forms of courses and coaching, and therefore the following analyses centralized around these formats. The use of an experimental language and emergent literacy curriculum, the use of children's assessment data to guide lesson planning, and any other unique factor that could affect quality were also coded.

The effects of courses. Figure 8 displays a series of subgroup analyses comparing pooled SMDs as a function of PD format characteristics. The first group of analyses investigated the effects of courses. Interventions including courses had a larger effect on process quality (SMD = 0.57 [95% CI 0.38, 0.77], *n* = 22) than did interventions that did not include courses (SMD = 0.23 [95% CI -0.10, 0.56], *n = 3*), but the difference was only approaching significance. A similar result was found for structural quality. Interventions including courses resulted in a larger pooled SMD of 1.11 (95% CI 0.70, 1.52) (*n* = 15), but the difference was not significance when compared to the one study that did not include courses (SMD .069 [95% CI 0.13, 1.26]).

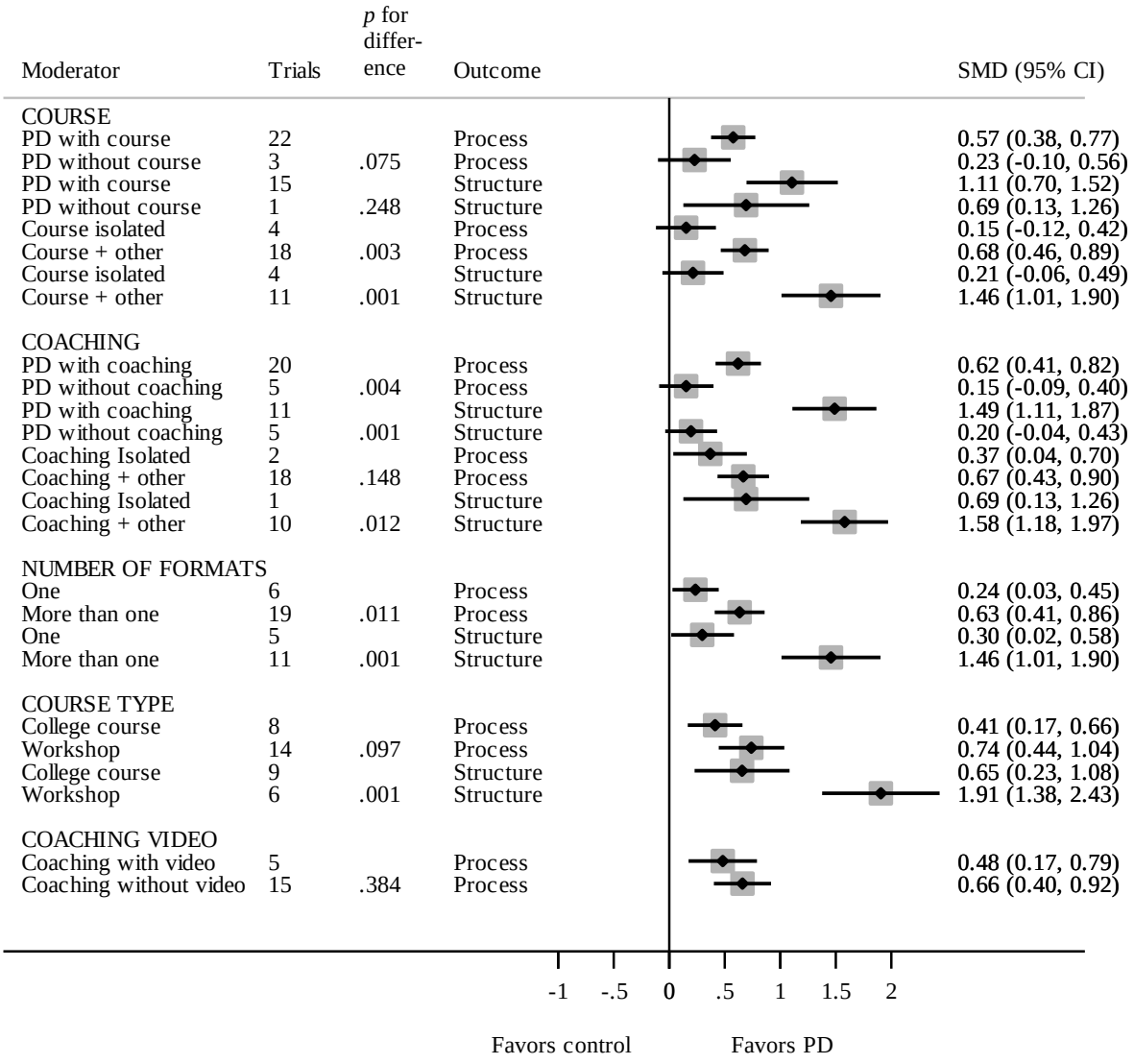

Figure 8. Effects of PD formats on process and structural quality.

Because four trials implemented PD interventions consisting only of courses, it was possible to evaluate the isolated effects of this format compared to interventions that combined courses with at least one other format of PD. For process quality, a non-significant pooled SMD of 0.15 (95% CI -0.12, 0.42) was estimated for courses in isolation. Interestingly, 18 trials combined courses with at least one other format, which resulted in a significantly larger effect (SMD 0.68 [95% CI 0.46, 0.89]). A similar pattern was revealed for structural quality. A pooled SMD of 0.21 (95% CI -0.06,

0.49) was estimated for the subgroup of four studies that isolated the effects of courses. However, the pooled SMD of the 11 trials that combined courses with at least one other format was substantially larger (SMD 1.46 [95% CI 1.01, 1.90]).

We expanded on our analysis of the effects of courses by further dividing courses into workshops and college-level courses. There were no significant differences between the contributions of workshops versus college courses on process quality. However, a significant difference was found for structural quality favoring workshops (SMD 1.91 [95% CI 1.38, 2.43], *n* = 6) versus college courses (SMD 0.65 [95% CI 0.23, 1.08], *n* = 9).

The effects of coaching. An advantage of including coaching in PD models was found. For process quality, the inclusion of coaching resulted in a SMD of 0.62 (95% CI 0.41, 0.82) (*n* = 20), whereas PD models without coaching resulted in a much smaller effect (SMD = 0.15 [95% CI -0.09, 0.40], *n* = 5), and the difference was significant. A similar, but more pronounced pattern was found for structural quality. Here the advantage of including coaching resulted in a SMD of 1.49 (95% CI 1.11, 1.87) (*n* = 11), which was significantly higher than the effect of not including coaching (SMD=0.20 [95% CI -0.04, 0.43], *n* = 5).

A subgroup of two trials isolated the effect of coaching and found a beneficial SMD of 0.37 (95% CI 0.04 0.70) on process quality, yet by adding at least one more format of PD, the SMD increased to 0.67 (95% CI 0.43, 0.90) (*n* = 18), but the difference was not significant. Only one trial isolated the effects of coaching on structural quality resulting in a pooled SMD of 0.69 (95% CI 0.13, 1.26). However, 10 other trials combined coaching with at least one other format of PD, which more than doubled the pooled SMD for structural quality to 1.58 (95% CI 1.18, 1.97) (*n* = 10).

Coaching trials were further divided into trials testing coaching with and without the use of video for feedback purposes. The comparison was only possible for process quality, and no significant differences were found between groups.

The effects of course and coaching intensity. The intensity of courses significantly predicted neither process quality (ß = -0.01 [95% CI -0.02, 0.01]), nor structural quality (ß = -0.03 [95% CI -0.08, 0.02]). Coaching intensity, however, did predict process quality (ß = 0.005 [95% CI 0.002, 0.007]), such that 100 hours of coaching, for example, predicted a SMD of approximately 0.50. Coaching intensity, however, did not predict structural quality (ß = 0.003 [95% CI -0.003, 0.002]).

Interestingly, when course and coaching intensities were combined into single predictors, significant slopes were found for both process (ß = 0.005 [95% CI 0.003, 0.007]) and structural qualities (ß = 0.005 [95% CI 0.001, 0.009]). Both models predict that 100 hours of PD are equivalent to a SMD of 0.50. The improvements in the models suggest that combined intensity is a better predictor than using only one of the formats.

The effects of number of formats. To explore whether an increasing number of PD formats (beyond courses and coaching) was predictive of higher quality, we entered the number of PD formats in each trial (range: 1-4) into a meta-regression analysis as the predictor variable. The analysis revealed a significant slope for both process quality (ß = 0.32 [95% CI 0.12, 0.52]), and structural quality (ß = 0.63 [95% CI 0.26, 1.00]), indicating that the addition of one format of PD contributed with an increase in SMD of 0.32 and 0.63 respectively.

Additional Analyses

The effects of duration and teacher work experience. In addition to variation due to PD format, we used meta-regression to investigate how effect sizes changed as a function of program duration and teacher work experience. Our interest in work experience stems from research that indicates that experience may not be a significantly related to process quality in particular (Justice, Mashburn, Hamre, & Pianta, 2008).

The length of intervention period significantly predicted structural quality (ß = 0.76 [95% CI 0.01, 1.52]), but not process quality (ß = 0.29 [95% CI -0.08, 0.66]). Teacher work experience predicted neither process quality (ß = 0.10 [95% CI -0.05, 0.25]), nor structural quality (ß = -0.01 [95% CI -1.95, 1.92]).

The effects of study quality. We used sub-group analysis to investigate variation in results due to study quality and risk of bias. The investigation of study quality revealed some statistically significant findings. Trial sizes of 50 or fewer participants had significantly larger gains (p = .012) in structural quality (SMD = 1.84 [95% CI 1.07, 2.60], n = 6) when compared to larger trials (SMD = 0.74 [95% CI 0.34, 1.13], n = 10) indicating benefit for smaller, more controlled studies. Effect sizes for trials including a majority of children at-risk due to poverty experienced significantly larger gains (p = .031) for process quality (SMD = 0.62 [95% CI 0.39, 0.85], n = 18) than did children not at risk (SMD = 0.26 [95% CI 0.02, 0.50], n = 7). Similarly, trials with children at-risk yielded significantly greater effects (p = .005) for structural quality (SMD = 1.24 [95% CI 0.79, 1.69], n = 13) than did trials with children not at risk (SMD = 0.42 [95% CI 0.07, .77], n = 3). In terms of study design, trials that randomized at the preschool level produced larger effects (p = .005) for structural quality (SMD = 2.01 [95% CI 1.34, 2.67], n = 4) versus studies that randomized at the classroom (SMD = 1.29 [95% CI 0.57, 2.02], n = 1), or teacher levels (SMD = 0.73 [95% CI 0.34, 1.13], n = 11). Trials that reported adequate procedural fidelity of their PD interventions resulted in a significantly smaller pooled SMD (p = .005) for structural quality (SMD = 0.67 [95% CI 0.26, 1.09], n = 9) than trials that failed to report procedural fidelity (SMD = 1.71 [95% CI 1.11, 2.31], n = 7).

Other quality indicators were found to have no relation to SMDs. These included the extent to which researchers ensured control and experiment groups were comparable and the degree to which post-test occurred reasonably close to the end of the implementation period. In terms of

outcome measures, significant differences were not found for commercial versus non-commercial measures. However, a borderline significant relation (p = .051) was found for process quality favoring instruments that were closely aligned with the content of the intervention (SMD = 0.70 [95% CI 0.40, 1.00], n = 13) versus instruments that were only partially aligned with PD content (SMD = 0.35 [95% CI 0.17, 0.53], n = 12).

The risk of bias. A series of subgroup analyses were also conducted using variables adapted from the Cochrane risk of bias assessment tool. Trials that demonstrated low risk of bias due to appropriate blinding procedures produced a significantly larger (p = .043) pooled SMD for process quality (SMD = 0.68 [95% CI 0.36, 1.00], n = 12) than did studies that had unclear (SMD = 0.44 [95% CI 0.26, 0.61], n = 12) or inadequate blinding procedures (SMD = -0.18 [95% CI -0.77, 0.42], n = 1). Furthermore, a risk of bias in process quality (p = .001) due to unaccounted for systematic attrition in a condition was detected favoring the one trial that had inadequate procedures for dealing with this issue (SMD = 2.56 [95% CI 1.63, 3.48]) versus the seven trials that had unclear procedures (SMD = 0.59 [95% CI 0.33, 0.84]) or the 17 trials that had adequate procedures (SMD = 0.39 [95% CI 0.23, 0.55]). All other variables were not significantly related to effect sizes.

Discussion

This systematic review estimated the effects of PD intervention studies that have a language and emergent literacy focus within a preschool context. The overall results indicate that current models of PD increase process and structural quality, but do not improve teachers' literacy knowledge. Our findings are partially in line with work by Fukkink and Lont (2007), who found that PD has a moderate effect on process quality. However, our results do not support their finding that that training improves knowledge. This discrepancy could be explained by our narrower focus on literacy knowledge – in terms of content knowledge, literacy knowledge may be more challenging for teachers (Cunningham et al., 2009). Our finding that PD produces larger effects for

structural quality than process quality suggests that structural quality may be considerably more malleable than teacher-child interactions. This finding is not necessarily surprising. Classroom processes are considered to be very complicated (Downer, Maier, Howes, Hamre, & Pianta, 2012), whereas changes to provision of literacy materials is likely easier to implement.

In terms of child outcomes, we found that PD had a statistically significant medium effect on children's phonological awareness, and a small significant effect on alphabet knowledge. A small effect was also detected for receptive vocabulary, but it was not significant. Recognizing that these results are based on a small subset of trials, this review finds some tentative support for the causal link between PD with a language and emergent literacy focus, and children's outcomes in these areas. However, it should be noted that we did not find that gains in process quality predicted child outcome gains.

One explanation for the lack of mediation of child outcomes by process quality may be that the measures of process quality did not capture the interactions responsible for children's language and literacy gains. An alternate explanation may be that the level of process quality achieved in the studies was not sufficient to facilitate and predict child outcomes. Burchinal and colleagues (2010) found that the quality of teacher-child interactions was a stronger predictor of academic outcomes (including language and literacy) in classrooms in which quality was high. It could be that that level of quality in the majority of included trials was too low to predict the child gains. Another possibility may be that the types of interactions that mediate children's learning of vocabulary, phonological awareness and alphabet knowledge did not improve as much as other interactions that were measured. Finally, it should be noted that the sample of trials that had both teacher and child level outcomes was fairly small. It is possible that a larger sample of trials would yield significant results.

The analysis of the effects of PD formats mainly centered around courses and coaching as these were the most common formats of PD. Two main findings emerged from the results of the various sub-group analyses that were performed. First of all, we found that although both courses and coaching appeared to contribute significantly to process and structural quality, coaching appeared to contribute more. This was particularly evident by the fact that courses – when employed in isolation – did not produce gains of statistical significance, whereas coaching did. However, it should be noted that the intensity of coaching tended to be higher than the intensity of courses. The average coaching intensity for a PD intervention was 73.6 hours, whereas it was only 32.3 hours for courses. Although the differences in effect sizes by function of PD format could be interpreted as an indication of coaching possessing more effective underlying processes, it could also simply be that intensity plays an important role, and coaching interventions were more intense. However, because some included trials did not report the intensity of coaching and/or courses, more research is required to understand the role of intensity with these two formats.

Our second main finding with regards to format was evidence that PD interventions employing more than one format produced significantly larger effects. This was particularly true for structural quality, where the effect of using more than one format of PD was found to be nearly five times greater than single-format interventions. The meta-regression analyses also revealed significant association between the total number of formats and quality, which suggests that three or four formats of PD are preferable to just two. One explanation to this finding may be that numerous learning formats offer multiple learning opportunities, and support different kinds of learning. For example, courses may provide a general orientation about project goals, and introduce teachers to new content knowledge, whereas coaching may provide teachers with the feedback needed to fine-tune new practices, and help them problem-solve in class obstacles. The use of an experimental curriculum may provide a fixed regiment that helps teachers come into a routine of new practice.

We find support for this "more-is-more" explanation from experimental studies such as Landry et al. (2009), who compared several PD interventions with varying degrees of learning supports, and found that the most comprehensive model benefited teachers most. In addition to the number of PD formats, the total intensity of PD was also significant for process and structural quality.

In addition to the importance of a multi-faceted and high intensity approach to PD, we found tentative support for the benefit of longer PD interventions. Although this finding was only significant for structural quality, it is noteworthy that all but one of the included trials were one school-year or shorter in duration. The exception, Grace et al. (2008) stood out in that it combined several formats of PD, and the intervention period was nearly three years. This study also yielded the largest effect for process quality and second largest for structural quality. This adds support to the notion that successful PD interventions are long-term, comprehensive endeavors in which teachers are afforded multiple learning platforms.

In terms of study characteristics, we found only a few indications of systematic variation due to study characteristics. Unsurprisingly, smaller studies tended to produce larger effect sizes but only for structural quality. We found that effect sizes were larger for teachers working in preschools serving at-risk children. This is a positive finding in light of other research that demonstrates that children in poverty are more likely to attend lower quality preschool (Justice et al., 2008). We also found that randomization at the preschool level rather than the classroom or teacher level was associated with higher gains in structural quality. This could be explained by the postulation that PD is more effective when all teachers from a single institution participate together (Birman, Desimone, Porter, & Garet, 2000). We found furthermore that measures that were more aligned with the content of the PD resulted in larger process quality effects. A similar finding was uncovered in Fukkink and Lont (2007). Finally, we found evidence of a negative bias associated with studies of less methodological rigor.

Applications for Researchers

Readers of this systematic review should be cognizant of the following limitations. First of all, several PD intervention studies were excluded from the analysis because they did not report the basic data needed to estimate effect sizes and/or standard errors. Others were excluded because they only reported a child outcome. This means that many studies that otherwise potentially could have altered the results of this review were excluded. Furthermore, the resultant number of studies that measured both teacher and child level outcomes was admittedly small. With a larger subset of studies, we may have found that gains in preschool quality predicted gains in child outcomes. Another limitation is the fact that we only included peer-reviewed studies, which increases the risk of publication bias. Finally, all but one of the included trials were conducted in either Canada or the United States. Although this may simply reflect that researchers in these nations are on the forefront of controlled research on the effect of PD on preschool quality, it nevertheless necessitates caution in interpreting the results into the contexts of nations outside North America.

These limitations in combination with our findings give way to new research directions. First of all, this systematic review indicates that PD is a valid method of improving preschool quality and thereby child outcomes. Although it does not directly disclose the underlying mechanisms that facilitate teachers' learning when receiving PD, it offers indications that formats such as courses and coaching can have effects on teacher outcomes, but identifying exactly what these mechanisms are, is an essential next research step. Some researchers have made steps towards proposing characteristics of successful courses (e.g., Scott-Little et al., 2011) or coaching models (e.g., Mraz, Kissel, Algozzine, Babb, & Foxworth, 2011), but future research needs to investigate empirically the impact of these internal characteristics, and how they can be implemented with high fidelity within a PD intervention context. Research in this direction would also facilitate the

paradigm shift away from the convention method of testing formats of PD, as advocated by Sheridan et al. (2009).

More research is also needed regarding the question of which teacher-child behaviours stimulate children's learning of specific language and emergent literacy skills. Knowing which specific teacher-child interactions stimulate skills such as vocabulary and phonological awareness could be better integrated into current measures of classroom quality such that quality indicators become more predictive of child outcomes. Our review found that teacher gains were not well aligned with child gains suggesting that some aspects of process quality may be too broad to function as reliable indicators of child literacy outcomes. This could be in-line with work by Hamre, Hatfield, Pianta, and Jamil (2013), who found evidence for general and domain-specific elements of quality interactions.

Finally, although the goal of PD is ultimately to improve child outcomes, studies of the effectiveness of PD interventions ought to measure the effects on child outcomes. Since gains on current measures of preschool quality do not automatically equate to gains on child outcomes, measuring children's response to the intervention is an important necessity – both in terms of evaluating the effectiveness of the intervention, but also as a contribution to our knowledge of the threshold of teacher change that facilitates meaningful child outcome changes.

References

Ahn, S., Ames, A. J., & Myers, N. D. (2012). A review of meta-analyses in education: Methodological strengths and weaknesses. *Review of Educational Research, 82*, 436-476.

Aikens, N., & Akers, L. (2011). Background Review of Existing Literature on Coaching. Washington, DC: Mathematica Policy Research: Mathematica Policy Research.

*Al Otaiba, S., Connor, C. M., Folsom, J. S., Greulich, L., Meadows, J., & Li, Z. (2011). Assessment data-informed guidance to individualize kindergarten reading instruction: Findings from a cluster-randomized control field trial. *The Elementary School Journal, 111*(4), 535.

*Algozzine, B., Babb, J., Algozzine, K., Mraz, M., Kissel, B., Spano, S., & Foxworth, K. (2011). Classroom Effects of an Early Childhood Educator Professional Development Partnership. *NHSA Dialog, 14*(4), 246-262. doi: 10.1080/15240754.2011.613125

Birman, B. F., Desimone, L., Porter, A. C., & Garet, M. S. (2000). Designing professional development that works. *Educational leadership, 57*(8), 28-33.

Bronfenbrenner, U. (1979). *The ecology of human development: Experiments by nature and design*: Harvard university press.

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 Development Science, 12*(3), 140-153.

Burchinal, M. R., Vandergrift, N., Pianta, R., & Mashburn, A. (2010). Threshold analysis of association between child care quality and child outcomes for low-income children in pre-kindergarten programs. *Early Childhood Research Quarterly, 25*(2), 166-176. doi: http://dx.doi.org/10.1016/j.ecresq.2009.10.004

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

Buysse, V., Winton, P. J., & Rous, B. (2009). Reaching consensus on a definition of professional development for the early childhood field. *Topics in early childhood special education, 28*(4), 235-243.

*Cabell, S. Q., Justice, L. M., Piasta, S. B., Curenton, S. M., Wiggins, A., Turnbull, K. P., & Petscher, Y. (2011). The impact of teacher responsivity education on preschoolers' language and literacy skills. *American Journal of Speech-Language Pathology, 20*(4), 315-330.

Card, N. A. (2011). *Applied meta-analysis for social science research*. New York: Guilford Press.

*Cassidy, D. I., Buell, M. I., Pugh-Hoese, S., & Russell, S. (1995). The effect of education on child care teachers' beliefs and classroom quality: Year one evaluation of the TEACH early childhood associate degree scholarship program. *Early Childhood Research Quarterly*, 10(2), 171-183. doi: http://dx.doi.org/10.1016/0885-2006(95)90002-0

Castro, D. (2005). *Early language and literacy classroom observation (ELLCO) addendum for English language learners*. Chapel Hill, NC: University of North Carolina, FPG Child Development Institute.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (Vol. 2nd Edition). Hillsdale: Lawrence Erlbaum.

Cooper, H. (1998). *Synthesizing research: A guide for literature reviews* (Vol. 2). Sage.

Cooper, H., & Hedges, L. V. (2009). Research synthesis as a scientific process. *The handbook of research synthesis and meta-analysis, 2*, 3-17.

Cunningham, A. E., Zibulsky, J., & Callahan, M. D. (2009). Starting small: Building preschool teacher knowledge that supports early literacy development. *Reading and Writing, 22*(4), 487-510.

da Costa, B. R., Nüesch, E., Rutjes, A. W., Johnston, B. C., Reichenbach, S., Trelle, S., . . . Jüni, P. (2013). Combining follow-up and change data is valid in meta-analyses of continuous outcomes: a meta-epidemiological study. *Journal of Clinical Epidemiology, 66*(8), 847-855. doi: http://dx.doi.org/10.1016/j.jclinepi.2013.03.009

DerSimonian, R., & Laird, N. (1986). Meta-analysis in clinical trials. *Controlled Clinical Trials, 7*(3), 177-188. doi: http://dx.doi.org/10.1016/0197-2456(86)90046-2

Dickinson, D. K. (2006). Toward a Toolkit Approach to Describing Classroom Quality. *Early Education and Development, 17*(1), 177-202. doi: 10.1207/s15566935eed1701_8

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

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

Downer, J. T., Maier, M., Howes, C., Hamre, B., & Pianta, R. (2012). Implications of information processing theory for professional development of early educators *Effective early childhood professional development: Improving teacher practice and child outcomes* (pp. 131-155). Baltimore, MD: Brookes.

Dowsett, C. J., Huston, A. C., Imes, A. E., & Gennetian, L. (2008). Structural and process features in three types of child care for children from high and low income families. *Early Childhood Research Quarterly, 23*(1), 69-93. doi: http://dx.doi.org/10.1016/j.ecresq.2007.06.003

Dunn, L. M., & Dunn, L. M. (1997). *PPVT-III: Peabody picture vocabulary test* (3rd ed. ed.). Circle Pines, MN: American Guidance Service.

Early, D. M., Bryant, D. M., Pianta, R. C., Clifford, R. M., Burchinal, M. R., Ritchie, S., . . . Barbarin, O. (2006). Are teachers' education, major, and credentials related to classroom quality and children's academic gains in pre-kindergarten? *Early Childhood Research Quarterly, 21*(2), 174-195. doi: http://dx.doi.org/10.1016/j.ecresq.2006.04.004

*Flowers, H., Girolametto, L., Weitzman, E., & Greenberg, J. (2007). Promoting early literacy skills: Effects of in-service education for early childhood educators. *Revue canadienne d'orthophonie et d'audiologie-Vol*, 31(1).

Fukkink, R. G., & Lont, A. (2007). Does training matter? A meta-analysis and review of caregiver training studies. *Early Childhood Research Quarterly, 22*(3), 294-311.

*Fukkink, R. G., & Tavecchio, L. W. (2010). Effects of video interaction guidance on early childhood teachers. *Teaching and Teacher Education*, 26(8), 1652-1659.

Gersten, R., Fuchs, L. S., Compton, D., Coyne, M., Greenwood, C., & Innocenti, M. S. (2005). Quality indicators for group experimental and quasi-experimental research in special education. *Exceptional children, 71*(2), 149-164.

*Girolametto, L., Weitzman, E., & Greenberg, J. (2003). Training day care staff to facilitate children's language. *American Journal of Speech-Language Pathology, 12*(3), 299-311.

*Girolametto, L., Weitzman, E., & Greenberg, J. (2012). Facilitating emergent literacy: Efficacy of a model that partners speech-language pathologists and educators. *American Journal of Speech-Language Pathology, 21*(1), 47-63.

*Girolametto, L., Weitzman, E., Lefebvre, P., & Greenberg, J. (2007). The Effects of In-Service Education to Promote Emergent Literacy in Child Care Centers: A Feasibility Study. *Language, Speech, and Hearing Services in Schools, 38*(1), 72-83. doi: 10.1044/0161-1461(2007/007)

*Grace, C., Bordelon, D., Cooper, P., Kazelskis, R., Reeves, C., & Thames, D. G. (2008). Impact of Professional Development on the Literacy Environments of Preschool Classrooms. *Journal of Research in Childhood Education, 23*(1), 52-81. doi: 10.1080/02568540809594646

Guo, Y., Justice, L. M., Kaderavek, J. N., & McGinty, A. (2012). The literacy environment of preschool classrooms: contributions to children's emergent literacy growth. *Journal of Research in Reading, 35*(3), 308-327. doi: 10.1111/j.1467-9817.2010.01467.x

Hamre, B. K., Hatfield, B., Pianta, R., & Jamil, F. (2013). Evidence for General and Domain-Specific Elements of Teacher–Child Interactions: Associations With Preschool Children's Development. *Child Development, 85*(3), 1257–1274.

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.

Higgins, J. P., & Green, S. (2008). *Cochrane handbook for systematic reviews of interventions* (Vol. 5): Wiley Online Library.

*Hindman, A. H., & Wasik, B. A. (2011). Exploring Head Start Teachers' Early Language and Literacy Knowledge: Lessons from the ExCELL Professional Development Intervention. *NHSA Dialog, 14*(4), 293-315.

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. doi: http://dx.doi.org/10.1016/j.ecresq.2007.05.002

Justice, L. M., Mashburn, A. J., Hamre, B. K., & Pianta, R. C. (2008). Quality of language and literacy instruction in preschool classrooms serving at-risk pupils. *Early Childhood Research Quarterly, 23*(1), 51-68. doi: http://dx.doi.org/10.1016/j.ecresq.2007.09.004

Keys, T. D., Farkas, G., Burchinal, M. R., Duncan, G. J., Vandell, D. L., Li, W., . . . Howes, C. (2013). Preschool Center Quality and School Readiness: Quality Effects and Variation by Demographic and Child Characteristics. *Child Development, 84*(4), 1171-1190. doi: 10.1111/cdev.12048

*Koh, S., & Neuman, S. B. (2009). The Impact of Professional Development in Family Child Care: A Practice-Based Approach. *Early Education and Development, 20*(3), 537-562. doi: 10.1080/10409280902908841

Landry, S. H., Anthony, J. L., Swank, P. R., & Monseque-Bailey, P. (2009). Effectiveness of comprehensive professional development for teachers of at-risk preschoolers. *Journal of Educational Psychology, 101*(2), 448-465. doi: 10.1037/a0013842

Lonigan, C. J., Farver, J. M., Phillips, B. M., & Clancy-Menchetti, J. (2011). Promoting the development of preschool children's emergent literacy skills: A randomized evaluation of a literacy-focused curriculum and two professional development models. *Reading and Writing, 24*(3), 305-337.

Markussen-Brown, J., Juhl, C. B., Piasta, S., Bleses, D., & Højen, A. (2014). *The effects of professional development (PD) on early childhood educators measured in terms of knowledge, process quality, and structural quality, and children's emergent literacy and language skills at 3-6 years of age: A meta-analysis*. Prospero database. www.crd.york.ac.uk/PROSPERO/display_record.asp?ID=CRD42014009361

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. doi: 10.1111/j.1467-8624.2008.01154.x

*McCollum, J. A., Hemmeter, M. L., & Hsieh, W.-Y. (2013). Coaching teachers for emergent literacy instruction using performance-based feedback. *Topics in early childhood special education*, *33*(1), 28-37.

*McCutchen, D., Abbott, R. D., Green, L. B., Beretvas, S. N., Cox, S., Potter, N. S., . . . Gray, A. L. (2002). Beginning Literacy: Links Among Teacher Knowledge, Teacher Practice, and Student Learning. *Journal of Learning Disabilities*, 35(1), 69-86. doi: 10.1177/002221940203500106

Moats, L. C. (1999). *Teaching reading is rocket science*. Washington D.C.: American Federation of teachers

Mraz, M., Kissel, B., Algozzine, B., Babb, J., & Foxworth, K. (2011). A collaborative professional development initiative supporting early literacy coaches. *NHSA Dialog*, *14*(3), 174-184.

National Early Literacy Panel. (2008). *Developing early literacy: Report of the National Early Literacy Panel: A scientific synthesis of early literacy development and implications for intervention*. Washington, D.C.: National Institute for Literacy.

Neuman, S. B. (1999). Books Make a Difference: A Study of Access to Literacy. *Reading Research Quarterly*, *34*(3), 286-311. doi: 10.1598/RRQ.34.3.3

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

Neuman, S. B., & Roskos, K. (1992). Literacy Objects as Cultural Tools: Effects on Children's Literacy Behaviors in Play. *Reading Research Quarterly, 27*(3), 203-225. doi: 10.2307/747792

Neuman, S. B., & Roskos, K. (1993). Access to Print for Children of Poverty: Differential Effects of Adult Mediation and Literacy-Enriched Play Settings on Environmental and Functional Print Tasks. *American Educational Research Journal, 30*(1), 95-122. doi: 10.2307/1163191

*Neuman, S. B., & Wright, T. S. (2010). Promoting language and literacy development for early childhood educators: A mixed-methods study of coursework and coaching. *The Elementary School Journal, 111*(1), 63-86.

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. doi: 10.1207/s1532480xads0903_2

Pianta, R. C., La Paro, K. M., & Hamre, B. K. (2008). *Classroom Assessment Scoring System Manual, Pre-K*. Baltimore, MD: Brookes.

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

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

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

Scammacca, N., Roberts, G., & Stuebing, K. K. (2013). Meta-Analysis With Complex Research Designs Dealing With Dependence From Multiple Measures and Multiple Group Comparisons. *Review of Educational Research*, 0034654313500826.

Scarr, S., Eisenberg, M., & Deater-Deckard, K. (1994). Measurement of quality in child care centers. *Early Childhood Research Quarterly, 9*(2), 131-151. doi: http://dx.doi.org/10.1016/0885-2006(94)90002-7

Schachter, R. E. (2015). An Analytic Study of the Professional Development Research in Early Childhood Education. *Early Education and Development*, 1-29. doi: 10.1080/10409289.2015.1009335

*Schwanenflugel, P. J., Hamilton, C. E., Neuharth-Pritchett, S., Restrepo, M. A., Bradley, B. A., & Webb, M.-Y. (2010). PAVEd for Success: An evaluation of a comprehensive preliteracy program for four-year-old children. *Journal of Literacy Research*, 42(3), 227-275.

Scott-Little, C., La Paro, K. M., Thomason, A. C., Pianta, R. C., Hamre, B., Downer, J., . . . Howes, C. (2011). Implementation of a Course Focused on Language and Literacy Within Teacher–Child Interactions: Instructor and Student Perspectives Across Three Institutions of Higher Education. *Journal of Early Childhood Teacher Education, 32*(3), 200-224.

Sheridan, S. M., Edwards, C. P., Marvin, C. A., & Knoche, L. L. (2009). Professional Development in Early Childhood Programs: Process Issues and Research Needs. *Early Education and Development, 20*(3), 377-401. doi: 10.1080/10409280802582795

Smith, M. W., & Dickinson, D. K. (2002). *Early Language & Literacy Classroom Observation (ELLCO)*. Baltimore, MD: Brookes.

Vygotsky, L. (1978). *Mind in society: The development of higher mental process*: Cambridge, MA: Harvard University Press.

*Wasik, B. A., Bond, M. A., & Hindman, A. (2006). The effects of a language and literacy intervention on Head Start children and teachers. *Journal of Educational Psychology, 98*(1), 63-74. doi: 10.1037/0022-0663.98.1.63

*Wasik, B. A., & Hindman, A. H. (2011). Improving vocabulary and pre-literacy skills of at-risk preschoolers through teacher professional development. *Journal of Educational Psychology*, *103*(2), 455.

Whitehurst, G. J., & Lonigan, C. J. (1998). Child Development and Emergent Literacy. *Child Development, 69*(3), 848-872. doi: 10.1111/j.1467-8624.1998.tb06247.x

Zaslow, M., Tout, K., Halle, T., Whittaker, J. V., & Lavelle, B. (2010). Toward the Identification of Features of Effective Professional Development for Early Childhood Educators. Literature Review *Office of Planning, Evaluation and Policy Development, US Department of Education*. Washington, D.C.: U.S. Department of Education.

Appendix A

Table 1A

Search terms of the systematic map

Language and communication

Comprehension, grammar, grammatical, language (only in title or keywords), lexic*, linguistic, literacy, morphology, phonetic, phonolog* , pragmatics , pre-literacy, print, reading, receptive, semantics, speech, syntax, syntactic*, verbal, vocabulary, writing, oral language, communicative development, language delay, language development, language acquisition, language disorder, language impairment*, late talk*, late language, delayed language, communication disorder*, communicative disorder*.

Developmental period of interest

Baby, early childhood, young children, infan*, preschool*, toddler, kindergarten, pre-kindergarten, pre-k, daycare, day care, day-care, nursery, nurseries, 1-year-old*, 2-year-old*, 3-year-old*, 4-year-old*, 5-year-old*, 6-year-old*, one-year-old*, two-year-old*, three-year-old*, four-year-old*, five-year-old*, six-year-old*, “age 1”, “age 2”, “age 3”, “age 4”, “age 5”, “age 6”.

Effect factors

effect*, environment*, predict*, rct, randomized control*, risk, screening, association, impact*, efficacy, treatment, outcome, caus*, influenc*, intervention, role.

Environmental factors

Book reading , chaos, chaotic, child-directed speech, contingency, conversation, dialogic reading, disadvantaged, family, father*, paternal, home literacy, infant-directed speech,

linguistic input, maternal input, paternal input, lexical input, interaction*, joint attention, language model, mother*, maternal, parent*, peers, poverty, responsivity, SES, Socio-econom*, siblings, stress, television, TV, Bilingual, Dual language, Ethnic, Language minority, Multilingual, coaching, curricul*, educator*, pediatric, head start, headstart, head-start, instruction, intervention, professional development, program*, teach*, training.

Appendix B

Table 1B

Characteristics of Included Trials

Trial, year	*N* teachers exp/cont	*N* children exp/cont	Unique Intervention Components	Intensity of coaching and workshops**	Test period for PD	Teacher outcomes			Child outcomes		
						Proc	Struc	Know	R.Voc.	Phon.	Let.
Al Otaiba et al. (2011)	23/21	305/251	Coaching and the usage of software for Response to Intervention individualized instruction	16 hours of coaching	Nov-Feb	x				x	x
Algozzine et al. (2011)	36/11		Workshops, coaching and usage of a special curriculum	unclear	Fall-Spring	x	x				
Buysse, Castro & Peisner-Feinberg (2010)	26/29	92/101	Workshops, coaching, and community of practice meetings	18 hours of workshops; 4 coaching sessions of unknown duration	Fall-Spring	x	x		x	x	x
Cassidy, Buell, Pugh-Hoese & Russel (1995)	19/15		College coursework	4 courses on average	6-9 months		x				
Dickinson & Caswell (2007)	30/40		A college course, performance-based assignments, coaching, and supervisor support.	45h course; coaching unclear	Oct-April/May	x	x				
Domitrovich et al. (2009)	43/44		Workshops, coaching and usage of a special curriculum	24h of workshops; 120 hours of coaching	12 months	x					

Flowers, Girolametto, Weitzman & Greenberg (2007)	8/8		Workshops and video-observation coaching	20h of workshops; 3h of coaching	4 months	x					
Fukkink & Tavecchio (2010)	52/43		Video-observation coaching	3.83 sessions	Unclear	x					
Girolametto, Weitzman & Greenberg (2012)	10/10	39/37	Workshops and combined coaching and video-observation coaching	18 hours of workshops; 3 sessions of coaching	Unclear	x				x	x
Girolametto, Weitzman & Greenberg (2003)	8/8		Workshops and video-observation coaching	20h of workshops; 3h of coaching	4 months	x					
Girolametto & Weitzman (2007)	8/7		Workshop	6h	1 month	x					
Grace et al. (2008)	20/15		Workshops, coaching, usage of a special curriculum, and funds for materials of own choice.	30 workshops over three years; 450h of coaching.	3 years	x	x				
Koh & Neuman (2009)--1	33/28*		College course (home-based care)	45h course	Fall-Spring (knowledge tested after 15 weeks)	x	x	x			
Koh & Neuman (2009)--2	40/27*		College course and coaching (home-based care)	45h course; 32h coaching	Fall-Spring (knowledge tested after 15 weeks)	x	x	x			
McCollum, Hemmeter & Ksieh (2011)	7/5		Workshops and coaching.	10h of workshops and 15 coaching sessions	Sept-May	x	x				

McCutchen et al. (2002)	24/13		Workshops and coaching.	102h of workshops and unspecified coaching.	Unclear			x			
Neuman & Cunningham (009)--1a	53/36*		College course	45h course	Fall-Spring (knowledge tested after 15 weeks)	x	x	x			
Neuman & Cunningham (009)--1b	53/35*		College course and coaching	45h course; 32h coaching	Fall-Spring (knowledge tested after 15 weeks)	x	x	x			
Neuman & Cunningham (009)--2a	32/23*		College course (home-based care)	45h course	Fall-Spring (knowledge tested after 15 weeks)	x	x	x			
Neuman & Cunningham (009)--2b	34/22*		College course and coaching (home-based care)	45h course; 32h coaching	Fall-Spring (knowledge tested after 15 weeks)	x	x	x			
Neuman & Wright (2010)--1	58/16*		Coursework	30h	Fall-Spring	x	x	x			
Neuman & Wright (2010)--2	58/16*		Coaching	30h	Fall-Spring	x	x	x			
Piasta & al. (2012); Cabell et al. (2011)	17/21	168/162	Workshops and video-observation coaching.	17h of workshops; 12 videos.	24 weeks	x			x		x
Powell, Diamond, Burchinal & Koehler (2010)	42/31	310/258	Workshops and coaching for some; workshops and video-observation coaching for others.	16h of workshops; 7 coaching sessions for both types	Sept/Oct - Dec/Jan	x			x	x	x

Schwanenflugel et al. (2010)	31/6		Workshops and coaching.	14h of workshops; 5 sessions of coaching	one school year	x	x			
Wasik, Bond & Hindman (2006)	10/6	139/68	Workshops, coaching and usage of a special curriculum	12h of workshops; 12h of coaching.	Sept-June	x		x		
Wasik & Hindman (2011)	19/11	358/183	Workshops and coaching.	42h of workshops; 10h of coaching.	Sept-May	x	x	x	x	x

Note. * indicates control samples that were split; ** intensity was calculated as dosage x number of occurences.

Paper 3

Training Awareness of Scaffolding Usage in Preschool Teachers

Justin Markussen-Brown and Dorthe Bleses

The University of Southern Denmark

Werner Vach

Freiburg University

Laura Justice

The Ohio State University

Author Note

Justin Markussen-Brown, Institute for Language and Communication, University of Southern Denmark; Dorthe Bleses, Institute for Language and Communication, University of Southern Denmark; Werner Vach, Center for Medical Biometry and Medical Informatics, Freiburg University, The Ohio State University; Laura M. Justice, Crane Center for Early Childhood Research and Policy, The Ohio State University.

Correspondence to concerning this article should be addressed to Justin Markussen-Brown, Institute for Language and Communication, University of Southern Denmark, Campusvej 55, Odense 5230, Denmark. E-mail: justinmb@sdu.dk.

Abstract

The scaffolding of children's language and emergent literacy skills in preschool is an important pedagogical skill, but some research indicates that preschool teachers may be unfamiliar with scaffolding, or be unaware of the extent to which they use scaffolding. In the current study, we aimed to increase teachers' self-awareness of skill performance using a self-coding task. Over three days of professional development (PD), teachers were taught to use six scaffolds that supported children's learning goals. Then teachers were taught to code their usage of these strategies from video. High coding accuracy was assumed to represent awareness of skill performance. The results of the self-coding task revealed that most teachers coded inaccurately, and in particular overrated their usage of high support strategies, which were in fact rarely used. Low support strategies were also coded incorrectly, but with teachers under and overrating their usage. However, in instances in which teachers coded accurately, it was often in recognition that no strategies were used. Furthermore, some scaffolding strategies appeared more difficult to code then. The PD did not appear sufficient to prepare teachers for the self-coding task. Future research is needed to determine if self-coding is a viable tool for learning to scaffold.

Keywords: scaffolding, professional development, early childhood education, social cognitive theory.

Training Awareness of Scaffolding Usage in Preschool Teachers

Awareness of one's skill performance helps promote learning during professional development (PD). When individuals are not cognizant of their skill strengths and weaknesses, they are less likely to orient towards learning goals, and self-regulate their own learning (Bandura & Cervone, 1983; Schunk, 1989). Reviewing video is widely viewed as one effective means of gaining awareness of the behaviour of oneself and others (Fuller & Manning, 1973). Research has demonstrated the positive benefits of video self-analysis in a number of fields such as nursing (Yoo, Son, Kim, & Park, 2009) and customer service (Brown, Malott, Dillon, & Keeps, 1980). Increasingly, video has been used as an effective tool for school teachers to self-evaluate and increase awareness of their own practice (see Tripp & Rich, 2012 for a review). One area that may benefit further from increased usage of self-analysis with regards to increasing awareness is the field of early childhood education.

Research has demonstrated that some preschool teachers are poorly *calibrated* with regards to their disciplinary knowledge of early literacy (Cunningham, Zibulsky, & Callahan, 2009), and their literacy instructional practices (Polk, 2013). Calibration refers to an individual's self-awareness of own skill performance or knowledge level (see Fischhoff, Slovic, & Lichtenstein, 1977; Hacker, Bol, & Keener, 2008). However, other research has demonstrated that when preschool teachers discover that their performance does not align with their expectations, they increase usage of pedagogical behaviours such as open-ended questions (Lynes, 2012), and praising (Wright, 1998). Thus it appears that video self-assessment can have a calibrating effect on teachers' awareness of skill performance, which may have useful applications for PD.

The potential effects of video self-assessment on teachers' calibration of pedagogical skill performance are also supported by social cognitive theory (Bandura, 1977, 1986; Schunk, 1989), which postulates that accurate self-awareness can motivate behavioural change. When individuals

discover that they perform lower than expected, they can experience a sudden focusing of attention on the cause of the disturbance, a state referred to as cognitive dissonance by Festinger (1962). The process can lead to a recalibration of one's awareness, which motivates behavioural changes that can restore beliefs of self-efficacy (Bandura & Cervone, 1983). Within in this framework, preschools teachers, when they are made aware of the fact that they do not perform as they thought, might strive more to adjust their practice to regain the feeling that they are effective teachers.

Scaffolding Children's Language and Literacy Skills

The language and literacy environment of preschools may be one area that could benefit from efforts to increase self-awareness of teacher skill performance. Research suggests that teacher-child interactions are the main locus of children's learning of language and literacy (Mashburn et al., 2008), but troublingly, other research has found low levels of instructional quality in the language and literacy practices of teachers serving children at risk for academic failure (e.g. Justice, Mashburn, Hamre, & Pianta, 2008). One interactional skill-set in particular, *scaffolding*, demonstrates considerable benefit for preschool-aged children's acquisition of language and literacy (Dieterich, Assel, Swank, Smith, & Landry, 2006), and therefore exploring avenues of increasing preschool teachers' usage of scaffolding may be advisable.

Scaffolding is an instructional technique whereby a proficient individual assists a novice in acquiring new knowledge or skills, and is often described in relation to supporting children's development of language and emergent literacy skills (Van de Pol, Volman, & Beishuizen, 2010). The metaphoric term "scaffolding" (referring to how a scaffold helps in the construction of a building) stems from work by Wood, Bruner, and Ross (1976), who observed and categorized natural patterns of how mothers scaffolded the efforts of their children to complete a novel task. The theoretic basis of scaffolding, however, stems from Lev Vygotsky's proposed zone of proximal development (1978). Vygotsky posited that a child's learning is to a large degree dependent on the

skilled adult's ability to identify and guide the child from one developmental level to the next. Similarly, Wood et al.'s study demonstrated that a child's task-solving is not merely tied to the age-appropriateness of the given task, but also to the amount and quality of individualized support that the child receives during the task-solving process.

Although definitions of scaffolding tend to vary to some degree, all descriptions of scaffolding involve (a) contingent responding and (b) fading of efforts by the teacher (Van de Pol et al., 2010). Contingent responding refers to responding to the child with an appropriate amount of support for the child, whereas fading refers to the gradual reduction in support as the child masters the skill. These two characteristics were clearly operationalized in a scaffolding model utilized by Pentimonti and Justice (2010). In their study, preschool teachers were taught to support four language and emergent literacy skills known to be precursors of conventional literacy (see National Early Literacy Panel, 2008) using six scaffolding strategies drawn from recommendations in O'Connor, Notari-Syverson, and Vadasy (2005). The strategies provided either high or low amounts of support, and are described with examples in Table 1.

In the study by Pentimonti and Justice (2010), the high support scaffolding strategies (*elicitation*, *co-participation*, and *reducing choices*) assisted children requiring help in answering questions related to an emergent literacy goal. For example, if a child struggled to think of a word that rhymed with "cat," the teacher could reduce the child's choices by presenting a correct answer and an incorrect alternative (e.g., "bat" versus "door"). Once the child demonstrated a growing understanding of the learning goal in question, the teacher could reduce support by using a low support strategy (*generalization*, *prediction*, and *reasoning*), which encouraged the child to extend the learning target in a decontextualized way. For example, if a child knew which letter his or her name started with, the teacher could use the generalization strategy, and ask the child to think of

other names that start with the same letter. In this way, the six strategies gave teachers concrete tools with which they could implement contingent responding and fading of efforts.

Table 1

The High and Low Support Scaffolding Strategies Used in Pentimonti and Justice (2010) and Adapted to the Current Study

High Support	Description	Example
Elicitation	When needed, the teacher provides the child with the correct answer to a question and then repeats the question.	Jesper, your name starts with the letter *J*. What letter does your name start with?
Co-participation	When in need of support, the teacher and the child complete a task together.	Freya, will you please lend me your finger and help me point to the all the words on the page?
Reducing choices	If needed, the teacher provides the child with the correct answer and one or more alternatives.	Bjørn, did the monster get sick at the start of the book or the end?
Low support		
Generalization	The teacher asks the child to generalize a newly mastered task to a broader context.	That's right, Jesper. Your name starts with the letter *J*. Do you know other children whose names also start with *J*?
Prediction	The teacher asks the child to predict the outcome of a certain event if a certain condition were changed.	Freya helped me read from left to right! But imagine there were no words on this page. Where would I start reading then?
Reasoning	The teacher asks the child to explain his or her thinking, or why something is like it is.	Bjørn, why do you think the monster got sick in the first place?

Preschool Teachers' Use of Scaffolding

Although scaffolding is considered to be beneficial to children's learning of language and emergent literacy, there is some evidence that it may be a technique unfamiliar to preschool teachers. McGee and Ukrainetz (2009) worked with preschool teachers who experienced difficulty teaching phonological awareness to children at-risk for academic failure due to poverty. Their observations of the teachers' instructional practices revealed, however, little usage of scaffolding when instructing on phonological awareness. The researchers subsequently trained teachers to

scaffold on a fading scale of intensity (intense, moderate, and minimal), and the resulting improvements in teacher practice transferred to the children's performance on phonological tasks. This study, if generalizable to other preschools, could indicate that some preschool teachers have little disciplinary knowledge of scaffolding.

Other research indicates that scaffolding may also be a challenging skillset to acquire. Pentimonti and Justice (2010) studied a small subgroup of teachers, who were participating in a larger early literacy intervention study (see Justice et al., 2010), with regards to their usage and awareness of scaffolding during shared-book reading. Specifically, teachers were trained over two days of workshops to use the six scaffolding strategies displayed in Table 1 as part of a story-book reading intervention for Head Start classrooms. The researchers found that teachers used high support strategies far less frequently than low support strategies, which was surprising since the preschools in the study served children at-risk for academic failure, who might be in need of extra learning support. In addition to this finding, teachers also demonstrated an inaccurate awareness of which and how many of the scaffolding strategies they used. When asked to recollect how many of each strategy they used following book-reading sessions, teachers significantly overestimated their usage of the high support strategies when in fact they used them rarely. Based on these results, Pentimonti and Justice suggested that teachers might benefit from PD opportunities (beyond the original two days of workshops) that could increase teachers' awareness of scaffolding.

Using Professional Development to Calibrate Teachers' Awareness of Skill Performance

One of the goals of PD is to enhance teachers' practice, but how PD fosters the acquisition of pedagogical skillsets such as scaffolding is difficult to discern since the moderating and mediating factors of PD processes go largely unobserved in research (Sheridan, Edwards, Marvin, & Knoche, 2009). This lack of insight into the inner workings of how teachers learn from PD experiences seems to align with the mixed results from research investigating the effects of PD

interventions with a language and literacy focus. For example, Grace et al. (2008) used the Early Language and Literacy Observation (ELLCO; Smith & Dickinson, 2002) to measure the effects of a three-year PD program consisting of workshops and coaching. The researchers found that the PD had large effects on process quality in terms of teachers' language and literacy practices, as well as structural quality in terms of teachers' provision of literacy materials for children's use. However, similar studies that also measured the effects of PD using the ELLCO found only benefits for structural quality (Neuman & Wright, 2010), while still other studies had no or negligibly effects on process and structural qualities (e.g., Clancy-Menchetti, 2006; Lonigan, Farver, Phillips, & Clancy-Menchetti, 2011). In light of these findings, it is unclear how effective common PD approaches can be expected to be with regards to increasing teachers' usage of interactional skillsets such as scaffolding.

One possible explanation as to why common approaches of PD do not result in more systematic changes in teacher-child interactions may be that some PD formats do not adequately calibrate teachers' self-awareness of PD learning targets. Often, PD approaches to adult learning depend on the ability of external actors such as instructors or coaches to infuse teachers with pertinent knowledge that should then be internalized and transferred to classroom practice over time (Sheridan et al., 2009). However, recent research by Hamre et al. (2012) has tested the mediating role of teacher knowledge on changes in classroom practices following a college course with a language and literacy focus, and found that the amount of knowledge preschool teachers possessed did not mediate improvements in classroom practice. Interestingly, the researchers found instead that teachers' ability to recognize effective practices from video partially mediated changes in classroom practices. Their study thus suggests that teachers' ability to differentiate between teacher-child interactions of high and low quality may be a skill that mediates changes in practice.

Tentative findings by Hamre et al. (2012) seem to support experimental studies that have investigated the effects that teacher self-assessment tasks have on awareness of skill performance, as these studies also suggested a link between self-awareness and improved practice. For example, Wright (1998) taught preschool teachers to code their usage of specific and general praise in short video-recorded segments. Wright found that when teachers were given the opportunity to view and code their videos in privacy for occurrences of praising, teachers increased their usage of praise in future interactions with children. Similarly, Lynes (2012) implemented a coaching model in which teachers were also taught to code their usage of language supporting strategies such as open-questions and expansions. Results indicated that the self-evaluation helped teachers to increase and generalize their usage of strategies to other contexts, especially when combined with coaching. In another study, preschool teachers used self-coding tasks to improve their vocabulary instruction in connection with storybook reading, and expressed satisfaction from the professional experience (Blamey, Beauchat, & Sweetman, 2012).

Self-Coding as a Means of Calibration

Using self-coding as method of self-evaluation is a promising application of the video-reflection methodology in teacher education. While many approaches to using video for teacher reflection exist (see Tripp & Rich, 2012), viewing videos without guidance or explicit instructions can result in more superficial reflections (Calandra, Gurvitch, & Lund, 2008). Self-coding potentially overcomes this challenge by aiding preschool teachers with the task of orienting towards the specific behaviours that are targeted during PD. In this way, self-coding can be used as a calibration tool, and simultaneously offer teachers a means of PD that can empower teachers to self-regulate their own learning. However, the benefits of self-coding are also dependent on the number and complexity of behaviours being coded, and the ability of teachers to code accurately (Prusak, Dye, Graham, & Graser, 2010). Prior research has found that teachers can be trained to code

accurately after relatively short trainings (Blamey et al., 2012; Prusak et al., 2010), but it is unclear whether a complex skillset such as scaffolding is more difficult to code accurately.

The Current Study

The primary goal of this study was to determine if a three day course could teacher preschool teachers to code accurately their use of scaffolding. More specifically, teachers were taught to use the six scaffolding strategies employed by Pentimonti and Justice (2010), and to code them from video-recorded story-book reading sessions. We assumed that if teachers coded accurately then they were calibrated with regards to their scaffolding skill performance.

As a secondary goal, we investigated associations between coding scores for the six strategies. This allowed us to compare whether some strategies were harder to code than others, or if coding patterns were general for all strategies. This goal was motivated by the finding of Pentimonti and Justice (2010) that teachers disproportionately used scaffolding strategies. Furthermore, other research has suggested that teachers direct less attention to skills they are less familiar (Powell, Steed, & Diamond, 2009), and more easily increase their usage of strategies they already use (Lynes, 2012).

As a final goal, we investigated whether teachers' coding accuracy systematically changed when coding the video of a colleague. Some researchers have found that participating teachers can feel self-conscious about being video-recorded (Sherin & Han, 2004). Other researchers have found that teachers tend to be less critical when viewing the practice of their colleagues in order to avoid conflict (Ball, 1995).

Method

Participants

The analysis in this study was conducted on coding sheets submitted by 73 teachers from 21 preschool centers in Denmark. Teacher background is presented in Table 2.

Table 2

Teacher characteristics

Age of teachers (n=62)		
	n	%
25-35	14	23
36-45	21	34
46-55	17	27
Over 55	10	16
Years of work experience (n=65)		
	n	%
under 5	12	18
6-10	11	17
11-15	17	26
16-20	7	11
over 20	18	28

Most participating teachers were female (58 women, 4 men, 11 gender unknown). Sixty of the 73 participants held a 3½ year college degree in early childhood education, which is a state-regulated post-secondary education that combines college level coursework with several pre-service internships, and is generally compulsory for being the lead teacher in a preschool classroom. Four other participants held a different post-secondary degree, four had no post-secondary education at all, and one had a shorter pedagogical education (four more did not report their educational background). Moreover, 23 of the 73 teachers had also participated in a four-day government sponsored intensive course on children's language and literacy development in 2011.

Procedure

Recruitment. The teachers in this study were members of an experimental condition receiving an expanded package of PD as part of a larger, multiple-group, randomized controlled trial that occurred in 2012-2013 (see Bleses et al., 2014). Entire school districts were recruited to the overarching project. Preschools from each district were randomly assigned to a condition.

Professional development. Teachers participating in the current investigation attended three days of workshops prior to conducting the self-coding task. Each workshop day was 6 hours in length, which included time for breaks and lunch. The first two workshop days occurred back-to-back. The first day focused on introducing teachers to the shared-book reading intervention and increasing teachers' knowledge of four key emergent literacy skills (vocabulary, phonological awareness, print concepts, and narrative competence). The second day focused on teaching participants to use the six scaffolding strategies in a shared-book reading context. During the workshops, teachers viewed multiple video examples, engaged in role-play, and brainstormed their own examples for each strategy. The third workshop day occurred one to two weeks after the first two days of PD. During this final workshop, teachers refreshed the six scaffolding strategies, and were given many opportunities to practice coding in preparation for conducting the self-coding task. Furthermore, teachers were informed that their videos would be master-coded by the project staff in order to learn more about adult-learning processes. Instructors emphasized that teachers' participation was a valuable contribution to science.

The self-coding task. Teachers' self-evaluation was operationalized using a self-coding task that was conducted in teachers' preschools. Using video-cameras provided by the project team, each teacher video-recorded a shared-book reading session with a group of four to six children. With few exceptions, teachers read the same storybook, and they were furthermore supplied with soft-scripted examples of the strategies, which they could use while reading to the children.

After video-recording a reading session, each teacher coded the first and last five minutes of the session resulting in a 10 minute total. As reading sessions varied in duration from approximately 20 to 30 minutes, this method minimized the time burden for teachers, and increased the comparability of videos across teachers. Teachers used a simple coding scheme to count and total their usage of each strategy. Besides coding their own videos, teachers were also asked to code the

video of a preschool colleague. Upon completion of the two coding tasks, teacher-pairs discussed what they observed in the videos using a list of discussion questions provided by the research team. Finally, teachers imputed their results into an online database, and returned the memory cards from the video-cameras so that their recordings could be master-coded.

Measures. A teacher background questionnaire was filled out by teachers prior to participating in the project. To measure teachers' usage of the six scaffolding strategies, we used a coding scheme identical to the one used by the teachers during the PD training and the coding tasks. Videos were master-coded in a coding laboratory by a single research staff member who was trained by the first author. The first author also determined reliability by double-coding 20 randomly selected master-coded videos (27% of videos in total). Agreement was achieved for each of the six scaffolding strategies if the first author observed the exact same number of strategies as the master-coder. Agreement between the first author and the master-coder was determined to be 98% for elicitation, 88% for co-participation, 88% for reducing choices, 93% for generalization, 83% for prediction, and 88% for reasoning. Thus the resulting reliability was 89%, which we considered satisfactory.

Analytic strategy. To explore the extent to which teachers correctly coded their usage of each scaffolding strategy, we began by visually inspecting scatterplots to see how well teacher and master codes aligned for each strategy. Next we tested whether teachers' mean scores were different from the master-coder's mean scores using the Wilcoxon signed-rank test. This was followed by estimating correlations between the teachers' and master-coder's scores. Finally, we calculated the mean absolute difference between the teachers' scores and those of the master-coder, which indicated the mean number of strategies by which the codes of the teacher deviated from the master score.

Our second research question asked whether there were associations between teachers' scores for each scaffolding strategy. We approached this question first by exploring the extent to which teachers' coding errors were strategy-specific or stable across strategies. This was done by grouping absolute differences into categories and estimating correlations between group memberships using the following three-point scale: inaccurate by three or more strategies, inaccurate by one to two strategies, and accurate. Secondly, we investigated the degree to which teachers' coding errors tended towards under- or overrating their actual scaffolding usage. This was achieved by grouping teachers' difference scores into categories and estimating correlations between group memberships using the following five-point scale: underrated by three or more strategies, underrated by two-three strategies, agreement, overrated by two-three strategies, and overrated by three or more strategies.

Our final question investigated how accurately teacher colleagues – acting as second-coders – coded the videos of the primary-coders. Using the Wilcoxon signed-rank test, we first tested the null hypothesis that the second-coders' scores and the master score did not differ significantly. Secondly, we estimated correlations between the scores of the second-coders and master-coder. Finally, the scores of the second-coders and the primary-coders were compared using the same analytic approach.

Results

Coding Accuracy

The primary goal of this study was to investigate how accurately teachers coded their usage of the six scaffolding strategies from video. Figure 1 visually represents the self-coding teachers' coding accuracy**.** In each scatterplot, a diagonal line represents agreement between teachers and the master-coder. From the individual graphs, it is evident that teachers highly overrated their usage of elicitation, co-participation, and to a certain extent reducing choices.

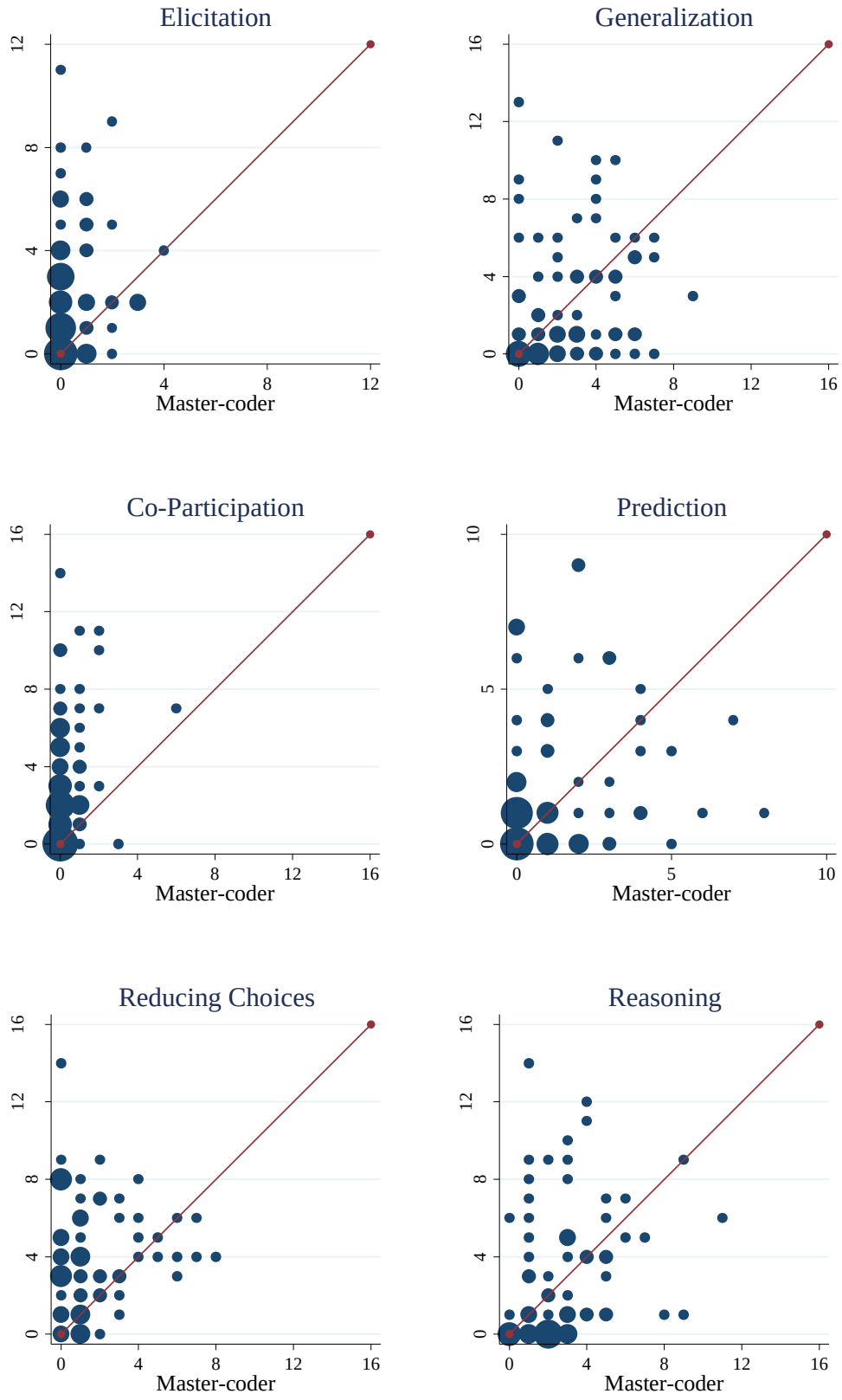

Figure 1. Teachers displayed inaccuracy in their ability to code their usage of the six scaffolding strategies.

The plots for the low support strategies, however, show that teachers both over- and underrated their strategy usage indicating coding inaccuracies with no directional tendency. These results demonstrate that many teachers perceived themselves to be using strategies that they were not in fact using, while in other occurrences, they failed to notice when they were using certain strategies. As can been seen in Table 3, Wilcoxon signed-rank tests revealed that teachers' mean perceived usage of the high support strategies was indeed significantly higher than the master score; however, means for the low support strategies were not significantly different. The prevalence of scoring inaccuracy by the teachers was further supported by correlational analyses, which revealed no significant relations between how teachers and the master-coder scored the videos with the exception of a weak relation for co-participation (r=0.240, p=.041) and reasoning (r = .241, p = .040). The mean absolute differences (which ignore whether inaccuracy is due to over- or underrating) revealed that teachers' scores generally diverged from the master score with two to three strategies on average.

Associations Between Coding Accuracy of Individual Strategies

To examine associations between teachers' scores for each scaffolding strategy, we first investigated the degree to which coding accuracy was strategy specific or stable across all six strategies. Accuracy scores were grouped using the three-point scale described above, and correlations were estimated between groups. As can be seen in Table 4, moderate correlations were found between the teachers' mean absolute differences for the three high support strategies, but they were not significantly correlated with the low support strategies.

Table 3

General Statistics for Comparison Between Self-Coding Teachers And the Master-Coder

	Self-coder	Master-coder					
n=73	Mean (SD)	Mean (SD)	*Z*	p-value	r	p-value	Mean absolute difference
High support strategies							
Elicitation	2.55 (2.47)	.56 (.91)	5.599	.001	.08	.483	2.26 (2.32)
Co-participation	3.42 (3.27)	.43 (.92)	6.730	.001	.24	.041	3.10 (3.07)
Reduce choices	4.00 (2.81)	1.67 (2.00)	5.021	.001	.07	.547	2.96 (2.78)
Low support strategies							
Generalization	2.95 (3.24)	2.77 (2.25)	-0.509	.612	.14	.247	2.62 (2.58)
Prediction	1.95 (2.35)	1.40 (1.83)	1462	.143	.12	.307	1.97 (2.05)
Reasoning	3.15 (3.51)	2.79 (2.28)	-0.324	.746	.24	.040	2.63 (2.61)

Table 4

Correlations Between Master and Self-coders

Correlations of Accuracy (3 Point Scale)						
Strategy	1	2	3	4	5	6
1. Elicitation						
2. Co-participation	.39*					
3. Reduce choices	.31*	.43*				
4. Generalization	.04	-.01	-.03			
5. Prediction	-.045	.12	.10	.21		
6. Reasoning	.16	.18	.04	.22	.04	
Correlations of Coding Tendencies (5 point scale)						
Strategy	1	2	3	4	5	6
1. Elicitation						
2. Co-participation	.54***					
3. Reduce choices	.15	.27*				
4. Generalization	-.22	-.03	.26*			
5. Prediction	-.07	-.06	.10	.24*		
6. Reasoning	.12	.06	.35**	.03	.15	
*p < .05, **p < .01, ***p < .001						

The three low support strategies were not significantly correlated with each other. These results indicate that a teacher's coding accuracy of one high support strategy tended to apply to the other high support strategies to some degree. However, teachers' coding accuracy of the low support strategies did not appear to generalize to other accuracy scores.

Next we investigated the nature of teachers' coding errors to determine whether inaccuracies tended towards over- or underrating the true values. To do this, we grouped teachers' difference scores using the five-point-scale described above, and estimated correlations between group memberships, the results of which are also given in Table 4. Of the high support strategies, teachers' scores for elicitation were found to correlate moderately with scores for co-participation, and a weak but significant correlation was found between scores for reduce choices and co-participation. For the low support strategies, a weak but significant correlation was estimated

between prediction and generalization. Furthermore, small significant correlations were also found between reducing choices and generalization and reasoning. The remaining correlation coefficients were small and insignificant thus suggesting that the coding tendencies of teachers transcended strategy boundaries only to a small degree.

Our final research question explored the extent to which coding accuracy changed as a function of whether teachers were coding a colleague instead of themselves. To study this possibility, we investigated the coding accuracy of the teacher colleagues who acted as second-coders for each video-recording, and examined whether it systematically differed from the scores of the original self-coders. Systematic differences could indicate a bias effect of the self-coding task. It should be noted that the analysis reported henceforth was performed on a smaller sample, as only 53 out of the original 73 video-recordings were double-coded by a teacher colleague.

The general statistics of the analysis of second-coders versus the master-coder are displayed in Table 5. As was observed with the self-coders, we found that second-coders also overrated usage of the high support strategies more than they underrated them, while a similar pattern of general inaccuracy was found for the low support strategies. Correlational analyses of the coding scores were small and mostly non-significant, but with two exceptions. There was a small negative correlation for elicitation (r = -.27, p = .050), and a medium correlation for reasoning (r = .36, p = .009), which could indicate higher accuracy. However, when we also compared the mean absolute differences and correlations of the self-coders to those of the second coders (see Table 6**),** we found no significant differences between the mean absolute differences of self-coders and second-coders. Furthermore, only two significant differences were detected in the degree to which the two teacher groups' scores correlated with the master scores (elicitation and co-participation), but these correlations were small. Combined, these results suggest little evidence of systematic differences in how the self-coders and second-coders scored the 53 video-recordings.

Table 5

General Statistics for Comparison Between the Second Coder and the Master-Coder

	Second-coder	Master					
n=53	Mean (SD)	Mean (SD)	Z	p-values	r	p-values	Mean absolute difference
High support strategies							
Elicitation	2.00 (2.28)	.66 (1.00)	3.27	.001	-.27	.050	2.06 (2.22)
Co-participation	3.22 (3.51)	.42(1.01)	5.16	.001	-.12	.394	3.19 (3.44)
Reduce choices	3.60 (2.67)	1.94 (2.19)	3.61	.001	.22	.108	2.60 (2.28)
Low support strategies							
Generalization	3.06 (2.82)	2.68 (2.20)	0.76	.450	.07	.612	2.64 (2.22)
Prediction	1.85 (1.92)	1.33 (1.75)	1.37	.171	.03	.848	1.87 (1.81)
Reasoning	2.72 (3.20)	2.98 (2.40)	-1.19	.235	.36	.009	2.53 (2.02)

Table 6

Significance of Differences Between Self- and Second Coders

	Absolute differences				Correlations with the master-coder			
n=53	Self-coders	Second-coders	*Z*	p-value	Self-coders	Second-coders	z	p-value
High support strategies								
Elicitation	2.26 (2.32)	2.06 (2.22)	0.01	.993	.14	-.27	2.08	.038
Co-participation	3.10 (3.07)	3.19 (3.44)	-1.30	.193	.27	-.12	1.97	.049
Reduce choices	2.96 (2.78)	2.60 (2.28)	0.27	.789	.06	.22	-0.86	.390
Low support strategies								
Generalization	2.62 (2.58)	2.64 (2.22)	-0.72	.473	.26	.07	0.95	.340
Prediction	1.97 (2.05)	1.87 (1.81)	-0.16	.874	.13	.03	0.53	.597
Reasoning	2.63 (2.61)	2.53 (2.02)	-0.91	.360	.36	.36	0.01	.995

As a final analysis, we investigated the extent to which the scores of the self-coders and second coders aligned with each other. The previous analysis demonstrated little evidence that second-coders were more (or less) accurate than the self-coders when compared to the master-coder, but the analysis did not elucidate the extent to which two different teachers coded the same video-recording similarly. To explore this, we tested whether the mean scores of the self-coders differed significantly from the second-coders, but no significant differences were detected.

Discussion

This study investigated the degree to which preschool teachers accurately coded their own usage of six scaffolding strategies following three days of PD. The coding task was intended to facilitate calibration of teachers' knowledge of how often they used each strategy.

In general, we found that teachers coded themselves and others inaccurately, but with two distinct patterns emerging. Teachers demonstrated a clear tendency towards overrating their usage of the high support strategies, which indicated that teachers believed themselves to be using these strategies when in fact they rarely used them. In contrast, teachers used the low support strategies more often, but both over- and underrated their usage of them. Furthermore, we found little evidence that coding accuracy changed as a function of whether teachers were coding themselves or a colleague. The findings mirror those of Pentimonti and Justice (2010), who similarly found that teachers rarely used high support strategies, and generally displayed a lack of calibration. Although the study by Pentimonti and Justice encompassed only a small sample, our study utilized a relatively large sample from Denmark indicating that the scaffolding of language and emergent literacy targets may also be a challenging skillset for Danish preschool teachers.

That teachers consistently overrated their usage of the high support strategies might indicate that these strategies were particularly novel to them, or perhaps harder to learn than the low support strategies. Pentimonti and Justice (2010) discussed that teachers' underuse of high support

strategies may be a troubling finding taking into consideration that these strategies are designed to support the children who are struggling most. We not only confirmed that teachers underused these strategies, but that they also believed themselves to be using them in larger quantities than they really were. It was also noteworthy that many teachers failed to observe occurrences of themselves using the low support strategies. This may indicate that teachers already use some of these strategies without being aware of it. This then raises the question of whether skill calibration is even important if teachers are already using important scaffolding strategies. We would argue, however, that awareness of how one uses a strategy is still important if one wishes to increase or optimize usage of the strategy in question.

It is difficult to ascertain whether the self-coding task in the current study helped teachers calibrate their scaffolding skill performance. The low coding accuracy seems to suggest that teachers – especially in the case of the high support strategies – were not able to derive accurate feedback from their coding videos, which we assumed to be a necessary prerequisite to the calibration process. However, it is entirely possible that low levels of coding inaccuracy may be less important in cases in which teachers are severely miscalibrated. For example, a teacher expecting to see that he or she had used ten elicitation strategies may still experience calibrating effects from the coding task if he or she "discovers" that only two strategies were used (when in fact none were used). Our conversations with teachers following the coding task did lend some anecdotal support to this possibility. In general, teachers reported that they learned that they used strategies less often than they thought they did.

Although teachers in general coded inaccurately, there were examples of accurate coding. Interestingly, in the occurrences in which teachers did agree with the master-coder, it was often in recognition of no strategies being used. For example, out of the 17 teachers who accurately coded their usage of elicitation, 12 of these instances were of zero strategy usages. Similarly for

prediction, 12 out of 19 teachers accurately observed that they did not use the strategy. According to our theoretical approach (i.e. Bandura, 1986; Bandura & Cervone, 1983), we expected that teachers who discovered that they used a particular strategy less than anticipated would be motivated to strive more to master the skill. We can thus speculate that the teachers who recognized that they did not use some strategies may have experienced recalibration effects for these particular scaffolds. However, despite this evidence of calibration in certain teachers, the majority of participants did not code accurately.

Another finding of the study revealed that teachers' coding accuracy of individual strategies did not generalize across strategies to a large degree. Although we did find that teachers who overrated their usage of one high support strategy were more likely to overrate their usage of another high rate strategy, the correlation was only strong for one pairing (elicitation and co-participation). In most cases, we found that coding accuracy tended to vary between strategies indicating that teachers may learn scaffolding strategies as discrete units. If this is the case, it may be advisable to rethink how scaffolding strategies are presented and taught to teachers. Rather than teaching the strategies as a total skillset, it may be more appropriate to teach strategies incrementally, as the strategies appear to vary in their learnability.

Incorporating Coding Tasks into Teacher Training

Self-coding tasks may be a tool that can help teachers learn to scaffold. By incorporating self-coding into teacher training programs (both as post-secondary education and PD), teachers can learn to take status of their own progress by intermittently calibrating their knowledge of how well they scaffold. McCullagh (2012), building upon Vygotskian theory, argued that video analysis offers teachers a *video supported zone of proximal development*, essentially giving teachers the means to scaffold their own PD. However, in order for self-coding to function in this regard, teachers must be able to code accurately. Previous research on teachers' coding accuracy (e.g.,

Blamey et al., 2012; Prusak et al., 2010) found that teachers could learn to code instructional practices from video with relatively little training. In contrast, we found that three days of PD were insufficient to prepare teachers for coding all six scaffolding strategies with equal accuracy.

This discrepancy in results might be due to variation in precursor skills such as observation skills, but another explanation may lie in the complexity of scaffolding. Scaffolding is a complex process that requires teachers to be cognizant of each child's zone of proximal development, and then to use an appropriate strategy that will help the child achieve the goal at hand. It may be that teachers require more time and practice to learn scaffolding strategies in comparison to other pedagogical practices. Coding in itself is also a skill that requires teachers to practice in order to reach proficiency. Teachers likely required more time for practicing coding than what they were allotted in the current study. Although we did allow teachers to view their own coding-videos as many times as they needed to ensure an accurate coding, time restraints may have forced some teachers to use less time on the coding than desired.

To improve the usefulness of self-coding exercises, some teachers may require expert feedback while learning to code. Such feedback would both help teachers learn to code accurately, but also likely deepen teachers' understanding of the scaffolding strategies. Teachers that code accurately may also be better situated to take charge of their own learning, which may be especially useful in preschools in which funds for PD activities are limited.

Limitations

This study is limited by the lack of longitudinal data. Although psychological research has demonstrated that individuals strive to change behaviour when they discover that there performance is lower than expected (Bandura & Cervone, 1983), we have not demonstrated that this is the case in the current study. Therefore, future research might investigate the extent to which gains in coding

accuracy predict increases in scaffolding usage. Such a finding would lend empirical support to the theory of calibration, and its importance in the learning processes of preschool teachers.

References

Ball, D. L. (1995). *Developing Mathematics Reform: What Don't We Know about Teacher Learning--but Would Make Good Working Hypotheses?* Paper presented at the Paper presented at the Conference on Teacher Enhancement in Mathematics K-6,, Arlington, VA.

Bandura, A. (1977). *Social learning theory*. Englewood Cliffs, NJ.: Prentice-Hall.

Bandura, A. (1986). *Social foundations of thought and action*: Englewood Cliffs, NJ Prentice Hall.

Bandura, A., & Cervone, D. (1983). Self-evaluative and self-efficacy mechanisms governing the motivational effects of goal systems. *Journal of personality and social psychology, 45*(5), 1017.

Blamey, K. L., Beauchat, K. A., & Sweetman, H. (2012). Vocabulary Instruction Through Storybook Reading. *NHSA Dialog, 15*(3), 260-265. doi: 10.1080/15240754.2012.666687

Bleses, D., Højen, A., Dybdal, L., Dale, P., Justice, L., Piasta, S., . . . Aktürk Ari, B. (2014). A randomized controlled trial of a large scale systematic and explicit language and literacy intervention in Danish daycares: The SPELL study. *British Journal of Education, Society & Behavioural science, 4*(11), 504-1524.

Brown, M. G., Malott, R. W., Dillon, M. J., & Keeps, E. J. (1980). Improving Customer Service in a Large Department Store Through the Use of Training and Feedback. *Journal of Organizational Behavior Management, 2*(4), 251-265. doi: 10.1300/J075v02n04_02

Calandra, B., Gurvitch, R., & Lund, J. (2008). An Exploratory Study of Digital Video Editing as a Tool for Teacher Preparation. *Journal of Technology and Teacher Education, 16*(2), 137-153.

Clancy-Menchetti, J. L. (2006). *Early literacy professional development: Exploring the effects of mentoring for preschool teachers.* (Dissertation). The Florida State University.

Cunningham, A. E., Zibulsky, J., & Callahan, M. D. (2009). Starting small: Building preschool teacher knowledge that supports early literacy development. *Reading and Writing, 22*(4), 487-510.

Dieterich, S. E., Assel, M. A., Swank, P., Smith, K. E., & Landry, S. H. (2006). The impact of early maternal verbal scaffolding and child language abilities on later decoding and reading comprehension skills. *Journal of School Psychology, 43*(6), 481-494. doi: http://dx.doi.org/10.1016/j.jsp.2005.10.003

Festinger, L. (1962). *A theory of cognitive dissonance* (Vol. 2): Stanford university press.

Fischhoff, B., Slovic, P., & Lichtenstein, S. (1977). Knowing with certainty: The appropriateness of extreme confidence. *Journal of Experimental Psychology: Human Perception and Performance, 3*(4), 552-564. doi: 10.1037/0096-1523.3.4.552

Fuller, F. F., & Manning, B. A. (1973). Self-Confrontation Reviewed: A Conceptualization for Video Playback in Teacher Education. *Review of Educational Research, 43*(4), 469-528. doi: 10.2307/1170077

Grace, C., Bordelon, D., Cooper, P., Kazelskis, R., Reeves, C., & Thames, D. G. (2008). Impact of Professional Development on the Literacy Environments of Preschool Classrooms. *Journal of Research in Childhood Education, 23*(1), 52-81. doi: 10.1080/02568540809594646

Hacker, D. J., Bol, L., & Keener, M. C. (2008). Metacognition in education: A focus on calibration. *Handbook of metamemory and memory*, 429-455.

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.

Justice, L. M., Mashburn, A. J., Hamre, B. K., & Pianta, R. C. (2008). Quality of language and literacy instruction in preschool classrooms serving at-risk pupils. *Early Childhood Research Quarterly, 23*(1), 51-68. doi: http://dx.doi.org/10.1016/j.ecresq.2007.09.004

Justice, L. M., McGinty, A. S., Cabell, S. Q., Kilday, C. R., Knighton, K., & Huffman, G. (2010). Language and Literacy Curriculum Supplement for Preschoolers Who Are Academically At Risk: A Feasibility Study. *Language, Speech, and Hearing Services in Schools, 41*(2), 161-178. doi: 10.1044/0161-1461(2009/08-0058)

Lonigan, C. J., Farver, J. M., Phillips, B. M., & Clancy-Menchetti, J. (2011). Promoting the development of preschool children's emergent literacy skills: A randomized evaluation of a literacy-focused curriculum and two professional development models. *Reading and Writing, 24*(3), 305-337.

Lynes, M. J. (2012). *The effects of self-evaluation with video on the use of oral language development strategies by preschool teachers.* (Dissertation). The University of Utah.

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. doi: 10.1111/j.1467-8624.2008.01154.x

McCullagh, J. (2012). How can video supported reflection enhance teachers' professional development? *Cultural Studies of Science Education, 7*(1), 137-152. doi: 10.1007/s11422-012-9396-0

McGee, L. M., & Ukrainetz, T. A. (2009). Using Scaffolding to Teach Phonemic Awareness in Preschool and Kindergarten. *The Reading Teacher, 62*(7), 599-603. doi: 10.1598/RT.62.7.6

National Early Literacy Panel. (2008). *Developing early literacy: Report of the National Early Literacy Panel: A scientific synthesis of early literacy development and implications for intervention*. Washington, D.C.: National Institute for Literacy.

Neuman, S. B., & Wright, T. S. (2010). Promoting language and literacy development for early childhood educators: A mixed-methods study of coursework and coaching. *The Elementary School Journal, 111*(1), 63-86.

O'Connor, R., Notari-Syverson, A., & Vadasy, P. F. (2005). *Ladders to Literacy: A kindergarten activity book* (2nd ed.). Baltimore, MD: Brookes.

Pentimonti, J. M., & Justice, L. M. (2010). Teachers' use of scaffolding strategies during read alouds in the preschool classroom. *Early Childhood Education Journal, 37*(4), 241-248.

Polk, K. (2013). *Language and literacy supports in early childhood classrooms: comparing ratings of teachers and observers.* Retrieved from http://scholar.utc.edu/theses/95

Powell, D. R., Steed, E. A., & Diamond, K. E. (2009). Dimensions of Literacy Coaching With Head Start Teachers. *Topics in early childhood special education*. doi: 10.1177/0271121409341197

Prusak, K., Dye, B., Graham, C., & Graser, S. (2010). Reliability of pre-service physical education teachers' coding of teaching videos using studiocode analysis software. *Journal of Technology and Teacher Education, 18*(1), 131-159.

Schunk, D. (1989). Social Cognitive Theory and Self-Regulated Learning. In B. Zimmerman & D. Schunk (Eds.), *Self-Regulated Learning and Academic Achievement* (pp. 83-110): Springer New York.

Sheridan, S. M., Edwards, C. P., Marvin, C. A., & Knoche, L. L. (2009). Professional Development in Early Childhood Programs: Process Issues and Research Needs. *Early Education and Development, 20*(3), 377-401. doi: 10.1080/10409280802582795

Sherin, M. G., & Han, S. Y. (2004). Teacher learning in the context of a video club. *Teaching and Teacher Education, 20*(2), 163-183. doi: http://dx.doi.org/10.1016/j.tate.2003.08.001

Smith, M. W., & Dickinson, D. K. (2002). *Early Language & Literacy Classroom Observation (ELLCO)*. Baltimore, MD: Brookes.

Tripp, T., & Rich, P. (2012). Using video to analyze one's own teaching. *British Journal of Educational Technology, 43*(4), 678-704. doi: http://dx.doi.org/10.1111/j.1467-8535.2011.01234.x

Van de Pol, J., Volman, M., & Beishuizen, J. (2010). Scaffolding in teacher–student interaction: A decade of research. *Educational Psychology Review, 22*(3), 271-296.

Vygotsky, L. (1978). *Mind in society: The development of higher mental process*: Cambridge, MA: Harvard University Press.

Wood, D., Bruner, J. S., & Ross, G. (1976). the role of tutoring in problem solving. *Journal of Child Psychology and Psychiatry, 17*(2), 89-100. doi: 10.1111/j.1469-7610.1976.tb00381.x

Wright, M. R. (1998). *The effects of video-based teacher self-evaluation on teacher instructional performance*. (Dissertation). University of South Alabama, Ann Arbor. ProQuest Dissertations & Theses Global database.

Yoo, M. S., Son, Y. J., Kim, Y. S., & Park, J. H. (2009). Video-based self-assessment: Implementation and evaluation in an undergraduate nursing course. *Nurse Education Today, 29*(6), 585-589. doi: http://dx.doi.org/10.1016/j.nedt.2008.12.008