Early Childhood Education and Care in the Netherlands

Quality, Curriculum, and Relations with Child Development

Pauline Louise Slot

ISBN 978 90 5335 965 5

Layout Ridderprint B.V., Ridderkerk, The Netherlands
Printing Ridderprint B.V., Ridderkerk, The Netherlands

The research reported in this thesis was supported by a grant from NWO PROO
project number 411-20-442 and by a grant from NWO BOPO project number 413-09-072

Early Childhood Education and Care in the Netherlands

Quality, Curriculum, and Relations with Child Development

Proefschrift

ter verkrijging van de graad van Doctor
aan de Universiteit Utrecht op gezag van de Rector Magnificus,
Prof. dr. G.J. van der Zwaan, ingevolge het besluit van het
College voor Promoties in het openbaar te verdedigen
op vrijdag 12 december 2014 des middags te 12.45 uur

door

Pauline Louise Slot
geboren op 25 februari 1980
te Nijmegen

Promotiecommissie:

Promotor:
Prof. dr. P.P.M. Leseman

Copromotoren:
Dr. H. Mulder
Dr. J. Verhagen

CONTENTS

CHAPTER 1

General Introduction

Introduction

Early childhood, spanning the age range from zero to six years, is in many respects the phase in life in which the basic functions of learning, communication, and cognitive and emotional control are laid out, with long-term consequences for educational achievement and psychosocial adjustment, and all the societal opportunities that are dependent on education and psychosocial health. Research in brain development has revealed sensitive periods that are marked by increased brain plasticity, making the brain maximally susceptible to environmental influences, but also maximally vulnerable to adverse circumstances (Shonkoff & Phillips, 2000). The quality of young children's environments is critical to the start children make in their lives, with *quality* referring to the emotional support, secure social relationships, cognitive stimulation, exposure to language models, and opportunities to gain control over activities and to develop self-regulation that are provided on a regular basis and in a consistent manner across the different contexts of child development throughout early childhood.

The quality of the environments in which young children grow up pertains to the family in the first place. The family, as a secure basis of affectionate and supportive social relationships and as a home learning environment that introduces young children to the language, literacy and math concepts and skills on which the school can build further, exerts a profound influence on child development in all domains. Early disparities in children's cognitive, language, self-regulation, and social development are strongly related to the socioeconomic status of the family, parents' education level and ethnic background, and to the style of parenting and the home learning environment that parents provide (Bradley & Corwyn, 2002; Brooks-Gun & Markman, 2005; Hammer, Farkas, & Maczuga, 2010; Hart & Risley, 1995; Hoff, 2006; Mistry, Benner, Biesanz, Clark, & Howes, 2010; Sektnan, McClelland, Acock, & Morrison, 2010). Already at an early age, children from disadvantaged family backgrounds show delays in development compared to children from more advantaged family backgrounds, especially in the areas of language, literacy, and math (Jordan & Levine, 2009; Magnuson, Lahaie, & Waldfogel, 2006; Mistry et al., 2010; Sektnan et al., 2010), but also in the social-emotional domain (Mistry et al., 2010). These early delays are referred to as the 'early education gap' and they tend to persist into long term disadvantages, including school drop-out, lower educational attainment, poorer mental health, and increased crime rates (Heckman & Masterov, 2007).

Next to the family, non-familial education and care provisions for zero to six-year-olds,

including center-based day care, playgroups and preschools, have become increasingly important as contexts of early child development in the past decades. Worldwide, but especially in Western industrialized countries, the vast majority of young children participate in some form of early childhood education and care (ECEC) for some hours per week and for a number of years before they start in primary school (OECD, 2010). This raises the question what the quality and developmental effects of ECEC are, which is the topic of the current dissertation.

ECEC and Children's Developmental and Educational Outcomes

Day care and preschool attendance have generally been shown to be beneficial for children's developmental and educational outcomes. Numerous studies have shown positive effects of ECEC attendance on various aspects of both short- and long-term cognitive development, including school achievement, IQ, grade retention, and high school graduation, self-regulation, executive functions, and socio-emotional development (Barnett, 1995; Bradley & Vandell, 2007; Burger, 2010; Gormley, Philips, & Gayer, 2008; Heckman, 2006; Love et al., 2005; Magnuson, Ruhm, & Waldfogel, 2007; Winsler et al., 2008). Effects tend to be stronger for disadvantaged children and for children who stayed longer in ECEC, or attended ECEC for more hours per week (Magnuson, Meyers, Ruhm, & Waldfogel, 2004; Magnuson et al., 2007; Wen, Leow, Hahs-Vaughn, Kormacher, & Marcus, 2012).

Given the potential benefits of ECEC, there is growing awareness of the potential role ECEC can play in compensating for the adverse effects of child poverty and in combating early educational disadvantages. Therefore, in many countries, increasing the provision and accessibility of ECEC is high on the national policy agendas (OECD, 2006). Investing in ECEC has been shown to be a promising means to partially compensate the effects of adverse early environmental circumstances. Moreover, investments in ECEC can have high economic returns to society at large, more than interventions later in life (Heckman, 2006). However, the benefits of ECEC depend on the quality of the education and care provided. In the UK, the *Effective Pre-school and Primary Education* (EPPE) project has demonstrated the beneficial role of ECEC in the longer-term educational and social development for all children, but the effects were strongest for children of low-income families (Sylva, Melhuish, Sammons, Siraj-Blatchford, & Taggart, 2011). However, these effects are moderated by the quality of the ECEC provisions, with higher quality provisions being related to greater benefits for children. Similarly, studies in the US have shown that high quality ECEC was related to increased cognitive outcomes (Dearing, McCartney &

Taylor, 2010; NICHD ECCRN, 2006; Vandell, Belsky, Burchinal, Steinberg, & Vandergrift, 2010). Furthermore, studies show that only high quality ECEC can protect children against the negative effects of low quality home environments (Hall et al., 2013), whereas low quality can increase the negative outcomes for children from disadvantaged homes (Watamura, Phillips, Morrissey, McCartney, & Bub, 2011).

The strongest evidence for the compensatory effects of ECEC comes from a number of experimental studies with randomized assignment, targeting socioeconomically disadvantaged children. These studies have have shown moderate to large effects on social-emotional, language, and academic outcomes in the short and long run, with in some studies effects persisting far into adulthood (Campbell, Pungello, Miller-Johnson, Burchinal, & Ramey, 2001; Ramey & Ramey, 2004; Nores, Belfield, & Barnett, 2005; Reynolds et al., 2011; Schweinhart & Weikart, 1997). However, findings from general population studies focusing on universal large scale ECEC systems have revealed much smaller effects on children's outcomes (Burchinal, Kainz, & Cai, 2011; Farran & Hofer, 2013; NICHD ECCRN, 2000, 2006; Pianta et al., 2009; Zaslow, 2010). The smaller magnitude of effects found in these studies probably points to lower overall quality in large scale ECEC systems compared to the small scale experimental studies evaluating model programs. Identifying the factors that contribute to quality and outcomes, and identifying effective strategies to improve the quality and impact of ECEC, is important for informing policy and practice.

Quality and Curriculum in Early Childhood Settings

Broadly defined, quality of ECEC concerns the day-to-day experiences children have while being involved in all kinds of activities and interactions and the conditions that are considered a prerequisite for these experiences, which are assumed to be beneficial for children's developmental and educational outcomes (Howes et al., 2008; Layzer & Goodson, 2006; Sylva et al., 2006; Thomason & La Paro, 2009). A distinction is commonly made between process quality and structural quality. Process quality refers to children's daily experiences and encompasses the physical, emotional, social, and educational aspects of children's interactions with teachers, peers, and materials (Howes et al., 2008; Pianta et al., 2005; Thomason & La Paro, 2009). Process quality aspects are seen as the proximal determinants of children's developmental and educational outcomes (Howes et al., 2008; Pianta et al., 2005; Thomason & La Paro, 2009). Structural quality aspects, such as group size, children-to-teacher ratio and teacher's qualifications, are considered the more distal, regulable aspects of quality which are assumed to be important preconditions

for process quality (Cryer, Tietze, Burchinal, Leal & Palacios, 1999; Philips, Mekos, Scarr, McCartney, & Abbott-Shim, 2000; Philipsen, Burchinal, Howes, & Cryer, 1997; Pianta et al., 2005; Vandell, 2004). An important question is whether structural characteristics are indeed related to process quality. As will be reviewed in this dissertation, the evidence is mixed, raising the question whether other aspects, previously not included in studies into the relationships between structural quality and process quality, can explain variance in process quality.

A relatively understudied topic in ECEC, particularly for children under three years of age, concerns the curriculum of activities children are provided with on a day-to-day basis, which are meant to serve particular developmental and educational goals. The curriculum as implemented can be considered to be part of process quality as it refers to children's actual experiences with materials and particular knowledge contents that influence the knowledge that children can gain and the skills they can develop. An important question is what constitutes a good curriculum for young children. There is an ongoing debate regarding the importance of play versus a stronger focus on pre-academic skills reflected in a school readiness tradition or pre-primary approach (Bennett, 2005; Bodrova, 2008; OECD, 2006). In some countries, the focus in ECEC is increasingly on school readiness skills such as language, literacy, and math in order to decrease the school achievement gap of disadvantaged children. However, recent research points to the importance of play for the development of executive functions (EF) and self-regulation skills, which have been shown to be strong predictors of later school achievement, social competence, behavioral adjustment and learning-related skills in many studies, (Berk, Mann, & Ogan, 2006; Diamond & Lee, 2011) as will be reviewed in this dissertation. Fostering EFs and self-regulation in children may require an approach that allows children to take initiative and exert control over the activity and the level of challenge and consequently require the provision of activities such as play and collaborative small group work (Leseman, 2012). More evidence is needed on what constitutes appropriate curricula for the early years.

In addition to defining quality at the conceptual level, there is the issue of measurement. Several measures exist that have been used in research in different countries allowing for comparative assessment and providing benchmarks to the evaluation of quality. Observation measures should be closely aligned to the conceptual definition of quality and for instance make a clear distinction between structural aspects and process aspects. Moreover, with regard to process quality, it is increasingly important to address both emotional and educational aspects of quality. For example, the Classroom Assessment

Scoring System (CLASS Toddler; La Paro, Hamre, & Pianta, 2011) used in the studies reported in this dissertation, seems particularly suited to fulfill these ends. However, whether instruments such as the CLASS and the Early Childhood Environment Rating Scale Revised (ECERS-R; Harms, Clifford, & Cryer, 1998), which were developed in the US, can be applied validly in other cultural contexts with different systems has still to be established. Related to this is the issue of psychometric quality, which is a neglected topic in ECEC evaluation research, yet highly relevant in view of the consequences of quality assessment for children, professionals, service providers, and society at large. Recent studies have shown that measurement quality of widely used assessment instruments is not unproblematic (Cassidy, Hestenes, Hegde, Hestenes & Mims, 2005; Colwell et al., 2013; Gordon et al., 2013; Perlman, Zellman & Le, 2004). Therefore, quality assessment of ECEC requires a thorough evaluation of the measurement properties of the assessment instrument used.

Early Childhood Education and Care in the Netherlands

In the Netherlands, over 80% of the two- to four-year-old children attend either formal center-based day care or preschool before they enter the primary school system at age four years (CBS, 2011; MOgroep, 2012). Regarding children under three years, the enrollment rate is among the highest in the world (OECD, 2010). Two main types of center-based ECEC provisions exist for children before they enter primary school. The first is center-based day care for children from birth to four years of age, with as its main function to support parents in combining work and care. Over 50% of the children under four attend center-based day care, however on average only for two full days a week (NCKO, 2011). The second type concerns preschools for two- to four-year-old children, which are attended, on average, for two to four half days a week. Socioeconomically disadvantaged children, with low educated parents or from non-Dutch speaking homes, are subjects of a targeted policy to combat early educational disadvantages and are offered an education program which is provided in preschools for four half days a week. Children who are not at risk usually attend preschools for two half days a week. Although the two types of ECEC provisions differ in the age range and populations served, and stem from different traditions (with a care and an education tradition, respectively), they are highly comparable in structural quality and increasingly so also in process quality. The similarity in quality between the two types can be attributed to successive legislations in the past decade. The recent OKE Act ("promoting development through quality and education") of 2010 brings day care

centers and preschools under the same statutory quality framework and emphasizes the equal importance of social, emotional, and cognitive outcomes for children.

Research into the quality and effectiveness of the Dutch ECEC system is limited. The quality of day care has been monitored since 1995 in observation studies with random samples of day care centers every three to four years. However, these studies have not investigated effects on children's developmental outcomes (Fukkink, Gevers-Deynoot-Schaub, Helmerhorst, Bollen, & Riksen-Walraven, 2013; Vermeer et al., 2008). Until recently, no studies have been conducted into the quality and effectiveness of the preschool system for children under four years of age. Results from day care centers show moderate overall quality regarding emotional aspects of care (Vermeer et al., 2008). Educational quality was only recently included in the measurement of quality, revealing low overall quality (Helmerhorst et al., 2014). Moreover, successive studies have shown a decline in quality over the years, which has been attributed to the enormous expansion of the day care system (De Kruif et al., 2009). In addition, quality decreased in spite of increased regulation of especially the structural quality of day care provisions over the years (NCKO, 2011), suggesting there may be other factors at stake in explaining process quality that have not been sufficiently regulated.

Since 2000, the Dutch national education policy has strongly focused on preschool education as a means to prevent academic difficulties in children from low-income and ethnic minority families. A targeted policy for pre- and early primary school (in Dutch: *Beleid Voor- en Vroegschoolse Educatie*, or *VVE-beleid*), became operative providing eligible children in the age range from two-and-a-half to six years with four half days preschool before they enter primary school at age four. Note that the first two grades of primary school, from age four to six years, encompass the former independent kindergarten. Targeted education programs (in Dutch: *VVE-programma's*) were developed, spanning the preschool and kindergarten ages, in order to provide a continuous and developmentally structured curriculum in collaborating preschool and kindergarten classrooms, focusing on holistic development, but with an emphasis on Dutch language development. Two quasi-experimental pilot studies with two of these programs, *Piramide* and *Kaleidoscoop* (the latter is a Dutch version of the High/Scope curriculum, developed in the US), indicated positive effects on cognitive and language outcomes over and beyond preschool and kindergarten education as usual (Veen, Roeleveld, & Leseman, 2000; De Goede & Reezigt, 2001). In the past years, this targeted policy has met increasing criticism and concerns have been raised regarding the implementation of the policy and its effectiveness (Doolaard &

Leseman, 2008; Napp-Kolhoff et al., 2008). Studies using retrospective designs, in which children's use of pre-school care and education provisions, either with or without a targeted education program, were reconstructed based on information provided by parents and primary schools and related to current school achievement, were not able to show any positive effects of targeted pre- and early primary school education on achievement in language, reading, and math in grade one (cf. Bruggers, Driessen, & Gesthuizen, 2014). Since the research design of these studies (using retrospective measurements without control of selective placement and variations in quality and program implementation) has been criticized, there is a clear need for more studies with a prospective design.

The Pre-COOL Study and the Current Dissertation

In 2009, in response to the need for more evidence regarding the quality and effectiveness of ECEC in the Netherlands, the Ministry of Education, Culture, and Sciences and the Netherlands Organization for Scientific Research commissioned a consortium of the Kohnstamm Institute of the University of Amsterdam, the Institute of Applied Social Sciences of the Radboud University Nijmegen, and the Department of Special Education of Utrecht University to conduct a large scale longitudinal cohort study into the quality and effectiveness of the Dutch day care and (targeted) preschool system, with a special focus on the compensatory effects on disadvantaged children, entitled pre-COOL (Veen et al., 2012). The aim of pre-COOL was to select children age two years and follow them up until age five, when part of them would enter the national cohort study COOL on students' educational careers in primary and secondary school, which follows children until age eighteen years.[1] The studies reported in the current dissertation are part of the pre-COOL study and specifically focus on the quality of education and care provisions in the Netherlands, the structural preconditions that influence quality, and the first indications of the possible effects of the provisions on children's development. Two sub-samples were distinguished at the start of the study: a center-based sample of children attending ECEC provisions who were recruited at their center, and a home-based sample of children recruited through municipal registration records. In the studies reported in this dissertation we focus on the center-based sample only. In total 162 centers, with 276 classrooms were involved in the current studies. In addition, an in-depth study was conducted for which a subsample of 37 classrooms participating in pre-COOL was selected.

1 Data collection of the final pre-COOL measurement wave was completed in the summer of 2014.

The main aim of the current study was to evaluate the quality of Dutch ECEC provisions and to identify the most important structural determinants of quality. A second aim was to examine the first indications of possible effects of ECEC quality on children's developmental outcomes. A third aim was to contribute to the debate about what constitutes an appropriate curriculum for young children in ECEC. Four main issues guided the research conducted for this dissertation.

The first issue concerned the measurement of quality. The choice within pre-COOL for the widely used Classroom Assessment Scoring System (CLASS) was made to ensure comparability with international research, which could provide benchmarks to evaluate Dutch ECEC. However, this raised the question of applicability of this observational measure to the Dutch context, which presents in many respects a different cultural context and a different tradition in ECEC than the US context in which this measure was developed. The second issue concerned the relationship between process quality and structural quality, a question which is a particularly important topic because structural features constitute the largest costs, whereas process quality is most strongly related to benefits for children, and, thus for society at large (Vandell et al., 2010). Therefore, a positive relation between structural and process quality is essential for the costs-efficiency of ECEC. The third issue concerned the effects of ECEC attendance on children's development. Although numerous studies have shown positive effects of ECEC attendance on child development (Burger, 2010; Gormley, Philips, & Gayer, 2008), particularly when the provided care was of high quality (NICHD ECCRN, 2000, 2006; Pianta et al., 2009; Sylva et al., 2011), the evidence concerning the Dutch ECEC system is still limited. The final issue concerned the debate about the early years curriculum, in particular the balance between play and academic activities. The four issues are addressed in the four empirical studies reported in this dissertation.

This Dissertation

This dissertation is structured in the following ways. Chapter 2 reports a study into the psychometric properties of the CLASS Toddler. Prior research on this topic was extended in two ways. First, the CLASS consists of several observation cycles for each classroom, which are usually aggregated to a single classroom score. However, in the study reported in Chapter 2 the analyses were conducted at the observation cycle level because this provides useful information regarding the variability within the classroom. A multilevel approach was used to assess the dimensional and domain factor structure of the CLASS, both at

the within classrooms level and the between classroom level. Second, in order to provide more detailed information on the measurement quality of the CLASS at the indicator level, a classical test theory (CTT) approach with confirmatory factor analyses was complemented with an Item Response Theory (IRT) approach, evaluating the difficulty level and discrimination of the indicators as primary sources of information. Thus, we investigated the psychometric quality of the CLASS in great detail, in order to enhance our understanding of the applicability of this classroom observation measure to the Dutch context.

In the study reported in Chapter 3, the associations between structural and process quality aspects were examined. A multi-method approach was employed, combining observation data with data from teacher reports to enrich the process quality concept. Classroom observations, measured with the CLASS Toddler, and the provision of developmental and educational activities, as reported by teachers, were used to construct two comprehensive process quality factors: emotional quality and educational quality. In addition, the usual set of structural quality aspects, with group size, teacher-child ratio, and teacher's qualifications, was extended with two additional structural characteristics, namely the use of an education program (*VVE-programma*) and the provision of professional development activities at the center, to determine the relative importance of all these structural aspects for process quality.

In the study reported in Chapter 4, the effects of ECEC quality on children's outcomes in two important domains of development were examined: vocabulary and attention, as these domains can be seen as important building blocks for later social and cognitive competence. A multilevel value-added approach was applied in which not only child and family background characteristics were controlled, but also the possible confounding effects of selective enrollment in ECEC. In addition, a distinction was made between domain-general and domain-specific quality aspects, including specific curriculum components, allowing the investigation of differential effects of these quality aspects on children's outcomes.

In Chapter 5, the results of an in-depth study into children's cognitive and emotional self-regulation in ECEC classrooms are reported. Observations of children's pretend play were used to assess children's self-regulation skills in a naturalistic context. A new observational measure was developed, based on extant literature, to evaluate children's cognitive and emotional self-regulation skills during a naturalistic play episode. Test-based measures of children's cognitive (cool) and affective (hot) executive functions (EF) were used in order

to examine to what extent children's EFs predict observed self-regulation. In addition, quality of play and general classroom quality were also studied as predictors of children's self-regulation.

Finally, in Chapter 6 we recapitulate and reflect upon the findings of the four empirical studies. We return to the main issues raised concerning the state of affairs of ECEC quality in the Netherlands, associations with structural quality characteristics, and relations with children's outcomes. Furthermore, several implications for policy and practice will be discussed, along with suggestions for future research.

CHAPTER 2

Psychometric Quality of the Classroom Assessment Scoring System Toddler

Slot, P.L., Boom, J., Verhagen, J., & Leseman, P. P. M., (2014)

Psychometric quality of the Classroom Assessment Scoring System Toddler.

Manuscript submitted for publication

Abstract

Classroom quality measures, such as the Classroom Assessment Scoring System (CLASS), are being extensively used in research and program evaluation studies and for accountability and professionalization purposes, but there is less evidence on the psychometric quality of the measures used. Particularly, for the recently developed CLASS Toddler the available evidence is limited and solely based on studies conducted in the United States. The present study investigated the psychometric quality of the Classroom Assessment Scoring System (CLASS) Toddler version, using data from an ongoing longitudinal Dutch cohort study in early childhood settings, and adds to the existing evidence in several ways. First, we examined both the domain and dimensional structure of the CLASS using the most basic information of the observation cycles allowing for comparisons of the factor structure both within and between classrooms. Second, we examined the quality of the CLASS indicators by evaluating the difficulty and discrimination parameters of the indicators using an Item Response Theory (IRT) approach. Finally, we assessed the criterion validity of the CLASS by investigating associations with classroom and curriculum characteristics. The findings generally support the measurement quality of the CLASS Toddler. Some differences regarding the domain structure were found, however, possibly pointing to cultural differences between the US and the Netherlands. Implications for practice are discussed.

Introduction

Measures of quality in early childhood settings have been extensively used in research and program evaluation studies, and are increasingly being used in professionalization programs at the center level and in quality monitoring and improvement systems at the policy level (Martinez-Beck, 2011). Yet, the quality of the measures used has recently become a topic of debate and some scholars have suggested that widely used measures do not seem to tap the construct of quality adequately (Burchinal, Kainz, & Cai, 2011; Colwell, Gordon, Fujimoto, Kaestner, & Korenman, 2013; Gordon et al., 2013; Layzer & Goodson, 2006; Zaslow et al., 2010). Although these widely used observational measures are based on well-established theoretical frameworks, the psychometric properties of these measures have not been extensively investigated. Moreover, the few studies to date that have comprehensively investigated psychometric properties of commonly used observation measures have revealed several problems regarding scaling of items, factor structure, and

criterion validity (Cassidy, Hestenes, Hegde, Hestenes, & Mims, 2005; Colwell et al., 2013; Gordon et al., 2013; Perlman, Zellman, & Le, 2004).

A recently developed observational measure is the Classroom Assessment Scoring System (CLASS; Hamre & Pianta, 2007) of which different versions have been developed for classrooms from infant and toddler care up until secondary school. The CLASS has been widely used in research and professional development in the United States (US), and also increasingly outside the US (Cadima, Leal, & Burchinal, 2010; Pakarinen et al., 2010; Slot, Leseman, Verhagen, & Mulder, under review; von Suchodoletz, Fäsche, Gunzenhauser, & Hamre, under review). However, the psychometric properties of the CLASS have been investigated mainly in US samples and consequently, the results of these evaluation studies may not be applicable to other countries. Moreover, research into the psychometric properties of quality measures has mostly been conducted adopting a classical test theory approach (e.g., correlational and factor analyses), which has a number of limitations. First, the classical test theory (CTT) approach is mainly focused on test-level information, that is, on the measure as a whole, but provides little information about the scalability and the psychometric quality of the individual items of the measure, including item statistics such as item difficulty and item discrimination (Bryant, Burchinal, & Zaslow, 2010; Hambleton & Jones, 1993; Lambert, Nelson, Brewer, & Burchinal, 2006). However, the measurement quality of an instrument is as good as the quality of its primary sources of information at the item level. Second, in the CTT approach, ability and item parameters are not invariant (de Ayala, 2013; Hambleton & Jones, 1993). This means that the instrument's characteristics affect the person's ability, in this case the observed quality scores, and sample characteristics, which, in turn, affect the item statistics, such as item difficulty and item discrimination. Therefore, in CTT, the scores derived from a specific measure are only representative of a person's ability on that particular measure (de Ayala, 2013; Hambleton & Jones, 1993). Item response theory (IRT) can deal with these issues. For quality observation measures, for example, IRT models assume that teachers' observed performance reflects a latent ability, which is independent of the items used. In addition, item statistics used in IRT, such as item difficulty and item discrimination, are assumed to be independent of the sample or the group in which they are estimated (de Ayala, 2013; Hambleton & Jones, 1993).

For one of the latest additions to the CLASS, the Toddler version (La Paro, Hamre, & Pianta, 2011), developed for evaluating process quality in toddler classrooms, there is only limited research on its psychometric quality available to date. To the best of our knowledge,

only two papers have investigated some psychometric properties of the CLASS Toddler in small US samples, such as the factor structure (La Paro, Williamson, & Hatfield, 2014) and the criterion validity (La Paro et al., 2014; Thomason & La Paro, 2009), but detailed information on the item quality, is lacking.

The CLASS is aimed at evaluating those aspects of process quality, which are hypothesized to be beneficial for children's developmental and educational outcomes and, as such, it is increasingly used for accountability and professional development purposes, requiring a more thorough investigation of the psychometric quality of the measure. The objective of the current study was to investigate the psychometric properties of the CLASS Toddler, using child care and preschool data from an ongoing longitudinal cohort study in the Netherlands. The study combined a CTT approach and an IRT approach, which can be considered complementary to each other and together to provide detailed information on the factor structure, item quality and validity of the CLASS Toddler.

Measuring Process Quality

Process quality of early childhood education and care (ECEC) provisions reflects the social-emotional and instructional features of teacher-child and child-child interactions that have been found to be positively related to children's development of self-regulation, pre-academic, and social skills (Curby et al., 2009; Howes et al., 2008; Mashburn et al., 2008; Rimm-Kaufmann, Curby, Grimm, Nathanson, & Brock, 2009). Although in many studies observational measures have been used to assess process quality in center-based child care and preschool settings, and to examine effects of quality on children's developmental outcomes, the psychometric quality of such measures is not frequently evaluated and if evaluated, psychometric problems are often found. For instance, the Early Childhood Environment Rating Scale (ECERS), its revised version (ECERS-R) and the Caregiver Interaction Scale, (CIS) have been used in many studies worldwide, but a number of studies has shown that these measures suffer from several psychometric problems concerning item scaling and factor structure of the measure (Cryer, Tietze, Burchinal, Leal, & Palacios, 1999; De Kruif, McWilliam, Ridley, & Wakely, 2000; Gordon et al., 2013), including disordered items and lack of dimensionality of items. Such problems posit a threat to the reliability and the validity of the results based on these instruments (Cassidy et al., 2005; Colwell et al., 2013; Perlman et al., 2004). Other problems relate to specific measures. For example, for the CIS, IRT analyses have revealed that it does not distinguish well between caregivers in the high versus the moderate range (Colwell et al., 2013). For

another measure assessing quality of teacher-child interactions in pre-kindergarten, the CLASS Pre-K (Pianta, La Paro, & Hamre, 2008), a recent study has found a different factor structure than the originally proposed factor structure (Hamre, Hatfield, Jamil, & Pianta, 2013).

The CLASS Pre-K has undergone some changes since its original release. The first version consisted of a two-factor structure (the domains Emotional Support and Instructional Support), with seven underlying dimensions (Clifford, 2005; Howes et al., 2008; Mashburn et al., 2008). A later version of the CLASS Pre-K was extended to a three-factor structure, distinguishing between the domains Emotional Support, Classroom Organization and Instructional Support, with ten underlying dimensions. The three-factor structure has been validated in a large US sample of preschool and elementary classrooms (Hamre, Pianta, Downer et al., 2013) and is now widely applied in US and European studies (Weiland, Ulvestad, Sachs, & Yoshikawa, 2013; Cadima et al., 2010; von Suchodoletz et al., under review). However, a recent Finnish study revealed that not all dimensions fit well in the three-factor model (Pakarinen et al., 2010). For instance, the dimension Negative Climate, originally part of the Emotional Support domain, also loaded on Classroom Organization, which led the researchers to exclude the dimension from the model altogether (Pakarinen et al, 2010). New analyses with US samples showed that, in fact, a bifactor structure provided a better fit to the data than the original three-factor model (Hamre, Pianta, Downer et al., 2013). For the newer CLASS Toddler (La Paro, et al., 2011), although theoretically well grounded, little evidence on the measurement quality is available. In one study, correlations between the CLASS Toddler and structural quality aspects were investigated to evaluate the criterion validity (Thomason & La Paro, 2009). Results indicated that classroom quality was higher in smaller groups, with a favorable teacher-child ratio and a higher teacher education level. In addition, another study showed correlations between the CLASS Toddler and another classroom quality measure, the Infant/Toddler Environment Rating Scale Revised and associations with children's behavior problems (La Paro et al., 2014). However, this evidence is based solely on relatively small US samples and scholars have suggested that more large-scale studies are necessary to establish reliability and validity of this particular CLASS version (Thomason & La Paro, 2009).

Previous studies on the psychometric quality of the CLASS have only looked at the overall factor structure of the measure, providing information about the measure as a whole, but not on the quality of its items, that is, of the indicators that constitute the primary observations. Since item quality is crucial, there is a clear need for more information on the items before

we can evaluate the reliability and validity of the CLASS Toddler. The CLASS consists of a hierarchical structure with, on the highest level, two or three comprehensive domains, for example Emotional Support, which are usually evaluated with confirmatory factor analysis and then used in subsequent analyses to investigate associations with structural quality aspects or effects on children's developmental outcomes. Each domain is based on several dimensions, such as for example Positive climate, which are scored on a seven-point rating scale. However, the ratings on the dimensions are actually based on the evaluation of three to four indicators per dimension that specify the concrete and observable behaviors of teachers that observers can rate during the selected observation period. According to the CLASS manual, these indicators help observers assigning a score to the dimensions, which are then aggregated to a domain score. Thus, the indicators of each dimension are the primary sources of information about the observed processes. However, it still has to be established whether pooling the observation units into a dimension score is warranted. Moreover, the regular CLASS observation procedure prescribes a total of four observation cycles for each classroom, yielding a nested data structure. Although it is common in ECEC research to aggregate detailed observation measures to the classroom level, this not only means a loss of potentially relevant information (Hox, 2010), but also raises the question whether the primary measurements at the observation cycle level are psychometrically equivalent to the aggregated measurements at the classroom level.

To summarize, research on the psychometric properties of the CLASS Pre-K has yielded mixed findings on its domain structure and underlying dimensions. Regarding the CLASS Toddler, thorough research on the reliability and validity of the measure is virtually absent, which in view of the increasing use of the measure in high-stakes assessment, also outside the US, is a cause of concern. Furthermore, the available studies on the CLASS Toddler and Pre-K have taken a CTT approach and do not provide detailed information on the quality of the primary observation and scoring units of the CLASS, that is, the indicators underlying the dimensions. Information on the quality of the indicators is essential as they determine the quality of the assigned scores on the dimensions.

Current Study

The current study combined IRT and CTT approaches to investigate the psychometric quality of the CLASS Toddler. As observations done with the CLASS Toddler were based on the lowest level of the observed concrete teacher and child behavior, captured in the

CLASS indicators, we started with an examination of the proposed dimensional structure of the CLASS, which specifies eight dimensions based on three to four indicators each. These indicators were in line with the CLASS manual rated on an ordinal five point-scale, ranging from 'low', 'low/mid', 'mid', 'mid/high', to 'high'. Given the ordinal nature of the indicators, IRT analyses were conducted while accounting for the nested nature of the multiple observation cycles within a classroom. By taking the observation cycle as the unit of analysis all information and *within* classroom variation can be taken into account, such as, for instance, within classroom variation that is related to the type of activity or setting. Next, the overarching domain structure of the CLASS dimensions was investigated, following a CTT approach by employing multilevel confirmatory factor analyses. The original two-factor model (La Paro et al., 2011) was tested against a one-factor model and a three-factor model. In addition, using the results of the previous step regarding the domain structure of the CLASS, the analysis focused again on the indicator level by evaluating item difficulty and discrimination. Finally, using the best fitting model of the domain-structure, the criterion validity of the CLASS Toddler was evaluated by investigating associations between observed process quality, on the one hand, and several structural quality aspects and curriculum, on the other hand. We focused on characteristics that have been examined in other studies and were found to be related to process quality (Burchinal, Cryer, Clifford, & Howes, 2002; Cryer et al., 1999; Kuger & Kluczniok, 2008; Mashburn et al., 2008).

The Dutch ECEC system consists of two main types of provision. The first type is center-based day care for children from birth until four years of age that children attend, on average, for two full days a week (NCKO, 2011). The second type concerns preschools for two- to four-year old children, which are attended for two to four half days a week, and often offer an education program targeted at children with lower socioeconomic backgrounds. At age four, almost all children in the Netherlands enter primary school (Van Tuijl & Leseman, 2007). The Dutch ECEC system is strongly regulated regarding structural quality aspects, such as group size and teacher-education level, but there is still considerable variance regarding process quality (Helmerhorst et al., 2014; De Kruif et al., 2009).

In summary, the main aim of the current study was to investigate the psychometric properties of the CLASS Toddler, including its factor structure and item quality. In addition, criterion validity was examined by relating the CLASS domains to structural teacher, classroom and curriculum characteristics. We expected associations between the CLASS domains and structural quality characteristics, including group size, children-to-teacher ratio, teacher's education level, work experience and ethnic group composition, based on

prior research (Burchinal et al., 2002; Cryer et al., 1999; Howes et al., 2008; Mashburn et al., 2008; Pianta et al., 2005; Thomason & La Paro, 2009). However, previous research from the Netherlands has shown mostly weak to moderate associations between structural aspects and process quality, presumably due to the strong regulations concerning structural quality characteristics in the Netherlands (De Kruif et al., 2009; Fukkink, Gevers-Deynoot-Schaub, Helmerhorst, Bollen, & Riksen-Walraven, 2013; Slot et al., under review; Vermeer et al., 2008). Therefore, we expected the magnitude of the associations to be small. We also expected weak to moderate associations between the CLASS domains and curriculum characteristics. Few studies to date have examined both observed process quality and curriculum, either with observations or teacher's self-reports, and correlations in these studies were typically weak. These studies have mostly investigated associations between process quality and academic activities, such as literacy and math (Howes et al., 2008; Kuger & Kluczniok, 2008), and found stronger associations when the two measures used addressed similar contents (Kuger & Kluczniok, 2008). In addition, play has been shown to be an important part of the ECEC curriculum both in the US and in the Netherlands (e.g., Chien, et al., 2010; Slot et al., under review). Therefore, we focused on the provision of play, literacy and math activities, and we expected associations with the CLASS domains, particularly with observed support for learning.

Method

Participants

The present study used data from the ongoing national cohort study pre-COOL, which investigates the effectiveness of preschool education and care provisions in the Netherlands. Pre-COOL was commissioned by the Dutch Ministry of Education, Culture and Sciences and the National Science Foundation. The cohort started in 2010, when children were about two years old. At age five, children will enter the national cohort study COOL on students' careers in primary and secondary education, and they will be followed-up until age eighteen. To increase the likelihood of pre-COOL children entering primary schools that take part in COOL, the sample was recruited in the following way. First, a random sample of 300 primary schools participating in COOL was drawn from the COOL cohort. Next, the 139 primary schools that agreed to participate (46.3%) were asked to identify the preschools and day care centers that were attended by most of their new students. Municipal records and the internet were used to identify additional preschools and day care

centers in the neighborhood of the COOL schools. About 500 centers were approached, of which 263 agreed to participate in pre-COOL (52.6%). For logistic reasons, observations were only conducted in centers with more than four children who had also participated in the child assessments of pre-COOL. This yielded 162 centers (61.6%) with a total of 276 classrooms, of which 155 preschools and 121 day care centers. In addition, for the purpose of investigating criterion validity we used teacher report data. 375 teachers of 182 centers (69.2% of the total sample) participated in the study by filling out a teacher questionnaire, providing information on 295 classrooms (170 preschool, 125 day care). Almost all teachers were women (99.2%) and predominantly Caucasian (89.4%). Regarding the evaluation of the criterion validity of the CLASS Toddler, complete observation data and self-reports were available for 110 classrooms (39.8%).

The present study focused on provisions for two- to four-year-old children, but in 63% of the day care classrooms, also a few younger children were present. The participating preschools and day care centers were geographically spread over all parts of the Netherlands and were located in urban, semi-urban and rural areas. Classroom composition with regard to children's age and ethnicity differed between day care centers and preschools, and is representative of the Netherlands.

Classroom Measures

Observed process quality.

The Classroom Assessment Scoring System Toddler (CLASS Toddler; La Paro et al., 2011) was used to assess classroom process quality. An officially approved Dutch translation of the CLASS manual was developed for the present study (Slot, Leseman, Mulder, & Verhagen, 2013). All observers were trained by a licensed CLASS trainer and achieved at least 80% agreement within one scale point for the dimensions with the CLASS trainer on an online test (average score was 86.4%; agreement by chance was 33%), as recommended by the developers of the CLASS. Following this online test, the trainer conducted live observations with all observers once, prior to data collection. Inter-observer agreement of the live observations was 89.9%. Each classroom was observed during one morning and all classrooms were observed within a three-month period after training. Following the CLASS manual, observers rated classroom processes and teacher behavior during four 15 to 20 minutes observation cycles on the observation morning.

Classroom quality was rated on eight dimensions, using 7-point scales ranging from 1 or 2 (*classroom is low on that aspect*); 3, 4 or 5 (*classroom is in the midrange*); and 6

or 7 (*classroom is high on that aspect*). Following the CLASS manual, two overarching domains were distinguished (La Paro et al., 2011). For the first domain, Emotional and Behavioral Support, the observed processes were evaluated on five dimensions: Positive Climate reflects the warmth, respect, and enjoyment displayed during interactions of the teacher and children; Negative Climate reflects the overall negativity expressed in the classroom by the teacher and the children (scores are reversed); Teacher Sensitivity is the extent to which the teacher is aware of and responsive to children's needs; Regard for Child Perspective captures the degree to which the teacher's interactions with children and classroom activities capture the children's interests, and the degree to which children's independence is encouraged. Behavior Guidance refers to the teacher's ability to promote positive behavior and redirect problem behavior. In the domain Engaged Support for Learning, observed processes were evaluated on three dimensions: Facilitation of Learning and Development considers how well the teacher facilitates activities to support children's learning and development; Quality of Feedback assesses the degree to which the teacher's feedback promotes learning and expands children's participation; Language Modeling refers to the extent to which the teacher fosters, models and encourages children's use of language.

Descriptive statistics on the CLASS dimensions are presented in Table 1.

Table 1. Descriptive Statistics of the CLASS Dimensions

CLASS dimension*	M	SD	Range
Positive Climate	5.42	1.17	1-7
Negative Climate	6.84	0.38	5-7
Teacher Sensitivity	5.34	1.08	2-7
Regard for Child Perspectives	4.24	1.34	1-7
Behavior Guidance	5.01	1.12	2-7
Facilitation of Learning and Development	3.73	1.35	1-7
Quality of Feedback	2.91	1.20	1-7
Language Modeling	3.22	1.29	1-7

*N= 1084 number of observation cycles with the CLASS

The CLASS dimension scores are based on the evaluation of three to four indicators. Each indicator is scored on an ordinal scale with five levels, which are scored separately as 'low'(1), 'low/mid' (2), 'mid' (3), 'mid/high'(4), to 'high' (5).These scores are then combined to assign a score on the aforementioned 7-point scale of each dimension. An overview of all the indicators is provided in Table 3 The CLASS manual provides clear guidelines on

how to weigh these indicator ratings in assigning a dimension score. For instance, when all indicators are in the low range, the final score will consequently also be in the low range, meaning a 1 or a 2. When one of the indicators is in the low range and the other indicator is in the mid range, this would result in a score at the low end of the mid range, or a 3. When all indicators are in the mid range, the final score would be exactly in the mid range, or a 4. The dimension *Positive Climate* consists of three indicators. For example, the indicator positive affect evaluates whether the teacher and children show enjoyment, enthusiasm and affection. *Negative Climate* consists of four indicators. An example is the indicator negative affect, which considers whether the teacher shows irritability, anger or uses a harsh voice in her interactions with the children. *Teacher Sensitivity* consists of three indicators. For instance, the indicator child comfort assesses whether the children feel comfortable in approaching the teacher and seeking support or help from the teacher. *Regard for Child Perspectives* consists of three indicators. An example is the indicator child focus, which reflects the teacher's provision of choices for children, following their lead and eliciting children's ideas. *Behavior Guidance* consists of three indicators. For instance, the indicator proactive evaluates the teacher's use of monitoring and communication of clear behavioral expectations for the children. *Facilitation of Learning and Development* consists of three indicators. An example is the indicator children's active engagement, which reflects children's physical and verbal involvement in activities. *Quality of Feedback* consists of three indicators. For example, the indicator scaffolding captures the teacher's use of hints, verbal or physical assistance and prompting children's thought processes. *Language Modeling* consists of four indicators. For instance, the indicator supporting language use captures the teacher's use of back-and-forth exchanges, contingent responding and open-ended questioning to elicit children's talk.

Self-reported developmental and educational activities.

A structured questionnaire for teachers was used to assess the developmental and educational activities provided to the children on a regular basis over a longer period of time. This questionnaire was carefully developed for the purposes of the current study, based on extant research into play, emergent literacy and emergent numeracy, and extensively tested in pilot research with teachers of two- and three-year-old children to ensure age-appropriateness of the listed activities (for more information, see Slot, Leseman, Verhagen, & Mulder, under review). Several scales were constructed covering a broad range of behaviors and activities. For the purpose of the current study, three types of activities were distinguished: free play activities, literacy activities and math activities.

The scale *Play* (9 items; Cronbach's alpha= .85) assesses the degree to which the teacher provides children with opportunities for free, self-managed play and enriches children's play, for instance by asking questions, making suggestions, or providing materials for richer play. Examples of items are: "I let the children play without interfering", "I ask children questions that stimulate their play". The scale ranges from 1 (*not applicable*) to 5 (*strongly applicable*).

The scale *Literacy activities* (4 items; Cronbach's alpha= .82) measures the average frequency with which activities are provided involving literacy and literacy materials. An example of an item is: "Asking the children questions about the content of the story during or after reading the story". Answers were rated on a 7-point scale, ranging from 1 (*never*) to 7 (*three or more times a day*).

The scale *Math activities* (12 items; Cronbach's alpha= .91) assesses the average frequency of several math activities, for instance counting and sorting activities, and activities exploring different shapes. An example of an item is: "Counting how many objects you have, for example counting till five and saying 'I have five marbles'". Answers were rated on a 7-point scale, ranging from 1 (*never*) to 7 (*three or more times a day*).

Structural classroom and center characteristics.

For the present purposes, the following structural quality variables were constructed based on observations and teacher reported background information.

Group size. CLASS observers recorded the number of children present during each observation cycle.

Children-to-teacher ratio. CLASS observers recorded the number of adults present during each observation cycle. The children-to-teacher ratio was calculated by dividing the number of children by the number of teachers.

Teacher's education. Teachers were asked to report their highest level of completed formal pre-service education on a scale representing the levels of Dutch secondary and tertiary education, ranging from 1 (*lower preparatory vocational track*) to 8 (*university*).

Work experience. Teachers were asked to report their work experience on a 7-point scale, ranging from 1 (*less than one year*) to 7 (*more than 30 years*).

Education program. Teachers were asked whether they used an education program. In the Netherlands, several educational programs, approved by the national Accreditation Committee for Child and Youth Interventions of the Netherlands' Youth Institute (www.nji.nl), are currently used in ECEC. Although these programs differ in how teachers are trained and monitored, they all aim at broad developmental and educational goals,

emphasize emotional support, sensitivity to children's needs, and provide a mixture of play and pre-academic activities with an emphasis on language and literacy. For the present purpose, a dummy variable was created with the values 1 = *yes* and 0 = *no*, indicating whether an education program was used, without further distinguishing between the programs.

Type of provision. The current study involved the two main types of ECEC in the Netherlands, namely day care and preschool. A dummy variable was created with the values 1 = *preschool* and 0 = *day care*.

Descriptive information on the center and classroom characteristics are shown in Table 2.

Table 2. Descriptive Statistics for Center and Teacher Characteristics

	Day care				Preschool			
	N	**M**	**SD**	**range**	**N**	**M**	**SD**	**range**
Children-to-teacher ratio[1]	485	5.3	2.1	1-14	597	5.0	2.3	1-16
Group size[1]	486	9.5	3.1	2-23	597	10.0	4.0	1-25
Professional development activities		2.91	.94	1.20-5.90		3.36	1.06	1.22-6.33
Frequency (N)/Percent (M)								
Educational program	149	55.6			205	97.6		
Classroom > 30% non-Dutch	36	23.5			110	55.0		
Age composition classrooms								
0 years	83	51.2						
1 year	99	61.1						
2 years	157	96.9			209	97.7		
3 years	102	63.0			92	43.0		
Educational level								
Lower preparatory vocational track	73	17.5			60	12.6		
4 years secondary vocational	174	41.8			197	41.5		
1 or 2 years intermediate vocational training	29	7.0			18	3.8		
3 or 4 years intermediate vocational training	60	14.4			78	16.4		
5 years secondary	55	13.2			95	20.0		
6 years secondary[2]	10	2.4			19	4.0		
Higher vocational	10	2.4			3	0.6		
University	5	1.2			5	1.1		
Teacher ethnicity								
Native Dutch	143	95.3			178	85.2		
Immigrant	6	4.7			31	14.8		

[1] based on the observation cycles
[2] entry level university

Procedures

Each classroom was observed during one morning with the CLASS Toddler. The observers rated classroom processes and teacher behavior during four 15 to 20 minutes cycles on the observation morning, resulting in a total of 1084 observation units. Visits were done for three to four hours during a regular morning in the classroom that was typical for the usual environment and routines (i.e., not during a day when a field trip was planned).

Analysis Strategy

Reliability and validity of the CLASS Toddler were examined by combining CTT and IRT approaches, analyzing the data both on the item level (referred to as indicators in the CLASS) and on the level of the usual subscales (referred to as dimensions in the CLASS). First, following IRT, a multilevel confirmatory factor analysis for categorical data was performed in order to examine the dimensional structure of the indicators with the observation cycles as level 1 and the classrooms as level 2 units, using Mplus (Mplus 7; Muthén & Muthén, 1998-2012). Second, following CTT, the structural validity of the overall domain structure of the CLASS dimensions was examined by means of multilevel confirmatory factor analyses. The originally proposed two-factor model was tested against an alternative one-factor model and an alternative three-factor model, using the observation cycles as the level 1 and the classrooms as level 2 units of analysis. Third, based on the best fitting domain structure, IRT analyses were conducted to assess the psychometric quality of the CLASS indicators. Two common item statistics were computed: item difficulty (by calculating the mean of all thresholds per item) and item discrimination (represented by the standardized factor loadings). Item difficulty is the average difficulty level that locates the indicator along the latent quality scale. Difficulty estimates are represented on a Logit scale, with the mean arbitrarily set to zero and with lower (negative) values indicating easier items (which means most teachers are most likely to be rated in the higher score categories of these indicators), values around 0 indicating average difficulty and higher (positive) values indicating harder items (which means that teachers are less likely to be rated in the higher score categories of these indicators) (de Ayala, 2013). Discrimination estimates, expressed as standard scores with values between 0 and 1, give an indication of the relative amount of information provided by each indicator.

Model fit in CTT was evaluated with several fit indicators: the ratio of the Chi-Square and the degrees of freedom (Chi-Square/df), the Comparative Fit Index (CFI), the Tucker Lewis Index (TLI), the Root Mean Square Error of Approximation (RMSEA) and

the Standardized Root Mean Square Residual (SRMR) at both the within and between classrooms level, with Chi-Square/df < 3, CFI/TLI > .95, RMSEA, $SRMR_{within}$ and $SRMR_{between}$ < .05, indicating good fit and < .08 indicating acceptable fit. The Chi-Square/df ratio was used as an indicator of model fit rather than the Chi-Square test, which is very sensitive to small violations of the multi-normality assumption with large samples (Kline, 2005). Improvement of the model fit was evaluated by testing the significance of the change in Chi-Square relative to the change in degrees of freedom.

All IRT analyses were run using a mean and variance-adjusted weighted least squares (WLSMV) estimator, which provides good approximations of parameter estimates with ordinal data (Hill et al., 2007). Limited evidence exists on the appropriateness of the usual fit indices, such as the Chi-square test or CFI, to evaluate model fit of IRT models (Reise, Widaman, & Pugh, 1993). In IRT analyses, generally, the number of degrees of freedom is so large that it is difficult to find a model that fits the data well according to the usual fit indices (Maydeu-Olivares, 2013). Furthermore, the Chi-square test is not applicable when five or more categories are used and the number of variables is more than six (Thissen & Steinberg, 1997), as in the current study. One fit index that has been shown to provide reliable results for IRT models is the RMSEA (Maydeu-Olivares, Cai, & Hernandez, 2011). Hence, model fit was evaluated based solely on the RMSEA with the usual cut-off points (i.e., a value below .08 indicating acceptable fit and below .05 indicating good fit).

A final analysis was conducted to determine criterion validity of the best fitting domain structure by examining relations between observed process quality, on the one hand, and report-based teacher, classroom, and curriculum characteristics, on the other hand. For continuous variables, Pearson correlations were computed; for dichotomous variables *T*-tests were calculated. Due to the design of the study and non-response, for this part of the analysis complete observational and self-report data were available for only 39.8% of the classrooms. Classrooms with and without self-reports did not differ significantly on any of the CLASS dimensions. However, classrooms with observations had slightly lower scores on self-reported math activities than classrooms for which no observations were available (standardized effect size is .10). There were no further missing data in the observation measures, and occasional missing data on activities and structural classroom characteristics in the self-reports were below 8%. Missing data were dealt with by using full information maximum likelihood (FIML) estimation in Mplus (Enders, 2010), in which the standard errors for the parameter estimates are computed using the complete observed information matrix (Muthén & Muthén, 1998-2012).

Results

As can be seen in Table 3, the scoring pattern is not evenly distributed across all categories of the CLASS dimensions. Notably, Negative Climate hardly revealed any variation, with all scores in the low-to-mid range. Furthermore, although all categories were observed, most scores were in the mid-to-high range for Positive Climate, Teacher Sensitivity, Regard for Child Perspectives and Behavior Guidance. Facilitation of Learning and Development, Quality of Feedback, and Language Modeling showed the reversed pattern with most scores in the low-to-mid range and few scores in the mid-to-high range.

Table 3. Assigned at Each Rating Category

Dimensions	Indicator	Low (1)	Low/Mid (1.5)	Mid (2)	Mid/High (2.5)	High (3)
Positive Climate	Relationships	1%	4%	34%	24%	37%
	Positive affect	1%	4%	34%	23%	37%
	Respect	1%	2%	30%	20%	47%
Negative Climate	Negative affect	92%	3%	5%	0%	0%
	Punitive control	94%	3%	4%	0%	0%
	Teacher negativity	98%	1%	1%	0%	0%
	Child negativity	90%	5%	4%	0%	0%
Teacher Sensitivity	Awareness	1%	5%	42%	24%	29%
	Responsiveness	1%	3%	38%	23%	35%
	Child comfort	0%	2%	27%	23%	48%
Regard for Child Perspectives	Child focus	12%	15%	40%	14%	19%
	Flexibility	5%	11%	43%	14%	27%
	Support of independence	14%	13%	53%	11%	8%
Behavior Guidance	Proactive	2%	4%	41%	24%	29%
	Supporting positive behavior	5%	7%	48%	20%	21%
	Problem behavior	1%	4%	37%	20%	39%
Facilitation of Learning and Development	Active facilitation	26%	16%	37%	11%	10%
	Expansion of cognition	38%	19%	30%	8%	6%
	Children's active engagement	2%	11%	46%	14%	27%
Quality of Feedback	Scaffolding	41%	16%	35%	5%	3%
	Providing information	40%	14%	39%	4%	3%
	Encouragement and affirmation	28%	18%	44%	5%	6%
Language Modeling	Supporting language use	21%	17%	42%	8%	13%
	Repetition and extension	30%	17%	40%	6%	7%
	Self- and parallel talk	38%	11%	40%	5%	6%
	Advanced language	31%	20%	38%	6%	6%

Structural Validity

First, a multilevel confirmatory factor analysis for categorical data was performed on all the indicators of the CLASS to test the eight-dimensional structure of the CLASS using an IRT approach. Details of this model are given in Table 4.

The RMSEA was .04, which indicates overall good fit. As can be seen in Table 4, most factor loadings were acceptable to good for all dimensions, particularly at the between classroom level. Regarding the dimension Negative Climate the indicator teacher negativity was problematic. The variance was very limited, as is also evident from Table 2, and could not be reliably estimated in Mplus, resulting in a low non-significant factor loading at the within level and a high factor loading exceeding 1 at the between level, which explained the negative residual variance for this indicator. The problem remained after fixing the residual variance to .001, indicating that this factor loading could not be readily interpreted. In addition, the correlations presented in Table 5, show that the correlation between Negative Climate and Behavior Guidance was high, indicating that these dimensions may actually represent one factor. Similarly, the correlations between a few other dimensions were also very high, particularly concerning Facilitation of Learning and Development, Quality of Feedback and Language Modeling (see Table 5), which again suggests that these dimensions may represent one overarching factor.

Next, the proposed two-factor model with the domains Emotional and Behavioral Support and Engaged Support for Learning as overarching domains was evaluated using a multilevel confirmatory factor analysis with the eight CLASS dimensions. Model testing revealed overall good fit (see Table 6). However, further inspection of the standardized factor loadings showed a relatively low factor loading for Negative Climate both at the *within* level (.35) and the *between* level (.45) on the Emotional and Behavioral Support domain (see Table 6). Next, the two-factor model was tested against two alternative models. The first alternative model was a one-factor model with one general classroom process quality factor. The second alternative model was a three-factor model. The multilevel one-factor model showed poor overall model fit and a significant deterioration of the model fit compared to the two-factor model, ($\Delta\chi^2(2)=310.02$, $p=.00$). Also, the factor loadings were generally lower for all dimensions at both the *within* and *between* level, particularly for the dimensions Negative Climate and Behavior Guidance. The second alternative model was a multilevel three-factor model based on the inter-correlations presented in the previous section, with the domains Emotional Support (indicated by the dimensions Positive Climate, Teacher Sensitivity, and Regard for Child Perspectives), Behavioral Support (indicated by Negative Climate

Table 4. Factor Loadings at the Within and Between Level for all Indicators in the Eight-Factor Model (Based on IRT Analyses)

Dimensions	Indicator	Factor loading Within level	Factor loading Between level
Positive Climate	Relationships	.57	.77
	Positive affect	.64	.85
	Respect	.56	.99
Negative Climate	Negative affect	.65	.85
	Punitive control	.47	.46
	Teacher negativity	.21	1.00[1]
	Child negativity	.58	.57
Teacher Sensitivity	Awareness	.61	.71
	Responsiveness	.60	.90
	Child comfort	.52	.88
Regard for Child Perspectives	Child focus	.83	.84
	Flexibility	.77	.73
	Support of independence	.48	.69
Behavior Guidance	Proactive	.58	.80
	Supporting positive behavior	.53	.65
	Problem behavior	.60	.74
Facilitation of Learning and Development	Active facilitation	.75	.91
	Expansion of cognition	.81	.85
	Children's active engagement	.51	.76
Quality of Feedback	Scaffolding	.54	.54
	Providing information	.72	.84
	Encouragement and affirmation	.51	.77
Language Modeling	Supporting language use	.75	.80
	Repetition and extension	.61	.79
	Self- and parallel talk	.36	.52
	Advanced language	.68	.65

[1] value after constraining the residual variance to 0.001

Table 5. Between-Level Intercorrelations Among CLASS Dimensions In the Eight-Factor Model (Based on IRT Analyses)

CLASS dimensions	1	2	3	4	5	6	7	8
1 Positive Climate		.56***	.83***	.79***	.68***	.81***	.83***	.92****
2 Negative Climate			.59***	.64***	.76***	.41***	.61***	.51***
3 Teacher Sensitivity				1.04***	.96***	.76***	.73***	.80***
4 Regard Child Pers.					.95***	.97***	1.01***	.90***
5 Behavior Guidance						.61***	.63***	.61***
6 Facilitation Learning							1.04***	1.06***
7 Quality of Feedback								1.10***
8 Language Modeling								

*** $p < .001$

and Behavior Guidance), and Engaged Support for Learning (indicated by Facilitation of Learning and Development, Quality of Feedback, and Language Modeling). The three-factor model showed good model fit and significantly improved model fit compared to the two-factor model ($\Delta\chi^2(4)=19.07$ $p=.00$) and the one-factor model ($\Delta\chi^2(6)=329{,}09$ $p=.00$). In addition, the factor loadings of both Negative Climate and Behavior Guidance were higher in the three-factor model compared with the factor loadings found in the two-factor model, both at the *within* and at the *between* level. To conclude, the three-factor model showed the best fit to the data and revealed a highly comparable factor structure at both the level of the observation cycles and the classroom level.

Following the previous analyses in which we established that the three-domain model revealed the best model fit and a robust factor structure within and between classrooms, we reexamined the three-domains model with IRT analyses to assess item difficulty and item discrimination parameters for all CLASS indicators, as shown in Table 7. The results for item difficulty show that almost all indicators of Positive Climate, Teacher Sensitivity, Regard for Child Perspectives and Behavior Guidance were negative or close to zero, indicating that teachers are likely to have high scores on these dimensions, also reflecting the skewness in the scores distribution shown in Table 2.

Table 6. Measures of Fit for the Multilevel One-Factor, Two-factor, and Three-Factor Models (Based on CTT)

Factor loadings at the *within (W)* and *between (B) level*	**Two-factor model**		**One-factor model**		**Three-factor model**	
	W	**B**	**W**	**B**	**W**	**B**
Positive Climate	.66	.81[1]	.49	.64	.67	.80[1]
Negative Climate	.35	.45[1]	.09	.09	.37	.51[2]
Teacher Sensitivity	.69	.90[1]	.49	.51	.70	.89[1]
Regard for Child Perspectives	.12	.89[1]	.17	.53	.13	.89[1]
Behavior Guidance	.59	.78[1]	.46	.43	.65	.94[2]
Facilitation of Learning and Development	.70	.95[2]	.71	.72	.71	.95[3]
Quality of Feedback	.67	.89[2]	.57	.65	.67	.88[3]
Language Modeling	.68	.95[2]	.53	.88	.68	.95[3]
Measures of fit information						
Chi-square value	101.57		411.59		82.50	
d.f.	38		40		34	
Ratio Chi-square/d.f.	2.67		10.29		2.43	
RMSEA	.04		.09		.04	
CFI	.97		.83		.98	
TLI	.96		.77		.96	
$SRMR_{within}$	.03		.06		.03	
$SRMR_{between}$	.05		.09		.04	

[#] values with the same superscript numbers belong to the same factor

The dimension Negative Climate showed the lowest item difficulty indicating that it was easy for teachers to receive a high score on this dimension. The dimensions Behavior Guidance, Facilitation of Learning and Development, Quality of Feedback, and Language Modeling revealed the reverse pattern, indicating that teachers were less likely to have high scores on the indicators underlying these dimensions. Discrimination values of all dimensions ranged from .412 to .803, indicating that the items showed moderate to good discrimination.

Table 7. Overall Item Difficulty and Item Discrimination Based on the Three-Factor Model

Domain	Dimensions	Indicator	Difficulty	Discrimination
Emotional Support	Positive Climate	Relationships	-.890	.703
		Positive affect	-.861	.800
		Respect	-1.093	.802
	Teacher Sensitivity	Awareness	-.692	.654
		Responsiveness	-.840	.775
		Child comfort	-1.160	.737
	Regard for Child Perspectives	Child focus	-.116	.755
		Flexibility	-.449	.877
		Support of independence	.140	.610
Behavioral Support	Negative Climate (reversely coded)	Negative affect	-.1.526	.905
		Punitive control	-1.667	.626
		Teacher negativity	-2.236	.659
		Child negativity	-1.523	.699
	Behavior Guidance	Proactive	-.692	.693
		Supporting positive behavior	-.417	.610
		Problem behavior	-.885	.651
Engaged Support for Learning	Facilitation of Learning and Development	Active facilitation	.338	.751
		Expansion of cognition	.676	.795
		Children's active engagement	-.495	.527
	Quality of Feedback	Scaffolding	.833	.504
		Providing information	.788	.734
		Encouragement and affirmation	.536	.591
	Language Modeling	Supporting language use	.235	.821
		Repetition and extension	.537	.707
		Self- and parallel talk	.670	.513
		Advanced language	.590	.704

Criterion Validity

Table 8 shows the correlations between the three CLASS domains and several teacher and classroom characteristics at the between groups level. There were a few small-

sized correlations between the CLASS factors and structural quality characteristics. For instance, a higher children-to-teacher ratio was related to both lower Emotional Support and Engaged Support for Learning. Teachers' work experience was related to Engaged Support for Learning. Furthermore, in classrooms with a lower proportion of children from monolingual Dutch families, teachers showed higher Engaged Support for Learning. Likewise, preschools had significantly higher levels of Engaged Support for Learning, which is in line with the targeted approach of providing ECEC for at-risk children in preschools. In addition, there were several expected, weak to moderate correlations between observed classroom quality and the self-reported type of activities provided by the teachers, particularly for the Engaged Support for Learning domain. The provision of literacy, and to a lesser extent the provision of play, was positively related to observed process quality. No correlations were found with the provision of math activities.

Table 8. Associations Between CLASS Domains and Teacher Characteristics

	Emotional Support	Behavioral Support	Engaged Support for Learning
Categorical classroom characteristics and CLASS domains and t-tests between subgroups			
Education program			
Yes	4.94	5.92	3.25
No	5.02	5.78	3.23
Type of provision			
Day care	5.04	5.88	3.08_a
Preschool	4.96	5.94	3.44_a
Continuous teacher and classroom characteristics and Pearson correlations			
Pre-service education level	-.03	.05	-.01
Group size	-.06	-.02	-.03
Children-to-teacher ratio	$-.12^{\dagger}$	-.10	-.16**
Work experience	.14	.16	.19*
% of children speaking little Dutch	.03	-.04	$.16^{\dagger}$
Provision of activities (based on self-reports) and Pearson correlations			
Play	.19*	.07	.27*
Literacy	$.17^{\dagger}$	.22*	.25**
Math	-.05	-.05	.06

Note. Values with the same subscript letters differ significantly at $p < .05$; $^{**}p < .01$, $^{*}p < .05$, $^{\dagger}p < .10$

Discussion

Quality measures for evaluating the process quality of ECEC provisions, such as the CLASS, are being increasingly used in research and for accountability and professionalization purposes with important implications for policy and practice. However, thorough research

on the reliability and validity of these measures, particularly for a recent version of the CLASS, the CLASS Toddler, is still lacking. The current study aimed to fill part of this gap by investigating the psychometric quality of the CLASS Toddler. We combined a CTT approach with an IRT approach in order to gain a better understanding of the measurement quality of the CLASS Toddler. Also, we used all available information at the level of the observation cycles by conducting multilevel analyses. IRT has several advantages compared to the CTT approach. Most importantly, IRT can provide more detailed information on the quality of a measure, including the item difficulty and item discrimination, which, in contrast to similar parameters in the CTT approach, are not dependent on the characteristics of a particular measure or a particular sample (de Ayala, 2013; Hambleton & Jones, 1993).

The CLASS Toddler has a hierarchical structure with on the highest level the domains. Domain scores are typically used in studies investigating associations with structural quality aspects and effects on children's developmental outcomes, and also for policy and accountability purposes. These domain scores are the average of several more specific dimension scores (e.g., Positive Climate), which themselves are derived from the ratings on three to four indicators per dimension. Considering this hierarchical structure, we first tested whether it is justified to assign CLASS dimension scores based on the indicators underlying these dimensions, by performing a multilevel confirmatory analysis with the indicators using an IRT approach to take into account the categorical data structure. The results indicated that the dimensional structure is generally robust and in accordance with the theoretically specified structure. Therefore, in a next step, we used multilevel confirmatory factor analyses in order to test the proposed two-domain structure against a one-domain and a three-domain structure. This analysis showed that the three-domain structure provided the best fit to the data and indicated that the domain structure was robust within and between classrooms. Hence, we reexamined the three-domain structure with IRT analyses to evaluate the quality of the indicators relative to the three-domain structure by estimating the difficulty and discrimination parameters. Finally, the criterion validity of the CLASS Toddler was explored by examining the concurrent relations between the three domains with relevant structural quality and curriculum aspects as reported by teachers.

The IRT analysis of the dimensional structure of the CLASS Toddler confirmed that all indicators fitted well into the dimensions. Although not all factor loadings within one dimension were equally high, we found that the overall dimensional structure of indicators

was robust. However, there were some problems with the dimension Negative Climate. The dimension Negative Climate revealed very limited variance, with most scores in the low range. This is most likely due to the way in which the behaviors captured in this dimension are operationally defined, resulting in a skewed distribution of scores. For instance, to receive a score in the mid or high range there must be evidence of severe instances of teacher or child negativity, including the use of physical force. One indicator in particular, teacher negativity, showed extremely restricted variance, which led to computational problems in estimating the factor loadings of this indicator on the dimension Negative Climate. Despite the problems in estimating the factor structure based on the indicators, this indicator may still be a conceptually relevant aspect of Negative Climate and, therefore, can be maintained in the dimension. Furthermore, the results also revealed very high correlations between some of the dimensions. For instance, Facilitation of Learning and Development, Quality of Feedback and Language Modeling showed very high inter-correlations, suggesting that these dimensions, although conceptually tapping different constructs, are psychometrically very similar. Also, we found that the correlation between Negative Climate and Behavior Guidance was high, in line with previous European research (Pakarinen et al., 2010; von Suchodoletz et al., under review).

Multilevel confirmatory factor analyses testing the domain structure showed that, although the proposed two-domain model (La Paro et al., 2011) fitted the data well, a three-factor structure provided the best fit to our data. Following findings of the first analysis of the dimensional structure of the CLASS Toddler, a third factor was distinguished consisting of Negative Climate and Behavior Guidance. A previous Finnish study using the CLASS Pre-K also reported problems with the dimension Negative Climate in a factor analysis, due to the pattern of correlations with other dimensions, including Behavior Guidance (Pakarinen et al., 2010). These findings might point to cultural differences between the US and European countries. For instance, in the Netherlands, there was little variation in the Negative Climate dimension, considerably less than in the US (La Paro et al., 2014; Thomason & La Paro, 2009), which is in line with the findings in Finnish and German kindergartens (Pakarinen et al., 2010; von Suchodoletz et al., under review). Most variation was found in the indicators child negativity and negative affect. Thus, Negative Climate in the Dutch context seems to reflect overall mild negativity in the classroom, but not the occurrence of more extreme teacher behavior such as teasing or humiliation, or the use of threats and physical force to establish control. Therefore, the correlation between Negative Climate and Behavior Guidance seems to reflect that less negativity was expressed in the

classroom as a result of teachers offering more behavioral support, through supporting and reinforcing positive behavior.

Our multilevel analyses revealed different factor loadings regarding the dimension Regard for Child Perspectives at the *within* and *between* level, which may point to cultural differences as well. The low factor loading at the *within* level indicated that the dimension did not fit well into the Emotional Support domain, whereas the factor loading at the *between* level was acceptable indicating a good fit into this domain. To understand this seemingly contradictory finding it is important to consider the observed settings in Dutch ECEC. Early childhood settings in the Netherlands are characterized by a rather fixed schedule of daily activities and routines (Slot et al, under review), including ample time for free play, mealtimes and whole group activities such as circle time or crafts. Previous research has shown that Regard for Child Perspectives received much higher ratings during free play compared to the more structured settings, such as whole group activities and mealtimes, resulting in large variation across the observation cycles, that is, *within* classroom variance (Slot et al., under review). Although, this dimension showed a better fit within the Emotional Support domain at the *between* level, it raises the question whether it is justified to incorporate a quality aspect that does not seem to capture the same concept at the *within* and *between* classroom level.

The quality of the indicators of the CLASS was found to be adequate overall. Although teachers were more likely to receive high scores on the indicators of the domains Emotional Support and Behavioral Support than on the indicators of the domain Engaged Support for Learning, there was still considerable variance. Moreover, the discrimination values indicated that, overall, the indicators were able to distinguish well between teachers in both the low and high quality range. Interestingly, the indicators of Emotional Support and Engaged Support for Learning had equally high discrimination values, indicating that the differences in item difficulties (the indicators of Emotional Support being 'easier' than the indicators of Engaged Support for Learning) most likely reflect that Dutch ECEC teachers are more able to provide high emotional process quality than they are able to do regarding educational process quality, and do not point to a methodological artifact.

Finally, the criterion validity of the CLASS Toddler was investigated by relating the three domains to structural quality aspects known to be related to process quality. In line with our expectations, several small-sized correlations between CLASS domains and structural aspects were found, confirming the criterion validity of the CLASS Toddler. Structural aspects, such as children-to-teacher ratio, teacher's work experience and teacher's pre-

service education level were all related to process quality, in line with previous findings with the CLASS Toddler (Thomason & La Paro, 2009) as well as with findings in Dutch studies using similar observation instruments (De Kruif et al., 2009; Fukkink et al., 2013; Slot et al., under review; Vermeer et al., 2008). Furthermore, several correlations between the CLASS factors and teachers' self-reported educational and play activities were found, particularly with Engaged Support for Learning. In addition, Engaged Support for Learning was higher in preschools with higher proportions of children learning Dutch as a second language, which is in line with the objective of these preschools to combat early arising educational disadvantages.

There are some limitations to the present study. A first limitation concerns the missing values in teachers' self-reports on data we used to assess the validity of the CLASS. However, we carefully checked whether there were systematic differences between centers with and without self-reports and this was generally not the case. Secondly, although the psychometric quality of the CLASS Toddler could be confirmed, more evidence is needed on the predictive validity of the CLASS Toddler for children's developmental outcomes, as the CLASS Toddler, like all other process quality measures, has been developed as an instrument to help improve classroom quality and, consequently, children's developmental outcomes. Future studies should incorporate children's socio-emotional and cognitive outcomes in assessing the validity of the CLASS Toddler. In this respect, it is noteworthy that most previous studies using the CLASS Pre-K found only modest effects of classroom quality on children's developmental and educational outcomes (Burchinal et al., 2011; Curby et al., 2009; Mashburn et al., 2008), raising the question whether the use of a classroom observation measure, like the CLASS Toddler or CLASS- Pre-K, is sufficiently sensitive to capture the processes that determine child outcomes. Classroom quality measures appear to be rather global measures and some scholars have pointed to the need to use domain-specific quality measures in addition to detect developmental effects (Bryant et al., 2010; Downer, Sabol, & Hamre, 2010; Zaslow et al., 2010). Moreover, observation measures are particularly suited to provide an assessment of real-time interactional quality, but may provide less information on the provision of developmental and educational activities over a longer period of time. Perhaps, process quality constructs can be strengthened by including domain-specific aspects of classroom quality relating to the curriculum (Sylva et al., 2006) and by using a multi-method approach to the assessment of quality, for instance when observations in real-time are combined with teacher reports over a longer period to integrate different time-scales in the quality assessment (Kuger & Kluczniok, 2008; Slot et al., under review).

To conclude, the current study adds to the existing evidence on the psychometric quality of the CLASS Toddler in several ways. To the best of our knowledge, only few studies to date have systematically examined the measurement properties of the CLASS Toddler. These studies and related studies on the CLASS Pre-K, however, have focused solely on the psychometric quality of the dimensions and domains of the CLASS, but did not evaluate the measurement properties of the primary sources of information from which the dimension and domain scores are derived, that is, the CLASS indicators. The present study replicated previous psychometric studies but in a different national context, and also examined the psychometric properties of the behavioral indicators. Our IRT and CFA analyses revealed good to excellent measurement properties of the indicators, confirmed the dimensional structure, and found an overarching three-domains structure of the CLASS Toddler as fitting the data best. Thus, the CLASS Toddler is an observational measure of ECEC classroom quality that can be used in a three-step procedure of first evaluating the concrete behavioral indicators, then assigning dimension scores based on these indicators, and finally averaging dimension scores to obtain overarching domain scores. Working according to this systematic procedure not only helps observers in assigning dimension scores more reliably, but also provides rather detailed information on classroom quality that can be used for providing feedback to teachers and other professionalization purposes.

Also, whereas most studies to date have aggregated information from several observation cycles to a single aggregated classroom score, we examined the dimension and domain structure of the CLASS Toddler at both the cycle and classroom level. Our findings showed reasonable invariance of the dimensional and domain structure of the CLASS Toddler across these levels, indicating that basically the same constructs are measured by the CLASS at both levels.

Finally, some of the issues that emerged in our analyses with regard to, for instance, the Negative climate dimension, may point to the effect of cultural differences in ECEC and not to inadequate measurement quality per se. Future research in other countries is needed to further explore cultural differences, but also to further confirm the cross-culturally common basis that is reflected in the CLASS Toddler. To conclude, the results of this study indicate good psychometric quality and validity of the CLASS Toddler.

CHAPTER 3

Associations Between Structural Quality Aspects and Process Quality in Dutch Early Childhood Education and Care Settings

Slot, P.L., Leseman, P. P. M., Verhagen, J., & Mulder, H. (2014)

Associations between Structural Quality Aspects and Process Quality in Dutch Early Childhood Education and Care Settings.

Manuscript under revision

Abstract

The relationship between structural quality and process quality in early childhood education and care (ECEC) has been addressed in several studies. However, the findings are not conclusive. The present study was conducted in the Netherlands, which has a strongly regulated mid-quality ECEC system regarding structural aspects, with still considerable variation in process quality. The study employed a multi-method approach and extended the existing research in two ways: first, by defining emotional and educational process quality as comprehensive constructs, including observations of the teacher-child interactions as well as teacher-reported developmental and educational activities offered on a larger time scale; and second, by extending the traditional structural quality characteristics with measures of professional development activities and the use of an education program as predictors of process quality. Results indicate that group size and children-to-teacher ratio are not related to emotional and educational process quality in the Dutch ECEC system. Teachers' formal pre-service education has a positive but small effect on emotional and educational process quality. The use of an education program and professional development activities at the center have the largest effects on emotional and educational process quality. Implications for policy and practice are discussed.

Introduction

Numerous studies have shown beneficial effects of high quality center-based early childhood education and care (ECEC) on children's social-emotional and cognitive development (Burchinal, Vandergrift, Pianta, & Mashburn, 2010; Campbell, Pungello, Miller-Johnson, Burchinal, & Ramey, 2001; Curby et al., 2009; Howes et al., 2008; Mashburn et al., 2008; NICHD & Duncan, 2003). ECEC quality is commonly defined by the structural and process characteristics that are thought to nurture child development (Howes et al., 2008; Layzer & Goodson, 2006; Sylva et al., 2006; Thomason & La Paro, 2009). Process quality refers to the child's day-to-day experiences in ECEC settings and encompasses the social, emotional, physical, and instructional aspects of children's activities and interactions with teachers, peers, and materials, that are seen as the proximal determinants of child development (Howes et al., 2008; Pianta et al., 2005; Thomason & La Paro, 2009). Structural characteristics of ECEC, such as group size, children-to-teacher ratio, and teachers' qualifications (e.g. Howes et al., 2008; Thomason & La Paro, 2009) are

the distal and regulable aspects of ECEC, and are regarded as important preconditions of proximal process quality (e.g. Cryer, Tietze, Burchinal, Leal, & Palacios, 1999; Philips, Mekos, Scarr, McCartney, & Abbott-Shim, 2000; Philipsen, Burchinal, Howes, & Cryer, 1997; Pianta et al., 2005; Vandell, 2004). Structural quality is the main objective of statutory quality regulations and national curricula (Bennett, 2005; Philipsen et al., 1997), and a major factor in the macroeconomic costs of ECEC, whereas the potential benefits for individuals and society are primarily dependent upon process quality (Vandell et al., 2010). A positive relationship between structural and process quality, therefore, is essential for the costs-efficiency of ECEC. In general, smaller classrooms, smaller children-to-teacher ratios and higher teachers' education levels are presupposed to lead to higher process quality, and, through process quality, to better child outcomes. However, the evidence for strong and consistent relationships between structural and process quality is far from conclusive, as will be reviewed below.

A possible explanation for the mixed findings concerns the effect of strong statutory regulations of structural quality at the state or country level that reduce the variance and, therefore, can lead to weak and inconsistent relationships with process quality (Love et al., 2003). Another possible explanation is that the set of structural quality characteristics commonly studied may not be sufficient to explain the variance in process quality. Most studies to date have focused on the so-called 'iron-triangle' of structural quality: children-to-teacher ratio, group size, and teacher formal pre-service education (Philipsen et al., 1997). Recent studies have included other structural quality aspects as well, in particular in-service professional development (Howes, James, & Ritchie, 2003; Zaslow, Anderson, Redd, Wessel, Tarullo, & Burchinal, 2010) and the use of a program of planned developmental and educational activities for children, including guided play, collaborative work, and age appropriate emergent literacy, mathematics and science activities as indicators of process quality and predictors of child outcomes (Assel, Landry, Swank, & Gunnewig, 2007; De Haan, Elbers, Hoofs, & Leseman, 2013; Sylva et al., 2007). Especially in contexts with strong regulation of the iron-triangle structural quality characteristics, the remaining variation in process quality may be largely dependent on other, less frequently studied structural quality aspects.

The objective of the current paper is to contribute to the existing evidence by extending traditional structural quality measures with measures of continuous in-service professional development and the use of an education program to predict process quality of center-based ECEC. The study was conducted in the Netherlands, which has a strongly regulated

ECEC system of average structural quality according to the comparative review by the Organisation of Economic Cooperation and Development of the statutory group sizes, children-to-teacher ratios and required teacher education level in 20 countries (OECD, 2006). Yet, despite strong national regulation and monitoring of structural quality, Dutch ECEC still shows considerable variation in process quality according to recent research (Helmerhorst et al., 2014; Leseman & Slot, 2013; NCKO, 2009).

Relations Between Structural and Process Quality

The relationships between children-to-teacher ratio, group size and process quality have been addressed in several studies in different countries. Smaller children-to-teacher ratios and smaller group sizes have been found to be associated with higher quality teacher-child interactions, as evidenced by responsive, warm, and supportive caregiving (Burchinal, Cryer, Clifford, & Howes, 2002; NICHD ECCRN, 2000a; Philipsen et al., 1997), and higher overall process quality (e.g. Barros & Aguiar, 2010; Burchinal et al., 2002; Mashburn et al., 2008; Philips et al., 2000; Philipsen et al., 1997; Thomason & La Paro, 2009). Yet, other studies have found only weak relations (Blau, 2000) or no relations at all between children-to-teacher ratio and process quality (Pessanha, Aguiar, & Bairrao, 2007; Pianta et al., 2005), and between group size and process quality (Barros & Aguiar, 2010; Blau, 2000; Pessanha et al., 2007; Philipsen et al., 1997). A cross-country comparison by Cryer et al. (1999) revealed that a smaller children-to-teacher ratio was related to higher process quality in Germany and the United States of America (USA), but not in Portugal and Spain. In addition, a negative relation was found between group size and overall process quality for Spain, but, remarkably, a positive relation was found for Germany. Note that the average group size in Spain was much bigger and showed stronger variation than in Germany, which may explain the contradictory results.

Dutch studies have shown similar mixed results. More favorable children-to-teacher ratios in Dutch day care centers were related to higher quality teacher-child interactions (De Kruif et al., 2009; De Schipper, Riksen-Walraven, & Geurts, 2006) and in classrooms with lower children-to-teacher ratios, children were provided with more learning opportunities (De Kruif et al., 2009), although the effects were rather small. In a study investigating day care for zero- to four-year-olds by De Schipper, Riksen-Walraven and Geurts (2006), group size was experimentally manipulated. Process quality was higher in smaller groups with a children-to-teacher ratio of 3:1 than in larger groups with a ratio of 5:1. However, other Dutch studies, with a correlational design, failed to replicate these

findings (Fukkink, Gevers-Deynoot-Schaub, Helmerhorst, Bollen, & Riksen-Walraven, 2013; Vermeer et al., 2008).

Several studies examined the relationship between teachers' formal pre-service education, specific (pre- or in-service) training in ECEC, and process quality. Higher levels of formal education have been found to be associated with higher overall classroom quality (Blau, 2000; Cryer et al., 1999; Philipsen et al., 1997), and, more specifically, with warmer, more supportive teacher-child interactions (Cryer et al., 1999; NICHD ECCRN, 2000a; Pianta et al., 2005; Thomason & La Paro, 2009). A comprehensive review by Tout, Zaslow and Berry (2006) showed that pre-service formal teacher education is more strongly associated with process quality if education includes ECEC content. However, also with regard to the effects of teacher pre-service education on process quality, the research findings are not consistent. In a large-scale multi-site and multi-state study in the USA, Early et al. (2006) found mixed effects of formal education on classroom quality in center-based daycare and pre-kindergarten. Teachers with more than a bachelor's degree had higher classroom quality than teachers with a degree below the bachelor level, but there were no differences between the bachelor and the below-bachelor degrees. Similarly, specific early childhood training mattered when teachers had lower formal education, but made no difference at or above the bachelor level. In a comparative review of seven large scale USA studies, Early et al. (2007) found contradictory effects of formal teacher education on process quality with effects varying from positive, null to negative. In a study in the United Kingdom on day care for infants and toddlers, Leach et al. (2006) found no clear associations between the qualifications and experience of the teachers and observed process quality. Likewise, recent Dutch studies on day care for zero- to four-year-olds did not find effects of teacher education on process quality either (De Kruif et al., 2009; Fukkink et al., 2013; Vermeer et al., 2008).

Increasing evidence indicates that in addition to formal pre-service education, in-service training, training- and coaching-on-the-job, and other strategies of continuous professional development with a focus on working with young children in ECEC settings contribute to process quality as well (for a comprehensive review, see Zaslow et al., 2010). For example, in a multi-state study in the USA, specialized training in ECEC with a focus on specific knowledge on child development was found to predict classroom process quality over and above formal education of the teacher (Philips et al., 2000). A meta-analysis by Fukkink and Lont (2007) revealed medium-sized average effects on caregivers' interaction competence of specialized training focusing on teacher-child interactions. Burchinal et al.

(2002) studied the effects of additional training on the job on quality of care and found that teachers who attended workshops for professional development were more sensitive in their interactions with children and provided higher overall quality of care compared to teachers who did not follow additional training. Several intervention studies have confirmed the benefits of in-service training (Hamre et al., 2012; LoCasale-Crouch et al., 2011). Also, other professional development activities such as consultation, mentoring and coaching on the job have been found to increase process quality (Campbell & Milbourne, 2005; Domitrovich et al., 2009; Howes et al., 2003; Pianta et al., 2008).

A relatively understudied aspect of quality in ECEC for zero- to four-year-old children is the provision of activities and materials that give particular content to children's experiences, often referred to as 'curriculum'. The relevance of these aspects was demonstrated in recent European studies involving preschoolers that incorporated the regular and systematic provision over a longer period of time of developmental and educational activities, such as guided play, and teacher-managed language, literacy and math activities, into the process quality construct (De Haan et al., 2013; Kuger & Kluczniok, 2008; Sylva et al., 2006). These studies revealed that higher process quality thus defined was associated with children's progress in pre-academic skills (De Haan et al., 2013; Sylva et al., 2006). Note that the activities that are actually provided can be considered aspects of process quality because they directly influence children's day-to-day experiences, whereas the plan or *education program* that more or less successfully regulates the provision of these activities can be considered an aspect of structural quality. Several studies, involving children from three to five years of age, have shown that the use of educational intervention programs to promote pre-academic skills by providing age-appropriate language, literacy and numeracy activities can be effective as far as the targeted skills are concerned (Clements & Sarama, 2007; Dickinson & Caswell, 2007; Domitrovich et al., 2009; Fantuzzo, Gadsden, & McDermott, 2011; Lonigan, Farver, Philips, & Clancy-Menchetti, 2011). Also, interventions focusing on social-emotional competences are found to be effective in the targeted social-emotional domain (e.g. Domitrovich, Cortes, & Greenberg, 2007). Comprehensive education programs, such as, for example, High/Scope and recently Tools of the Mind, Creative Curriculum, and others, addressing a broad range of developmental and educational goals and providing an education program to ensure implementation of activities that serve these goals, have also been found to be effective for broad developmental outcomes (Barnett et al., 2008; Bierman, Nix, Greenberg, Blair, & Domitrovich, 2008; Diamond, Barnett, Thomas, & Munro, 2007; Fantuzzo et al., 2011; Lambert & Abott-Shim, 2008; Schweinhart & Weikart, 1997).

To summarize, children's activity patterns and the balance between developmental and pre-academic educational content present in these activities are an important aspect of process quality. The way in which the provision of these activities is regulated by an explicit education program can be considered an important aspect of structural quality.

Measuring Process Quality

Process quality refers to the daily experiences of children in ECEC settings while engaging in activities and social interactions that drive development. Previous studies have mostly used systematic observation of the activities and interactions children engage in to assess process quality. However, despite the intuitive appeal of observation methods to assess process quality, there are also some limitations calling for the use of other assessment methods as well to supplement observations. Whereas observation measures of process quality are particularly suited to assess interaction quality and children's engagement in activities in real-time, they are less suited to assess how often and how consistently particular developmental and educational activities are provided over a longer stretch of time. In view of this, several studies have used teacher self-reports to assess process quality (Charlesworth et al., 1993, Walston & West, 2004; Xue & Meisels, 2004), occasionally in combination with observation measures, revealing small to moderate correlations between both methods (Kuger & Kluczniok, 2008). Self-reports can suffer from response bias due to social desirability, but yield more stable results over time than global observations of classroom quality (Pianta & Hamre, 2009). Combining different measures to assess process quality, thus, can increase the comprehensiveness of the quality assessment and strengthen the reliability and stability of the measurements by reducing method-bound error variance (Douglas, 2009).

Current Study

The current study combined two approaches to assess process quality of Dutch center-based ECEC in a comprehensive way. The Classroom Assessment Scoring System Toddler (CLASS; La Paro, Hamre, & Pianta, 2011) was used to evaluate process quality by means of observations. The CLASS framework reflects the social-emotional and educational features of teacher-child and child-child interactions that have been found to be positively related to children's development of self-regulation, pre-academic, and social skills (e.g. Curby et al., 2009; Howes et al., 2008; Mashburn et al., 2008; Rimm-Kaufmann, Curby, Grimm, Nathanson, & Brock, 2009). In addition, a teacher self-report questionnaire was used to

obtain information about the type of activities provided by the teachers on a larger time scale, focusing on the balance between affective behavior, play, pre-academic activities, and activities promoting self-regulation. Following the structure of the CLASS (La Paro et al., 2011), two comprehensive process quality constructs, Emotional Quality and Educational Quality, were defined using both observational measures and teacher reports as indicators. Emotional Quality included the CLASS dimensions Positive climate, Negative climate, Teacher sensitivity, Regard for child perspectives and Behavior guidance, and in addition the teacher-reported occurrence of emotionally supportive activities, support and enrichment of children's free play, and activities intended to promote self-regulation. In Dutch ECEC, children's free play is considered especially important for social-emotional development (e.g., to acquire autonomy and social skills), and hence teacher's play support was included in the overarching Emotional Quality construct. Teacher reported activities to promote children's behavioral self-regulation are conceptually related to the CLASS dimension Behavior guidance and, therefore, also included in the Emotional Quality construct. The construct Educational Quality included the CLASS dimensions Facilitation of learning and development, Quality of feedback, and Language modeling, and teacher-reported curriculum measures reflecting the provision of language, literacy, math, and pretend play activities. Pretend play in the current study, in contrast to support of children's free play, represented teachers' deliberate guidance of children's symbolic play to enhance cognitive development, and, therefore, was considered part of Educational Quality.

The Dutch ECEC system consists of two main types of provision. The first type is center-based day care for children from birth until four years of age, on average attended for two full days a week (NCKO, 2011). The second type concerns preschools for two- to four-year old children, which are attended for two to four half days a week. At age four, almost all children in the Netherlands enter full-day primary school. The Dutch ECEC system is strongly regulated. The Dutch Childcare Act of 2005 prescribes a children-to-teacher ratio of 7:1 for two- and three-year-old children and a maximum group size of 12 for two- to three-year-old and 16 for three- to four-year-old children (Convenant Kwaliteit Kinderopvang, 2008). Also, teachers are required to have completed a minimum of three-year vocational training in a relevant subject. The OKE (Promoting Development through Quality and Education) Act of 2010 brings day care centers and preschools under the same statutory quality framework and emphasizes the equal importance of social, emotional and cognitive outcomes for children. Although the two types of ECEC differ in the age range and socioeconomic background of the children served and stem from different traditions

in ECEC (with a care and education orientation, respectively), differences in structural quality have largely disappeared due to new legislation. Differences in quality between the two types of provision are not the main focus of the current study, but will be controlled for in the main analysis.

The aim of the current study is twofold. First, classroom process quality of a representative sample of the Dutch ECEC system will be determined using the CLASS Toddler and teachers' self-reported activities. Second, the relationships between structural and process quality will be examined, including as structural characteristics the implementation at the center of strategies of professional development and the use of an education program.

Method

Participants

The present study used data from the ongoing national cohort study pre-COOL, which investigates the effectiveness of preschool education and care provisions in the Netherlands. Pre-COOL was commissioned by the Dutch Ministry of Education, Culture and Sciences and the National Science Foundation. The cohort started in 2010, when children were two years old. At age five, children enter the national cohort study COOL on students' careers in primary and secondary education, and they will be followed-up until age eighteen. To increase the likelihood of pre-COOL children entering primary schools that take part in COOL, the sample was recruited in the following way. First, a random sample of 300 primary schools was drawn from the COOL cohort, of which, 139 schools (46.3%) agreed to participate. Next, the participating primary schools were asked to identify the preschools and day care centers that were attended by most of their new students. Municipal records and the internet were used to identify additional preschools and day care centers in the neighborhood of the schools. About 500 centers were approached, of which 263 agreed to participate in pre-COOL (52.6%). A total of 375 teachers of 182 centers (69.2%) participated in the study by filling out the teacher questionnaire, providing information on 295 classrooms (170 preschool, 125 day care). Almost all teachers were women (99.2%) and predominantly Caucasian (89.4%). For logistic and methodological reasons, observations were only conducted in classrooms with at least four children participating in the child assessments of pre-COOL (not the topic of the present study), resulting in 162 centers (61.6% of the entire pre-COOL sample) with a total of 276 classrooms (155 preschool and 121 day care classrooms). The participating preschools and day care

centers were geographically spread over all parts of the Netherlands, were located in urban, semi-urban and rural areas, and did not differ significantly on these characteristics from non-participating preschools and day care centers (Pre-COOL Consortium, 2012). The present study focused on provisions for two- to four-year-old children, but in 63% day care classrooms, also younger children were present. Classroom composition with regard to children's age and ethnicity differed between day care centers and preschools, as is representative for the Netherlands. Descriptive statistics of the final sample of 276 classrooms and 375 teachers are presented in Table 1.

Table 1. Descriptive Statistics for Center and Teacher Characteristics

	Day care				Preschool			
	N	**M**	**SD**	**Range**	**N**	**M**	**SD**	**range**
Children-to-teacher ratio	156	5.4	1.2	3-8	208	7.0	2.5	3-16
Group size	159	13.6	1.7	6-17	212	14.6	2.8	7-17
Professional development activities	157	2.91	0.94	1.20-5.90	213	3.36	1.06	1.22-6.33
Frequency (F)/Percent (P)	F	P			F	P		
Educational program	149	55.6			205	97.6		
Classroom > 30% non-Dutch	36	23.5			110	55.0		
Age composition classrooms								
0 years	83	51.2						
1 year	99	61.1						
2 years	157	96.9			209	97.7		
3 years	102	63.0			92	43.0		
Educational level								
Lower preparatory track	73	17.5			60	12.6		
4 yrs secondary vocational	174	41.8			197	41.5		
1-2 yrs intermediate vocational	29	7.0			18	3.8		
3-4 yrs intermediate vocational	60	14.4			78	16.4		
5 yrs secondary	55	13.2			95	20.0		
6 yrs secondary*	10	2.4			19	4.0		
Higher vocational	10	2.4			3	0.6		
University	5	1.2			5	1.1		
Teacher ethnicity								
Native Dutch	143	95.3			178	85.2		
Immigrant	6	4.7			31	14.8		

* entry level university

Measures and Procedures

Observed process quality.

The *Classroom Assessment Scoring System Toddler* (CLASS Toddler; La Paro, Hamre, & Pianta, 2011) was used to assess classroom process quality. An officially approved Dutch

translation of the CLASS manual was developed for the present study. All observers were trained by a licensed CLASS trainer and achieved at least 80% agreement within one scale-point deviation with the trainer on an online test before they were admitted to the study (average agreement was 86.4%; agreement by chance was 33%), as recommended by the developers of the CLASS. Following the online test, the trainer conducted live observations with all observers once, prior to the data collection. Inter-observer agreement of the live observations within one scale-point deviation was 89.9%. Each classroom was observed during one morning and all classrooms were observed within a three-month period in the spring of 2011. Following instructions in the CLASS manual, observers rated classroom processes and teacher behavior during four 15 to 20 minutes cycles on the observation morning, resulting in a total of 1092 observation units. Observers were instructed to observe all regular activities as they occurred, except outdoor play, in line with the CLASS manual. For each observation unit, also the main type of activity at stake was registered.

Classroom quality was rated on eight dimensions in two broad domains, using 7-point scales ranging from 1 or 2 (*classroom is low on that aspect*); 3, 4 or 5 (*classroom is in the midrange*); and 6 or 7 (*classroom is high on that aspect*). Descriptive statistics of the scores on the CLASS dimensions are displayed in Table 2.

Regarding Emotional Quality, the observed processes were evaluated on five dimensions: *Positive Climate* reflects the warmth, respect, and enjoyment displayed during interactions of the teacher and children; *Negative Climate* reflects the overall negativity expressed in the classroom by the teacher and the children (scores are reversed); *Teacher Sensitivity* is the extent to which the teacher is aware and responsive to the children's needs; *Regard* for *Child Perspectives* captures the degree to which the teacher's interactions with children and classroom activities capture the children's interests, and the degree to which children's independence is encouraged; *Behavior Guidance* refers to the teacher's ability to promote positive behavior and redirect problem behavior.

Regarding Educational Quality, observed processes were evaluated on three dimensions: *Facilitation of Learning and Development* considers how well the teacher facilitates activities to support children's learning and development; *Quality of Feedback* assesses the degree to which the teacher's feedback promotes learning and expands children's participation; *Language Modeling* refers to the extent to which the teacher fosters, models and encourages children's use of language.

Observed type of activity

Observers registered the type of activity during each observation cycle. If several different activities were observed during a cycle, all activities were listed. In total, 15 different types of activity were observed, which were independently recoded by the first and second author into four main categories based on consensus between the two researchers: indoor free play, including play in centers and construction play, educational activities, such as circle time and book reading, creative activities, such as drawing, painting and crafts, and care routines, such as snack- and mealtimes, toileting and clearing up. When an observation cycle covered more than one of these main activity types, priority was given to the educational activity over the other activities, to creative activity over play and care routines, and to play over care routines, respectively, in order to obtain unambiguous codes for each cycle. For the main analysis, dummy variables were created to represent each observed type of activity (scored 1) versus all other activities (scored 0): *Free Play, Educational Activity, Creative Activity,* and *Care Activity*. About 9% of all observation cycles could not be unequivocally categorized and were combined with the reference category to avoid loss of data.

Self-reported developmental and educational activities

A structured questionnaire for teachers was used to assess the developmental and educational activities provided to the children on a regular basis during the year. The list of affective, play-supporting, self-regulation promoting and academic activities presented to the teachers was carefully developed, based on extant research into social-emotional (e.g., Domitrovich et al., 2007) and self-regulation development (e.g., Barnett et al., 2008; Bierman et al., 2008; Domitrovich et al., 2007; Diamond et al., 2007), and emergent academic skills (e.g., Clements & Sarama, 2007; Dickinson & Caswell, 2007; Klibanoff et al., 2006; Leseman & Van Tuijl, 2006), and extensively tested in pilot research to ensure age-appropriateness of the listed activities. Eight scales were constructed covering a broad range of behaviors and activities. The scales Emotional support, Play, and Self-regulation are considered indicators of emotional process quality, while the scales Pretend play, Language activities, Literacy activities, and Math activities are considered indicators of educational process quality. Descriptive statistics are presented in Table 2

Emotional support (8 items; Cronbach's alpha= .88) reflects the degree to which the teacher provides emotional support and comfort to children and shows verbal and physical affection. An example of an item is: "I hug the children or give them a pat on the head". The scale ranges from 1 (*never*) to 7 (*always*).

Play (9 items; Cronbach's alpha= .85) assesses the degree to which the teacher provides children with opportunities for free, self-managed play and enriches children's play, for instance by asking questions, making suggestions, or providing materials for richer play. Examples of items are: "I let the children play without interfering", "I ask children questions that stimulate their play". The scale ranges from 1 (*not applicable*) to 5 (*strongly applicable*).
Self-regulation (12 items; Cronbach's alpha= .88) assesses the extent to which the teacher uses play and other activities to enhance children's behavioral self-regulation, for instance talking about feelings and emotions, helping them resolve peer conflicts or playing games in which children have to take turns. An example of an item is: "When children have a conflict I let them express their own opinion so they better understand what the other thinks". The scale ranges from 1 (*never*) to 7 (*always*).
Pretend play (8 items; Cronbach's alpha= .91) represents to what extent the teacher stimulates cognitive distancing, symbolizing and pretend in children by modeling behavior and encouraging children to participate in symbolic and pretend play. An example of an item is: "I show children how to use an object for something else then intended, for instance driving a wooden block *as if* it is a car". The scale ranges from 1 (*never*) to 7 (*always*).
Language activities (8 items; Cronbach's alpha= .89) assesses the average frequency of activities involving several forms of language use, including singing, rhyming, conversations and language instruction. An example of an item is "Having elaborate conversations about children's personal experiences, for instance what they did in the weekend". Answers were rated on a 7-point scale, ranging from 1 (*never*), 2 (less than twice a month), 3 (twice or thrice a month), 4 (weekly), 5 (two to four times a week), 6 (daily) and 7 (*three or more times a day*).
Literacy activities (4 items; Cronbach's alpha= .82) measures the average frequency with which activities are provided involving literacy and literacy materials. An example of an item is: "Asking the children questions about the content of the story during or after reading the story". Answers were rated on the same scale as Language activities.
Math activities (12 items; Cronbach's alpha= .91) assesses the average frequency of several math activities, for instance counting and sorting activities, and activities exploring different shapes. An example of an item is: "Counting how many objects you have, for example counting till five and saying 'I have five marbles'". Answers were rated on the same scale as the Language and Literacy activities.

Table 2. Descriptives of the Process Quality Measures

	M	SD	Range	N
CLASS dimension				
Positive Climate	5.42	1.17	1.00-7.00	1084*
Negative Climate (recoded)	5.84	0.38	5.00-7.00	1084
Teacher Sensitivity	5.34	1.08	2.00-7.00	1084
Regard for Child Perspectives	4.24	1.34	1.00-7.00	1084
Behavior Guidance	5.01	1.12	2.00-7.00	1084
Facilitation of Learning and Development	3.73	1.35	1.00-7.00	1084
Quality of Feedback	2.91	1.20	1.00-7.00	1084
Language Modeling	3.22	1.29	1.00-7.00	1084
Self-report				
Play	3.23	0.55	1.43-4.61	371
Emotional support	6.10	0.65	4.00-7.00	371
Self-regulation	4.15	0.88	1.64-6.91	372
Pretend play	4.05	0.99	1.25-7.00	370
Language activities	5.18	1.11	2.00-7.00	370
Literacy activities	4.86	1.06	1.80-7.00	369
Math activities	3.79	1.05	1.58-6.92	373

* number of observation cycles with the CLASS

Structural classroom and center characteristics

Teachers filled out a questionnaire addressing group size, number of adults present in the classroom, and their own professional training and demographic background. In addition, teachers reported on the professional development activities provided by their center and the use of an education program. For the present purpose, the following structural quality variables were constructed; descriptive statistics are given in Table 1:

Group size was computed as the teachers' reported maximum number of children in the classroom during regular days of the week.

Children-to-teacher ratio was computed by dividing group size by the number of licensed professionals present during regular days, as reported by the teachers, not including student-teachers on an internship, household personnel, center managers or, occasionally, volunteering parents.

Teacher's education was defined as the highest level of completed formal pre-service education by the teachers and was measured on a scale representing the levels of the Dutch secondary and tertiary education system, ranging from 1 (*lower preparatory vocational education*) to 8 (*university education*).

Education program reflects the use by teachers of a structured education program. In the Netherlands, several education programs, approved by the national Accreditation Committee for Child and Youth Interventions of the Netherlands' Youth Institute (www.nji.nl), are currently used in ECEC, in both day care centers and preschools. Although these programs differ in how teachers are trained and monitored, they all aim at broad developmental and educational goals, emphasize emotional support, sensitivity to children's needs, and provide a mixture of play and pre-academic activities with an emphasis on language and literacy. All programs use manuals listing activities that can be provided, work with themes (e.g. the seasons of the year, important feasts) and specify a year-schedule for providing the themes and activities. Some programs also use specific materials as part of the program, such as picture and storybooks, experiential Montessori materials, or puppets to elicit talk during circle time. For the present purpose, a dummy variable was created, indicating whether an education program was used with the values 1 = *yes* and 0 = *no*, without further distinguishing between the programs.

Professional development (8 items; Cronbach's alpha= .91) assesses the implementation of several strategies of continuous professional development within the team of teachers and at the center. A questionnaire listing several professional development activities was presented. Teachers rated how frequently these activities occurred, with a scale ranging from 1 (*never*), 2 (less than once a month), 3 (once a month), 4 (twice or thrice a month), 5 (weekly), 6 (two to four times a week), and 7 (*every day*). Examples of professional development activities included in the list were: having regular staff meetings to discuss the developmental and educational goals of working with young children, discussing children with special developmental and educational needs, using collegial observation and feedback to improve practice, opportunities for in-service training and personal coaching, team-based reading of professional literature, and visiting professional conferences.

Type of provision represents whether the classroom was part of a day care center or preschool. To control for possible confounding of type of provision with structural quality characteristics, a dummy variable was created and included in the main analysis, with the values 1 = *preschool* and 0 = *day care*.

Other measures

For the purpose of sample description, additional information was obtained about the age- and ethnic composition of the classrooms, and about teachers' gender, age, work experience, and ethnic background. These measures were not included in the main analysis.

Analysis Strategy

Following most studies into ECEC quality, the present study focused on the classroom and in addition, within the classroom, on the observed activity setting as the units of analysis. Quality was considered a multifaceted construct and assumed to be represented by a set of observed and teacher-reported variables covering both emotional and educational process quality in real-time as well as on a larger time scale. Observations were conducted in four distinct situations within each classroom, yielding a nested data structure. Although it is common in ECEC research to aggregate detailed observation measures to the classroom level, a clear disadvantage is the loss of potentially relevant information (Hox, 2010). In order to combine the multiple indicators in a single model of process quality and to take the nested data structure into account, Multilevel Structural Equation Modeling was applied using the Mplus statistical package (Version 7; Muthén & Muthén, 1998-2012).

Due to the design of the study and non-response, complete observational and self-report data were available for 113 classrooms (40.9% of all observed classrooms). Classrooms with and without teacher self-reports did not differ significantly on any of the CLASS dimensions. Classrooms with and without observations differed significantly on three of fourteen self-reported measures. Group size was bigger in classrooms with observations due to the selection of classrooms with at least four children eligible for child assessment. Further, classrooms with observations had slightly lower scores on self-reported language and math activities than classrooms without (all standardized effect sizes < .25). There were no further missing data in the observation measures, and missing data on activities and structural classroom characteristics in the self-reports were below 8%. As recommended, missing data were dealt with by using full information maximum likelihood (FIML) estimation in Mplus (Enders, 2010), in which the standard errors for the parameter estimates are computed using the complete observed information matrix (Muthén & Muthén, 1998-2012).

The main analysis was carried out in four steps. First, the intraclass correlations (ICC's) were calculated separately for each CLASS dimension to determine the proportions of *within* and *between classrooms* variance. Second, the measurement model of process quality was estimated at the within and the between classrooms level using a non-restricted baseline model that included all observed and self-reported process quality indicators. Following the proposed structure of the CLASS (La Paro et al., 2011), a two-factor model was examined with the CLASS dimensions and teacher-report scales as indicators of two

latent factors representing emotional and educational process quality, respectively. Finally, all predictors at the within level (types of observed activity) and between level (all structural quality characteristics) were entered into the resulting measurement model to test the relationships between the structural quality aspects and both process quality constructs. Note that at the within groups level, Care Activity was the reference category and, therefore, not included as predictor. Model fit was evaluated with several fit indicators: the ratio of the Chi-Square and the degrees of freedom (Chi-Square/df), the Comparative Fit Index (CFI), the Root Mean Square Error of Approximation (RMSEA) and the Standardized Root Mean Square Residual (SRMR) at both the within and between classrooms level, with Chi-Square/df < 3, CFI > .90, RMSEA < .08, and SRMR < .08 indicating acceptable to good fit. The Chi-Square/df ratio was used as indicator of model fit rather than the Chi-Square test, because the latter is very sensitive to small violations of the multi-normality assumption with large samples (Kline, 2005). Improvement of the model fit was evaluated by testing the significance of the change in Chi-Square relative to the change in degrees of freedom.

Results

Descriptive statistics of the sample, the structural characteristics of the classrooms and centers, the process quality observation measures and the self-reported quality measures are presented in Tables 1 and 2. Regarding the process quality measures based on the CLASS observations, the results reveal moderate to high emotional process quality and low to moderate educational quality. The pattern of activities that was observed is in line with the highly regular day schedule that is typical of ECEC classrooms in the Netherlands with Free Play as the predominant type of activity, occurring in 26.6% of all observation cycles, and an Educational Activity, Care Activity and Creative Activity occurring in 25.7%, 24.3%, and 14.3% of the observation cycles, respectively. There were slight differences between day care and preschool classrooms. In day care classrooms, for instance, more time was spent on care routines due to the care for younger children present in the groups, and in preschools more time was spent on educational activities reflecting the slightly older age of the children and the stronger educational orientation of this type of provision.

The teachers' self-reports show a comparatively high level of emotional support and much lower levels of educational support through providing pretend play and academic activities, with the reported support of children's self-managed play falling in-between.

Table 3 shows the bivariate correlations between all process quality indicators. The

inter-correlations of the CLASS dimensions were moderate to strong. Also, the inter-correlations between the self-reported process quality indicators were moderate to strong. The correlations between the CLASS dimensions and the self-reported activities were mostly significant and generally in the expected direction, but much smaller in magnitude. Table 4 presents the intraclass correlations for all CLASS dimensions, revealing significant within and between classrooms variance, indicating that multilevel modeling is indeed required.

Table 3. Bivariate Correlations Between Self-Reported Activities and Observed CLASS Dimensions (observation cycles)

			Teachers' self-reports									CLASS dimensions						
			1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Self reports	1	Pretend play		.56**	.56**	.32**	.42**	.34**	.41**	.22**	.10**	.02	.05	.04	.03	.26**	.16**	.21**
	2	Play			.54**	.40**	.46**	.32**	.37**	.20**	.10**	.06	.10**	.08**	.05	.21**	.17**	.23**
	3	Self-regulation				.42**	.44**	.32**	.53**	.29**	.11**	.10**	.09**	.05	.06†	.20**	.07†	.16**
	4	Emotional support					.26**	.30**	.25**	.17**	.11**	.07†	-.02	.01	.04	.12**	.08*	.11**
	5	Language						.68**	.62**	.40**	.11**	.18**	.07†	.00	.07†	.18**	.15**	.14**
	6	Literacy							.40**	.44**	.15**	.18**	.06	-.04	.04	.20**	.13**	.16**
	7	Math								.39**	.09**	.11**	.05	-.07	.04	.18*	.05	.14**
	8	Science									.05	.07	-.05	-.03	.01	.04	-.00	.03
CLASS	9	Positive Climate										.28**	.58**	.29**	.44**	.46**	.45**	.54**
	10	Negative Climate											.27**	.15**	.32**	.20**	.21**	.19**
	11	Teacher Sensitivity												.26**	.56**	.42**	.38**	.44**
	12	Regard for Child Perspectives													.23**	.30**	.24**	.28**
	13	Behavior Guidance														.38**	.33**	.35**
	14	Facilitation of Learning															.60**	.64**
	15	Quality of Feedback																.63**
	16	Language Modeling																

**$p < .01$, * $p < .05$, † $p < .10$

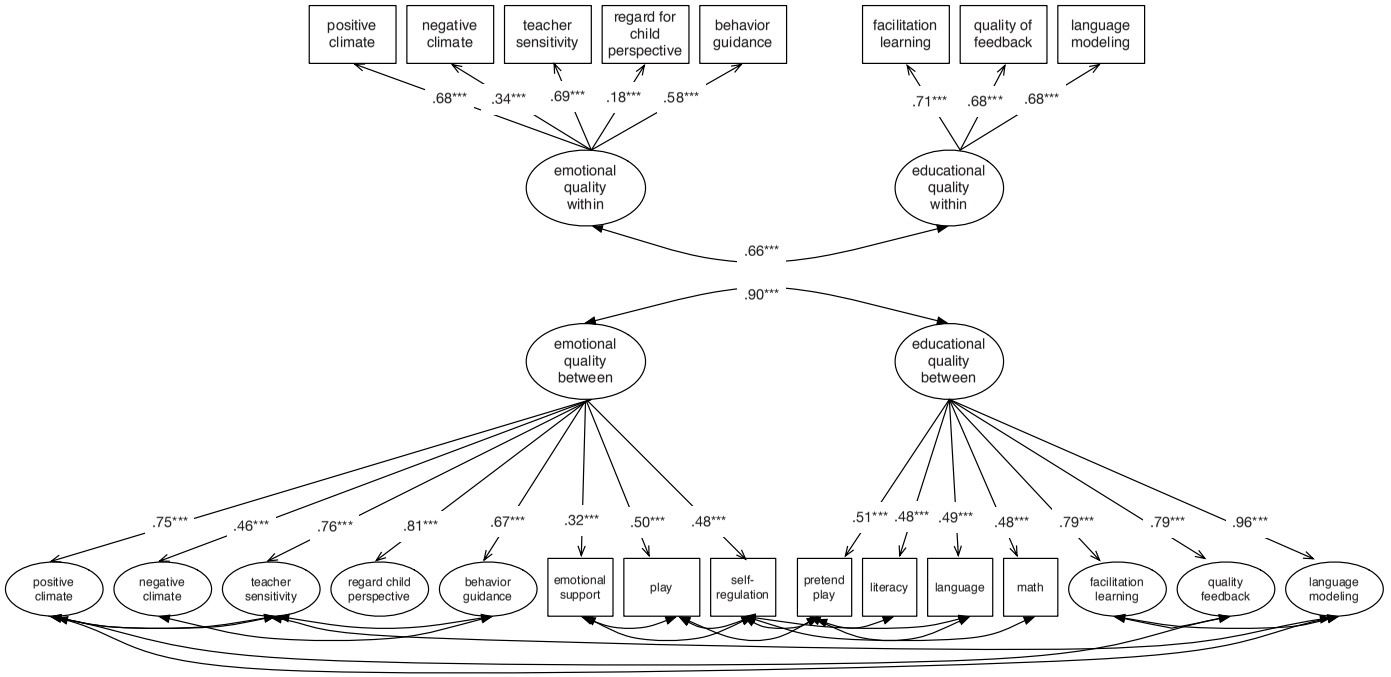
Table 4. Intraclass Correlations (ICC) for the CLASS Dimensions in the Model

Variable	ICC	Between- variance (SE)	Within-variance (SE)
Positive Climate	.51*	.35 (.04)*	.61 (.08)*
Negative Climate	.36*	.08 (.02)*	.20 (.07)*
Teacher Sensitivity	.37*	.29 (.04)*	.61 (.08)*
Facilitation Learning	.30*	.46 (.04)*	.73 (.08)*
Quality of Feedback	.37*	.38 (.04)*	.58 (.08)*
Language Modeling	.54*	.40 (.04)*	.58 (.07)*

N_{within} = 1313 , $N_{between}$ = 398, *p < .05

As the first step of the main analysis, the measurement model of process quality was examined at the within and the between classrooms level. Based on theoretical

Figure 1. Final Quality Model with Within and Between Predictors of Quality.



considerations, a two-factor model was estimated specifying a latent emotional support and a latent educational support factor, both with multiple indicators based on observations and teacher reports. The saturated model showed poor model fit ($\chi^2(122) = 686.101$, $p = .00$; $\chi^2/df = 5.62$; RMSEA = .06; CFI = .82; $SRMR_{within} = .10$, $SRMR_{between} = .24$). As suggested by the modification indices (MI) provided by Mplus, we allowed the error variances of the self-reported practices and the observation measures to correlate (indicating the presence of method-bound covariances not captured by the two process quality factors), which improved the model fit significantly ($\Delta\chi^2(33) = 423{,}25$ $p = .00$, with resulting model fit indices showing acceptable fit: $\chi^2(89) = 262.850$, $p = .00$; $\chi^2/df = 2.95$; RMSEA = .04; CFI = .94; $SRMR_{within} = .03$, $SRMR_{between} = .16$). Figure 1 shows the final measurement model.

Next, all predictors were added to the model: the observed type of activity at the within classrooms level and the structural quality characteristics at the between classrooms level. Model fit was not satisfactory (χ^2 (185) = 668.39, $p = .00$; $\chi^2/df = 3.61$; RMSEA = .06; CFI = .82; $SRMR_{within} = .08$, $SRMR_{between} = .15$). The modification indices revealed an estimation problem related to the dimension Regard for child perspectives (MI = 178.15). Upon closer examination of the data, Regard for child perspectives was found to be rated especially high for free play, but relatively low for the other activity types, whereas most other quality measures were found to be rated relatively low for free play and high for the other activity types, indicating misfit. Therefore, we decided to remove Regard for child perspectives from the model. We will return to this issue in the discussion section. After removing this dimension, the model fit was acceptable χ^2 (157) = 343.261, $p = .00$; $\chi^2/df = 2.19$; RMSEA = .04, CFI = .92; $SRMR_{within} = .05$, $SRMR_{between} = .12$).

The analysis was conducted on the whole sample, pooling the two types of ECEC provision. To check whether this was appropriate, a multigroup analysis was performed splitting the sample into two subsamples of day care and preschool groups, respectively, to examine measurement invariance across the two types of provision. The results indicated complete measurement invariance or scalar invariance with acceptable fit of the multigroup model to the data (χ^2 (319) = 569.224, $p = .00$; $\chi^2/df = 1.78$; RMSEA = .04, CFI = .90; $SRMR_{within} = .05$, $SRMR_{between} = .15$) .

At the *within classrooms* level, different types of activities were related to higher observed emotional and educational process quality compared to the reference category Care Activity (see Table 5). Providing an educational activity to children was associated with higher emotional process quality as observed with the CLASS, with a small effect size

according to Cohen (1992). The provision of educational and creative activities as well as free play was associated with higher educational process quality, compared to quality in care routines, with the educational activities showing the strongest association.

At the *between classrooms* level, four predictors were significantly related to emotional and educational process quality (see Table 5). Teacher pre-service education positively predicted emotional and educational process quality, but the effect size was small.

Table 5. Within Level and Between Level Predictors of Emotional and Educational Quality (N=849 classrooms)

	Emotional quality			Educational quality		
Predictor	**B**	**SE B**	**β**	**B**	**SE B**	**β**
Within level						
Educational activity	.20	.06	.15***	.90	.12	.42***
Creative activity	.13	.08	.07	.34	.13	.13**
Free play activity	-.05	.07	-.04	.21	.10	.10*
Between level						
Teacher education level	.03	.02	.18†	.03	.02	.13*
Group size	-.02	.02	-.14	.02	.01	.07
Children-to-adult ratio	.00	.01	.03	.01	.01	.04
Education program	.15	.09	.19†	.21	.09	.20*
Professionalization	.13	.06	.41*	.19	.05	.46***
Type of provision	.05	.06	.09	.14	.08	.17†

***$p < .001$**$p < .01$, *$p < .05$, †$p < .10$

Type of ECEC provision only explained variance in educational process quality, indicating slightly higher educational quality in preschool compared to day care classrooms, but the effect was also rather small. Using an education program was related to both emotional and educational process quality, with a small to medium effect size. The provision of professional development activities was the strongest predictor of emotional and educational process quality, with medium to large effect sizes. Group size and children-to-teacher ratio did not explain any variance in emotional nor educational process quality.

Discussion

The present study examined the relationship between structural and process quality in Dutch day care and preschool classrooms using a multi-method approach. Emotional and educational process quality were defined as comprehensive, multifaceted constructs, which include observed social-emotional and educational aspects of teacher-child interactions,

as well as the self-reported activities teachers provide to children. Furthermore, to predict emotional and educational process quality, the present study combined frequently studied structural quality characteristics, such as group size, children-to-teacher ratio and teacher pre-service education, with less frequently studied structural quality aspects, in particular the implementation of professional development activities and the use of an education program.

The results of the observations with the CLASS show that process quality in Dutch ECEC was on average moderate to high regarding emotional and behavioral support, but low to moderate regarding educational support, which corresponds with findings in other studies in both the Netherlands and other countries (Helmerhorst, Riksen-Walraven, Vermeer, Fukkink, & Tavecchio, 2014; Pakarinen et al., 2010; Pianta & Hamre, 2009; Thomason & La Paro, 2009; Weiland, Ulvestad, Sachs, & Yoshikawa, 2013). From the teachers' self-reports a similar overall pattern appeared. Teachers reported to frequently provide children with emotional supportive activities, such as comforting children and showing children verbal and physical affection, and to be moderately inclined to support children's self-managed play. However, especially pretend play and academically focused activities were less frequently provided. The present results are consistent with findings in previous Dutch research and in other European studies, in which the provision of activities was assessed with an observation method (Anders et al., 2012; De Kruif et al., 2009; Hall et al., 2009) or with a combination of observations and teachers' self-reports (Kuger & Kluczniok, 2008).
As reviewed in the introduction section, previous evidence on the relationship between structural quality and process quality is not conclusive (e.g. Cryer et al., 1999; Early et al., 2007). The findings of this study add to this evidence. Group size and children-to-teacher ratio were not significantly related to emotional and educational process quality, contrary to findings in several studies in the USA (Burchinal et al., 2002; Mashburn et al., 2008; NICHD ECCRN, 2000a; Philips et al., 2000; Thomason & La Paro, 2009), but, at least partly, in line with findings of previous Dutch studies (De Kruif et al., 2009; De Schipper et al., 2006; Vermeer et al., 2008) and other European studies (Barros & Aguiar, 2010; Pessanha, et al., 2007). The absence of effects of group size and children-to-teacher ratio in the current study is likely due to the limited variation in these structural characteristics (cf. Love et al., 2003). Teacher formal education level was positively related to emotional and educational process quality, consistent with other studies (Blau, 2000; Cryer et al., 1999; NICHD ECCRN, 2000a; Philipsen et al., 1997; Pianta et al., 2005), but the size of the relationship was rather small. Previous Dutch studies did not find effects of teacher formal

education level on process quality (De Kruif et al., 2009; Vermeer et al., 2008). Note that the variation in education level was restricted, as in previous Dutch research (De Kruif et al., 2009), which can explain the lack of stronger effects.

Three main aspects of structural quality were found to be associated with process quality. At the *within classrooms* level, this included the provision of activities with an educational focus. At the *between classrooms* level, this included the use of an education program and implementation of professional development activities at the team and center level.

First, at the within classrooms level, emotional process quality as observed with the CLASS was highest during educational activities in smaller subgroups, including circle time, book reading and making puzzles, compared to care routines, which is consistent with other research (De Schipper et al., 2006; Pianta et al., 2005). In addition, educational process quality was higher during creative and educational activities, and to a less extent during free play, compared to educational quality during care routines, with educational activities being most strongly associated with educational process quality.

This relates to the second major finding of the present study. At the between classrooms level, the use of an approved high quality comprehensive education program was found to be moderately related to emotional and educational process quality. The results of the present study are in line with evidence from intervention research in ECEC revealing positive effects of comprehensive programs on classroom quality and child outcomes (e.g. Barnett et al., 2008; Bierman et al., 2008; Fantuzzo et al., 2011; Lambert & Abott-Shim, 2008).

The third major finding was that providing professional development activities at the center is most strongly associated with emotional and educational process quality, which is consistent with findings in other studies in ECEC and with research on concepts of professional development such as reflective practice and team learning (e.g. Howes et al., 2003; Pianta et al., 2008; Zaslow et al., 2010).

Overall, the results of the current study only partially replicated findings from previous Dutch studies on ECEC quality (De Kruif et al., 2009; Vermeer et al., 2008). Most importantly, we did not find significant associations of group size and children-to-teacher ratio with process quality. A possible explanation is that the previous Dutch studies were conducted before the recent changes in the national ECEC policy. With new legislation in 2005 and 2010 concerning the maximum group size and children-to-teacher ratio, and the introduction of joint quality monitoring by the Municipal Health Authorities and the Inspectorate of Education, the range in structural quality may have become even more restricted.

There are several limitations to the present study. First, the study was conducted in the context of a strongly regulated ECEC system in the mid-range of structural quality (OECD, 2006). The conclusion that using an education program and providing professional development activities can raise process quality, therefore, only applies if group size, children-to-teacher ratio and teacher education are within the boundaries of mid-range structural quality. Second, although the current sample of classrooms was recruited in relation to a random sample of primary schools, there was considerable non-response and missing data. The presence of selection effects cannot be ruled out, which limits the possibilities of generalizing the present findings. Note, however, that the sample was geographically well spread and that the overall findings concerning the average levels of both structural and process quality are largely in line with previous research in the Netherlands, suggesting the findings are representative for current Dutch ECEC. Third, the CLASS dimensions Regard for child perspectives did not fit well in the model after including the predictors. Differences in ECEC contexts between the Netherlands and the USA, in which the CLASS was developed and validated, may be at stake. For example, further inspection of the data showed that teachers' scores on the Regard for child perspectives were especially high during free play but not during other activities (on average one SD difference), whereas the scores on the other quality dimensions, such as teacher sensitivity and behavior guidance, were relatively low during free play compared to other activities. An explanation is that free play in the Dutch context is typically characterized by rather limited teacher involvement compared to other activity settings, allowing children much initiative but also providing them with little support and guidance (De Haan et al., 2013; Leseman, Rollenberg & Rispens, 2001). A third limitation of the current study is that one of the predictors of process quality, the use of an education program, was correlated with type of ECEC provision. In the current sample, the majority of preschools but only just over half of the day care centers used an education program. However, possible confounding of type of provision (preschool or day care center) with the use of an education program was controlled by adding type of provision as a predictor to the model, with only a small effect on educational process quality, while a medium effect of using an education program remained. In addition, multigroup analyses showed complete measurement invariance across the two types of provision, confirming that both types of provision can indeed be pooled. A final limitation concerns the correlational design of the current study, which limits the possibilities of drawing causal conclusions. Future studies using a longitudinal design can provide a stronger basis for conclusions about the direction of effects.

To conclude, the current study adds to the existing evidence on the relationships between structural and process quality by providing a comprehensive view on process quality and by including additional structural quality characteristics as predictors of process quality. The study revealed that within an ECEC system in the mid-range of structural quality the commonly examined structural quality aspects matter less for process quality. Rather, aspects of structural quality not traditionally included in research on ECEC quality were found important for process quality. The present results need to be replicated and examined in other contexts and other ECEC systems. Moreover, it is important to examine whether the quality characteristics that were the topic of this study are indeed related to child wellbeing and child development in short and long term. If the present results are sustained, there may be important implications for policy and practice. For example, many countries are considering to raise the required education level of ECEC workers to the bachelor's level in order to enhance the quality, impact and economic benefits of ECEC (OECD, 2006). Whereas there may be several advantages of such a policy, it will raise the costs of ECEC considerably. Yet it may not be the most costs-effective way of improving process quality. Following the present results, it may be more efficient to concentrate on continuous professional development to enhance the quality of ECEC (Campbell & Milbourne, 2005; Domitrovich et al., 2009; Pianta et al., 2008). In addition, the use of an education program can enhance process quality by supporting teachers in providing children with appropriate developmental and educational activities.

CHAPTER 4

General and Specific Quality of Early Childhood Education and Care Predict Growth of Two-Year-Olds' Vocabulary and Attention Skills

Slot, P. L., Mulder, H., Verhagen, J., & Leseman, P. P. M., (2014)

Domain-General and Domain-Specific Quality Characteristics of Early Childhood Education and Care Predict Growth of Two-Year-Olds' Vocabulary and Attention Skills Over One Year.

Manuscript under revision

Abstract

The present study investigated the effects of domain-general and domain-specific process quality of early childhood education and care (ECEC) classrooms on the development of two-year old children's vocabulary and attention skills over the course of one year using a value-added approach. Process quality was evaluated with the Classroom Assessment Scoring System (CLASS) and with teachers' self-reports on the provision of play, pre-academic activities, and activities aimed at promoting self-regulation. We distinguished between domain-general emotional support, and domain-specific behavioral support and instructional support based on observations and the self-reported curriculum. Data were used of about 850 children, with child assessments of vocabulary and attention at ages two and age three years, attending 185 preschool and child care classrooms. Results show both domain-general and domain-specific effects of ECEC quality. Emotional support was related to growth in vocabulary whereas observed support for learning was related to growth in children's attention skills. The provision of play was negatively related to growth in both vocabulary and attention skills, which possibly can be explained by the fact that play in Dutch ECEC settings is mostly unguided play. The findings show the importance of evaluating process quality as a multi-faceted construct.

Introduction

Most young children in western countries participate in some form of center-based early childhood education and care (ECEC) before they enter the formal schooling system (OECD, 2013). In the Netherlands, over 80% of the two- to four-year-old children attend either formal center-based day care or preschool (CBS, 2011; MOgroep, 2012). Day care and preschool attendance have been shown to be beneficial for children's developmental outcomes in various domains, especially when the provided education and care is of high quality (NICHD ECCRN, 2000, 2006; Melhuish, 2011; Pianta, Barnett, Burchinal, & Thornburg, 2009). However, the effects of ECEC quality on children's developmental outcomes are generally weak to moderately strong at best (Burchinal, Kainz, & Cai, 2011). This raises the question if particular aspects of quality matter more than others and if domain-specific quality measures will show stronger effects on child outcomes than domain-general quality measures (Bryant, Burchinal, & Zaslow, 2010; Zaslow et al., 2006).

Most studies to date have used rather global measures of ECEC quality involving either

structural quality characteristics such as group size and children-to-teacher ratio (e.g., Bauchmüller, Gørtz, & Rasmussen, 2011; Broberg, Wessels, Lamb, & Hwang, 1997) or global process quality constructs. The latter typically comprise emotional support measures or a single composite of emotional and instructional support measures (Herrera, Mathiesen, Merino, & Recart, 2005; Melhuish et al., 2013; Votruba-Drzal, Coley, & Chase-Lansdale, 2004). However, some recent studies have found stronger associations between ECEC quality and children's outcomes when domain-specific quality measures were used which were more aligned with particular developmental domains (for reviews, see Burchinal et al., 2011; Downer, Sabol, & Hamre, 2010; Zaslow et al., 2010). For example, measures of instructional support were found to be related to language and academic outcomes, whereas measures of emotional support were found to be related to socioemotional outcomes (Burchinal et al., 2011; Curby et al., 2009; Mashburn et al., 2008). Other studies have shown that the types of activities provided, referred to as the *implemented curriculum* in this study, are specifically related to developmental outcomes. For example, the provision of language, literacy and mathematics activities was found to specifically predict children's development in the targeted domains (Anders et al., 2012; Anders, Grosse, Rossbach, Ebert, & Weinert, 2013; Bowers & Vasilyeva, 2011; De Haan, Hoofs, Elbers, & Leseman, 2013; Hamre et al., 2010; Howes et al., 2008; Justice, Mashburn, Pence, & Wiggins, 2008; Klibanoff et al., 2006; Kuger & Kluczniok, 2008; Landry, Swank, Smith, Assel, & Gunnewig, 2006; Sylva et al., 2006). Likewise, providing play and work activities that were specifically designed to foster executive functions (Barnett et al., 2008; Diamond et al., 2007; Diamond & Lee, 2011) and social and emotional competences (Bierman et al., 2008; Domitrovich, Cortes, & Greenberg, 2007) predicted the development of children's abilities in these domains.

Studies investigating the effects of domain-specific quality measures on child outcomes mostly concern children older than three years of age. Studies on children below age three have mostly used a domain-general quality approach with a predominant orientation on emotional support. Therefore, little is known about possible domain-specific effects of differentiated quality and curriculum aspects of ECEC on the development of younger children. However, domain-specific quality may be a relevant issue in ECEC for younger children as well. The period before age three years is increasingly recognized as a sensitive period for the development of basic language and attention skills, which serve as building blocks for later social and cognitive competence (Diamond & Lee, 2011; Shonkoff & Phillips, 2000; Rose, Feldman, Jankowski, & van Rossem, 2008). A host of studies attests to the importance of early vocabulary acquisition as a gateway to acquiring

other language, literacy and cognitive skills (Hoff, 2006; Lee, 2011; Whitehurst & Lonigan, 1998). For example, building up a sufficiently large store of lexical items in the third year of life is thought to be a prerequisite for subsequent morpho-syntactic development (Hoff, 2006; Tomasello, 2003) and the vocabulary acquired at age three is a strong predictor of vocabulary and literacy skills at school age (Hart & Risley, 1995). In addition, recent studies point to the importance of early attention and related information processing capacities in the development of cognitive functions across the first years of life (Rose et al., 2008). Early attention underlies the development of executive functions such as cognitive flexibility, inhibitory control, and working memory (Cuevas & Bell, 2013; Garon, Bryson, & Smith, 2008) and is an important predictor of (pre-)academic skills such as literacy and numeracy (Duncan et al., 2007; Rhoades, Warren, Domitrovich, & Greenberg, 2011; Welsh et al., 2010). In sum, language and attention in very young children are foundational for the development of a wide array of other cognitive functions and skills later in life. However, it is currently unknown how general and specific aspects of ECEC quality can contribute to the development of these skills in very young children.

The present study examined the effects of ECEC quality on two-year-olds' development in Dutch day care centers and preschools and contributes to the issues outlined above in several ways. First, a differentiated approach to process quality in ECEC was adopted to examine both domain-general and domain-specific effects of ECEC quality. We distinguished between domain-general emotional support, and domain-specific behavioral instructional support, based on observations. In addition, we included domain-specific measures of three core elements of the early childhood curriculum, namely free play, activities fostering self-regulation, and pre-academic activities, based on teacher reports. Second, we included children's vocabulary and selective attention as outcome measures, and we followed the development of these key skills between age two and three years and we determined the effects of quality and curriculum. Before turning to the aims of the present study, we review findings from previous studies on the effects of domain-general and domain-specific ECEC quality on vocabulary and attention development.

Domain-General and Domain-Specific Quality Effects on Vocabulary and Attention

Several studies examining the effects of domain-general ECEC quality have revealed the importance of close and warm teacher-child relationships for preschoolers' and kindergarteners' cognitive and language development, including vocabulary (Burchinal,

Peisner-Feinberg, Pianta, & Howes, 2002; Cadima, Leal, & Burchinal, 2010; Howes et al., 2008; Peisner-Feinberg et al., 2001). However, the effects of general quality on vocabulary and related language outcomes tend to be small, according to review studies (Burchinal, Kainz, & Kai, 2011; Zaslow et al., 2010). In a recent Canadian study involving two- to four-year-old children, stronger effects of domain-general ECEC quality on language outcomes at age four years were found, but the comprehensive quality construct in this study consisted of both global aspects of emotional support and specific aspects of language stimulation (Côté et al., 2013). Another recent study involving two-year-old children also revealed medium-sized associations between domain-general quality and children's cognitive development, including language, but this study used a cross-sectional design, hence not allowing conclusions about the effects on development (Ruzek, Burchinal, Farkas, & Duncan, 2014). The NICHD Early Child Day Care study in the USA revealed small effects of global positive caregiving on children's language and cognitive development from 15 to 36 months of age (NICHD ECCRN, 2000, 2006). Interestingly, most of these effects of domain-general quality disappeared when a specific measure of teachers' degree of language stimulation was included as a predictor in the analysis, suggesting a stronger effect of domain-specific quality. Likewise, several studies with preschoolers and kindergarteners found effects of domain-specific stimulation of children's language development in addition to effects of domain-general quality (Cadima, Leal, & Burchinal, 2010; Curby et al., 2009; Howes et al., 2008; Mashburn et al., 2008). Studies that included measures of the systematic provision of domain-specific language activities in the curriculum revealed that higher domain-specific quality so defined was associated with stronger progress of children in the targeted language skills (De Haan et al., 2013; Howes et al., 2008; Phillips, Gormley, & Lowenstein, 2009; Sylva et al., 2006). Also, domain-specific language and literacy intervention programs involving children from three to five years of age (Bierman et al., 2008; Dickinson & Caswell, 2007; Fantuzzo, Gadsden, & McDermott, 2011; Justice et al., 2008; Lonigan, Farver, Philips, & Clancy-Menchetti, 2011) and toddlers (Girolametto, Weitzman, & Greenberg, 2003) were found to be effective as far as the targeted skills were concerned. Typically, these intervention programs frequently provide structured language and literacy activities to children, such as circle time group discussion, shared book reading with group discussion, or explicit vocabulary instruction (for an overview, see Dickinson, 2011).

Only a few studies to date have related ECEC quality to children's development of attention and attention-related cognitive and affective control functions. The NICHD

Early Child Day Care study investigated the cumulative effects of domain-general ECEC quality based on teachers' positive caregiving involving children attending day care centers between 15 and 54 months old. Children's attention was measured when children were 54-months old using a standard assessment. No effects of domain-general quality were found when child and family background factors were controlled for (NICHD, 2006). However, in one longitudinal study with four-year-olds, teacher-child emotional closeness was found to predict children's attention four years later (Peisner-Feinberg et al., 2001). A number of studies examined domain-specific relationships between classroom quality and children's attention-related self-regulation skills, involving four-to-six-year-old children, showing associations between emotional support, classroom organization, and instructional support, and children's self-regulation development over the course of one year (Rimm-Kaufman, Grimm, Curby, Nathanson, & Brock, 2009; Weiland, Ulvestad, Sachs, & Yoshikawa, 2013). Interestingly, classroom organization, which refers to teachers' strategies to regulate children's behavior in the classroom, showed stronger effects than emotional and instructional support (Rimm-Kaufman et al., 2009). Further evidence for the role of ECEC in the development of attention-related control functions comes from a number of intervention studies that specifically targeted attention, executive functions and self-regulation. The Tools of the Mind (ToM) curriculum for three- to four-year-olds aims to promote broad foundational cognitive and social-emotional skills, including children's attention, working memory, and their ability to regulate their social and cognitive behavior (Barnett et al., 2008; Diamond, Barnett, Thomas, & Munro, 2007). In ToM guided play is considered the leading activity for developing these skills, with a major role for guided pretend play in which children are stimulated to plan, monitor and sustain a complex pretend scenario. In a randomized controlled trial, three- and four-year-old children who participated in ToM showed stronger growth in attention-based executive function skills and social behavior regulation (fewer behavior problems) than children who attended a conventional academic literacy-focused curriculum (Barnett et al., 2008; Diamond et al., 2007). Another curriculum focusing on enhancing emotional self-regulation is the Promoting Alternative Thinking Strategies curriculum (PATHS; Domitrovich, Cortes, & Greenberg, 2007). This curriculum has been shown to improve three- and four-year-old children's executive functioning skills, particularly the ability to sustain attention in performing tasks (Bierman, Nix, Greenberg, Blair, & Domitrovich, 2008; Domitrovich et al., 2007). The PATHS intervention consists of several components: the establishment of clear and predictable classroom rules and routines; the provision of activities specifically

fostering socio-emotional development in the daily program; the inclusion of lessons and activities to educate children in recognizing and regulating emotions; teacher modeling of social conflict-solving strategies and supporting children in applying these strategies themselves.

To summarize, there is solid evidence that both domain-general and domain-specific ECEC classroom quality matters for children's vocabulary. However, there is less evidence on the effects of ECEC quality on children's development of attention-related executive functions and self-regulation. Moreover, most of these studies concern children older than three years, whereas studies involving younger children have mostly used global emotional process quality measures. Therefore, the question is whether the same pattern of findings holds for younger children who are in the midst of a period of rapid development of vocabulary and attention skills.

Current Study

The current study aimed to investigate the effects of domain-general and domain specific process quality on children's development across the third year of life, focusing on two key cognitive skills: vocabulary and attention. Domain-general quality comprised of global emotional and behavioral support provided by the teacher, whereas domain-specific quality included several measures of teachers' support for children's cognitive and language development in ECEC provisions in the Netherlands. Observational measures were used to assess real time interactional process quality, distinguishing between emotional support, behavioral support and instructional support. In addition, teacher self-reports were used to assess the provision of free play activities, guided play activities to support executive function and self-regulation development, and teacher-led pre-academic activities. Children's attention and vocabulary were assessed at ages two and three years. Multilevel value-added structural equation modeling was applied to determine the net effects of process and curriculum quality on attention and vocabulary development, controlling for child and family characteristics.

The Dutch ECEC system consists of two main types of provision. The first type is center-based day care for children from birth until four years of age, which is on average attended for two full days a week (NCKO, 2011). The second type concerns preschools for two- to four-year old children. At age four, almost all children in the Netherlands enter primary school (Van Tuijl & Leseman, 2007). Following new legislation and changes in financing

(Child Care Act, 2005; OKE Act, 2010), the two types of preschool provision have become highly comparable in their structural and process quality characteristics (Slot et al., under review; Leseman & Slot, 2013), and, therefore no distinction was made between the two in the current analyses (for a similar approach, see NICHD ECCRN, 2006; Melhuish et al., 2013).

In the current study we addressed the following research question: What are the effects of observed process quality, including emotional, behavioral and instructional support, and the implemented curriculum, including the provision of play, self-regulation and pre-academic activities, on the development of children's vocabulary and attention skills between age two and three years? Based on previous research reviewed above, we expected that both domain-general and domain-specific quality predict growth in children's vocabulary and attention. With regard to domain-specific quality, we specifically expected (i) instructional support, including language modeling, and curriculum activities focused at language and literacy to predict vocabulary growth and (ii) behavioral support, instructional support and curriculum activities aimed at enhancing children's self-regulation to predict children's attention development. Regarding the effects of (free) play in the curriculum we used an exploratory approach.

Method

The present study used data from the first and second wave of the ongoing Dutch national cohort study pre-COOL, which is aimed at investigating the effectiveness of preschool education and care provisions in the Netherlands. Children's development will be tracked longitudinally across four measurement waves, from age two up until age five years. By that time the children will enter the national cohort study COOL on students' educational careers in primary and secondary school and they will be followed-up until age eighteen. To increase the likelihood of pre-COOL children entering primary schools that take part in the COOL study, sample recruitment followed a number of steps. First, a random sample of 300 primary schools participating in COOL was drawn from the COOL cohort and was asked to participate in pre-COOL. Next, the 139 schools that agreed to participate were asked to identify preschools and day care centers that were attended by most of their new students. In addition, municipal records and the internet were used to identify preschools and day care centers in the neighborhood of the participating COOL schools. About 500 centers were approached, of which 263 agreed to take part in pre-COOL (52.6%). A total of 375 teachers of 182 centers (69.2%) participated in the study by filling out a teacher

questionnaire, providing information on 295 classrooms (170 preschool, 125 day care). For logistic and methodological reasons, observations were only conducted in classrooms containing at least four children who participated in the child assessments of pre-COOL, resulting in 162 centers (61.6% of the entire pre-COOL sample) with a total of 276 classrooms (155 preschool and 121 day care classrooms). The participating preschools and day care centers were geographically spread and were located in urban, semi-urban or rural areas in all parts of the Netherlands. Children were eligible to enroll in pre-COOL when they were born between April 1 and November 1, 2008. Parents of eligible children were personally informed by their child's teacher about the pre-COOL study and were given a letter containing information about the study, explicitly giving them the opportunity to withdraw their child from participation by notifying the teacher. A total of 1878 children were eligible for participation. For 91.2% of them, parents agreed with participation in the study, resulting in a final sample of 1819 children at wave 1. An additional 610 children enrolled in the study at wave 2, because they had not yet entered the ECEC center at wave 1.

Participants

For the present study, data were used of children for whom classroom observations and child assessment data of at least one study wave were available (N = 850 children, 46.7% of the entire pre-COOL children's sample and N = 185 classrooms, 67.0% of the entire pre-COOL classroom sample, of which 99 day care classrooms). Most children in day care centers enrolled in the center at age two years, but children in preschools enrolled between age two and two-and-a-half years, and, therefore, enrolled into the study when they were slightly older (mean difference 2.1 months, p <.001) and a number of children (22.7%) enrolled in the study at wave 2, because they were not yet enrolled in the center during wave 1. Information on important child and family background characteristics was obtained from both parent and teacher reports. Family socio-economic background (SES) was based on the highest level of educational attainment of children's parents. In the Netherlands, two main educational tracks exist, namely the vocational track and the general academic-professional track. The vocational track was considered low to middle SES and the general track was considered high SES. However, SES was missing for more than half of the children (53.4%) due to a low response rate on the parent questionnaire. Table 1 provides descriptive statistics for the background variables age, gender, SES, and language(s) spoken at home.

Table 1. Children's Background Variables (N=850)

	Age in months M (SD)	(months) Range	Gender % male	SES % low	Home language non-Dutch %
Wave 1	28.3 (3.0) (N=645[1])	20-37	53.4% (N=626[1])	45.3% (N=393[2])	26.3% (N=533[1])
Wave 2	42.0 (2.7) (N=789[1])	34-49	52.5% (N=764[1])	44.6% (N=401[2])	26.4% (N=511[1])
Time (in months) between wave 1 and 2	13.6 (2.5) (574[1])	6-23			

[1] N based on information collected by the test assistants during the assessments at waves 1 and 2 respectively
[2] N based on parents' responses to the parent questionnaire

Child Measures

Receptive vocabulary

Receptive vocabulary was measured with the Dutch version of the Peabody Picture Vocabulary Test (PPVT-III-NL, Dunn, Dunn, & Schlichting, 2005). In this test, children select one out of four picture drawings after an orally presented word. Whereas this task is usually performed as a paper-and-pencil task, stimuli presentation in the current study was controlled by the experimental software E-Prime 2.0 (Schneider, Eschman, & Zuccolotto, 2002), and administered through a laptop computer to facilitate administration and scoring. The shortened version used in our study contained eight items per test set, instead of the usual twelve items, due to testing time constraints. At the first wave, sets 1, 2 and 3 were presented. At the second wave, sets 3, 4, and 5 were presented with more difficult vocabulary items. As each set contained eight items, there were 24 items in total at each wave, with eight items overlapping between waves. Pilot research with 111 two-year-olds and 97 three-year-olds established that the items that were removed did not differentiate well among children, as they were either very easy or very difficult (i.e., mean accuracy scores on these items were either below 30% or above 70%). Scores were calculated as the percentage of correct responses for each child. The PPVT had good internal consistency at both waves (Cronbach's alpha is .88 for wave 1 and .79 for wave 2).

Attention

Attention was measured with a computerized visual search task, which was designed for the purposes of the present study (Mulder, Hoofs, Verhagen, & Leseman, 2014) based on work by Gerhardstein and Rovee-Collier (2002) and Scerif, Cornish, Wilding, Driver, and Karmiloff-Smith (2004). In this task, children were shown a structured display containing

pictures of three different types of animals on the laptop screen using E-prime 2.0 (Schneider et al., 2002). The goal was to locate as many targets (elephants) as possible while ignoring distractors (bears and horses) that looked highly similar in terms of color and shape. When the child pointed to a target, it was crossed off with a line by the assessor through pressing a key, to minimize demands placed on working memory. Throughout the task, children were encouraged to search as fast as possible. Following three practice trials, children were given three test items, which lasted 40 seconds each. Each trial contained eight targets. At the first wave of the study, when children were aged two years, each of the three items contained a total of 48 animals in a structured 8 x 6 grid with a target to distractor ratio of 1:5. At the second wave of the study, when children were aged three years, the first two test items were the same as those given at the previous wave. The third item was replaced by a more difficult item, which contained a total of 72 animals in a structured 8 x 9 grid with a target to distractor ratio of 1:8 to make the task challenging enough for three-year-olds. The number of correctly located targets was summed across trials to achieve a total accuracy score. The test showed good internal consistency (Cronbach's alpha .90 for wave 1 and .77 for wave 2) and satisfactory convergent and criterion validity (Mulder et al., 2014). The task captures a number of different key attentional processes (Chun et al., 2011), in particular, the ability to focus on relevant information while ignoring distracting information (selective attention) and to remain vigilant over time (sustained attention).

Classroom Measures

Observed process quality

The *Classroom Assessment Scoring System Toddler* (CLASS Toddler; La Paro, Hamre, & Pianta, 2011) was used to assess classroom process quality. An officially approved Dutch translation of the CLASS manual was developed for the present study (Slot, Leseman, Mulder, & Verhagen, 2013). All observers were trained by a licensed CLASS trainer and achieved at least 80% inter-observer reliability (average agreement was 86.4%) on an online test, as recommended by the developers of the CLASS. Following the online test, the trainer conducted live observations with all observers once, prior to the data collection. Inter-observer reliability of the live observations was 89.9%.

Classroom quality was rated on eight dimensions, using 7-point scales ranging from 1 or 2 (*classroom is low on that aspect*); 3, 4 or 5 (*classroom is in the midrange*); to 6 or 7 (*classroom is high on that aspect*). Based on theoretical considerations (La Paro et al., 2011) and supported by confirmative factor analysis of the observation data of the current

sample (Slot, Boom, Verhagen, & Leseman, submitted), three domains were distinguished and included as predictors of child outcomes in the main analysis. *Emotional Support* was computed as the average of the scores on three dimensions (Cronbach's alpha = .75): Positive Climate, reflecting the warmth, respect, and enjoyment displayed during interactions between the teacher and children; Teacher Sensitivity, evaluating the extent to which the teacher is aware of and responsive to children's needs; and Regard for Child Perspectives, capturing the degree to which the teacher's interactions with children and the classroom activities provided are attuned to children's interests, and the degree to which children's independence is encouraged. The second domain *Behavioral Support* was the average of the scores on two dimensions (alpha = .47): Negative Climate, reflecting the overall negativity expressed in the classroom by the teacher and the children (scores are reversed), and Behavior Guidance, referring to the teacher's ability to promote positive behavior and redirect problem behavior. Negativity was hardly observed in the Dutch classrooms. As a result, the dimension Negative Climate had limited variation, which could explain the lower alpha. The third domain *Engaged Support for Learning* was based on three dimensions (alpha = .88): Facilitation of Learning and Development, considering how well the teacher facilitates activities to support children's learning and development; Quality of Feedback, assessing the degree to which the teacher's feedback promotes learning and expands children's participation, and Language Modeling, referring to the extent to which the teacher fosters, models and encourages children's use of language.

Teacher-reported developmental and educational activities

A structured questionnaire for teachers was used to assess the developmental and educational activities provided to children on a regular basis over a longer period of time. This questionnaire was carefully developed for the purposes of the current study, based on extant research into vocabulary development, emergent literacy and emergent numeracy, and extensively tested in pilot research with teachers of two- and three-year-old children to ensure age-appropriateness of the listed activities (for more information, see Slot, Leseman, Verhagen et al., submitted). Several scales were constructed covering a broad range of behaviors and activities. For the purpose of the current study, three types of activities were distinguished: free play activities, self-regulation activities and academic activities.

The scale *Play* (17 items; Cronbach's alpha= .89) assesses the degree to which the teacher encourages and enriches children's free play, for instance by asking questions, making suggestions, or providing materials for richer play or by encouraging pretend play. Examples of items are: "I let the children play without interfering", "I ask children

questions that stimulate their play" and "I encourage children to engage in role-play with other children, for instance by letting them take the role of 'a father' or 'cook' in the play". The scale ranges from 1 (*never*) to 7 (*always*).

The scale *Self-regulation activities* (12 items; Cronbach's alpha= .88) measures the extent to which the teacher provides play and work activities and uses guiding strategies that can foster several aspects of children's self-regulation. The questionnaire included items concerning the promotion of emotion recognition and regulation (e.g., "I talk to children about feelings and emotions, for instance when reading a book or when addressing children's personal experiences"), planning, monitoring and evaluating play (e.g., "Before children start a task or an activity, I ask them how they are going to tackle it, what the plan is"), turn taking and waiting (e.g., "I provide activities in which children learn to take turns, such as memory or a board game"), and attention (e.g., "When children's attention fades during an activity, I try to reengage them, for instance by asking them questions or to bring in a new idea"). Items were based on principles of the ToM and PATHS curricula (Barnett et al., 2008; Bierman et al., 2008; Domitrovich et al., 2007; Diamond et al., 2007). The scale ranges from 1 (*never*) to 7 (*always*).

The scale *Academic activities* (25 items; Cronbach's alpha= .92) assesses the average frequency of pre-academic activities including language, involving several forms of language use, such as singing, rhyming, conversations and language instruction (e.g., "Having conversations about informative subjects, for instance about animals and plants or the seasons"), literacy and literacy materials (e.g., "Asking the children questions about the content of the story during or after reading the story"), and math activities, for instance counting and sorting activities, and activities exploring different shapes (e.g., "Counting how many objects you have, for example counting till five and saying 'I have five marbles'"). Answers were rated on a 7-point scale, ranging from 1 (*never*) to 7 (*at least three times a day*).

Procedures

Children's vocabulary and attention were measured twice. During the first wave of the study, children in day care centers were assessed when they were about two years old and preschool children were tested as soon as they entered the preschool, mostly around two-and-a-half years of age. The second measurement took place about one year later. Children were tested individually by trained research assistants in a quiet room in the center in test sessions that lasted approximately 45 minutes. Tasks were intermixed with tasks not reported on in this study and were given in a fixed order. The vocabulary and attention

tasks were always given as second and third in the battery.
Classroom observations were conducted in a four-month period between the two child assessments. Each classroom was observed during one regular morning. All classrooms were observed within three months after the observers were trained. The observers rated classroom processes and teacher behavior during four observation cycles on one morning, each lasting between 15 to 20 minutes, in accordance with the guidelines of the CLASS. Scores per cycle were aggregated to a single score for each dimension.

Analysis Strategy

The child measures for vocabulary and attention were adapted in difficulty for wave 2 with partially overlapping items between wave 1 and 2. First, we applied Item Response Theory (IRT) modeling to the dichotomous vocabulary item scores at waves 1 and 2, using the overlapping items (33%) to calibrate all items. Test equating is common for IRT models and using non-parallel forms combined with test equating is considered to be a distinct advantage of IRT over the Classical Test Theory approach (Embretson & Reise, 2000). We exported the estimated ability scores for all children as input for subsequent analyses. This resulted in two ability scores for each participant (one for wave 1 and one for wave 2). We used both a Maximum Likelihood estimator and a Weighted Least Squares estimator: the first is assumed to give slightly more accurate item estimates (which were used in the current study), but gives no absolute fit measures, whereas the second estimator does give absolute fit measures, which showed satisfactory fit (RSMEA = .020, CFI = 0.954, TLI = 0.953). The correlation between results of both estimation methods was very high.

The scores on the selective attention test items ranged from 0 to 8 correctly identified targets and were approximately normally distributed. Therefore, IRT modeling was not considered appropriate in this case. However, we copied the calibration approach for vocabulary and computed factor scores for wave 1 and wave 2 using Confirmatory Factor Analysis with means structure. Since the third item of the attention test used at wave 2 was comparable in many ways to the items used at wave 1, but more difficult than its counterpart at wave 1, we allowed the intercept for this item to be different by the same constant value for all participants. The model fit was good (RSMEA = .036; CFI = .990; TLI = .988) and we used the two estimated factor scores as input for subsequent analyses.
Second, to determine effects of quality and curriculum, a multilevel value-added approach was applied, using multilevel structural equation modeling. On average five children per classroom participated in the study, yielding a nested data structure. All analyses were done

with Mplus (Version 7; Muthén & Muthén, 1998-2012). Intraclass correlations (ICC's) were calculated separately for attention and vocabulary to determine the proportions of classroom level variance, revealing significant classroom level variance (see Table 2).

To be able to determine the net contribution of quality and curriculum to children's development at the classroom level, child level effects were controlled for by specifying a structural model with cross-lagged effects of wave 1 vocabulary and attention on wave 2 vocabulary and attention, and with several child level predictors of wave 2 vocabulary and attention. Three dichotomous variables were constructed to represent gender (0=male and 1=female), SES (0 = low to middle SES and 1 = high SES) and home language (0= children exposed to other language(s) than Dutch (as well) and 1= exposed to only Dutch at home) respectively. Furthermore, age at wave 1 was controlled for as well as differences in the time elapsed between wave 1 and wave 2. All background variables were used as covariates of the wave 1 child assessments. Likewise, at the classroom level, the cross-lagged effects of wave 1 vocabulary and attention on wave 2 vocabulary and attention, and all covariances of the classroom quality and curriculum measures with wave 1 vocabulary and attention were specified in the model.

Following the recommendations of Gollob and Reichardt (1987), the net added value of the predictors for children's vocabulary and attention development can be determined by, first, specifying the autoregression effects of wave 1 on wave 2 vocabulary and attention, reflecting endogenous development of these skills, and second, by partialling out the shared variances of predictors and the child assessments at wave 1, at both the child and the classroom level. The coefficients of the autoregression paths reflect the mean slope of the developing skills, whereas the residual variances reflect individual deviations from the mean slope: some children gain more, others less. The degree to which the residual variances in wave 2 vocabulary and attention can be predicted by model variables at the child and classroom level, controlled for covariances with the wave 1 vocabulary and attention measures, is a direct estimate of their added value.

Table 2. Descriptive Statistics at Waves 1 and 2 and Intraclass Correlations (ICCs) for the Child Measures in the Model Based on IRT (N_{within} = 850, $N_{between}$ = 185, average cluster size 4.60)

Variable	ICC	Mean	SD	Range	N
Vocabulary wave 1	.17*	1.09	1.04	-.176 – 4.19	850
Vocabulary wave 2	.20*	3.64	1.68	-.87 – 8.41	850
Attention wave 1	.14*	3.43	1.33	-.55 – 7.04	835
Attention wave 2	.13*	6.08	.75	1.86 – 7.72	835

* p<.05

Missing data

Data were only used for the children participating in the observation study (*N*=850). There were no differences in children's vocabulary at wave 1 between the final sample for whom classroom observations were available, compared to the remainder of the sample without classroom observations. Children in the final sample had significantly higher attention scores at wave 1 compared to the remainder sample, but this effect was rather small (*d* = .14). Due to the design of the study, a number of children did not participate in the first data collection wave, because they were not yet enrolled in the participating center. Children who enrolled at wave 2 did not differ significantly on child and family background characteristics (gender, age at wave 2, SES and home language) from the children who had been enrolled at wave 1. For the first wave of the study, data were available for 77.2% of the total sample. Complete observational data were available for all 180 classrooms and complete teacher self-report data were available for 81 classrooms (45%) of the current sample. Comparisons between the current subsample and the total pre-COOL sample revealed no significant differences on the teacher report measures, but the scores on the CLASS domains were significantly lower in the current sample compared to the total sample (all *ds* < .30). As recommended, missing data were dealt with by using full information maximum likelihood (FIML) estimation in Mplus (Enders, 2010), in which the standard errors for the parameter estimates are computed using the complete observed information matrix (Muthén & Muthén, 1998-2012).

Structural equations model

The main analysis was carried out in two steps. First, the structural equation model for both vocabulary and attention was estimated at the child and classroom level using a non-restricted baseline model. At the child level, the model included the wave 1 vocabulary and attention scores and background variables as predictors of wave 2 vocabulary and attention scores. At the classroom level, the model included the wave 1 vocabulary and attention scores aggregated to the classroom level and all quality aspects as predictors of the classroom level wave 2 vocabulary and attention scores. Furthermore, all classroom level covariances were specified. Second, the model was trimmed by eliminating non-significant paths with $p > .10$ and with a $|\beta| < .05$ (Wuensch, 2012) in a step-by-step fashion to obtain the most parsimonious model. Model fit was evaluated with several fit indicators: the Chi-Square test of goodness of fit, the Comparative Fit Index (CFI), the Tucker Lewis Index (TLI), the Root Mean Square Error of Approximation (RMSEA) and the Standardized Root Mean Square Residual (SRMR) at both the child and classroom level,

with CFI and TLI > .95, RMSEA < .05, and SRMR < .05 indicating good fit. Standardized regression coefficients were used as measures of the effect size with β < .10 indicating a small effect, a β of around .30 a medium-sized effect and β >.50 indicating a large effect (Kline, 2005). Mplus by default standardizes to the variance on the *within* level for *within*-level relationships, and to the variance of the *between* level for *between*-level relationships (Hox, 2010). To interpret and compare between *within* and *between* level effects we chose to standardize the regression coefficients to the total variance according to the following formula: b * ($SD_{predictor}/SD_{outcome}$) in which the *SD* is the square root of the total variance of the dependent variable concerned (Hox, 2010).

Results

Table 2 presents the descriptive statistics and the intraclass correlations for children's vocabulary and attention skills at wave 1 and 2, revealing significant classroom level variance. Both vocabulary and attention showed a considerable average growth over the period of about one year between age two and three years, with Cohen's *d* of 1.9 and 2.3 respectively, based on the IRT scaled scores at wave 1 and 2 for children with data at both measurements waves. Descriptive statistics of the quality measures are presented in Table 3. Regarding the process quality measures based on the CLASS observations, the results revealed moderate to high quality in the domains of Emotional Support and Behavioral Support and low to moderate quality in the domain of Engaged Support for Learning according to the CLASS standards. The teachers' self-reports on provided activities showed a moderate level of provision of play, self-regulation and academic activities, indicating that, on average, these activities were offered on a weekly basis.

Table 4 shows the correlations between children's wave 1 attention and vocabulary measures and the quality and curriculum measures at the classroom level obtained between wave 1 and wave 2. The inter-correlations of the CLASS domains as well as the inter-correlations between the self-reported process quality indicators are moderate to strong. The correlations between the CLASS dimensions and the self-reported activities are much smaller in magnitude, but mostly significant and in the expected direction. Note that several significant, moderately strong correlations between wave 1 vocabulary and attention and the classroom quality and curriculum measures are observed, indicating effects of selective placement, which were controlled for in the model. We will return to this issue in the Discussion section.

Table 3. Descriptive Statistics for the Observed and Self-Reported Quality Aspects

Quality aspect	M	SD	Range	N
Emotional Support	4.94	0.70	2.67-6.50	924
Behavioral Support	5.85	0.48	4.38-7.00	924
Engaged Support for Learning	3.23	0.78	1.83-5.92	924
Play	4.19	0.74	2.43-6.05	397
Self-regulation activities	4.03	0.85	2.64-6.91	416
Academic activities	4.34	0.87	2.29-6.83	397

Table 4. Correlations (Standardized Covariances) Between Wave 1 Child Measures Aggregated to the Classroom Level and Quality and Curriculum Measures at the Classroom Level

			Child measures		CLASS			Activities		
		Wave 1	1	2	3	4	5	6	7	8
Child measures	1	Vocabulary		.02	.21	.21	.10	-.36*	-.30*	-.49**
	2	Attention			-.20	.09	-.35**	.01	.25	-.04
Observed quality	3	Emotional Support				.63**	.70***	.10	.06	.08
	4	Behavioral Support					.48***	.13	.15†	.25**
	5	Support for Learning						.27**	.09	.22†
Self-reported activities	6	Play							.63***	.49***
	7	Self-regulation								.50***
	8	Academic								

***p < .001, **p < .01, *p < .05, †p < .10

As a first step of the main analysis, the non-restricted baseline multilevel model was estimated. This was a saturated model, hence no fit indices could be obtained. However, this model contained several non-significant covariances and structural paths. To obtain a more parsimonious model, the structural paths at the child and the classroom level that were not significant at $p < .10$ and with a $|\beta| < .05$ were constrained to zero, yielding a good model fit: $\chi^2(9) = 7.06$, $p = .63$; RMSEA = .00; CFI = 1.00; TLI = 1.01; $SRMR_{within} = .01$, $SRMR_{between} = .01$. The final model is presented in Figure 1.

At the child level, vocabulary and attention showed strong stability from age two to three years, as can be seen in Figure 1 and Table 5. Attention at wave 1 significantly predicted growth in vocabulary, whereas vocabulary at wave 1 significantly predicted growth in attention, both with small effect sizes. The time between wave 1 and wave 2, due to differences in age of enrollment, was also associated with growth in vocabulary and attention, as was expected: the longer the time between first and second assessment, the larger the gain. Furthermore, children from monolingual Dutch homes showed stronger growth in vocabulary than children from multilingual homes. There were no unique net

Table 5. Child Level and Classroom Level Predictors of Children's Gains in Vocabulary and Attention From Age Two to Three Years

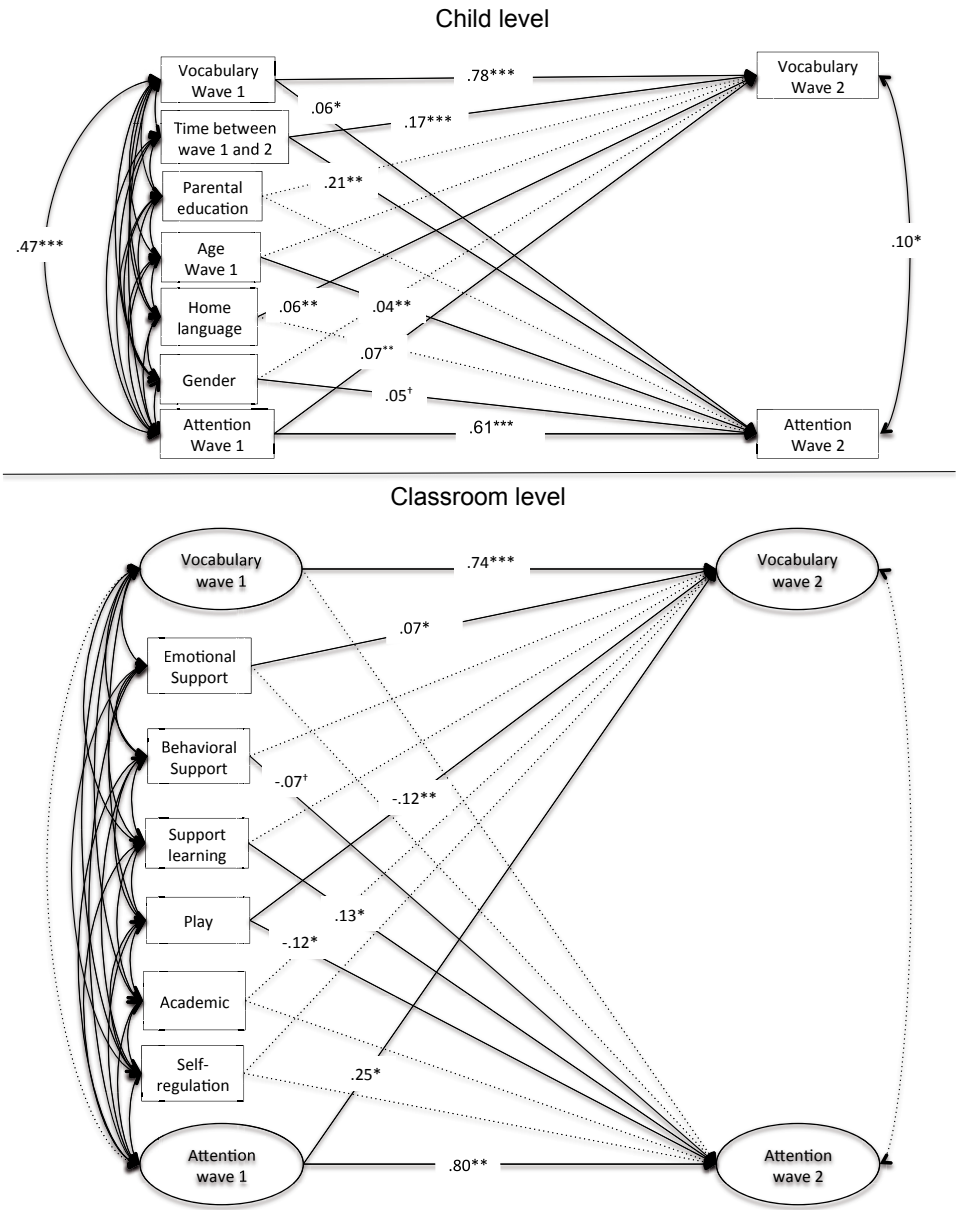
	Vocabulary			Attention		
	B	SE	β	B	SE	β
Child level						
Gender	#			.07	.04	.05†
Age	#			.03	.01	.04**
Time between wave 1 and 2	.11	.01	.17***	.06	.01	.21***
SES	#			.08	.06	.06
Home language	.17	.02	.06**			
Vocabulary wave 1	1.28	.01	.78***	.05	.05	.06*
Attention wave 1	.08	.03	.07**	.35	.03	.61***
Classroom level						
Vocabulary wave 1	1.20	.31	.74***	#		
Attention wave 1	.31	.16	.25*	.47	.11	.80***
Emotional Support	.17	.07	.07*	#		
Behavioral Support	-.11	.13	-.03	-.13	.07	-.07†
Support for Learning	#			.12	.04	.13**
Play	-.24	.10	-.12*	-.11	.05	-.12*
Academic activities	#			-.06	.04	-.07
Self-regulation activities	#			.07	.05	.07

***p < .001, **p < .01, *p < .05, †p < .10

effects of gender, age and SES on vocabulary development. Age was positively related to growth in attention from age two to three years, with a small effect size. There were no significant unique effects of gender, home language and SES on attention skills.

At the classroom level, vocabulary and attention showed strong stability between age two and three years. Attention significantly predicted classroom-level growth in vocabulary, with a small effect size, whereas there was no significant cross-lagged classroom-level effect of vocabulary on attention. Regarding classroom quality, a significant positive, but small effect of observed global emotional support on vocabulary growth was found. Note that the effect sizes were standardized to the total variance at both the child and classroom level. In addition, observed support for learning had a small positive and statistically significant effect on children's growth in attention. No effects were found of observed behavior support and support for learning on children's vocabulary. Teacher reported support of free play was negatively associated with children's gains in both vocabulary and attention at age three years. Also, no statistically significant effects were found for the teacher reported provision of academic and self-regulation activities on children's vocabulary and attention

Figure 1. Multilevel Model with Classroom Quality Measures as Predictors of Children's Growth in Attention and Vocabulary. Non-Significant Paths at the .05-Alpha Level are Represented by Dashed Lines.



***p < .001, **p < .01, *p < .05, †p < .10
indicating paths that were constrained in the model.

development. The between-level variance in vocabulary was 20% and 12% in attention, both with non-significant residual variances, indicating that almost all variance at the classroom level was explained by the predictors in the model.

Discussion

The current study examined domain-general and domain-specific effects of ECEC quality and curriculum on children's vocabulary and attention development during the sensitive period between age two and three years. By applying a value-added multilevel design, controlling for child level effects and effects of selective placement, the net effects of quality and curriculum on children's gains in these basic skills could be estimated. We distinguished between emotional support, behavioral support and instructional support observed with the CLASS Toddler and included three different measures of the early childhood curriculum, based on teacher self-reports: the provision of free play activities, activities fostering self-regulation, and pre-academic activities.

Our results indicate that a differentiated approach to assessing effects of quality and curriculum of early childhood education and care on children's outcomes can be recommended, as the effects of the quality and curriculum measures on children's outcomes differed by developmental outcome. Emotional support was related to gains in children's vocabulary, whereas support for learning was associated with growth in children's attention. Moreover, the relevance of focusing on vocabulary and attention development in the third year of life was confirmed by the finding that children made substantial gains in vocabulary, as has been found in other studies involving children of this age (Dickinson, 2011; Duncan et al., 2007), and in attention, in line with brain research that shows enhanced plasticity of the attention-related prefrontal cortex in this period (Shonkoff & Phillips, 2000).

Although most of the variance in vocabulary and attention was located at the child level, substantial and statistically significant additional variance was associated with the classroom level. Children's gains in vocabulary from age two to three years were predicted most strongly by child level variables, including children's initial vocabulary and attention skills, their home language, and the time between the two measurement waves, which reflected differences in enrollment age and was included as a control variable. At the classroom level, a positive small effect of observed domain-general emotional support on vocabulary growth was found, in line with previous studies (Burchinal, Peisner-Feinberg, Pianta, & Howes, 2002; Cadima, Leal, & Burchinal, 2010; Côté et al., 2013; Peisner-Feinberg et al.,

2001). Although the effect size was small, based on the total variance, general emotional support predicted a relevant part of the classroom level variance in vocabulary at wave 2. However, we found no effects of domain-specific observed support for learning, which included the dimension language modeling by the teacher, nor for the teacher reported provision of pre-academic activities, which included language and literacy activities such as circle time discussion and shared reading. In addition, no effect of the provision of activities intended to support children's self-regulation was found on vocabulary growth. Remarkably, however, there was a small *negative* effect of provision of free play activities on vocabulary growth.

Gains in children's attention were predicted by both child and classroom level variables. Children's gains in attention skills from age two to three years were also most strongly predicted by child level variables, including children's initial attention skills and vocabulary at wave 1, gender, age, and the time between the two measurement waves. At the classroom level, the strongest effect was found for domain-specific observed support for learning on children's attention development, although small in magnitude based on the total variance at both child and classroom level. This CLASS domain emphasizes the facilitation of cognitive development by the teacher through providing planned activities that elicit thinking and reasoning, focused play and work, requiring the teacher to actively stimulate children's play and to scaffold their thinking and learning. The effect of support for learning on attention development is in line with studies involving older children (Weiland et al., 2013). Contrary to our expectations we did not find significant effects of observed behavioral support, the provision of (pre)academic activities and activities aimed at enhancing self-regulation. Also, as for vocabulary, we found a negative effect of teachers' reported support of free play on attention development.

A possible explanation for the partly unexpected findings regarding the effects of the observed support for learning and the self-reported provision of (pre) academic activities on vocabulary concerns the way vocabulary was assessed. The vocabulary test used in the current study was a shortened version of the PPVT, which is a widely used test of *general* vocabulary, meaning that the test items were sampled from a large corpus of age-appropriate words used mainly in informal situations (e.g., daily routines at home) at this young age, which may explain why no effects of support for learning and the provision of academic activities were found in the current study. For example, a high score on support for learning, one of the CLASS domains, is likely to indicate exposure of children to conceptual talk, explanations and relatively advanced language use during group discussions, dialogical

reading, discovery and problem-solving activities. Possible effects of these activities on children's vocabulary may not have been adequately captured by the PPVT. In contrast, the domain emotional support strongly focuses on the personal interactions of teachers and children to establish and maintain close affectionate relationships, and emphasizes a verbal style of engaging with children in activities throughout the day and particularly during routines such as mealtimes. A high level of emotional support, thus, is likely to indicate a high degree of exposure of children to rather general (informal) language throughout the day, the effect of which might have been well-captured by the PPVT as a general vocabulary measure. In addition, previous research suggests that warm, affectionate relations may also directly influence word learning as a consequence of better attunement of adult and child in verbal interactions (Van IJzendoorn, Dijkstra, & Bus, 2006). In support of this explanation, a recent review shows that specific vocabulary interventions have medium to strong effects on trained and closely related vocabulary (within the same register, indicating near transfer), but only small effects, or no effects at all, when general vocabulary measures are used (Marulis & Neuman, 2010). A recent intervention study involving two- and three-year old children, which also focused on the provision of academic activities, revealed no effects on children's general vocabulary development either (Landry et al, 2014).

In view of what is considered developmentally appropriate for young children, the negative effects of the provision of free play activities in the curriculum are counter-intuitive. To explain these negative effects, it is important to consider the nature of play in the Dutch context. Play in Dutch early childhood education and care predominantly means free play activities without much teacher involvement. Leseman, Rollenberg and Rispens (2001), studying four-year-olds in Dutch preschools, found that only during 5% of the time teachers were verbally interacting with children in play situations. In a previous study involving the present sample, we found that on a regular morning, children spent about one third of the observed time to free play, mostly without teacher involvement, and thus not constituting a context involving scaffolding and language modeling (Slot, Leseman, Verhagen, et al. submitted). Indeed, the separate CLASS scores for free play indicated relatively low support for learning in comparison to observations of situations involving teacher guided educational activities, such as circle time or shared reading, and creative activities. Although our questionnaire assessing the occurrence of play activities included a number of items that implied teacher involvement (e.g., asking children questions during play), we suspect that the scores on this scale mainly reflected unguided play, which may explain the negative relation with vocabulary at wave 2. Likewise, we presuppose that

during free play, children are less likely to show focused attention, because usually no goals are set by the teacher, nor are children encouraged to complete a task or to persevere in performing a task or play, possibly explaining the negative effect on attention at wave 2. In support of this, a recent study in a sample of Dutch three- to six-year-old preschoolers showed that children tended to wander around a lot during free play activities (De Haan et al., 2013). The provision of activities aimed at promoting self-regulation showed a small positive, but contrary to our expectations, statistically not significant effect on the growth of children's attention skills. A possible explanation is that the self-regulation activities measure was rather broad, covering several different aspects of self-regulation, and therefore not sensitive enough to reveal associations with the specific measure of children's attention skills. Another possible explanation is that using teachers' self-reports is not suited to capture the actual promotion of self-regulation in the classroom.

A final note concerns the moderate to strong negative correlations found at the classroom level of children's wave 1 vocabulary and, to a lesser extent, attention skills with classroom quality, particularly observed engaged support for learning and the provision of academic activities. These negative correlations suggest *positive* selection effects that are likely due to the targeted early education policy in the Netherlands. In ECEC classrooms with a comparatively big share of socioeconomically disadvantaged children, classroom quality tends to be higher and teachers tend to report to provide play, self-regulatory and pre-academic activities more frequently than in classrooms with a comparatively small share of disadvantaged children. In the past decades, day care centers and preschools serving mainly disadvantaged children have been provided with additional teacher training activities and education programs to raise the emotional and educational quality of ECEC for disadvantaged groups, albeit with partial success (Doolaard & Leseman, 2008).

There are a number of limitations to the present study. In non-experimental field studies, like the current study, selection bias is a common issue (NICHD, 2006; Sylva et al., 2011). The present study dealt with this issue by employing a value-added multilevel approach, separating child level effects from classroom level effects and controlling for all covariances of children's prior vocabulary and attention with the quality and curriculum measures. Another limitation concerns missing data. Due to logistic considerations, differences in enrollment age and non-response, complete child data were only available for a subset of the current sample. Following methodological recommendations (Enders, 2010), we used full information maximum likelihood estimation to deal with missing data. Another limitation is that the current sample differed in ECEC quality from the total sample,

limiting generalizability. This difference in quality might be due to selective response on a questionnaire that was sent out to obtain information of the centers needed for merging child level and classroom level data. A final limitation of the present study concerns the measurement of background characteristics. Only rough indicators were available of the home language environment and SES, with substantial missing data for the latter. Although we did find an effect of home language on vocabulary growth, a more sophisticated measure, including other aspects of children's home learning environment as well, could have revealed stronger influences of the home environment on children's development (Melhuish, 2010).

Despite these limitations, the current study contributes to our knowledge of ECEC effects on children's development by revealing effects of ECEC quality and curriculum characteristics on children's vocabulary and attention development from age two to three years. These effects concern domain-general as well as domain-specific characteristics of ECEC quality, including effects of the curriculum of developmental and educational activities provided to children. As such, the current findings point to the importance of evaluating quality as a multi-faceted construct.

CHAPTER 5

Preschoolers' Cognitive and Emotional Self-Regulation in Pretend Play: Relations with Executive Functions, Quality of Play, and Classroom Quality

Slot, P. L., Mulder, H., & Leseman, P. P. M., (2014)

Preschoolers' cognitive and emotional self-regulation in pretend play: Relations with executive functions, quality of play, and classroom quality.

Manuscript under revision

Abstract

The preschool period is marked by rapid growth of children's self-regulation and related executive functions (EF). Self-regulation is considered an important aspect of school readiness and is related to academic and social-emotional outcomes in childhood. Early childhood programs and curricula, including play-based curricula, are increasingly aiming at improving children's self-regulation and some have shown to be effective. However, how particular curriculum activities affect self-regulation is still not clear and calls for the use of observational methods. The current study observed three-year-olds' cognitive and emotional self-regulation in a naturalistic play setting. In addition, the relation between test-based measures of EF and children's observed cognitive and emotional self-regulation were examined. Moreover, associations between self-regulation, the quality of pretend play, and global classroom quality were investigated to explore the contribution of contextual factors to children's self-regulation. The results indicated that children at this age already show several aspects of cognitive and emotional self-regulation. Furthermore, children's test-based EFs predicted observed cognitive and emotional self-regulation, but not entirely as expected. Cognitive, or 'cool', EF was significantly related to emotional self-regulation but not to cognitive self-regulation, whereas affective, or 'hot', EF was not significantly related to either cognitive or emotional self-regulation. The quality of pretend play was strongly associated with cognitive self-regulation and, to a lesser extent, with emotional self-regulation. Finally, global emotional classroom quality as assessed with the Classroom Assessment Scoring System was not related to cognitive and emotional self-regulation. Possible implications of the findings for early childhood programs are discussed.

Introduction

Self-regulation, defined as the ability to control or direct one's attention, thoughts, emotions and actions optimally to situational demands in order to reach personal goals (Baumeister, Schmeichel, & Vohs, 2007; Carver & Scheier, 2011; Hofmann, Schmeichel, & Baddeley, 2012; McClelland & Cameron, 2012; McClelland, Ponitz, Messersmith, & Tominey, 2010), begins to develop already in the first years of life (Blair & Diamond, 2008; Bronson, 2000). Developmentally appropriate self-regulation is considered a core aspect of children's school readiness (e.g. Blair & Diamond, 2008; Calkins & Williford, 2009) and is found to predict academic achievement, social competence, and positive classroom behavior in several studies

(Calkins & Williford, 2009; McClelland, Cameron, Connor et al., 2007; McClelland, Acock, & Morrison, 2006; McClelland, Morrison, & Holmes, 2000; Morrison, Ponitz, & McClelland, 2010; Raver et al., 2012; Rimm-Kaufmann et al., 2009). Most young children in western countries participate in some form of center-based early childhood education and care (ECEC) before they enter the formal school system (OECD, 2013). However, it is currently not clear in what ways ECEC can contribute to children's self-regulation development.

Despite consensus on the importance of self-regulation, there is a lack of conceptual clarity about the construct and its underlying components (McClelland & Cameron, 2012). In some studies self-regulation is treated as a broad overarching concept, referring to the strategic use of attention, effort, verbalizations, and metacognitive and meta-emotional knowledge (such as expressing ideas about feelings) in everyday situations (Whitebread et al., 2009; Zimmerman, 2000). Within this tradition, self-regulation is usually studied as situated in actual social settings such as the preschool classroom, and assessed with observation and interview methods. In other studies self-regulation is closely related to the set of skills referred to as executive functions (EF), including working memory, inhibitory control and cognitive flexibility (Blair & Ursache, 2004; Blair, Zelazo, & Greenberg, 2005; Miyake et al., 2000). Definitions of EFs show convergence with the broad construct of self-regulation outlined above by emphasizing the role EFs play in flexibly adapting behavior to the demands in everyday situations (Blair & Ursache, 2004; Blair, Zelazo, & Greenberg, 2005; Hofmann et al., 2012; Morrison, Ponitz, & McClelland, 2009), yet EFs in these studies are usually assessed with standard neuropsychological assessment tools in a test setting. EFs have been found to predict school achievement, social competence and behavioral adjustment in many studies (Blair & Razza, 2007; Espy et al., 2004; Rhoades, Greenberg, & Domitrovich, 2009; Riggs, Jahromi, Razza, Dillworth-Bart, & Mueller, 2006; Van der Ven, Kroesbergen, Boom, & Leseman, 2011). However, these studies are correlational, relating test-based measures of EFs to cognitive and behavioral outcome measures, but do not reveal how children's EF skills operate in children's self-regulation behavior in actual educational contexts, and how these educational contexts, in turn, can contribute to self-regulation development. Moreover, the two approaches diverge with regard to recommendations for practice. Whereas the first approach is focusing on possible supports that can be provided in the social context of, for instance, the ECEC classroom, the second approach has given rise to specific EF training programs for individual children with a focus on children showing delayed development of self-regulation.

The current study attempts to bring the two approaches together in order to increase

insight in what can be done in ECEC to support self-regulation development of all children. Focusing on pretend play in the preschool classroom context, we examine the relationships between observed self-regulation during play and children's performance on tests of EF, and we look for situational factors, in particular the quality of play and the general quality of the classroom that are hypothesized to co-determine children's observed self-regulation beyond test-based EF.

Cognitive and Emotional Self-Regulation

Self-regulation in preschool contexts has been shown to comprise of cognitive as well as emotional aspects (Bodrova & Leong, 2006). Cognitive self-regulation includes children's explicit metacognitive knowledge about thinking and learning processes, metacognitive strategies to regulate task behavior, such as planning, monitoring, and control of ongoing cognitive processes, and motivation-related factors such as persistence and sustained attention (Bodrova & Leong, 2006, Pintrich, 2002; Whitebread et al., 2009). Cognitive self-regulation has been found to be related to efficient learning behavior, task engagement, and academic achievement (Brock et al., 2009; Willoughby et al., 2011) and to social-emotional outcomes such as compliance, social competence and behavioral adjustment in the classroom (Cadima, Leal, Ferreira, Vieira, & Matos, 2014; Eisenberg, Smith, Sadovsky, & Spinrad, 2004; Fabes et al., 1999). Emotional self-regulation includes children's explicit knowledge about emotions, their strategies to control and modulate the expression of emotions, and their ability to meet the social expectations of the situation, to get along with peers and to resolve conflicts with peers (Bodrova & Leong, 2006; Denham et al., 2012). Emotional self-regulation has been found to be related to children's ability to conform to classroom behavior rules and academic achievement in kindergarten (Howse, Calkins, Anastopoulos, Keane, & Shelton, 2003) and primary school (Denham et al., 2012). Emotion regulation has also been shown to be negatively related to externalizing and internalizing behavior problems (Blair, Denham, Kochanoff, & Whipple, 2004; Cadima et al., 2014). Likewise, a distinction is frequently made between executive control of cognitive behavior and executive control of affective behavior, or between 'cool' and 'hot' executive functions (Brock, Rimm-Kaufman, Nathanson, & Grimm, 2009; Denham, Warren-Knot, Hamada Bassett, Wyatt, & Perna, 2012; Willoughby et al., 2011), for which different assessment tools have been developed (Carlson, 2005). Evidence indicates that cool EFs are related to academic skills (Brock et al., 2009; Hongwanishkul et al., 2005; Welsh et al., 2010; Willoughby et al., 2011), whereas hot EFs are associated with social-behavioral skills, such

as emotion regulation, and behavioral adjustment (Carlson & Wang, 2007; Kochanska, Murray, & Harlan, 2000; Willoughby et al., 2011). In the present study, therefore, we will examine if cognitive and emotional self-regulation in play can be reliably distinguished and whether test-based measures of children's hot and cool EFs show the expected differential relations with observed cognitive and emotional self-regulation in play.

Self-Regulation and Classroom Quality

Although self-regulation and EF development show marked individual differences already at a young age (Rothbart, Ahadi, & Evans, 2000; Mulder et al., 2014; Slot et al., under review), it has been well established that self-regulation is malleable and develops from an interplay of biological and social factors (Blair & Diamond, 2008; Diamond, 2007; McClelland et al., 2010; Ursache, Blair, & Raver, 2012; Raver et al., 2012). Emotionally supportive caregiver-child relationships have been shown to be beneficial for children's self-regulation (Silva et al., 2011). Correspondingly, in the context of preschool education, an emotionally positive classroom climate, marked by secure social relationships, low negativity and stress, and high teacher sensitivity, has been shown to be beneficial for children's self-regulation development (Merritt, Wanless, Rimm-Kaufman, Cameron, & Peugh, 2012; Morris, Millenky, Raver, & Jones, 2013). Acccording to Ursache and colleagues (2012), a positive classroom environment reduces children's and teachers' stress levels and supports children in learning to regulate their attention, task-orientation, and emotions.

The increasing awareness of the importance of optimal self-regulation development in early childhood and the malleability of self-regulation has led to the implementation of intervention programs to promote EF and self-regulation as a means to prevent academic and behavioral difficulties later in life. Interventions consisting of short intensive training of specific EFs such as inhibitory control and working memory (see for instance Holmes et al., 2009; Thorell et al., 2009) have shown that sizeable training effects on EF can be obtained with clear transfer to untrained domain-general cognitive skills. However, according to recent systematic reviews, transfer to regular education contexts, academic achievement, and social behavior is limited (Diamond & Lee, 2011; Melby-Lervåg & Hulme, 2013). Alternative approaches have attempted to improve early education and care settings in a comprehensive way by introducing a curriculum that is thought to foster executive functions and self-regulation development. Recent review studies have shown benefits of these curricula on children's cognitive and social-emotional outcomes, including self-regulation and its underlying EFs (McCabe & Altamura, 2011; Ursache et al., 2012).

One of these curricula, the Promoting Alternative Thinking Skills (PATHS), focused on enhancing children's social-emotional skills by teaching and modeling emotion regulation, self-control of arousal and behavior, and conflict resolution strategies while establishing a positive classroom climate with clear behavioral rules and expectations (Domitrovich, Cortes, & Greenberg, 2007). PATHS has been shown to improve children's social-emotional skills and executive functioning in task behavior in preschool and kindergarten (Bierman, Domitrovich, et al., 2008; Nix, Bierman, Domitrovich, & Gill, 2013). In another curriculum, Tools of the Mind (Tools), activities included in particular collaborative small group work and sociodramatic pretend play (Bodrova & Leong, 2007; Bodrova, Leong, & Akhutinia, 2011). Tools has been shown to have positive effects on EFs and (pre)academic skills (Barnett et al., 2008; Diamond, Barnett, Thomas, & Munro, 2007). However, although effects on EFs were found, none of these studies has provided direct observation-based information on children's self-regulation and executive functioning during curriculum activities, for example during pretend and role play. Therefore, how the particular curriculum activities that were provided actually triggered children's self-regulation and supported self-regulation development is not known.

Pretend Play as a Context for Promoting Self-Regulation

The current study focuses specifically on children's cognitive and emotional self-regulation behavior during pretend play, as pretend play is since long considered an important context for children to develop self-regulation (Berk, Mann, & Ogan, 2006; Vygotsky, 1967). In pretend play children create symbols by their imaginative use of objects, actions, persons, and verbalizations, which facilitates separation of thought and action from external stimuli and enables children to rely on internal ideas to guide their behavior (Stambak & Sinclair, 1993; Vygotsky, 1967). In play, according to Vygotsky (1967), children make the transition from other-regulation to self-regulation, as they monitor their play partners, become aware of the rules of the play, and follow directions implied by these rules or issued by the other players. In addition, pretend play is hypothesized to provide children with the opportunity to enact and modify emotional experiences in a safe environment, which is thought to contribute to the development of emotional self-regulation (Berk, et al., 2006).

Several studies have shown associations between pretend play and children's self-regulation (Elias & Berk, 2002; Lillard, et al., 2013; Nader-Grosbois & Vieillevoye, 2012; Vieillevoye & Nader-Grosbois, 2008). In a study with three- to six-year-old children, concurrent associations between complexity of children's pretend play and self-regulation were found,

using an observational measure to assess several aspects of self-regulation, such as planning, focused attention, and self-evaluation (Nader-Grosbois & Vieillevoye, 2012; Vieillevoye & Nader-Grosbois, 2008). Furthermore, in another study involving four- to six-year-old children who were observed during free play, complexity of pretend play was concurrently related to children's attention shifting skills (Matthews, 2008). Also, associations between pretend play and emotional aspects of self-regulation have been reported. For instance, in a study with three- and four-year-old children, Elias and Berk (2002) found that the complexity of children's pretend play was predictive of their self-regulation a few months later. Self-regulation, in this study, included children taking responsibility to clean up and help other children. Studies involving three- to five-year-old children have also shown relations between pretend play and emotion regulation as assessed either by observing children's response to an emotionally negative event during pretend play introduced by the researcher (Galyer & Evans, 2011) or by parent or teacher ratings (Lindsey & Colwell, 2003).

To conclude, several studies have investigated associations between children's self-regulation and pretend play. Some of these studies have included observational measures to assess children's self-regulation during play. However, to date and to the best of our knowledge no studies have adopted a broad view on self-regulation by integrating cognitive and emotional self-regulation in a single framework and examining the relationships of both cognitive and emotional self-regulation with pretend play. Moreover, in none of these studies observed self-regulation in play has been related to test-based measures of children's trait EFs that are thought to underlie observed behavioral self-regulation. Therefore, these studies do not provide conclusive evidence regarding the joint contribution of child-related and context-related factors to observed self-regulation in play behavior.

Current Study

The current study examined three-year-old children's cognitive and emotional self-regulation in center-based education and care provisions using observational measures to gain more detailed insight in the behavioral manifestations of children's self-regulation in a naturalistic play setting. We addressed the following research questions and related hypotheses:

1) To what extent do children show cognitive and emotional self-regulatory skills in their play at this young age, and how are these interrelated? Based on the extant literature, we expected children of this young to show already cognitive and emotional self-

regulation, but to what extent and with regard to which aspects of the broad construct of self-regulation, is an exploratory question. We furthermore expected cognitive and emotional self-regulation to be interrelated, but clearly distinguishable.

2) To what extent are children's cool and hot executive functions as assessed with neuropsychological tests predictive of their observed cognitive and emotional self-regulation? Based on theoretical considerations, we expected test-based cool EFs to be related to observed cognitive self-regulation and test-based hot EFs to be related to observed emotional self-regulation.
3) How is the quality of children's pretend play related to their observed cognitive and emotional self-regulation? Following previous studies into pretend play and self-regulation, we expected the quality of pretend play, defined as the complexity of symbolizing and role-play, to be positively related to observed cognitive and emotional self-regulation, also when differences between children in test-based measures of cognitive and emotional EFs and other potentially relevant characteristics are controlled.
4) How is global classroom quality related to children's cognitive and emotional self-regulation? In line with recent research results on the effects of comprehensive curricula addressing the classroom's emotional climate, the provided behavior guidance and support for learning on self-regulation development, we expected positive relations between classroom quality thus defined and observed cognitive and emotional self-regulation, while controlling for differences in children's test-based EFs and other child characteristics.

To the best of our knowledge, no observational measures were available at the time of the study to assess both children's cognitive and emotional self-regulation skills in a naturalistic setting. For instance, the extensive coding framework developed by Whitebread and colleagues (2009) mainly focuses on the cognitive aspects of self-regulation and to some extent on motivational aspects, but does not include measures of children's emotional self-control or ability to resolve peer conflicts. Another framework, the Individualized Classroom Assessment Scoring System (inCLASS; Downer, Booren, Lima, Luckner, & Pianta, 2010), does include aspects of peer interactions relevant for the current purpose, such as resolving conflicts, and also assesses children's task engagement, but does not include aspects of cognitive self-regulation. Therefore, we developed a new observation scheme, based on the existing literature, to assess both cognitive and emotional self-regulation in play.

Method

Sample

The present study reports data from an observational in-depth study within the Dutch national pre-COOL cohort study into the developmental effects of early childhood education and care provisions (Slot, Leseman, Verhagen, & Mulder, submitted manuscript). In pre-COOL, a large cohort of children attending preschool education and day care centers in the Netherlands is followed from age two to five years. At the first measurement wave, 1819 children across 289 centers participated in pre-COOL. For the present study, 87 centers were selected using a purposive sampling procedure to ensure a balanced mix of centers from rural and urban areas and type of provision. From the centers that were approached, 44 centers (51%), with 65 classrooms, agreed to participate. Next, classrooms were selected with at least two children who had participated in the first wave of child assessments of the pre-COOL study, and for whom parents had consented to participation in the in-depth study, resulting in 37 classrooms. All children participating in pre-COOL child-assessment who were present during the video data collection were included in the current sample (N = 95). Additional children were randomly selected to increase the number of target children per classroom minimally necessary for the planned multilevel analyses (N = 18), resulting in a total of 113 children, three to four per classroom, of which 59 (52.2%) were boys. For 95 (84%) children test-based EF data were available from the first child-assessment wave of pre-COOL. Children's mean age at the time of the in-depth study was 37 months (*SD*=3.5; range 28-45 months). The majority of children, 71 (62.8%), was monolingual with Dutch as home language. About half of the classrooms provided a half-day program for two-to-four-year old children (preschools) and the other half provided a full-day program for zero- to four-year-old children (day care centers). The classrooms varied in cultural diversity. In 17 classrooms (45.9%) the vast majority of children were native Dutch. In 12 classrooms (32.4%) the majority of children were non-native Dutch children with immigrant background. The remaining eight classrooms were culturally mixed. All teachers (N = 37) were female and the vast majority of the teachers (75.9%) was native Dutch. The majority of the teachers (62.1%) completed five years post-primary vocational training. The rest had a Bachelor's degree. Most teachers had worked in the ECEC field for more than 5 years (79.3%).

Procedures

Children's executive functions were assessed when children were on average 28 months old (*SD*=2.7; range 23-35 months), a few months before the video recordings were made. Children were tested individually by trained research assistants in a quiet room in the center in test sessions that lasted approximately 45 minutes using a test battery that included tests of EF and receptive vocabulary along with a number of other measures not reported in this study (Mulder, Hoofs, Verhagen, Van der Veen, & Leseman, 2014; Verhagen, De Bree, Mulder & Leseman, submitted manuscript).

For the in-depth study, classrooms were visited twice during a regular morning. Teachers and children were videotaped for 15 to 20 minutes in four different situations. Two of these situations were regular daily recurring situations, i.e., mealtime and free play. The other situations were guided play situations for which the researchers provided the standard sets of play materials to all classrooms to ensure comparability between classrooms. The current study focuses on play with kitchen toys, such as pots, pans, plates and cutlery, and different kinds of toy food. This situation was likely to elicit pretend play. The teacher was asked to select a number of children, with a minimum of four and including all the pre-COOL children and additional target children present that day, and to arrange a play session with these materials as she usually would do. No further instruction was provided. After 15 minutes of videotaping the teacher was told she could end the play and the research assistant stopped videotaping.

Child Observational Measures

For the purpose of the current study new observational measures were developed to assess children's cognitive and emotional self-regulation. In addition, the existing Smilansky Scale for Evaluation of Dramatic and Sociodramatic Play of Smilansky and Shefataya (1990) was adapted to fit the observational procedure of the current study. Each observational measure consisted of several behavioral indicators and the target children, three to four per classroom, were for each indicator rated on 5-point scales with scores ranging from *low* (1) to *high* (5). A high score reflects a child who showed the specified behavior (e.g., metacognitive regulation) frequently, a medium score reflects a child who showed the specified behavior occasionally or only when the teacher stimulated this, and a low score indicates a child who hardly showed the specified behavior if at all. The observations were conducted by trained research-assistants, who scored all three scales and two other scales not used in the current study, in one session. The research assistants were blind to the objectives of the current study.

Cognitive self-regulation

Cognitive self-regulation in thus study referred to the verbal and non-verbal ways in which children regulate their behavior towards a particular goal and continue to stay involved in their play. Three indicators of cognitive self-regulation were scored. *Metacognitive knowledge* indicates the knowledge the child expresses verbally about his or her own and other children's thinking, learning, and problem solving, and includes knowledge about strategies and the effectiveness of these strategies (Pintrich, 2002; Whitebread et al., 2009). *Metacognitive regulation* involves the degree to which the child uses planning, monitoring, control, and evaluation of behavior during play, which includes both verbal (a child stating "I am going to make soup") and non-verbal behavior (a child performing a systematic sequence of actions with a clear goal) (Whitebread et al., 2009). *Persistence* captures the child's degree of commitment and concentration during his or her play, indicating how long a child can sustain an activity and how much effort a child is willing to invest when encountering difficulties (Egeland, Erickson, Clemenhagen-Moon, Hiester, & Korfmacher, 1990).

Emotional self-regulation

Emotional self-regulation was defined as the verbal and non-verbal ways in which children regulate their emotions and social behavior in (peer) play. Four indicators of emotional self-regulation were distinguished. *Knowledge of emotions* refers to the knowledge the child verbally expresses of his or her own and other children's emotions (Whitebread et al., 2009). *Emotion regulation* involves deliberate attempts of the child to change the nature, intensity and time-course of emotions that are disruptive for the play in order to continue playing (Eisenberg & Spinrad, 2004; Eisenberg & Sulik, 2012). *Resolving conflicts* refers to the child's ability to resolve a peer conflict in a socially acceptable way (De Haan & Singer, 2003; De Haan & Singer, 2010). *Behavioral self-control* reflects the degree in which a child is able to control his or her behavior and to meet the social expectations of the play situation, such as sharing toys or waiting for a turn (Kopp, 1982).

Pretend play

The Smilansky Scale for Evaluation of Dramatic and Sociodramatic Play of Smilansky and Shefatya (1990) was adapted for the present study as a measure of the quality of pretend play in terms of the complexity of role-play and symbolization. The original scale is designed for observations based on time-sampling and was redefined to fit the whole-session rating procedure of the current study. The adapted scale included four indicators of children's pretend play to be scored on 5-point scales. *Role-play* captures the degree in which a child enacts a role by imitative action and/or verbalization and the degree of

persistence in role-play during the play episode. A high score reflects a child showing sustained elaborate role-play. *Make believe* reflects the level of object substitution (using a toy for something else than intended) and verbal substitution of actions and situations (verbal descriptions of an action in an imaginary situation without performing the action; e.g. "I am going to the supermarket"). A high score reflects a child using object substitution on multiple occasions while verbally describing actions or situations. *Interaction* assesses the degree in which a child directs his words or actions to others (peers and the teacher) in the play and the use of communication within the play episode (within-frame talk, thus communication that is part of the play). A high score reflects a child who has reciprocal interactions with others and uses within-frame talk. Finally, an additional indicator not present in the original scale of Smilansky and Shefataya, was *meta-communication* which reflects the degree of outside-frame talk necessary to direct and sustain a satisfactory play episode, such as assigning roles and discussing the course of the play as it evolves. Meta-communication, by definition, assumes interaction between two or more children and is considered a mature form of mutual play regulation (Whitebread & Sullivan, 2012).

Training and Inter-Observer Reliability

Seven research assistants were trained on the self-regulation and pretend play scales by the first author. Following a training of two half-days, two videos were coded independently by the assistants to determine reliability prior to data collection. Six assistants passed the pre-set reliability criterion of 80% agreement within one scale point difference for all indicators with the first author (chance level agreement is 52%), and were allowed to continue with the data collection. In addition, part of the data (at least 18%) was coded independently by both the first author and each assistant to determine inter-observer reliability by calculating the percentage agreement for all indicators and the intraclass correlation coefficient (ICC) for the total scales. Inter-observer agreement for each indicator within one scale point difference ranged from 83% to 100% for cognitive self-regulation, from 67% to 89% for emotional self-regulation and from 79% to 100% for pretend play. The *average measures ICCs* for the total scales, using a two-way mixed effects model with absolute agreement, averaged across assistants, were .81, .76, and .81 for cognitive self-regulation, emotional self-regulation and quality of pretend play respectively.

Children's Executive Functions

Two separate cool and hot EF measures were constructed using the factor scores of a confirmatory factor analysis involving the full pre-COOL sample. The tasks used

as indicators for each of the two measures are described briefly below. Detailed task descriptions and the results of a psychometric analysis of the EF test battery are reported elsewhere (Mulder et al., 2014). Psychometric quality was found to be satisfactory.

Cool EF

For the cool EF construct, children's scores on a selective attention task, visuo-spatial short-term memory task, and visuo-spatial working memory task were used as indicators. *Selective attention* was assessed with a visual search task in which children had to identify targets amidst a display of distractors as fast as possible. This task was administered on a laptop computer and designed for the purposes of the pre-COOL study, based on previous work by Gerhardstein and Rovee-Collier (2002), and Scerif, Cornish, Wilding, Driver, and Karmiloff-Smith (2004). The average number of targets identified across three trials was scored. *Visuospatial short-term memory* was assessed with a memory for location task in which children had to remember the location of hidden toys (Oudgenoeg-Paz, Boom, Volman, & Leseman, submitted manuscript; Pelphrey et al., 2004; Vicari et al., 2004). Six identical white boxes were used as hiding locations. The task was given in an adaptive fashion, and the number of toys hidden ranged from one to four. The number of locations children could remember simultaneously was counted to obtain a measure of their short-term memory span. *Visuospatial working memory* was measured with the Six-Boxes Task (Diamond, Prevor, Callender, & Druin, 1997). In this task, children were shown how six toys were hidden in six identical white boxes. Children were asked to search for the toys by opening one box at a time, with a six second delay between each two consecutive search attempts. The total number of toys obtained in six search attempts was scored.

Hot EF

For the hot EF construct, children's scores on a *snack delay* and a *gift delay* task were used as indicators (Kochanska, Murray, & Coy, 1997; Kochanska, Murray, Jaqcues, Koenig, & Vandergeest, 1996; Kochanska, Murray, & Harlan, 2000). Children were shown an attractive object, i.e., a snack and a gift respectively, and asked to try not touch the object until the research assistant finished another task. After the instruction, the research assistant turned and moved away to a distant corner of the room, supposedly to make notes on an unrelated topic. The delay time was one minute. The assistant scored whether children touched the object and recorded their specific actions (e.g., eating the raisins, tearing the wrapping paper).

Time Between EF Assessment and Observations

Due to differences in enrolment age, the age at which the first assessment of EFs was

conducted differed between children (ranging from 23 to 35 months). To control for differences in the time elapsed between the EF assessment and the observation of self-regulation, the variable *time between test and observatio*n was constructed representing the difference between children's age at the time of the video observation and their age at the time of EF assessment.

Other Child and Background Characteristics

Several child background characteristics were taken into account in the analyses for control purposes. Two dichotomous variables were constructed to represent *gender* (1=female; 0=male) and *home language* (1=only Dutch; 0=other language(s) as well) based on parent reports. Children's *age* during the observation was calculated based on their date of birth. Finally, as prior research has shown that children's vocabulary skills are related to EFs (Fuhs & Day, 2011; Vallotton & Ayoub, 2011), we included receptive Dutch *vocabulary* measured with the Dutch version of the Peabody Picture Vocabulary Test (PPVT-III-NL, Dunn, Dunn, & Schlichting, 2005) as a covariate. A shortened version consisting of in total 24 items was used, with good internal consistency (Cronbach's α= .88). Scores were calculated as the percentage of correct responses for each child (for more details on the vocabulary measure, see Verhagen et al., submitted manuscript).

Classroom Quality

Observed process quality

The Classroom Assessment Scoring System Toddler (CLASS Toddler; La Paro, Hamre, & Pianta, 2011) was used to assess global classroom process quality. An approved Dutch translation of the CLASS manual was developed for the present study (Slot, Leseman, Mulder, & Verhagen, 2013). All observers were trained by a licensed CLASS trainer and achieved at least 80% agreement within one scale point on 7-point scales on an online test, as recommended by the developers of the CLASS (average agreement was 86.4%; agreement by chance was 33%). The videos of the play sessions were rated using the CLASS by assistants who were not involved in the evaluation of children's self-regulation behavior and quality of pretend play during the play session.

Classroom quality was rated on eight dimensions, using 7-point scales ranging from 1 or 2 (classroom is low on that aspect); to 3, 4 or 5 (classroom is in the midrange); and to 6 or 7 (classroom is high on that aspect). Based on theoretical considerations (La Paro et al., 2011) and previous confirmative factor analysis of the observation data of the larger

pre-COOL sample (Slot, Boom, Verhagen, & Leseman, submitted), three domains were distinguished. *Emotional Support* was computed as the average of the scores on three dimensions: Positive Climate, reflecting the warmth, respect, and enjoyment displayed during interactions between the teacher and children; Teacher Sensitivity, evaluating the extent to which the teacher is aware of and responsive to children's needs; and Regard for Child Perspectives, capturing the degree in which the teacher's interactions with children and the classroom activities provided are attuned to children's interests, and the degree in which children's independence is encouraged. The second domain, *Behavioral Support*, was the average of the scores on two dimensions: Negative Climate, reflecting the overall negativity expressed in the classroom by the teacher and the children (scores are reversed), and Behavior Guidance, referring to the teacher's ability to promote positive behavior and redirect problem behavior. However, because of the very limited variance in Negative Climate in this particular sample, we decided to exclude this dimension and to use the Behavior Guidance dimension separately, because of its hypothesized association with self-regulation (Merrit et al., 2012). The third domain, *Engaged Support for Learning,* was based on three dimensions: Facilitation of Learning and Development, considering how well the teacher facilitates activities to support children's learning and development; Quality of Feedback, assessing the degree in which the teacher's feedback promotes learning and expands children's participation, and Language Modeling, referring to the extent to which the teacher fosters, models and encourages children's use of language.

Structural classroom characteristics

The following structural quality variables were used as covariates. *Group size* during the play activity was based on registration by CLASS observers who counted the number of children present during the activity. *Cultural classroom composition* was based on teachers' reports on the number of children with a non-Dutch background in their classroom on a scale ranging from 1 (0-10%) to 10 (90-100%).

Analysis Strategy

To answer the first research question, descriptive statistics were calculated to examine the degree of and variation in children's self-regulation and pretend play for each of the indicators. For descriptive purposes, also the correlations between the indicators of cognitive and emotional self-regulation and the quality of pretend play were examined. Next, confirmatory factor analysis was performed in Mplus (version 7; Muthén & Muthén, 1998-2012) to evaluate the factor structure of cognitive and emotional self-regulation, and

to test whether these measures represent distinguishable constructs, using a maximum likelihood robust (MLR) estimator to deal with non-normality in some of the indicators. The factor scores extracted from these analyses instead of the complete measurement models were used in the subsequent analyses in order to reduce the number of parameters in the final model, which was deemed necessary given the small sample size.

Not all children received a score on the indicators emotion regulation and resolving conflicts, as in these cases situations requiring emotion regulation or conflict resolution did not occur during the play episode. The missing values can be considered planned, because they are inherent to the way in which both indicators were defined. Although these missing values, therefore, cannot be considered to be completely at random, missing data were dealt with by using full information maximum likelihood (FIML) estimation in Mplus (Enders, 2010), in which the standard errors for the parameter estimates are computed using the complete observed information matrix (Muthén & Muthén, 1998-2012). Children with and without missings on these indicators did not systematically differ in scores on the other indicators. Several steps were taken to assure robustness of the findings. Both cognitive and emotional self-regulation were estimated in the same model, allowing Mplus to estimate values for the missing scores in both indicators based on all other indicators of the two constructs. Moreover, to check robustness of the findings, correlations were calculated between the extracted factor scores with the missings imputed and the mean scale scores based on the observed values of the indicators only. Correlations were good (r= .97 for cognitive self-regulation and r= .89 for emotional self-regulation; all p's < .001).

Pretend play was the context in which children's self-regulation was observed and the quality of pretend play was regarded as a potential determinant of children's self-regulation in the current study. Therefore, a separate model was estimated to test the factor structure of the pretend play construct. Again, the estimated factor scores were extracted and used in the subsequent analyses.

To address the second and third research question, structural equations modeling was applied to investigate the multivariate relations of children's observed cognitive and emotional self-regulation with hot and cool executive functions and the complexity of pretend play, respectively. The models that were examined are similar to ordinary multiple regression analysis, in which estimates of the unique effects of the independent variables on the dependent variable are obtained after specifying all covariances between the independent and control variables. We controlled for children's vocabulary and background characteristics (age, gender, home language) and for the nesting of children within classrooms.

Finally, to answer the fourth research question concerning associations between children's observed self-regulation and global classroom quality measured at the classroom level as predictor of observed self-regulation, multilevel modeling was applied as is recommended when variables measured at different levels of aggregation are analyzed simultaneously (Hox, 2010). Given the small sample size at the classroom level, we examined the three classroom quality measures in separate models, with group size and cultural classroom composition as covariates at the classroom level.

Model building proceeded in a number of steps. First, the non-restricted baseline models were estimated using the Maximum Likelihood Robust (MLR) estimator. Second, the models were trimmed by eliminating non-significant paths with $p > .10$ or with $| \beta | < .05$ (Wuensch, 2012) in a step-by-step fashion to obtain the most parsimonious model. Model fit was evaluated with several fit indicators: the Chi-Square test of goodness of fit, the Comparative Fit Index (CFI) and the Standardized Root Mean Square Residual (SRMR) at both the child and classroom level, with a non-significant Chi-square, CFI > .95, SRMR < .05 indicating good fit. The Root Mean Square Error of Approximation (RMSEA) was not used to evaluate the model fit, because recent evidence suggests this index too often falsely indicates poor model fit in small samples or in models with a small number of degrees of freedom (Kenny, Kaniskan, & McCoach, 2014). Standardized regression coefficients β were used as measures of effect size with β around .10 indicating a small effect, β around .30 a medium-sized effect and $\beta > .50$ indicating a large effect (Kline, 2005).

Mplus, by default, standardizes to the variance on the *within* level for *within*-level relationships, and to the variance of the *between* level for *between*-level relationships (Hox, 2010). To interpret and compare between *within* and *between* level effects we chose to standardize the regression coefficients to the total variance according to the following formula: $b * (SD_{predictor}/SD_{outcome})$ in which the *SD* is the square root of the total variance of the dependent variable concerned (Hox, 2010).

Results

Descriptive statistics of the classroom quality and structural characteristics are presented in Table 1.

Descriptive statistics for all indicators of cognitive and emotional self-regulation and pretend play are shown in Tables 2 and 3. Regarding cognitive self-regulation, the three-year-old children hardly showed any explicit metacognitive knowledge during their play.

Table 1. Classroom Characteristics and Classroom Quality

Classroom quality aspects	M	SD	Range	N
Group size during activity	5.66	1.42	3-10	113
Cultural classroom composition	4.80	3.68	1-5	113
Positive Climate	5.20	.93	3-7	113
Negative Climate	6.70	.46	6-7	113
Teacher Sensitivity	4.92	.76	3-6	113
Regard for Child Perspectives	4.22	1.05	2-6	113
Behavior Guidance	4.61	.94	2-6	113
Facilitation of Learning and Development	4.30	1.00	2-6	113
Quality of Feedback	2.74	.94	1-5	113
Language Modeling	3.52	.88	2-5	113

The scores on this indicator were very low on average with limited variation. However, about 6% of the children received a score higher than the lowest score, indicating that at least some children expressed explicit metacognitive knowledge. Children showed some metacognitive regulation, such as planning and monitoring of their play, however the distribution of the scores was skewed towards the lower end of the scale. Children's persistence during play yielded a higher mean score, with scores mainly varying between the low/mid and high range. Concerning emotional self-regulation, children did not show explicit knowledge of emotions during the play episode, with only one exception. Furthermore, about 65% of the children obtained a score on emotion regulation. The remaining children did not show emotions that threatened the continuity of the play session and needed to be regulated. For the children who did show disruptive emotions, the scores were in the mid to high range, indicating they were, on average, quite able to regulate their emotions and to continue their play, but sometimes needed help from the teacher. Peer conflicts were quite common, given that almost 78% of the children received a score on this indicator. Children scored in the mid to high range on average, indicating that they were able to resolve conflicts themselves most of the time, but occasionally needed help from the teacher. Finally, children's behavioral self-control was in the mid to high range on average, indicating that children were mostly able to adapt to the situational demands of the play setting.

Regarding pretend play, children showed medium levels of role-play, make believe actions, and they had some interactions with either peers or the teacher during the play. The use of meta-communication occurred much less frequently, but did have some variation, with 12% of the children occasionally engaging in meta-communication during play. Note that there is an expected dependency between the indicators interaction and

meta-communication, as children could only engage in meta-communication if they were interacting with peers.

Table 2. Descriptive Statistics for the Observed Self-Regulation and Children's Pretend Play

Self-regulation indicator	M	SD	Range	N
Metacognitive knowledge	1.11	.43	1-3	113
Metacognitive regulation	2.58	.97	1-5	113
Persistence	3.24	1.16	1-5	113
Knowledge of emotions	1.02	.19	1-3	113
Emotion regulation	3.66	1.08	1-5	73
Resolving conflicts	3.53	1.16	1-5	88
Behavioral self-control	3.54	1.04	1-5	113
Role-play	2.81	1.05	1-5	113
Make believe	2.69	.85	1-5	113
Interaction	2.69	1.01	1-5	113
Meta-communication	1.19	.53	1-3	113

Table 3. Frequency Distribution of Children's Self-Regulation During Play

Self-regulation indicator	Low	Low/mid	Mid	Mid/high	High
Metacognitive knowledge	106	2	5	0	0
Frequency (Percent)	(93.8%)	(1.8)	(4.4)		
Metacognitive regulation	13	45	35	17	3
Frequency (Percent)	(11.5)	(39.8)	(31.0)	(15.0)	(2.7)
Persistence	7	21	46	16	23
Frequency (Percent)	(6.2)	(18.6)	(40.7)	(14.2)	(20.4)
Knowledge of emotions	112	1	0	0	0
Frequency (Percent)	(99.1)	(0.9)			
Emotion regulation	2	6	29	14	22
Frequency (Percent)	(2.7)	(8.2)	(39.7)	(19.2)	(30.1)
Resolving conflicts	5	9	32	18	24
Frequency (Percent)	(5.7)	(10.2)	(36.4)	(20.5)	(27.3)
Behavioral self-control	2	16	39	31	25
Frequency (Percent)	(1.8)	(14.2)	(34.5)	(27.4)	(22.1)
Role-play	13	27	49	16	8
Frequency (Percent)	(11.5)	(23.9)	(43.4)	(14.2)	(7.1)
Make believe	5	43	52	8	5
Frequency (Percent)	(4.4)	(38.1)	(46.0)	(7.1)	(4.4)
Interaction	12	38	42	15	6
Frequency (Percent)	(10.6)	(33.6)	(37.2)	(13.3)	(5.3)
Meta-communication	99	7	7	0	0
Frequency (Percent)	(87.6)	(6.2)	(6.2)		

Table 4. Bivariate Correlations Between Indicators of Self-Regulation and Pretend Play

	2	3	4	5	6	7	8	9	10	11
1. metacognitive knowledge	.11	.16	.42**	-.08	.06	.01	.10	.14	.20*	.23*
2. metacognitive regulation		.58**	-.08	.15	.16	.01	.66**	.60**	.48**	.26**
3. persistence			-.02	.18	.22*	.14	.67**	.56*	.47**	.25**
4. knowledge of emotions				-.07	-.05	.04	.11	.04	.12	-.03
5. emotion regulation					.49**	.45*	.29**	.13	-.03	.13
6. resolving conflicts						.34**	.24**	.13	.13	.09
7. behavioral self-control							.04	.06	-.15	-.02
8. role-play								.68**	.57**	.34**
9. make believe									.56*	.27**
10. interaction										.43**
11. meta-communication										

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

The inter-correlations of children's cognitive and emotional self-regulation and pretend play were examined at the level of the indicators, which revealed different patterns for cognitive and emotional self-regulation (see Table 4). Overall, the indicators of cognitive self-regulation were moderately to strongly related to the pretend play indicators, whereas the indicators of emotional self-regulation were moderately related to the role-play indicator only. Metacognitive knowledge was moderately strongly correlated with knowledge of emotions and meta-communication in play, which is not surprising as all these indicators point to more mature levels of regulation.

Next, the factor structures of the observation measures were examined. A confirmatory factor analysis was conducted in Mplus to test whether cognitive and emotional self-regulation were indeed two related but distinct constructs. The indicator knowledge of emotions was excluded from the factor analysis, because of its low occurrence and severely limited variation. Factor analysis revealed good model fit ($\chi^2(8) = 11.29$, $p = .19$; CFI = .97; SRMR = .04) and factor loadings were satisfactory, except for metacognitive knowledge, which had a non-significant factor loading of .18. Therefore, also this indicator was excluded from the final model. Fit of the final model was acceptable ($\chi^2(4) = 9.43$, $p = .05$; CFI = .95; SRMR = .04) and confirmed two moderately inter-related but distinct constructs of self-regulation. The extracted factor scores were used in further analyses.

The same procedure was followed for pretend play (see Figure 2). At first, model fit was not satisfactory ($\chi^2(2) = 9.35$, $p = .01$; CFI = 0.95; SRMR = .04). Based on the modification indices provided by MPlus we allowed the error variances of the indicators interaction and meta-communication to correlate, resulting in significant improvement of model fit

Figure 1. Factor Structure of Cognitive and Emotional Self-Regulation

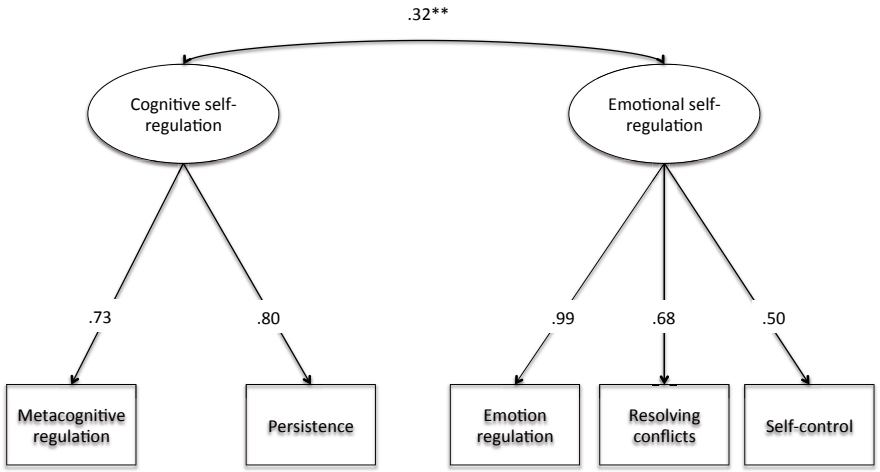
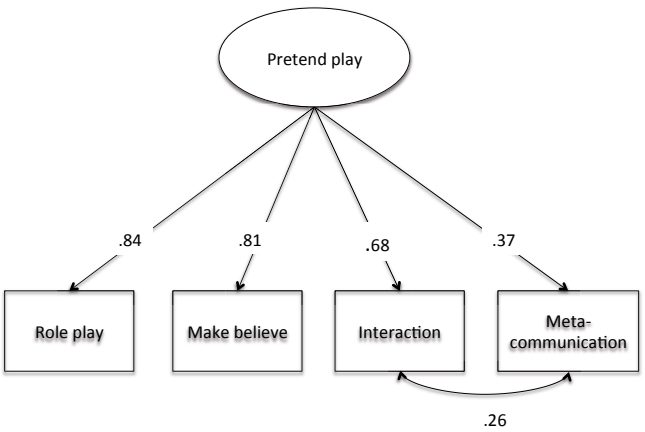


Figure 2. Factor Structure of Pretend Play



(($\Delta\chi^2$ (1) = 6.14, p < .001). Fit of the final model was satisfactory (χ^2(1) = .72, p = .40; CFI = 1.00; SRMR = .01). All factor loadings were satisfactory. Although the factor loading of meta-communication was relatively low, it was above the suggested cut-off value of .32 (Tabachnick & Fidell, 2007). The extracted factor scores were used in the subsequent analyses.

As children were nested within classrooms, an unconditional model was specified, partitioning the total variance in child level and classroom level variance. Most of the

variance was located at the child level (σ^2_{v0} = .368 for cognitive self-regulation, σ^2_{v0} = .852 for emotional self-regulation, and σ^2_{v0} = .600 for pretend play). A smaller portion of the variance was located at the classroom level (σ^2_{u0} = .004 for cognitive self-regulation, σ^2_{u0} = .037 for emotional self-regulation and σ^2_{u0} = .045 for pretend play). The ICCs were ρ = .007 for cognitive self-regulation, ρ = .049 for emotional self-regulation and ρ = .095 for pretend play, respectively.

Next, the multivariate relationships of children's observed self-regulation with test-based measures of cool and hot executive functions, and the observed quality of pretend play were investigated, controlling for children's vocabulary and background characteristics. Bivariate correlations between child characteristics, children's self-regulation, EFs and quality of pretend play are shown in Table 4. The structural equations model was estimated with observed cognitive and emotional self-regulation as outcomes measures, while specifying all covariances between the independent and control variables. In the first model, the associations between cognitive and emotional self-regulation and cool and hot EF, respectively, were examined. As the model was saturated, no fit indices could be computed. Model trimming was performed to obtain a more parsimonious model by constraining non-significant paths and paths with $|\beta|$ <.05 in a stepwise fashion to zero. The trimmed model fitted the data well ($\chi^2(18) = 8.87$, p = .96; CFI = 1.00; SRMR = .06). The results are presented in Table 6. Cool EF was positively related to emotional self-regulation, but not to cognitive self-regulation. Hot EF was not related to observed cognitive self-regulation and showed a trending negative relation with observed emotional self-regulation.

In the second model, the quality of pretend play was entered, resulting in a saturated model. To obtain a more parsimonious model, non-significant paths were again constrained to zero. This model fitted the data well ($\chi^2(24) = 14.56$, p = .93; CFI = 1.00; SRMR = .06). In this model the quality of pretend play was strongly related to children's cognitive self-regulation. Pretend play was also significantly associated with children's emotional self-regulation, but the effect size was small. Note that cool EF was still positively related to emotional self-regulation, whereas the previous negative trend-level relation of hot EF with emotional self-regulation disappeared upon including the quality of pretend play in the model.

Finally, the associations between children's observed self-regulation and global classroom quality were investigated, while controlling for EFs, vocabulary, and background characteristics at the child level and for group size and cultural diversity at the classroom level. Three classroom quality aspects, Emotional Support, Behavior Guidance, and Engaged

Table 5. Bivariate Correlations Between Child Characteristics, Executive Function, Self-Regulation and Pretend Play

	2	3	4	5	6	7
1. Age at test	.21*	.23*	-	-	-	-
2. Cool EF		.80***	-	05	.28**	.08
3. Hot EF			-	-.06.	.17	-.06
4. Age at observation				.14	.10	.20*
5. Cognitive self-regulation					.38**	.75**
6. Emotional self-regulation						.28**
7. Pretend play						

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

Table 6. Associations Between Cognitive and Emotional Self-Regulation with Cool and Hot EF and Complexity of Pretend Play

	Cognitive SR			Emotional SR		
	B	SE B	β	B	SE B	β
Model 1						
Age	#			.05	.03	.17
Gender	#			#		
Home language	.14	.14	.11	#		
Time between test and observation	.06	.02	.27**	-.03	.03	-.08
Vocabulary	.42	.23	.24†	-.40	.25	-.14†
Cool EF	.19	.15	.21	.68	.22	.47**
Hot EF	-.30	.22	-.22	-.48	.29	-.23†
Model 2						
Age	-.02	.01	-.09	.04	.03	.13
Gender	#			#		
Home language	.11	.07	.09	#		
Time between test and observation	.03	.02	.15	-.04	.03	-.12
Vocabulary	.29	.13	.16*	-.48	.23	-.17*
Cool EF	#			.58	.20	.40**
Hot EF	#			-.32	.28	-.15
Pretend play	.55	.05	.74***	.32	.10	.27*

***p < .001, **p < .01, *p < .05, †p <.10, # paths constrained to zero

Support for Learning respectively, were examined separately. In all models, covariances between all independent child and classroom variables were specified. The results are shown in Table 7. In the first model, associations between cognitive and emotional self-regulation and Emotional Support were investigated, resulting in a saturated model. To obtain a more parsimonious model, non-significant paths were constrained in a stepwise fashion, resulting in good model fit ($\chi^2(19) = 10.66$, $p = .93$; CFI = 1.00; $SRMR_{within} = .05$,

$SRMR_{between}$ = .06). No significant associations were found between children's self-regulation and observed Emotional Support, but group size during play was negatively related to children's emotional self-regulation. The second model examined the relations between cognitive and emotional self-regulation and Behavior Guidance, resulting in a saturated model. Model trimming was applied to obtain a more parsimonious model, resulting in a good model fit ($\chi^2(17) = 9.17$, $p = .93$; CFI = 1.00; $SRMR_{within}$ = .05, $SRMR_{between}$ = .07). The results showed that teachers' observed Behavior Guidance was not related to either cognitive or emotional self-regulation. In the third model, the relations between cognitive and emotional self-regulation with Engaged Support for Learning were investigated, revealing a saturated model. Model trimming resulted in good model fit ($\chi^2(17) = 9.05$, $p = .94$; CFI = 1.00; $SRMR_{within}$ = .05, $SRMR_{between}$ = .08). Teachers' Engaged Support for Learning was also not related to observed self-regulation.

Discussion

Children's self-regulation skills develop rapidly in the first years of life (Blair & Diamond, 2008; Bronson, 2000) and are important for school achievement and positive social-behavioral outcomes (Calkins & Williford, 2009; McClelland, Cameron, Connor, et al., 2007; McClelland, Acock, & Morrison, 2006; McClelland, Morrison, & Holmes, 2000; Morrison, Ponitz, & McClelland, 2010; Raver et al., 2012; Rimm-Kaufmann, et al., 2009). In view of supporting children's self-regulation development in early childhood education and care provisions, it is important to gain more insight into how children use these skills in actual behavior in the classroom, and how the classroom context and particular activities such as pretend play can contribute to self-regulation development.

The results of the present study indicate that children as young as three years already showed elements of effective cognitive and emotional self-regulation during pretend play. In line with previous research with the same age group, the children in this study displayed metacognitive regulation of their play behavior, as evidenced by verbal and non-verbal indications of planning, monitoring, and control (Nader-Grosbois & Vieillevoye, 2012; Vieillevoye & Nader-Grosbois, 2008; Whitebread et al., 2007). The children also showed medium to high levels of persistence during play. However, only a few children showed elaborate and explicit metacognitive regulation, and only a few children expressed explicit metacognitive knowledge, indicating that these skills are still developing at this age, corroborating findings from previous studies (Whitebread et al., 2007; Whitebread et al., 2009).

Table 7. Associations Between Cognitive and Emotional Self-Regulation with Classroom Quality[1]

	Cognitive SR			Emotional SR		
	B	SE B	β	B	SE B	β
Model 3 Emotional Support						
Age	-.03	.02	-.11†	#		
Gender	#			#		
Home language	.18	.15	.10	-.07	.20	.02
Time between test and observation	.08	.03	.25**	#		
Cool EF	.18	.15	.14	.64	.20	.33**
Hot EF	-.27	.22	-.14	-.47	.29	-.16
Vocabulary	.42	.22	.17†	-.19	.25	-.05
Classroom level						
Emotional Support	.11	.07	.08	#		
Cultural classroom diversity	#			.02	.02	.05
Group size	-.04	.04	-.07	-.11	.06	-.12*
Model 4 Behavior Guidance						
Age	-.03	.02	-.11†	#		
Gender	#			#		
Home language	.15	.17	.08	-.10	.22	-.04
Time between test and observation	.09	.03	.28**	.02	.03	.04
Cool EF	.21	.15	.17	.66	.20	.34**
Hot EF	-.31	.23	-.16	-.47	.29	-.16
Vocabulary	.44	.24	.17†	-.23	.25	-.06
Classroom level						
Behavior Guidance	.05	.05	.06	-.06	.09	-.04
Cultural classroom diversity	.01	.01	.04	.03	.02	.08
Group size	-.05	.03	-.08†	-.11	.06	-.12†
Model 5 Engaged Support for Learning						
Age	-.03	.02	.11	#		
Gender	#			#		
Home language	.13	.16	.07	-.16	.21	.06
Time between test and observation	.08	.02	.25**			
Cool EF	.21	.15	.17	.71	.21	.36**
Hot EF	-.31	.23	-.16	-.55	.36	-.19†
Vocabulary	.42	.24	.17 †	-.37	.24	-.09
Classroom level						
Engaged Support for Learning	.02	.05	.02	.02	.11	.01
Cultural classroom diversity	.01	.02	.04	.03	.03	.08
Group size	-.05	.03	-.08	-.10	.06	-.11†

***$p < .001$, *$p < .05$, ,†$p < .10$, # paths constrained to zero

[1]Note that the parameter estimates were standardized to the total variance at both the child and classroom level.

Concerning emotional self-regulation, the results indicated that, if needed, children on average are quite well able to regulate their emotions by modulating and managing the intensity and expression of their emotions that were potentially disruptive to their play. Furthermore, children were quite able to solve (mild) conflicts with peers, which as such were rather common but seldom disruptive (De Haan & Singer, 2010). Occasionally, help of the teacher was needed to resolve conflicts. Finally, children on average successfully adapted their behavior to social-situational demands, as evidenced by their ability to wait for a turn and to share toys. However, children did not show evidence of explicit meta-emotional knowledge at this age. Although previous research has shown that children at age three years are able to recognize and label emotions, this is only the case when this type of knowledge is specifically asked for (Denham et al., 2012). The current play situation apparently did not provide strong enough triggers in this regard. Furthermore, about one third of the children did not show emotion regulation because the need to regulate emotions was absent. These children did not experience intense and possibly disruptive emotions during the play session. Note, that in most research on young children's emotion regulation special paradigms are used in which strong emotions are deliberately elicited (Calkins et al, 1999; Galyer & Evans, 2001). Finally, our findings confirmed that cognitive and emotional self-regulation indeed can be considered two interrelated but distinct constructs, in line with findings in previous research (Brock et al., 2009; Willoughby et al., 2012).

The second aim of the study was to examine the relations between observed self-regulation in a naturalistic setting and children's cool and hot EF skills as assessed with a neuropsychological test battery. Based on theoretical considerations, we expected test-based cool EF to be related to observed cognitive self-regulation and test-based hot EF to observed emotional self-regulation. These expectations were not confirmed and the results surprisingly indicated a reverse pattern of relations. We found a relatively strong positive association between cool EF and emotional self-regulation, while controlling for children's gender, age, home language and vocabulary, but no association between cool EF and cognitive self-regulation. Regarding the relation between cool EF and emotional self-regulation, the results fit in with theoretical models of (emotional) self-regulation that emphasize the role of cool EF, in particular attention (part of the factor structure of the cool EF measure used in this study; Mulder et al., 2014), in effortful control of affect-driven behavioral impulses (Kochanska et al., 2000; Posner & Rothbart, 2000; Rothbart et al., 2011).

The lack of a relationship between cool EF and observed cognitive self-regulation is more difficult to explain. Exploratory inspection of the correlations of cool EF with the separate indicators of cognitive self-regulation revealed no significant correlations. This suggests that the concepts of cool EF, involving attention, visuospatial short-term memory, and visuo-spatial working memory in this study, and cognitive self-regulation as defined here, with an emphasis on meta-cognitive functioning in a play setting are in fact unrelated constructs, despite the fact that at an abstract level both refer to optimal behavioral adaptation to situational demands. This underscores the need for precise definitions of constructs that seemingly address the same behavioral phenomena. The current findings fit in with conclusions on the effects of specific EF training interventions that do show transfer to EF-related cognitive abilities, including intelligence, but not to behavior in naturalistic education contexts (Melby-Lervåg & Hulme, 2013).

Hot EF was unexpectedly not associated with children's observed emotional self-regulation during play. The delay of gratification tasks that were used to measure hot EF, required children to resist temptation and exert behavioral self-control and, therefore, are conceptually clearly related to the indicators behavioral self-control and conflict resolution of the emotional self-regulation scale, which both involve control of affect-driven behavioral impulses. Indeed, the separate correlations between hot EF and these indicators were $r = .22$ with self-control and $r = .27$ with resolving conflicts (all p's $< .05$). Yet, the broader construct of emotional self-regulation used in this study also included an indicator addressing the expression, instead of inhibition, of affect in a regulated, socially acceptable way, which indeed was not correlated with hot EF. Taken together, the present results suggest that the constructs of hot EF based on affect-inhibition and emotional self-regulation involving a play setting and interaction with peers only partly overlap, and that emotional self-regulation in a naturalistic context depends on other skills as well, including cool EFs as was found in this study.

The third aim of this study was to examine the associations between observed cognitive and emotional self-regulation and the quality of children's pretend play. Overall, associations between cognitive self -regulation and pretend play were strong, even when controlling for children's EFs, vocabulary and other child (background) characteristics. However, it should be noted that these associations can at least in part be attributed to method-bound shared variance as the same observers scored cognitive and emotional self-regulation and the quality of pretend play in single observation sessions. However, note that the operational definitions of the indicators of the observation scales were semantically distinct. Especially

the indicators of the pretend play scale, which showed the strongest correlations with cognitive self-regulation, were referring to entirely different aspects of children's behavior (e.g., uptake of a role, use of symbolic substitution; see Table 4), whereas the indicators with a content which was more similar to the indicators of cognitive self-regulation showed weaker associations (e.g., interaction, meta-communication), a pattern which would not be expected in the case of method-bound correlation. Moreover, whereas the associations between complexity of play and cognitive self-regulation were quite strong overall, the associations with emotional self-regulation were weak to moderately strong at best, both at the scale and at the indicator level. Altogether, this suggests that the associations between complexity of pretend play and cognitive and emotional self-regulation cannot be attributed to method-bound shared variance only. Therefore, the current findings can be regarded as supporting the hypothesis that the complexity of pretend play is related to the level of cognitive and, to a lesser extent, emotional self-regulation that children display. Moreover, because child-related factors, including children's cool and hot EFs and vocabulary, were controlled in the analysis, the findings suggest that pretend play can indeed be a context that can contribute to self-regulation development beyond the self-regulation skills children already possess as play requires children to coordinate their goals, negotiate, monitor and update their plan as the play progresses, and to adapt their behavior accordingly to sustain a satisfactory pretend play episode, as has been found in previous research (Barnett et al., 2008; Diamond et al., 2007).

The relation between the quality of pretend play and emotional self-regulation was overall weaker. At the level of the indicators (see Table 4), the only significant relations were observed between role-play and emotion regulation and resolving conflicts. This result is in line with previous research (Fantuzzo et al., 2004; Galyer & Evans, 2011) and provides additional support to the hypothesis that role-play in particular can help children to learn to express and manage (imitated and imagined) emotions in socially desirable ways.

The fourth aim concerned investigating the relationships between global classroom quality based on the CLASS and children's cognitive and emotional self-regulation. The findings revealed no significant associations between global classroom quality and observed cognitive and emotional self-regulation, when controlling for the number of children during the activity and the group composition at the classroom level, and children's EFs, vocabulary and background characteristics at the child level. This finding is not in line with previous research that suggests that especially an emotionally positive classroom climate, free of stress and negativity, and a high level of teacher sensitivity,

provides a context for self-regulation development (Domitrovich et al., 2007; Ursache at al., 2012). Interestingly, group size during the play activity was found to be negatively related to children's emotional self-regulation. A possible interpretation is that larger groups are more chaotic and stressful, and may thus provide a less optimal context to support children to learn to control their emotional expressions, to regulate their emotions when they have to share toys and to reconcile conflicts, suggesting that group size is important to consider when organizing a pretend play activity.

Overall, general classroom quality was not associated with observed cognitive and emotional self-regulation. There are several possible explanations that can account for this result. First, in general, associations between classroom quality and child outcomes tend to be small (Burchinal, Kainz, & Cai, 2011; Zaslow et al., 2006), which has raised the question whether currently widely used classroom quality measures are specific enough to show effects on children's outcomes (Bryant, Burchinal, & Zaslow, 2010; Burchinal, Kainz, & Cai, 2011; Slot et al., under review; Zaslow et al., 2006). Second, classroom quality measures reflect general quality based on the experiences of the average child, whereas individual children within the same classroom can have quite different experiences (Bulotsky-Shearer, Fantuzzo, & McDermott, 2008; Williford, Vick Whittaker, Vitiello, & Downer, 2013). Children's individual experiences have been shown to be more strongly related to children's outcomes than global emotional classroom climate (Birch & Ladd, 1997). Note also that the classroom level variance in the self-regulation measures in this study was rather small, leaving little variance to be explained by global classroom quality.

The current study contributes to the evidence that contextual factors may support children's self-regulation in early childhood education and care provisions. Especially the quality of pretend play seems important in this regard. Although the present study does not allow for causal conclusions, the results of the multivariate multilevel analyses applied in this study, which included several control variables at the child and classroom level, lend support to the hypothesis following from previous experimental research that contextual factors such as the quality of play may contribute to the self-regulation that children actually show and, thereby, provide children with opportunities to develop and practice self-regulation skills. The findings of this study, together with the evidence from other studies, including randomized experimental studies, can have important implications for early education and care practice. Several scholars have emphasized the importance of play for children's broad development (Berk et al., 2006; Bodrova, 2008; Bodrova & Leong, 2010; Whitebread & Sullivan, 2012). However, in view of enhancing children's school readiness

skills, early childhood programs increasingly tend to emphasize academic content, which can be at the expense of enhancing children's self-regulation through pretend play (Leseman, 2012).

It has been suggested that only complex pretend play involving an imaginary situation in which children enact roles and follow the rules determined by this role, is beneficial for children's self-regulation development and that not all children reach these high levels of pretend play on their own (Bodrova, 2008). The current results indicate that the quality of pretend play seems to matter for children's self-regulation, with quality referring to the degree of role enacting, symbolizing, collaboration and meta-communication. Teachers' guidance should be focused on supporting children to reach high levels of pretend play regarding all these aspects, but in particular role-play, to create optimal learning opportunities for children (Bodrova, 2008; Whitebread & Sullivan, 2012). This is an important implication for current practice in early childhood education and care provisions in which children are often given ample time for free play, however with little or no teacher guidance and support, as opposed to academically focused activities.

There are a number of limitations to this study. First, the study was small-scale and involved a deliberately selected sample. Therefore, we are not able to draw strong conclusions beyond the current sample. Also, as only concurrent relationships between self-regulation and pretend play were investigated, we cannot make causal inferences. Newly developed observation instruments for self-regulation were used that need independent evaluation of their validity in future research and further adaptations to be applicable to a wider range in developmental levels. Yet, the present results, especially the relations with cool EF and the quality of play, attest to the validity of the new measures. Additionally, children's self-regulation and the quality of their pretend play were simultaneously evaluated by the same research-assistant. Although the assistants were blind to the study objectives, it is likely that there is method-bound shared variance. However, upon closer scrutiny of the pattern of correlations (see above), we are confident that the statistical relations that were found reflect true relations. Finally, post-hoc power analyses revealed that the power was sufficient to detect medium or large effects (power > .80). However, the power to detect small effects was lower (.50), which might be another explanation for the null relations between observed self-regulation and global classroom quality.

To conclude, the current study adds to the existing evidence on young children's self-regulation as displayed in a naturalistic play setting. Cognitive and emotional self-regulation as observed in a naturalistic setting are related but distinct constructs. Three-year-old

children already show important elements of cognitive and emotional self-regulation in play, which can be partly related to EFs as measured by neuropsychological tests and partly to the quality of play. Global classroom quality, however, does not relate to observed self-regulation. This calls for further research to investigate the specific aspects of classroom quality that are important for children's self-regulation development.

Chapter 6

Summary and General Discussion

The studies reported in this dissertation are part of the large scale, ongoing, national cohort study, pre-COOL, to evaluate the developmental and educational effects of early childhood education and care (ECEC) provisions in the Netherlands. Within pre-COOL, the studies reported in this dissertation specifically focused on the quality of a large representative sample of ECEC provisions and the potential effects of ECEC quality on child development in two-to-three-year-olds. ECEC quality refers to children's daily experiences in interacting with teachers and peers while children participate in all kinds of activities and routines, referred to as process quality, and the structural and organizational characteristics that are considered important preconditions of these experiences, which are hypothesized to be beneficial for children's development (Howes et al., 2008; Layzer & Goodson, 2006; Sylva et al., 2006; Thomason & La Paro, 2009). Process quality concerns the physical, emotional, social, and instructional aspects of children's interactions with teachers, peers and materials (Howes et al., 2008; Pianta et al., 2005; Thomason & La Paro, 2009), whereas structural quality refers to classroom or teacher characteristics, such as group size, children-to-teacher ratio, and teacher's qualifications, which have been shown to be associated with process quality (Cryer, Tietze, Burchinal, Leal, & Palacios, 1999; Philips, Mekos, Scarr, McCartney, & Abbott-Shim, 2000; Philipsen, Burchinal, Howes, & Cryer, 1997; Pianta et al., 2005; Vandell, 2004). In the present studies, the systematic provision of particular activities, referred to as the implemented curriculum, was added as a third aspect to the quality concept, acknowledging that children's experiences have particular knowledge contents and are meant to serve particular developmental and educational goals.

Four major issues stood out at the start of the present studies. The first issue concerned the measurement of quality. The choice within pre-COOL for the widely used Classroom Assessment Scoring System (CLASS) was made to ensure comparability with international research, which could provide benchmarks to evaluate Dutch ECEC, but raised the question of applicability of this observational measure to the Dutch context, which presents in many respects a different cultural context and a different tradition in ECEC than the US context in which this measure was developed. The second issue concerned the relationship between process quality and structural quality, an issue which is particularly important because structural features constitute the largest costs of ECEC, whereas process quality is most strongly related to effects on children's development, and, thus to the potential benefits for society at large (Vandell et al., 2010). Therefore, a strong positive relation between structural and process quality is essential for the costs-efficiency of ECEC. The evidence on the relations between the traditional structural quality characteristics, such

as group size, children-to-teacher ratio, and teacher education level, and process quality has been shown to be inconclusive (Cryer et al., 1999; Early et al., 2007) and, therefore, may not provide optimal starting points for improving process quality. In this regard, potentially more effective and economically more efficient starting points for improving process quality may be the use of well-structured education programs and, particularly, strategies of continuous professional development (Domitrovich et al., 2009; Zaslow, Anderson, Redd, Wessel, Tarullo, & Burchinal, 2010). The third issue concerned the effects of ECEC, and ECEC quality in particular, on children's development. Numerous studies have shown positive effects of ECEC attendance on child development (Burger, 2010; Gormley, Philips, & Gayer, 2008), especially when the provided care and education was of high quality (NICHD ECCRN, 2000, 2006; Pianta et al., 2009; Sylva et al., 2011). However, evidence concerning the effectiveness of the Dutch ECEC system is still limited. In the past decades, only a few quasi-experimental pilot studies and a number of retrospective studies have been conducted, but no major prospective study was carried out to investigate the effects of ECEC quality on children's developmental and educational outcomes. The final issue related to a core debate about the early years curriculum, in particular the balance between play and academic activities. This issue is especially important in view of recent insights into the role a playful curriculum can have in supporting children's self-regulation development, which is considered an important learning-related skill. These four issues were addressed in the four empirical studies reported in this dissertation.

Below, we first summarize the main findings of each of the studies. Next, we integrate the findings and discuss the theoretical issues emerging from the studies. Then, we discuss implications for policy and practice, and identify directions for future research.

Summary of the Main Findings

Prior to investigating the quality of early childhood provisions in the Netherlands and the effects on children's developmental outcomes, we investigated whether the measure used to assess ECEC quality has good psychometric properties. In Chapter 2 we presented findings concerning the measurement quality of this instrument, the Classroom Assessment Scoring System Toddler (CLASS, La Paro et al., 2012). The main motivation of this study was twofold. First, the evidence on the psychometric properties of the CLASS Toddler, the particular version we used in our studies, is still limited. Although, the available evidence supports the reliability and validity of the CLASS Toddler, the evidence is solely based on studies

conducted in the US, raising the question whether the results apply to contexts outside the US and to the Dutch ECEC provisions in particular. Second, the available evidence on the measurement quality of the CLASS Toddler (and also the Pre-K version) is based on the measure as a whole within a classical test-theory approach. Information on the separate observation items (called indicators) and their factor structure is lacking.

In our study reported in Chapter 2, we extended prior research on the psychometric properties of the CLASS in several ways. We conducted multilevel confirmatory factor analyses, which allowed for comparison of the measurement properties *within* and *between* classrooms, for both the domain-structure and the dimension-structure of the CLASS. The findings indicated a few problems with the dimensions Negative Climate and Regard for Child Perspectives, which may point to cultural differences between the US and Dutch context. In the Netherlands, overall, the mean and variance in Negative Climate were very low, indicating only mild expressed negativity and few occasions of negative teacher-child and peer interactions. Furthermore, the Negative Climate dimension appeared to be related to the Behavior Guidance dimension, together constituting a separate domain in the pre-COOL data. This finding seems to reflect that in classrooms with less expressed negativity, teachers were better able to support children's behavior by stating positive behavioral expectations and reinforcing positive behavior, resulting in less wandering around and problem behavior. For Regard for Child Perspectives we found differences in the measurement structure at the *within* and *between* classrooms level, indicating this dimension did not fit well in the overarching domain at the *within* level, whereas it did fit well at the *between* level. This finding suggests there is substantial variation in teachers' child-centeredness between the different activity settings that were observed within the classrooms. Further inspection of the data revealed that Regard for Child Perspectives was comparatively high during free play (with low teacher involvement), but comparatively low in other settings (with high teacher involvement), such as during educational and creative group activities and during care routines, suggesting that teachers' child-centeredness may not be well-balanced during the day.

Furthermore, we investigated the measurement properties of the CLASS indicators, which are the primary sources of information on which the dimension scores and overall domain evaluations are based, by evaluating item difficulty and item discrimination using an Item Response Theory (IRT) approach. Overall, the indicators appeared to have satisfactory measurement properties, based on the difficulty and discrimination parameters. The results furthermore revealed that teachers were more likely to receive a high score on the indicators

of Emotional Support and Behavioral Support than on the indicators of Engaged Support for Learning. However, the item discrimination parameters were equally high across all indicators, revealing that the indicators distinguished equally well between teachers, both in the low and high quality range. Taken together, this suggests that the differences in item difficulty most likely reflect actual differences in teachers' abilities and are not due to a methodological artifact.

Finally, we investigated the validity of the CLASS Toddler by relating the CLASS domains to commonly used structural quality and curriculum aspects. In line with our expectations and prior research, we found several associations between the domains and structural quality aspects, such as children-to-teacher ratio, teacher's work experience, and teacher's qualifications, and curriculum characteristics, including the provision of educational and play activities, confirming the construct validity of the CLASS. Overall, the findings indicate good psychometric quality and validity of the CLASS Toddler in the Dutch context.

In Chapter 3, we investigated the associations between structural quality and process quality, extending the existing literature in three ways. First, we used a multi-method approach to assess quality by combining two types of quality measures, observations and teacher reports, and by defining two comprehensive quality constructs, Emotional and Educational Quality, including observed teacher-child interactions as well as the reported provision of activities as part of the implemented curriculum. Second, we included the usual structural quality aspects, group size, children-to-teacher ratio, and teachers' pre-service education level as potential determinants of process quality, but added two additional characteristics, namely the use of an education program (in Dutch: *VVE programma*) and the provision of professional development activities to evaluate their relative contribution to process quality when assessed simultaneously. Third, we applied multilevel modeling using the information of the four observation cycles at the *within classrooms* level and of both the observations and the reported curriculum at the *between classrooms* level to examine the relations with type of setting (at the within level) and structural quality characteristics (at the between level).

The results showed that emotional quality was moderate to high in Dutch ECEC, whereas educational quality was low to moderate, with preschools scoring significantly higher on the latter compared to day care centers. These results are in line with previous Dutch and international research (Helmerhorst, Riksen-Walraven, Vermeer, Fukkink, & Tavecchio, 2014; Pakarinen et al., 2010; Pianta & Hamre, 2009; Thomason & La Paro, 2009; Weiland, Ulvestad, Sachs, & Yoshikawa, 2013). Data from the teachers' self-reports revealed a similar

pattern. Teachers reported being emotionally supportive oftentimes and were moderately inclined to support children's self-managed play, whereas they reported providing pretend play and academically focused activities less frequently. The type of activity setting, situated *within* the classroom, was related to process quality. Emotional process quality was highest during educational activities, such as book reading, making puzzles, and circle time, compared to care routines, including toileting and mealtimes, which is consistent with prior research (De Schipper et al., 2006; Pianta et al., 2005). Observed educational process quality was higher during educational and creative activities, and to a lesser extent during free play compared to care routines, with educational activities being most strongly related to observed educational quality. At the *between* classrooms level the results indicated that average group size and children-to-teacher ratio were not related to process quality, probably due to the low variation in these variables, given that these structural characteristics are strongly regulated in the Netherlands, whereas teachers' pre-service education level only had a small effect. The strongest associations were found between process quality and the use of an education program and, particularly, the provision of professional development activities at the center.

In the last two studies of this dissertation, Chapters 4 and 5, we focused on the relations between classroom quality and children's developmental outcomes. Building on the findings reported in Chapter 2, we applied the three-domain-structure of the CLASS to investigate associations with children's language and self-regulation development. In Chapter 4, we presented findings from a large sample of children attending day care and preschools and investigated the effects of process quality on children's growth in vocabulary and attention skills from age two to three years using a value-added approach. Research on effects of ECEC quality on young children's developmental and educational outcomes tends to be focused on global classroom quality, while research with preschoolers has shown that domain-specific aspects of quality and curriculum can have larger effects than global quality, overall. Therefore, we combined global and specific measures of classroom quality to determine their contribution to children's outcomes one year later. As outcome measures, we selected vocabulary and attention because these skills have been shown to be strong predictors of later social-emotional and academic development. The value-added approach allowed us to control for possible selection mechanisms that are common in non-experimental studies. The findings revealed a small positive effect of Emotional Support on the development of children's vocabulary skills over one year and a stronger positive effect of Engaged Support for Learning on the development of children's attention skills. Also, negative effects of the

provision of free play activities were found for both children's vocabulary and attention skills, while no effects of the provision of academic and self-regulation activities were observed. The results confirm that both global and specific process quality aspects can have impact on children's outcomes, at least on their receptive vocabulary and selective attention skills.

In Chapter 5, we used an in-depth approach to assess children's self-regulation skills during pretend play and related this to test-based measures of children's cognitive (cool) and affective (hot) executive functions (EF). We furthermore examined the relations between children's self-regulation and contextual factors, including the quality of pretend play and global classroom quality as assessed with the CLASS. For this purpose, an observational study using video-recordings was conducted in a subsample nested within the larger pre-COOL sample, focusing on pretend play as a setting that may support self-regulation development. We developed an observational measure to assess children's cognitive and emotional self-regulation skills to gain better understanding of how children use these skills in daily practice. The findings revealed that the new observational measure was useful for assessing children's self-regulation skills, as there was quite some variation between children in the extent to which they showed self-regulation skills in their play. We also found differential associations between observed self-regulation and test-based measures of children's cognitive (cool) and affective (hot) executive functions (EF), with the strongest relations occurring between cool EF and emotional self-regulation. The quality of pretend play was significantly related to children's emotional and cognitive self-regulation skills during play, when children's background characteristics and their EFs were controlled for, with the quality of play being strongest related to cognitive self-regulation. Children's level of role-play, as part of the quality of the play construct, in particular was quite strongly associated with children's emotional self-regulation. Global classroom quality did not contribute to children's self-regulation skills. The findings revealed that three-year-old children already show important elements of cognitive and emotional self-regulation in play, which can be partly attributed to EFs as measured with neuropsychological tests and partly to the quality of play.

Integration of the Findings

The findings presented in this dissertation raise several topics for further discussion that will be taken up in this section.

Quality of Dutch ECEC

This dissertation reports the findings of a large-scale study into the quality of day care and preschool provisions for children under four years in the Netherlands. The findings show that process quality of Dutch ECEC is on average in the mid to high range for the emotional support and behavioral guidance domains, and in the low to mid range for the support for learning domain. However, there is considerable variability in all domains and preschools tend to have higher educational process quality than day care centers. Similar results, indicating moderate emotional and behavioral support and low educational support, were found in previous research in Dutch day care centers (De Kruif et al., 2009; De Schipper, Riksen-Walraven, & Geurts, 2006; Fukkink, Gevers-Deynoot-Schaub, Helmerhorst, Bollen, & Riksen-Walraven, 2013; Helmerhorst, Riksen-Walraven, Vermeer, Fukkink, & Tavecchio, 2014) and in international research in day care centers and preschools (Bandel, Aikens, Vogel, Boller, & Murphy, 2014; OECD, 2006; Hamre et al., 2013; La Paro et al. 2009; La Paro, Williamson, & Hatfield, 2014; Philips & Lowenstein, 2011; Shonkoff & Philips, 2000; Tietze & Cryer, 2004). Regarding the Dutch situation, previous research has shown that the quality of Dutch day care has steadily declined between 1995 and 2008. This can be partly attributed to the enormous growth of day care usage, which more than doubled since 1995 (De Kruif et al., 2009). Moreover, with the introduction of the Child Care Act in 2005, the sector has been completely reorganized from a publicly funded supply-oriented system into a partly non-profit and partly for-profit demand-driven system. Although at the same time structural quality regulations were made more rigorous, quality monitoring was strengthened, and global process quality aspects were introduced as part of the statutory quality regulations, this apparently has not been sufficient to maintain and increase the quality of day care in the context of rapid expansion of its capacity. The aim of the Child Care Act was to increase the capacity and accessibility of ECEC provisions within a context of enhanced market competition, which in turn was thought to lead to high overall quality (Noailly & Visser, 2009). The current findings, along with findings of other Dutch studies, suggest that the policy did not work out as intended since overall quality declined between 2005 and 2008 and can currently be considered low regarding educational quality and mid to high regarding emotional quality.

Enhancing ECEC Quality

Enhancing the quality and impact of ECEC is a topic of high priority in Dutch policy, like in many other countries (Kamerman, 2007). Following numerous studies on the quality

of ECEC, a number of structural quality aspects have been put forward as important preconditions for process quality (Cryer, Tietze, Burchinal, Leal, & Palacios, 1999; Philips, Mekos, Scarr, McCartney, & Abbott-Shim, 2000; Philipsen, Burchinal, Howes, & Cryer, 1997; Pianta et al., 2005; Vandell, 2004). Therefore, in many countries all over the world these structural quality characteristics have been included in national statutory quality regulations. However, research has revealed mixed results regarding the importance of these structural quality characteristics as determinants of process quality (Barros & Aguiar, 2010; Blau, 2000; Burchinal et al., 2002; Burchinal, Cryer, Clifford, & Howes, 2002; Cryer et al., 1999; De Kruif et al., 2009; De Schipper, Riksen-Walraven, & Geurts, 2006; Fukkink, Gevers-Deynoot-Schaub, Helmerhorst, Bollen, & Riksen-Walraven, 2013; Mashburn et al., 2008; NICHD ECCRN 2000a; Pessanha, Aguiar, & Bairrao, 2007; Philips et al., 2000; Philipsen et al., 1997; Pianta et al., 2005; Thomason & La Paro, 2009; Vermeer et al., 2008). The findings of this dissertation confirm that the relationships between structural quality characteristics and process quality are not as straightforward as has been thought. This, however, does not imply that these structural characteristics are not important, but merely that it is difficult to detect effects when there is limited variation. As noted before, the Dutch ECEC system is strongly regulated concerning structural characteristics, reducing the variance in structural quality and thus the likelihood of revealing possible effects. Yet, considerable variation in process quality in Dutch ECEC remains, as was found in this dissertation, suggesting that other aspects might play an important role in determining process quality. In this regard, two major findings have emerged from this dissertation.

The first finding concerns the importance of continuous professional development for process quality, as was shown in Chapter 3. Internationally, there is also increasing attention for professional development beyond pre-service education as a means to increasing quality in ECEC (CoRe, 2011; OECD, 2012). Several strategies exist to support teachers' professional development, including in-service training, consultation, mentoring and coaching on the job, and all these strategies have been shown to positively affect process quality (Campbell & Milbourne, 2005; Domitrovich et al., 2009; Fukkink & Lont, 2007; Howes et al., 2003; Hamre et al., 2012; LoCasale-Crouch et al., 2011; Pianta et al., 2008) and children's outcomes (Bierman et al., 2014; Bierman, Nix, Greenberg, Blair, & Domitrovich, 2008; Dickinson & Caswell, 2007; Domitrovich, Cortes, & Greenberg, 2007). Our measure of professional development also involved aspects like in-service training, the use of observation and feedback to improve practices, and regular staff meetings to discuss children's development, the goals of working with young children, and the development and implementation of a

curriculum, which was found to be moderately to strongly related to process quality.

Relatedly, our second finding concerns the use of a well-structured education program (*VVE programma*). Using such a program was found to be positively related to emotional and educational process quality, as defined in the current dissertation in Chapter 3. This finding may be relevant for the current debate in the Netherlands on the usefulness and effectiveness of structured education programs for early years settings. In the Netherlands, education programs for preschools are part of a targeted policy in combating early educational disadvantages and are provided to children from lower educated parents or with a migration background. Although originally developed for targeted groups of disadvantaged children in preschools, today also regular preschools and day care centers serving a general population are increasingly using these programs. These education programs have been criticized because of their strong emphasis on children's cognitive and language skills, and relative lack of attention to socio-emotional and motor skills (Onderwijsraad, 2008). Moreover, the effectiveness of these programs has been questioned. For instance, a recent Dutch study (Bruggers, Driessen, & Gesthuizen, 2014) has shown null effects of the use of education programs on children's language and math skills in the second year of kindergarten (*groep 2*). Note however, that this study suffered from several methodological limitations, which possibly explain these null findings. First, the researchers used a retrospective research design, asking parents and primary schools to provide information on preschool use several years earlier. Secondly, although the researchers controlled for important family background characteristics, they did not control for children's skills prior to preschool participation, thus a baseline measurement was lacking. Third, the *quality* of the ECEC provisions used was not taken into account in this study while, as previous research and the findings in this dissertation have shown, effects on children's outcomes critically depend on the quality of the ECEC provisions (Melhuish, 2011; Pianta, Barnett, Burchinal, & Thornburg, 2009; Sylva et al., 2011). Finally, and perhaps most importantly, the study did not control for possible selection effects in preschool or day care enrolment. In our study we addressed all the issues outlined above, by using a prospective, value-added approach, controlling for the autoregressive effects of children's development, and controlling for (possible) selection effects. Our findings revealed that the use of an education program contributed to emotional and educational process quality, more than the usual structural quality characteristics did (Chapter 3), while emotional process quality and especially educational process quality, in turn, were found to be related to children's developmental outcomes (Chapter 4). Moreover, in our large sample of children in a representative variety of day care and preschool centers,

rather strong selection effects were evident. We found that children's vocabulary and attention skills at the first wave of the pre-COOL study, when children were aged two years, were moderately to strongly negatively related to process quality, indicating that children most in need for high quality indeed received higher quality care in targeted day care centers and preschools, as intended by educational policy. However, as a consequence, the average skill level of children in targeted preschool centers working with an education program, was that far below average right at the start that, in a retrospective analysis without adequate control of children's entry-level, possible effects of the program may have been cancelled out.

The use of an education program in itself does not automatically imply high quality and improved outcomes. The potential effects on children's developmental and educational outcomes depend on the way in which the program activities are implemented, embedded in a meaningful context, and attuned to children's experiences, interests, and prior knowledge. Despite the fact that the quality of Dutch ECEC was found to be generally low regarding the educational aspects and mid to high regarding the emotional aspects, we did find first indications for positive effects of quality on children's development. Although these effects were small, as these are in line with previous studies reporting on effects of ECEC quality (Burchinal et al., 2011; Pianta et al., 2009; Zaslow et al., 2010), they can be regarded as promising.

Early years curriculum

There is an ongoing debate about what the focus of ECEC provisions should be regarding the curriculum provided, with play and academics often placed at opposite ends of the continuum. Our findings can have implications for this debate. First, we found that observed process quality was highest during educational activities, such as circle time, book reading and making puzzles, and creative activities and lowest during care routines, including mealtimes and toileting, which is in line with previous studies (De Schipper et al., 2006; Pianta et al., 2005). In free play, which in the current study was observed to take about one third of the time children spent during a regular morning in the classroom, educational quality was found to be lower than during educational and creative group activities (Chapter 3). In Dutch ECEC provisions, play usually concerns a free, child-directed context characterized by little teacher involvement. Other studies have reported that children tend to wander around a lot during free play (De Haan et al., 2013) and only engage in verbal interactions with teachers to a limited extent (Leseman, Rollenberg, &

Rispens, 2001; Powell, Burchinal, File, & Kontos, 2008). These findings may also provide an explanation for the negative associations we found between the teacher-reported provision of play activities with children's vocabulary and attention development, assuming that the provided play activities are usually free play activities with low teacher involvement (Chapter 4). Based on the findings of this dissertation, from the video observations conducted in the in-depth study, with a subsample of the classrooms involved in pre-COOL, we argue that teacher involvement in order to guide children and to support the quality of their play is critical, which is also in line with the findings of other studies with experimental designs (Barnett et al., 2008; Diamond, Barnett, Thomas, & Munro, 2007). In the observation study reported in Chapter 5, teachers were asked to set up a pretend play activity involving a set of kitchen toys. Although no further instructions were given, this elicited more teacher involvement and guidance. The findings from this study indicated furthermore that the quality of play was related to children's observed self-regulation, while controlling for test-based executive functions, vocabulary and background characteristics, suggesting that merely providing play opportunities may indeed not be sufficient, but that rather the quality of play and the guidance teachers provide to raise the quality of play can contribute to children's development in self-regulation.

Concerning the provision of academic activities, our findings suggest that these activities are not provided very frequently in Dutch day care centers and preschools involved in pre-COOL; on average one or a couple of times a week. Whether this is cause for concern, is disputable, because in the current study we did not find effects of the provision of academic activities on the development of children's vocabulary and attention skills between age two and three years, suggesting that a too strong focus on academics might not yet be appropriate, at least not for children in this age range. We did find a significant effect of educational quality, as measured with the CLASS, on children's attention development, however, which suggest that not so much a focus on 'teaching' academic skills, but providing learning opportunities in play and discovery activities, and scaffolding children's exploration and learning in meaningful and language-rich contexts, is beneficial for children's development.

Altogether, in view of the debate on what constitutes an appropriate curriculum for the early years, play and academics are probably unfairly contrasted at opposite ends of the continuum. The key seems to be providing children with intentional learning opportunities, embedding knowledge content into a meaningful context, with ample opportunities for children's self-directed exploration. Play seems an excellent context for this, assuming

that teachers guide children's play and intentionally create learning opportunities in play (Bodrova, 2008), but given the relatively high educational quality of activities in educational and creative settings, as we found in the study reported in Chapter 3, other activity settings may be equally suitable for promoting cognitive development.

Classroom quality measures

As classroom observation measure we used the CLASS to ensure comparability with international research. The results from the analyses of the psychometric properties of the CLASS indeed revealed some cultural differences, which are important to keep in mind when interpreting the current results. Moreover, the effects of classroom quality, as measured with the CLASS, on children's developmental outcomes were small. Although this finding is in line with previous research (Burchinal, Kainz, & Cai, 2011), it raises the question whether the use of a classroom measure, such as the CLASS, is specific enough to detect effects on children's development. Perhaps, in addition, we need more fine-grained measures identifying those particular aspects that affect children's development. To date, such specific quality measures are lacking (Burchinal et al., 2011). For instance, there seems to be a lacuna in the currently available classroom quality measures to assess the conditions that are specifically relevant for children's self-regulation development (Hyson et al., 2011).

Furthermore, most research on the association between ECEC quality and children's outcomes has assessed process quality at the classroom level by focusing on the classroom as a whole (Farran & Hofer, 2013; Hallam, Fouts, Bargreen, & Caudle, 2009), thus investigating quality from a "top down" perspective. A top down approach evaluates what the teacher is providing to all children or reflects the experiences of an "average" child. However, this does not necessarily mean all children profit to the same extent from the quality provided, and differential associations may exist depending, for instance, on children's temperament (Pluess & Belsky, 2009; Vitiello et al., 2012) or self-regulation (Broekhuizen, van Aken, Dubas, Mulder, & Leseman, under review). Hence, this might partially explain why effects of overall classroom quality on individual children's outcomes tend to be small. In contrast, a "bottom up" approach, which considers the child's perspective, might shed a different light on classroom quality (Hallam et al., 2009; Howes & Smith, 1995; Powell et al., 2008). A bottom up approach evaluates children's behavior as reflective of quality instead of the teacher's behavior, and complements a broader view on classroom processes and how they are related to children's outcomes.

Implications for Policy and Practice

The findings of the studies reported in this dissertation can have implications for policy and practice. The importance of quality in ECEC provisions has been demonstrated and we have also noted that the quality of Dutch ECEC is not yet satisfactory, particularly regarding educational quality aspects aimed at supporting children's cognitive and language development. Therefore, enhancing process quality should be an important goal for both policy and practice. As noted previously, education programs specifically developed for improving the educational quality of ECEC can provide support to teachers to enhance educational process quality. Most important, however, may be investments in continuous professional development, as this appeared to be the most promising strategy to enhance ECEC quality. Professional development can consist of a variety of activities and strategies, which are most likely not equally successful in improving quality. Therefore, the use of programs, training or other forms of professional development should be guided by evidence-based research on what works best. A recent comprehensive review by Zaslow and colleagues (2010) identifies a number of approaches to effective professional development in ECEC, which can provide useful starting points for developing interventions or programs targeted at professional development. An important recommendation following from this work is that the focus of professionalization should be specific in the content addressed and closely aligned with the areas of practice for which improvement is needed.

The use of an observational measure can support teachers and practitioners in providing specific goals for improvement, especially if the measure used is strongly related to specific aspects of teacher-child interactions, which in turn are known to be related to particular developmental outcomes for children (Pianta, Hamre, & Downer, 2011). The CLASS, for instance, has been used in professional development interventions in the US and has been proven to be effective in enhancing teacher's interaction skills (Hamre et al., 2012) and children's development (Cabell & Downer, 2011; Mashburn, Downer, Hamre, Justice, & Pianta, 2010). The CLASS is constructed in a way such that it addresses very concrete and recognizable teacher behaviors, providing direct insights in the areas that are in need of improvement. The use of an observational measure, such as the CLASS, might also prove useful for professional development interventions in the Dutch context. Another important aspect in professionalization emerging from Zaslow et al.'s review study concerns the link between knowledge and practice. Specifically, a combination of a course, training program or workshop with onsite or web-based consultation and feedback on practices seems a promising

strategy to enhance quality as well (see also: Domitrovich et al., 2009; Pianta et al., 2008). Also, the use of video feedback can be recommended. A recently developed intervention program targeting teachers' sensitive responsiveness during interactions with children showed that video feedback is effective in this particular domain (Werner, Vermeer, Linting, & Van IJzendoorn, under review), in line with previous research (Fukkink & Lont, 2007).

In considering what is a developmentally appropriate curriculum for children, attention should be focused on providing a comprehensive set of activities attuned to children's interests and experiences using a playful, child-centered approach. From the perspective of combating early developmental and educational inequalities, increasing emphasis has been placed on school readiness for which education programs have been developed and implemented in ECEC provisions. However, the use of an education program can lead to too rigidly adhering to a manual with instructions (Doolaard & Leseman, 2008), often at the expense of using unplanned, informal opportunities for learning. This calls for strong professionals who closely observe children and their needs, who are capable of flexibly adjusting their plans to children's interests and experiences, and who can turn any situation into an interesting learning opportunity. Professional development activities can support teachers in developing these skills.

Limitations and Future Directions

There are a number of limitations that should be noted. First, the findings reported on in this dissertation only used data from the first two measurement waves of the pre-COOL study as the data collection is still ongoing. Furthermore, the effects of ECEC quality on children's development examined in this dissertation were limited to only two, although important, developmental domains. Further analyses of the pre-COOL data are planned to evaluate effects in other domains of development as well. In addition, there are two methodological aspects that should be noted. The first concerns the missing data in the teacher, parent and, to a lesser extent, child data. Although statistical analyses are increasingly sophisticated in dealing with missing data, this remains a concern. The second concern is that the data on the quality of ECEC used in the studies of this dissertation are primarily based on the classroom level, except for the in-depth video observation study that was conducted in a subsample of the pre-COOL classrooms. Child level data of actual behavior and experiences in the classroom could have added to our knowledge of how children perceive actual classroom processes.

The findings reported in this dissertation suggest a number of directions for future research. First, and foremost, future research could focus more on the effects of curriculum on children's outcomes. Relatedly, we recommend a stronger alignment of the quality and curriculum assessment measures to the specific child outcome measures of interest, in order to be able to detect stronger effects (Bryant, Burchinal, & Zaslow, 2011). Ideally, classroom measures should be complemented with measures of children's behavior in the classroom to gain a more comprehensive view of how actual classroom processes relate to children's development and educational outcomes.

Another recommendation concerns the need to investigate the nature of associations between structural quality, process quality, and children's outcomes. We have not directly tested the hypothesized path of structural quality aspects affecting children's outcomes *through* process quality in one model. This would add to the existing literature, as only a few studies have directly tested a mediation model, which, however, were limited to the so-called iron-triangle structural aspects (NICHD Early Child Care Research Network, 2002). An interesting line of research to pursue would be to investigate the direct and indirect effects of professional development on children's development, as in the present dissertation this aspect was found to be most strongly related to process quality.

Finally, there is a clear need for more longitudinal studies on the effects of ECEC quality on children's outcomes in the Dutch context with strong methodological designs, preferably using an experimental, if possible, or, as next best, a value-added approach. Several Dutch reports and articles have been published revealing null effects of ECEC attendance, the use of an education program, or ECEC quality on children's outcomes, but these studies have relied on retrospective research designs, which cannot sufficiently control for selection bias. Therefore, the results of these studies can be seriously confounded (e.g. Bruggers et al., 2014; Karssen, van der Veen, Veen, van Daalen, & Roeleveld, 2013). Policy makers increasingly tend to base their decisions regarding early childhood education and care on the available evidence. Therefore, it is critical that the available evidence is based on strong research.

General Conclusion

The findings presented in this dissertation have revealed that emotional quality in Dutch ECEC provisions is moderate to high, while educational quality is low to moderate, with large variation between classrooms and between centers. Our findings provide a first

indication that ECEC quality is related to children's developmental outcomes in important areas. Given the potential benefits for children's development, raising the quality of ECEC is an important goal for the near future. Our findings have revealed that the use of an education program and, particularly, the provision of activities for continuous professional development can provide starting points for enhancing ECEC quality. However, more research is needed to corroborate the current findings.

References

Anders, Y., Grosse, C., Rossbach, H.-G., Ebert, S., & Weinert, S. (2013). Preschool and primary school influences on the development of children's early numeracy skills between the ages of 3 and 7 years in Germany. *School Effectiveness and School Improvement: An International Journal of Research, Policy and Practice, 24*, 195-211. doi: 10.1080/09243453.2012.749794

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

Assel, M. A., Landry, S. H., Swank, P. R., & Gunnewig, S. (2007). An evaluation of curriculum, setting, and mentoring of children enrolled in pre-kindergarten. *Reading and Writing, 20*, 463-494. doi: 10.1007/s11145-006-9039-5

Barnett, W. S. (1995). Long-term effects of early childhood programs on cognitive and school outcomes. *The Future of Children, 5*, 25-50.

Barnett, W. S., Jung, K., Yarosz, D. J., Thomas, J. Hornbeck, A., Stechuk, R., & Burns, S. (2008). Educational effects of the Tools of the Mind curriculum: A randomized trial. *Early Childhood Research Quarterly, 23*, 299-313. doi: 10.1016/j.ecresq.2008.03.001.

Barros, S. & Aguiar, C. (2010). Assessing the quality of Portuguese child care programs for toddlers. *Early Childhood Research Quarterly, 25*, 527-535. doi: 10.1016/j.ecresq.2009.12.003

Bauchmüller, R., Gørtz, M., & Rasmussen, A.W. (2011). *Long-Run Benefits from Universal High-Quality Pre-Schooling Denmark*. AKF Working Paper, 2011

Bennet, J. (2003). Starting Strong: The persistent division between care and education. *Journal of Early Childhood Research, 1*, 21-48. doi: 10.1177/1476718X030011006

Baumeister, R. F., Schmeichel, B. J. & Vohs, K. D. (2007). Self-regulation and the executive function: The self as controlling agent. In A. W. Kruglanski & E.T. Higgins (Eds.), *Social psychology: Handbook of basic principles* (Second edition). New York: Guilford.

Bennet, J. (2005). Curriculum issues in national policymaking. *European Early Childhood Education Research Journal, 13*, 5-23. doi: 10.1080/13502930585209641

Berk, L.E., Mann, T.D., & Ogan, A.T. (2006). Make-believe play: Wellspring for development of self-regulation. In D.G. Singer, R.M. Golinkoff, & K. Hirsh-Pasek (Eds.), *Play=learning: How play motivates and enhances children's cognitive and social-emotional growth* (pp. 74-100). New York: Oxford University Press.

Bierman, K. L., Nix, R. L., Greenberg, M. T., Blair, C., & Domitrovich, C. E. (2008). Executive functions and school readiness intervention: Impact, moderation, and mediation in the Head Start REDI program. *Development and Psychopathology, 20*, 821-843. doi: 10.1017/S0954579408000394

Blair, C. (2002). School readiness: integrating cognition and emotion in a neurobiological conceptualization of child functioning at school entry. *American Psychologist, 57*, 111-127. doi: 10.1037//0003-066X.57.2.111

Blair, K. A., Denham, S. A., & Kochanoff, Whipple, B. (2004). Playing it cool: Temperament, emotion regulation, and social behavior in preschoolers. *Journal of School Psychology, 42*, 419-443. doi: 10.1016/j.jsp.2004.10.002

Blair, C., & Diamond, A. (2008). Biological processes in prevention and intervention: The promotion of self-regulation as a means of preventing school failure. *Development and Psychopathology, 20*, 899-911. doi: 10.1017/S0954579408000436

Blair, C. & Razza, R. P. (2007). Relating effortful control, executive function, and false belief understanding to emerging literacy ability in kindergarten. *Child Development, 78*, 647-663. doi: 0009-3920/2007/7802-0018

Blair, C., & Ursache, A. (2004). A bidirectional model of executive functions and self-regulation In R. F. Baumeister & K. D. Vohs (Eds.) *Handbook of self-regulation: Research , theory, and applications* (second edition, pp. 300-320). New York: Guilford.

Blair, C., Zelazo, P. D., & Greenberg, M. T. (2005). The measurement of executive function in early childhood. *Developmental Neuropsychology, 28*, 561-571. doi: 10.1207/s15326942dn2802_1

Blau, D. M. (2000). The production of quality on child-care centers: Another look. *Applied Developmental Science, 4*, 136-148. doi: 10.1207/S1532480XADS0403

Bodrova, E. (2008). Make-believe play versus academic skills: A Vygotskian approach to today's dilemma of early childhood education. *European Early Childhood Education Research Journal, 16*(3), 357-369.

Bodrova, E., & Leong, D.J. (2007). *Tools of The Mind: The Vygotskian approach to early childhood early childhood education* (2nd edition). New York: Prentice Hall.

Bodrova, E., & Leong, D. J. (2006). Self-regulation as key to school readiness: How early childhood teachers promote this critical competency. In M. Zaslow & I. Martinez-Beck (Eds.), *Critical issues in early childhood professional development* (pp. 203–224). Baltimore: Paul H. Brookes Publishing.

Bodrova, E., Leong, D.J., & Akhutinia, T.V. (2011). When everything new is well-forgotten old: Vygotsky/ Luria insights in the development of executive functions. *New Directions for Child and Adolescent Development, 133*, 11-28. doi:10.1002/cd

Bono, K. E., & Bizri, R. (2013). The role of language and private speech in preschoolers' self-regulation. *Early Child Development and Care,* doi:10.1080/03004430.2013.813846

Bowers, E. P., & Vasilyeva, M. (2011). The relation between teacher input and lexical growth of preschoolers. *Applied Psycholinguistics, 32*, 221-241. doi: 10.1017/S0142716410000354

Bradley, R.H., & Corwyn, R.F. (2002). Socioeconomic status and child development. *Annual Review of Psychology, 53*, 371-399. doi: 10.1146/annurev.psych.53.100901.135233

Bradley, R. H., & Vandell, D. L. (2007). Child care and the well-being of children. *Archives of Pediatrics & Adolescent Medicine, 161*, 669-676. doi: 10.1001/archpedi.161.7.669

Brock, L. L., Rimm-Kaufmann, S. E., Nathanson, L., & Grimm, K. J. (2009). The contributions of 'hot' and 'cool' executive function to children's academic achievement, learning-related behaviors, and engagement in kindergarten. *Early Childhood Research Quarterly, 24*, 337-349. doi: 10.1016/j.ecresq.2009.06.001

Bronson, M. B. (2000). *Self-regulation in early childhood*. New York: The Guilford Press.

Brooks-Gunn, J., & Markman, L. B. (2005). The contribution of parenting to ethnic and racial gaps in school readiness. *Future of Children, 15*, 139-168

Bruggers, I., Driessen, G., & Gesthuizen, M. (2014). Voor- en vroegschoolse voorzieningen, effectief of niet? De samenhang tussen deelname aan voor-en-vroegschoolse educatie en de taal-en rekenprestaties van leerlingen op de korte en langere termijn. [Preschool education: effective or not? The relations between participation in preschool education and language and math achievement in primary school in short and long term] *Mens & Maatschappij, 89*, 1170150. doi: 10.5117/MEM2014.2.BRUG

Bryant, D. M., Burchinal, M., & Zaslow, M. (2011). Empirical approaches to strengthening the measurement of quality: Issues in the development and use of quality measures in research and applied settings. In M. Zaslow (Ed.). *Quality measurement in early childhood settings*. (pp. 33-51). Baltimore, MD: Paul H. Brookes

Bulotsky-Shearer, R. J., Fantuzzo, J. W.,& McDermott, P. A. (2008). An investigation of classroom situational dimensions of emotional and behavioral adjustment and cognitive and social outcomes for Head Start children. *Developmental Psychology, 44*,139-154. doi: 10.1037/0012-1649.44.1.139

Burchinal, M. R., Cryer, D., Clifford, R. M., & Howes, C. (2002). Caregiver training and classroom quality in child care centers. *Applied Developmental Science, 6*, 2-11. doi: 10.1207/S1532480XADS0601_01

Burchinal, M., Kainz, K., & Cai, Y. (2011). How well do our measures of quality predict child outcomes? A meta-analysis and coordinated analysis of data from large-scale studies of early childhood settings. In M. Zaslow (Ed.). *Quality measurement in early childhood settings*. (pp. 11-33). Baltimore, MD: Paul H. Brookes

Burchinal, M. R., Peisner-Feinberg, E., Pianta, R., & Howes, C. (2002). Development of academic skills from preschool through second grade: Family and classroom predictors of developmental trajectories. *Journal of School Psychology, 40*, 415-436. doi: 10.1016/S0022-4405(02)00107-3

Burchinal, M., 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, 166-176. doi: 10.1016/j.ecresq.2009.10.004

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

Byrne, B. M. (2012*). Structural equation modeling with Mplus: basic concepts, applications, and programming*. New York, London: Routledge, Taylor & Francis Classroom

Cabell, S. Q., & Downer, J. T. (2011). Improving preschoolers' language and literacy skills through web-mediated professional development. *NHSA Dialog: A Research-to-Practice Journal for the Early Childhood Field, 14*, 316-322. doi: 10.1080/15240754.2011.613130

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*, 457-482. doi: 10.1016/j.jsp.2010.09.001

Cadima, J. Leal, J., Ferreira, T, Vieira, J. M., & Matos, P. M. (2014). *Self-regulation in young children and teacher-child interactions in Portugal: testing differential effects*. Manuscript under preparation.

Calkins, S. D., Gill, K. L., Johnson, M. C., & Smith, C. L. (1999). Emotional reactivity and emotional regulation strategies as predictors of social behavior with peers during toddlerhood. *Social Development, 8*, 310-334. doi: 10.1111/1467-9507.00098

Calkins, S. D., & Williford, A. P. (2009). Taming the terrible twos: Self-regulation and school readiness. In O.A. Barbarin & B. H. Wasik (Eds.) *Handbook of child development and early education* (pp. 172-199*)*. New York: Guilford

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

Campbell, P. H., & Milbourne, S. A. (2005). Improving the quality of infant-toddler child care through professional development. *Topics in Early Childhood Special Education, 25*, 3–14. doi: 10.1177/02711214050250010101

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*, 231-242. doi: 10.1037/0012-1649.37.2.231

Carlson, S. M. (2005). Developmentally sensitive measures of executive function in preschool children. *Developmental Neuropsychology, 28*, 595-616. doi: 10.1207/s15326942dn2802_3

Carlson, S. M., & Wang, T. S. (2007). Inhibitory control and emotion regulation in preschool children. *Cognitive Development, 22*, 489-510. doi:10.1016/j.cogdev.2007.08.002

Carver, C. S., & Scheier, M. F. (2004). Self-regulation of action and affect. In R. F. Baumeister & K. D. Vohs (Eds.) *Handbook of self-regulation: Research , theory, and applications* (second edition). New York: Guilford.

Cassidy, D. J., Hestenes, L. L., Hegde, A., Hestenes, S., & Mims, S. (2005). Measurement of quality in preschool child care classrooms: An exploratory and confirmatory factor analysis of the early childhood environment rating scale-revised. *Early Childhood Research Quarterly, 20*, 345-360. doi: 10.1016/j.ecresq.2005.07.005

Centraal Bureau voor Statistiek (CBS) (2011). Emancipatiemonitor 2012. Retrieved from http://www.cbs.nl/NR/rdonlyres/0510AAAE-2465-4F69-86C4-8DD7626F5EF9/0/pb12n072.pdf at 7-5-2013.

Charlesworth, R. Hart, C. H., Burts, D. C., Thomasson, R. H.,Mosley, J., & Fleege, P. O. (1993). Measuring the developmental appropriateness of kindergarten teachers' beliefs and practices, *Early Childhood Research Quarterly, 8*, 255-276. doi: 10.1016/S0885-2006(05)80067

Chien, N. C., Howes, C., Burchinal, M., Pianta, R. C., Ritchie, S., Bryant, D. M., . . . Barbarin, O. A. (2010). Children's classroom engagement and school readiness gains in prekindergarten. *Child Development, 81*, 1534-1549. doi: 10.1111/j.1467-8624.2010.01490.x

Clements, D. H., & Sarama, J. (2007). Effects of preschool mathematics curriculum: summative research in the Building Blocks Project. *Journal for Research in Mathematics Education, 38*, 136-163.

Clifford, R. M., Barbarin, O., Chang, F., Early, D., Bryant, D., Howes, C., . . . Pianta, R. C. 2005). What is prekindergarten: Characteristics of public pre-kindergarten programs. *Applied Developmental Science*, *9*, 126–134. doi: 10.1207/s1532480xads0903_1

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*, 155-159. doi: 10.1037/0033-2909.112.1.155

Colwell, N., Gordon, R. A., Fujimoto, K., Kaestner, R., & Korenman, S. (2013). New evidence on the validity of the Arnett Caregiver Interaction Scale: Results for the Early Childhood Longitudinal Study-Birth Cohort. *Early Childhood Research Quarterly, 28*, 218-233. doi: 10.1016/j.ecresq.2012.004

Convenant Kwaliteit Kinderopvang (2008). Retrieved on 11-01-2012 from http://www.rijksoverheid.nl/onderwerpen/kinderopvang/documenten-en-publicaties/kamerstukken/2008/11/02/bijlage-convenant-kwaliteit-kinderopvang.html

CoRe (2011). *Competence Requirements in Early Childhood Education and Care: Final Report*. Brussels: European Commission, Directorate-General for Education and Culture.

Côté, S. M., Japel, C., Séguin, J. R., Mongeau, C., Xu, Q., & Tremblay, R. E. (2013). Child care quality and cognitive development: Trajectories leading to better preacademic skills. *Child Development, 84*, 752-766. doi: 10.1111/cdev.12007

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

Cuevas, K., & Bell, M. A. (2014). Infant attention and early executive function. *Child Development, 85*, 397-404. doi: 10.1111/cdev.12126

Curby, T. W., LoCasale-Crouch, J., Konold, T. R., Pianta, R. C., Howes, C., Burchinal., M., . . . Barbarin, O. (2009). The relations of observed Pre-K classroom quality profiles to children's achievement and social competence. *Early Education and Development, 20*, 346-372. doi: 10.1080/10409280802581284.

Dearing, E., McCartney, K., & Taylor, B. A. (2009). Does higher quality early child care promote low-income children's math and reading achievement in middle childhood? *Child Development, 80*, 1329-1349. doi: 10.1111/j.1467-8624.2009.01336.x

De Ayala, R. J. (2013). *Theory and practice of item response theory* (pp.1-38, 99-122). New York, NY: Guilford Press

De Goede, D., & Reezigt, G. J. (2001). *Implementatie en effecten van de Voorschool in Amsterdam*. Groningen: GION

De Haan, A., Elbers, E., Hoofs, H. & Leseman, P. (2013). Targeted versus mixed preschool and kindergartners: effects of classroom composition and teacher-managed activities on disadvantaged children's emergent academic skills. *School Effectiveness and School Improvement: An International Journal of Research, Policy, and Practice, 24*, 177-194. doi: 10.1080/0924353.2012.74792

De Haan, A., Leseman, P.P.M., & Elbers, E. (2011). Pilot gemengde groepen 2007-2010 [Pilot mixed classrooms 2007-2010]. Retrieved from http://www.utrecht.nl/images/BCD/Bestuursinformatie/publicaties/2011/rapport_pilot_gemengde_groepen.pdf Accessed June, 19th, 2013.

De Haan, D. & Singer, E. (2003). 'Use your words': A sociocultural approach to the teacher's role in the transition from physical to verbal strategies of resolving peer conflicts among toddlers, *Journal of Early Childhood Research, 1*, 95-109. doi: 10.1177/1476718X030011001

De Haan, D., & Singer, E. (2010). The relationship between young children's linguistic ability, home language, and their adaptive modifying strategies in peer conflicts. *First Language*, 30, 421-439. doi: 10.1177/0142723710370546

De Kruif, R. E. L., Riksen-Walraven, J. M., Gevers Deynoot-Schaub, M. J. J. M., Helmerhorst, K. O. W., Tavecchio, L. W. C., & Fukkink, R. G. (2009). *Pedagogogische kwaliteit van de opvang voor 0- tot 4-jarigen in Nederlandse kinderdagverblijven in 2008* [Pedagogical quality of Dutch child care for 0- to 4- years-olds in 2008]. Amsterdam/Nijmegen: NCKO

De Kruif, R. E. L., McWilliam, R. A., Maher Ridley, S., & Wakely, M. B. (2000). Classification of teachers' interaction behaviors in early childhood classroom. *Early Childhood Research Quarterly, 15*, 247-268. doi: 10.1016/S0885-2006(00)00051-X

De Schipper, E. J., Riksen-Walraven, M., & Geurts, S. A. E. (2006). Effects of child-caregiver ratio on the interactions between caregivers and children in child-care centers: An experimental study. *Child Development, 77*, 861-874. doi: 10.1111/j.1467-8624.2006.00907.x

Denham, S. A., Hamada Bassett, H., Thayer, S. K., Mincic, M. S., Sirotkin, Y. S., & Zinsser, K. (2012). Observing preschoolers' social-emotional behavior: Structure, foundations, and prediction of early school success. *The Journal of Genetic Psychology: Research and Theory on Human Development, 173*, 246-278. doi: 10.1080/00221325.2011.597457

Denham, S. A., Warren-Knot, H. K. Hamada Bassett, H., Wyatt, T., & Perna, A. (2012). Factor structure of self-regulation in preschoolers: Testing models of a field-based assessment for early school readiness. *Journal of Experimental Child Psychology, 111*, 386-404. doi: 10.1016/j.ecp.2011.10.002

Diamond, A. (2007). Interrelated and interdependent. *Developmental Science, 10*, 152-158. doi: 10.1111/j.1467-7687.2007.00578.x

Diamond, A., Barnett, W. S., Thomas, J., & Munro, J. (2007). Preschool program improves cognitive control. *Science, 318*, 1387-1388. doi: 10.1126/science.1151148

Diamond, A., & Lee, K. (2011). Interventions shown to aid executive function development in children 4 to 12 years old. *Science, 333*, 959-964. doi: 10.1126/science.1204529

Diamond, A., Prevor, M., Callender, G., & Druin, D. (1997). Prefrontal cortex cognitive deficits in children treated early and continuously for PKU. *Monographs of the Society for Research in Child Development, 62*, 1-208. doi: 10.2307/1166208

Dickinson, D. K. (2011). Teachers' Language Practices and Academic Outcomes of Preschool Children, *Science 333*, 964-967. doi: 10.1126/science.

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*, 243-260. doi: 10.1016/j.ecresq.2007.03.001

Domitrovich, C. E, Cortes, R. C., & Greenberg, M. T. (2007). Improving young children's social and emotional competence: A randomized trial of the preschool "PATHS" curriculum. *The Journal of Primary Intervention, 28*, 67-91. doi: 10.1007/s10935-007-0081-0

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*, 567-597. doi: 10.3102/0002831208328089Doolaard, S., & Leseman, P.P.M. (2008). *Versterking van het fundament. Integrerende studie n.a.v. de opbrengsten van de onderzoekslijn sociale en institutionele context van scholen uit het onderzoeksprogramma beleidsgericht onderzoek primair onderwijs 2005-2008* [Strengthening the foundation. Integrating study following the returns of the researchline social and institutional context of schools from the Research Programme Primary Education 2005-2008]. Groningen: GION

Downer, J. T., Booren, L. M., Lima, O. K., Luckner, A. E., & Pianta, R. C. (2010). The Individualized Classroom Assessment Scoring System (inCLASS): Preliminary reliability and validity of a system for observing preschoolers' competence in classroom interactions. *Early Childhood Research Quarterly, 25*, 1-16. doi: 10.1016/j.ecresq.2009.08.004

Duncan,G. J., Claessens, A., Huston, A. C., Pagani, L. S., Sexton, C. J., Magnuson, K., . . . Japel, C. (2007). School readiness and later achievement. *Developmental Psychology, 43*, 1428-1446. doi: 10.1037/0012-1649.43.6.1428

Douglas, K. (2009). Sharpening our focus in measuring classroom instruction. *Educational Researcher, 38*, 518-521. doi: 10.3102/0013X09350881

Downer, J. T., Booren, L. M., Lima, O. K., Luckner, A. E., & Pianta, R. C. (2010). The Individualized Classroom Assessment Scoring System (inCLASS): Preliminary reliability and validity of a system for observing preschoolers' competence in classroom interactions. *Early Childhood Research Quarterly, 25*, 1-16. doi: 10.1016/j.ecresq.2009.08.004

Downer, J. T., & Pianta, R. C. (2006). Academic and cognitive functioning in first grade: Associations with earlier home and child care predictors and with concurrent home and classroom experiences. *School Psychology Review, 35*, 11-30.

Downer, J., Sabol, T. J., & Hamre, B. (2010). Teacher-child interactions in the classroom: Toward a theory of within- and cross-domain links to children's developmental outcomes. *Early Education and Development, 21*, 699-723. doi: 10.1080/10409289.2010.497453.

Dunn, M., Dunn, L. M., & Schlichting, L. (2005). *Peabody Picture Vocabulary Test-III-NL*. Amsterdam: Harcourt Test Publishers

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*, 174-195. doi: 10.1016/j.ecresq.2006.04.004

Early, D., Maxwell, K. L., Burchinal, M., Alva, S., Bender, R. H., Bryant, D., … Zill, N. (2007). Teacher's education, classroom quality, and young children's academic skills: Results from seven studies of preschool programs. *Child Development, 78*, 558-580. doi: 10.1111/j.1467-8624.2007.01014.x

Egeland, B., Erickson, M. F., Clemenhagen-Moon, J., Hiester, M. K., & Korfmacher, J. (1990). *24 months tools coding manual: Project steep—revised 1990 from Mother–Child Project scales*. Minneapolis: University of Minnesota, Institute of Child Development.

Eisenberg, N. & Spinrad, T. L. (2004). Emotion-related regulation: Sharpening the definition. *Child Development, 75*, 334-339. doi: 10.1111/j.1467-8624.2004.00674.x.

Eisenberg, N., Spinrad, T. L., Fabes, R. A., Reiser, M., Cumberland, A., Shepard, S. A. . . . Murphy, B. (2004). The relations of effortful control and impulsivity to children's resiliency and adjustment. *Child Development, 75*, 25-46. doi: 10.1111/j.1467-8624.2004.00652.x

Eisenberg, N., Smith, C. L., Sadovsky, A., & Spinrad, T. L. (2004). Effortful control: Relations with emotion regulation, adjustment, and socialization. In R. F. Baumeister & K. D. Vohs (Eds.). *Handbook of self-regulation: Research, theory and applications* (pp. 259-282). New York: Guilford.

Eisenberg, N., & Sulik, M. J. (2012). Emotion-related self-regulation in children. *Teaching of Psychology, 39*, 77-83. doi: 10.1177/0098628311430172.

Elias, C.L., & Berk, L.E. (2002). Self-regulation in young children: Is there a role for sociodramatic play? *Early Childhood Research Quarterly, 17*, 216-231. doi: 0.1016/S0885-2006(02)00146-1

Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Mahwah, NJ: Lawrence Erlbaum.

Enders, C. K. (2010). *Applied missing data analysis*. New York, NY: Guilford Press.

Espy, K. A., McDiarmid, M. M., Cwik, M. F., Meade Stalets, M. M., Hamby, A., & Senn, T. (2004). The contribution of executive functions to emergent mathematic skills in preschool children. *Developmental Neuropsychology, 26*, 465-486 doi:10.1207/s15326942dn2601_6

Fabes, R. A., Eisenberg, N., Jones, S., Smith, M., Guthrie, I., Poulin, R., . . . Friedman, J. (1999). Regulation, emotionality, and preschoolers' socially competent peer interactions. *Child Development, 70*, 432-442. doi: 10.1111/1467-8624.00031

Fantuzzo, J. W., Gadsden, V. L., & McDermott, P. A. (2011). An integrated curriculum to improve mathematics, language, and literacy for Head Start children. *American Educational Research Journal, 48*, 763-793. doi: 10.3102/003831210385446

Farran, D. C., & Hofer, K. (2011). Evaluating the quality of early childhood education programs. In O. Saracho & B. Spodek (Eds.), *Handbook of Research on the Education of Young Children* (pp. 426-437). New York, NY: Routledge/Taylor & Francis.

Fuhs, M. W., & Day, J. D. (2011). Verbal ability and executive functioning development in preschoolers at Head Start. *Developmental Psychology, 47*, 404-416. doi: 10.1037/a0021065.

Fukkink, R. G., Gever Deynoot-Schaub, M. J. J. M., Helmerhorst, K. O. W., Bollen, I. , & Riksen-Walraven, J. M. A. (2013). *Pedagogische kwaliteit van de kinderopvang voor 0-4 jarigen in Nederlandse kinderdagverblijven in 2012*. [Pedagogical quality of Dutch child care for 0- to 4- years-olds in Dutch day care centers in 2012]. Amsterdam/Nijmegen: NCKO.

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

Galyer, K. T., & Evans, J. M. (2001) Pretend Play and the Development of Emotion Regulation in Preschool Children, *Early Child Development and Care, 166*, 93-108, doi: 10.1080/0300443011660108

Garon, N., Bryson, S. E., & Smith I. M. (2008). Executive function in preschoolers: A review using an integrative framework. *Psychological Bulletin, 134*, 31-60. doi: 10.1037/0033-2909.134.1.31

Gerhardstein, P., & Rovee-Collier, C. (2002). The development of visual search in infants and very young children. *Journal of Experimental Child Psychology, 81*, 194-215. doi: 10.1006/jecp.2001.2649

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

Gollob, H. F., & Reichardt, C. S. (1987). Taking account of time lags in causal models. *Child Development, 58*, 80-92.

Gordon, R. A., Fujimoto, K., Kaestner, R., Korenman, S., & Abner, K. (2013). An assessment of the validity of the ECERS-R with implications for measures of child care quality an relations for child development. *Developmental Psychology, 49*, 146-160. doi: 10.1037/a002789

Hall, J., Sylva, K., Melhuish, E., Sammons, P., Siraj-Blatchford, I., & Taggart, B. (2009). The role of pre-school quality in promoting resilience in the cognitive development of young children. *Oxford Review of Education, 35*, 331-352. doi: 10.1080/03054980902934613

Hallam, R., Fouts, H., Bargreen, K., & Caudle, L. (2009). Quality from a toddler's perspective: A bottom-up examination of classroom experiences. *Early Childhood Research & Practice, 11*, 2.

Hambleton, R. K., & Jones, R. W. (1993). Comparison of classical test theory and item response theory and their applications to test development. *Educational Measurement: Issues and Practice, 12(3),* 38-47

Hammer, C. S., Farkas, G., & Maczuga, S. (2010). The language and literacy development of Head Start children: A study using the Family and Child Experiences Survey Database. *Language, Speech, and Hearing Services in Schools, 41*, 70-83

Hamre, B. K., Hatfield, B. E., Jamil, F., & Pianta, R. C. (2013). Evidence for general and domain specific elements of teacher-child interactions: Associations with preschool children's development. *Child Development*. Advanced online publication. doi:10.1111/cdev.12184

Hamre, B. K., Justice, L. M., Pianta, R. C., Kilday, C., Sweeney, B., Downer, J. T., & Leach, A. (2010). Implementation fidelity of My Teaching Partner literacy and language activities: Association with preschoolers' language and literacy growth. *Early Childhood Research Quarterly, 25*, 329-347. doi: 10.1016/j.ecresq

Hamre, B. K., & Pianta, R. C. (2007). Learning opportunities in preschool and early elementary classrooms. In R. C. Pianta, M. J. Cox, & K. Snow (Eds.) *School readiness, early learning and the transition to kindergarten*. (pp.49-84). Baltimore: Brookes Publishing.

Hamre, B., Pianta, R., Hatfield, B., & Jamil, F. (2013). Evidence for general and domain-specific elements of teacher-child interactions: Associations with preschool children's development. *Child Development, 85*, 1257-1274. doi: 10.1111/cdev.12184

Hamre, B., Pianta, R., Downer, J., DeCoster, J., Mashburn, A. J., Jones, S. M., . . . Hamagami, A. (2013). Teaching through interactions: Testing a developmental framework of teacher effectiveness in over 4,000 classrooms. *Elementary School Journal, 113*, 461-487. doi: 10.1086/669616

Harden, B. J., Sandstrom, H., & Chazan-Cohen, R. (2012). Early Head Start and African American families: Impacts and mechanisms of child outcomes. *Early Childhood Research Quarterly, 27*, 572-581. doi: 10.1016.j.ecresq.2012.07.006

Harms, T., Clifford, R., & Cryer, D. (1998). *Early Childhood Environment Rating Scale (Rev. ed.)*. New York: Teachers College Press.

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

Heckman, J. (2006). Skill formation and the economics of investing in disadvantaged children. *Science, 312*, 1900-1902. doi: 10.1126/science.1128898

Heckman, J. J., & Masterov, D. V. (2007). *The Productivity Argument for Investing in Young Children*. IZA Discussion Paper No. 2725.

Helmerhorst, K. O.W., Riksen-Walraven, M. J., Vermeer, H. J., Fukkink, R. G., & Tavecchio, L. W. C. (2014). Measuring the interactive skills of caregivers in child care centers: Development and validation of the Caregiver Interaction Profile Scales. *Early Education and Development*, *25*, 1-21. doi: 10.1080/10409289.2014.840482

Herrera, M., O., Mathiesen, M. E., Merino, J. M., & Recart, I. (2005). Learning contexts for young children in Chile: Process quality assessment in preschool centres. *International Journal of Early Years Education, 13*, 13-27. doi: 10.1080/09669760500048253

Hill, C. D., Edwards, M. C., Thissen, D., Langer, M. M., Wirth, M. A., . . . Varni, J. W. (2007). Practical issues in the application of item response theory. *Medical Care, 45*, 39-47.

Hoff, E. (2006). How social contexts support and shape language development. *Developmental Review, 26*, 55–88. doi:10.1016/j.dr.2005.11.002

Hofmann, W., Schmeichel, B. J., & Baddeley, A. D. (2012). Executive functions and self-regulation. *Trend in Cognitive Sciences, 16*, 174-180. doi: 10.1016/j.tics.2012.01.006

Holmes, J., Gathercole, S.E., & Dunning, D.L. (2009). Adaptive training leads to sustained enhancement of poor working memory in children. Developmental Science, 12, F9-F15. doi: 10.1111/j.1467-7687.2009.00848.x

Hongwanishkul, D., Happaney, K. R., Lee, W. S. C., & Zelazo, P. D. (2005). Assessment of hot and cool executive function in young children: Age-related changes and individual differences. *Developmental Neuropsychology, 28*, 617-644. doi: 10.1207/s15326942dn2802_4

Howes, C., James, J. & Ritchie, R. (2003). Pathways to effective teaching. *Early Childhood Research Quarterly, 18*, 104-120. doi: 10.1016/S0885-2006(03)00008-5

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*, 27-50. doi: 10.1016/j.ecresq.2007.05.002

Howse, R. B., Calkins, S. D., Anastopoulos, A. D., Keane, S. P., & Shelton, T. L. (2003). Regulatory contributors to children's kindergarten achievement. *Early Education & Development, 14*, 101-119. doi: 10.1207/s15566935eed1401_7

Hox, J. J. (2010). *Multilevel analysis. Techniques and applications. 2nd Edition*. New York: Routledge

Hsieh, W.-Y., Hemmeter, M. L., McCollum, J. A., & Ostrosky, M. M. (2009). Using coaching to increase preschool teachers' use of emergent literacy teaching strategies. *Early Childhood Research Quarterly, 24*, 229-247. doi: 10.1016/j.ecresq.2009.03.007

Hyson, M., Vick Witthaker, J. E., Zaslow, M., Leong, D., Bodrova, Hamre, B., & Smith, S. (2011). Measuring the quality of environmental supports for young children's children's social and emotional competence. In M. Zaslow (Ed.). *Quality measurement in early childhood settings*. (pp. 105-135). Baltimore, MD: Paul H. Brookes

Jordan, N. C., & Levine, S. C. (2009). Socioeconomic variation, number competence, and mathematics learning difficulties in young children. *Developmental Disabilities Research Reviews, 15,* 60-68. doi: 10.1002/ddrr.46

Justice, L. M., Mashburn, A., Pence, K. L., & Wiggins, A. (2008). Experimental evaluation of a preschool language curriculum: Influence on children's expressive language skills. *Journal of Speech, Language and Hearing Research, 51*, 983-10001

Klibanoff, R. S., Levine, S. C., Huttenlocher, J., Vasilyeva, M., & Hedges, L. H. (2006). Preschool children's mathematical knowledge: The effect of teacher "math talk". *Developmental Psychology, 42*, 59-69. doi: 10.1037/0012-1649.42.1.59

Karssen, A.M., Veen, I. van der, Veen, A., Daalen, M.M. van, & Roeleveld, J. (2013). *Effecten van deelname aan en kwaliteit van voor- en vroegschoolse educatie op de ontwikkeling van kinderen*. Amsterdam: Kohnstamm Instituut. Rapport 894

Kenny, D. A., Kaniskan, B., & McCoach, D. B. (2014). The performance of RMSEA in models with small degrees of freedom. *Sociological Methods & Research*, in press

Kline, R. B. (2005). *Principles and practice of structural equation modeling* (2nd ed.). New York: Guilford Press.

Kochanska, G., Murray, K., & Coy, K. C. (1997). Inhibitory control as a contributor to conscience in childhood: From toddler to early school age. *Child Development, 68*, 263-277. doi:10.1111/j.1467-8624.1997.tb01939.x

Kochanska, G., Murray, K., Jacques, T. Y., Koenig, A. L., & Vandergeest, K. A. (1996). Inhibitory control in young children and its role in emerging internalization. *Child Development, 67*, 490-507. doi:10.1111/j.1467-8624.1996.tb01747.x

Kochanska, G., Murray, K. T., & Harlan, E. T. (2000). Effortful control in early childhood: Continuity and change, antecedents, and implications for social development. *Developmental Psychology, 36*, 220-232. doi:10.1037//0012-1649.36.2.220

Kopp, C. B. (1982). Antecedents of self-regulation: a developmental perspective. *Developmental Psychology, 18*, 199-214.

Kuger, S., & Kluczniok, K. (2008). Prozessqualität im Kindergarten – Konzept, Umsetzung und Befunde. [Process quality in kindergarten – Concept, implementation and findings] *Zeitschrift für Erziehungswissenschaft, special issue, 11*, 159-178. doi: 10.1007/978-3-531-91452-7_11

Lambert, R. G., Nelson, L., Brewer, D., & Burchinal, M. R. (2006). Measurement issues and psychometric methods in developmental research. *Monographs of the Society for Research in Child Development, 71*, 24-41. doi: 10.1111/j.1540-5834.2006.00403.x

Landry, S. H., Swank, P. R., Smith, K. E., Assel, M. A., & Gunnewig, S. B. (2006). Enhancing early literacy skills for preschool children: Bringing a professional development model to scale. *Journal of Learning Disabilities, 39*, 306-324. doi: 10.1177/00222194060390040501

Landry, S. H., Zucker, T. A., Taylor, H. B., Swank, P. R., Williams, J. M., Assel. M., . . . Klein, A. (2014). Enhancing early child care quality and learning for toddlers at risk: The Responsive Early Childhood Program. *Developmental Psychology, 50*, 526-541. doi: 10.1037/a0033494

La Paro, L. M., Pianta, R. C., & Stuhlman, M. (2004). Classroom Assessment Scoring System. Findings from the pre-K year. *Elementary School Journal, 105*, 409–426. doi: 10.1086/499760

La Paro, K. M., Hamre, B. K. , Locasale-Crouch, J. , Pianta, R. C. , Bryant, D. , Early, D., . . . & Burchinal, M. (2009). Quality in Kindergarten Classrooms: Observational Evidence for the Need to Increase Children's Learning Opportunities in Early Education Classrooms. *Early Education & Development, 20*, 657-692 doi: 10.1080/10409280802541965

La Paro, K. M., Hamre, B. K., & Pianta, R. C. (2011). *Classroom Assessment Scoring System Toddler Manual*. Charlottesville, VA: Teachstone.

La Paro, K. M., Williamson, A. C., & Hatfield, B. (2014). Assessing quality in toddler classrooms using the CLASS-Toddler and the ITERS-R. *Early Education and Development, 25*, 875-893. doi: 10.1080/10409289.2014.883586

Layzer, J. I., & Goodson, B. D. (2006). The "quality" of early care and education settings: definitional and measurement issues. *Evaluation Review, 30*, 556-576. doi: 10.1177/0193841X06291524

Leach, P., Barnes, J, Malmberg, L.-E., Sylva, K., Stein, A., & FCCC team (2008). The quality of different types of child care at 10 and 18 months: A comparison between types and factors related to quality. *Early Child Development and Care, 178*, 177-209. doi: 10.1080/03004430600722655

Leseman P. (2012). Preschool and learning-related skills. Melhuish E, topic ed. In: Tremblay RE, Boivin M, Peters RDeV, eds. *Encyclopedia on Early Childhood Development* [online]. Montreal, Quebec: Centre of Excellence for Early Childhood Development and Strategic Knowledge Cluster on Early Child Development; 2012:1-8. Available at: http://www.child- encyclopedia.com/documents/LesemanANGxp1.pdf. Accessed [1-6-2014].

Leseman, P.P.M., Rollenberg, L., & Rispens, J. (2001). Playing and working in kindergarten: Cognitive co-construction in different educational situations. *Early Childhood Research Quarterly, 16*, 3, 363-384

Leseman, P. P. M., & Slot, P. L. (2013). *Kwaliteit en curriculum van voorschoolse opvang en educatie in Nederland* [Quality and curriculum of early childhood education and care in the Netherlands]. Utrecht: the Netherlands: Utrecht University Department of Special Education

Lillard, A. S., Lerner, M. D., Hopkins, E. J., Dore, R. A., Smith, E. D., & Palmquist, C. M. (2013). The impact of pretend play in children's development: A review of the evidence. *Psychological Bulletin, 139*, 1-34. doi: 10.1037/a0029321

Lindsey, E. W., & Colwell, M. J. (2003). Preschoolers' emotional competence: Links to pretend play and physical play. *Child Study Journal, 33*, 39-52.

LoCasale-Crouch, J., Kraft-Sayre, M., Pianta, R. C., Hamre, B. K., Downer, J. T., Leach., A., . . . Scott-Little, C. (2011). Implementing an early childhood professional development course across 10 sites and 15 sections: Lessons learned. *NHSA Dialog, 14*, 275-292. doi: 10.1080/15240754.2011.617527

Lonigan, C. J., Farver, J. M., Philips, B. M., & Clancy-Menchetti, J. (2011). Promoting the development of preschool children's emergent literacy skill: a randomized evaluation of a literacy-focused curriculum and two professional development models. *Reading and Writing, 24*, 305-337. doi: 0.1007/s11145-009-9214-6

Love, J. M., Harrison, L., Sagi-Schwartz, A., van IJzendoorn, M.H., Ross, C., Ungerer, J. A., . . . Chazan-Cohen, R. (2003). Child care quality matters: How conclusions may vary with context, *Child Development, 74*, 1021-1033. doi: 0009-3920/2003/7404-0004

Magnuson, K., Lahaie, C., & Waldfogel, J. (2006). Preschool and school readiness of children of immigrants. *Social Science Quarterly, 87*, 1241-1262.

Magnuson, C. A., Meyers, M. K., Ruhm, C., & Waldfogel, J. (2004). Inequality in preschool education and school readiness. *American Educational Research Journal, 20*, 115-157. doi: 10.3102/00028312041001115

Magnuson, C. A., Ruhm, C., & Waldfogel, J. (2007). Does prekindergarten improve school preparation and performance? *Economics of Education Review, 26*, 33-51. doi: 0.1016/j.econedurev.2005.09.008

Martinez-Beck, I. (2011). Introduction: Why strengthening the measurement of quality in early childhood settings has taken on new importance In M. Zaslow (Ed.). *Quality measurement in early childhood settings*. (pp. xviii-xxiv). Baltimore, MD: Paul H. Brookes

Marulis, L. M., & Neuman, S. B. (2010). The effects of vocabulary intervention on young children's word learning. *Review of Educational Research, 80*, 300-335. doi: 10.3102/0034654310377087

Mashburn, A. J., Downer, J. T., Hamre, B., Justice, L. M., & Pianta, R. C. (2010). Consultation for teachers and children's language and literacy development during pre-kindergarten. *Applied Developmental Science, 14*, 179-196. doi: 10.1080/10888691.2010.516187

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*, 732-749. doi: 10.1111/j.1467-8624.2008.01154.x

Matthews, S. B. (2008). The relationship among self-regulation, sociodramatic play, and preschoolers' readiness for kindergarten. *Counseling Psychology Dissertations.* Paper 8. http://hdl.handle.net/2047/d10017045

Maydeu-Olivares, A. (2013) Goodness-of-Fit assessment of Item Response Theory models. *Measurement: Interdisciplinary Research and Perspectives, 11*, 71-101, doi: 10.1080/15366367.2013.831680

Maydeu-Olivares, A., Cai, L., & Hernandez, A. (2011). Comparing the fit of item response theory and factor analysis models. *Structural Equation Modeling, 18*, 333-356. doi: 10.1080/10705511.2011.581993.

McCabe, P., & Altamura, M. (2011). Empirically valid strategies to improve social and emotional competence of preschool children. *Psychology in the Schools, 48*, 513-540. doi: 10.1002/pits.20570

McClelland, M. M., Acock, A. C., & Morrison, F. J. (2006). The impact of kindergarten learning-related skills on academic trajectories at the end of elementary school. *Early Childhood Research Quarterly, 21*, 471-490. doi: 10.1016/j.ecresq.2006.09.003

McClelland, M. M., & Cameron, C. E. (2012). Self-regulation in early childhood: Improving conceptual clarity and developing ecologically valid measures. *Child Development Perspectives, 6,* 136-142. doi: 10.1111/j.1750-8606.2011.00191.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*, 947-959. doi: 10.1037/0012-1649.43.4.947

McClelland, M. M., Morrison, F. J., & Holmes, D. L. (2000). Children at risk for early academic problems: The role of learning-related social skills. *Early Childhood Research Quarterly, 15*, 307-329. doi: 10.1016/S0885-2006(00)00069-7

McClelland, M. M., Ponitz, C. C., Messersmith, E., & Tominey, S. (2010). Self-regulation: The integration of cognition and emotion. In R. Lerner (Series Ed.) & W. Overton (Vol. Ed.), *Handbook of Lifespan human development* (pp. 509-553). Hoboken, NJ: Wiley and Sons.

Melby-Lervåg, M., & Hulme, C., (2013). Is working memory training effective? A meta-analytic review. *Developmental Psychology , 49*, 270–291. doi: 10.1037/a0028228

Melhuish, E. (2010). Why children, parents and home learning are important. In K. Sylva, E. Melhuish, P. Sammons, I. Siraj-Blatchford & B. Taggart (Eds.), *Early Childhood Matters: Evidence from the Effective Pre-school and Primary Education Project* (pp. 44-69). London: Routledge.

Melhuish, E. C. (2011). Preschool matters. *Science, 333*, 299-300. doi: 10.1126/science.1209459

Melhuish, E., Quinn, L., Sylva, K., Sammons, P., Siraj-Blatchford, I., & Taggart, B. (2013). Preschool affects longer term literacy and numeracy: results from a general population longitudinal study in Northern Ireland. *School Effectiveness and School Improvement: An International Journal of Research, Policy and Practice, 24*, 234-250. doi: 10.1080/09243453.2012.749796

Merritt, E. G., Wanless, S. B., Rimm-Kaufman, S. E., Cameron, C., & Pugh, J. L. (2012). The contribution of teachers' emotional support to children's social behaviors and self-regulatory skills in first grade. *School Psychology Review, 41*, 141-159.

Mistry, R., Benner, A. D., Biesanz, J. C., Clark, S. L., & Howes, C. (2010). Family and social risk, and parental investments during the early childhood years as predictors of low-income children's school readiness outcomes. *Early Childhood Research Quarterly, 25*, 432-449. doi: 10.1016/j.ecresq.2010.01.002

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., & Howerter, A. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive Psychology, 41*, 49-100. doi: 10.1006/cogp.1999.0734

Mogroep (2012). Update toekomstscenario integratie peuterwerk en dagopvang. Retrieved from http://www.mogroep.nl/categorieen/thema/peuterspeelzaalwerk/nieuws/2310-ed-buitenhek-schetst-scenario-van-toekomst-peuterspeelzalen on 7-8-2013

Morris, P., Millenky, M., Raver, C., & Jones, S. (2013). Does a preschool social and emotional learning intervention pay off for classroom instruction and children's behavior and academic skills? Evidence from the Foundations of Learning Project. *Early Education and Development, 24*, 1020-1042. doi: 10.1080/10409289.2013.825187

Morrison, F. J., Ponitz, C. C., & McClelland, M. M. (2010). Self-regulation and academic achievement in the transition to school. In M. Posner (Series Ed.) & S. Calkins & M. Bell (Vol. Eds.), *The developing brain: Development at the intersection of emotion and cognition* (pp. 203-224). Washington, D. C.: American Psychological Association.

Mulder, H., Hoofs, H., Verhagen, J., van der Veen, I., & Leseman, P. P. M. (2014). Psychometric properties and convergent and predictive validity of an executive function test battery for two-year-olds. *Frontiers in Developmental Psychology*. Advance online publication. doi: 10.3389/fpsyg.2014.0073.

Mulder, H., Verhagen, J., & Leseman, P.P.M. (2012). *Executieve functietaken.* In: Pre-COOL cohortonderzoek. Technisch rapport tweejarigencohort, eerste meting 2010 – 2011 [Executive function tasks. In: Pre-COOL cohort study. Technical report two-year-olds cohort, first measurement 2010-2011]. Pre-COOL consortium. Amsterdam, Kohnstamm Instituut.

Muthén, L. K. , & Muthén, B. O. (1998-2012). *Mplus User's Guide*. Seventh Edition. Los Angeles, CA: Muthén & Muthén

Muthén, B. O., & Satorra, A. (1989). Multilevel aspects of varying parameters in structural models. In R. D. Bock (Ed.), *Multilevel analysis of educational data* (pp.87–99). San Diego, CA: Academic Press.

Muthén, B., & Satorra, A. (1995). Complex sample data in structural equation modeling. *Sociological Methodology, 25*, 267-316. doi: 10.2307/271070

Nader-Grosbois N., & Vieillevoye S. (2012). Variability of self-regulatory strategies in children with intellectual disability and typically developing children in pretend play situations. *Journal of Intellectual Disability Research, 56*, 140-156. doi: 10.1111/j.1365-2788.2011.01443.x

Nederlands Consortium Kinderopvang Onderzoek (NCKO)[Dutch Consortium for Child Care Research] (2011). *Pedagogische kwaliteit van de kinderopvang en de ontwikkeling van jonge kinderen: een longitudinale studie* [Pedagogical quality of Dutch child care en de development of young children: a longitudinal study]. Amsterdam/Nijmegen: NCKO

NICHD Early Child Care Research Network. (2000a). Characteristics and quality of child care for toddlers and preschoolers. *Applied Developmental Science, 4*, 116-135.

NICHD Early Child Care Research Network (2000b). The relation of child care to cognitive and language development. *Child Development, 71*, 960

NICHD Early Child Care Research Network (2002). Child-care structure -> Process -> Outcome and indirect effects of child-care quality on young children's development, *Psychological Science, 13*, 199-206.

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*, 99-116. doi: 10.1037/0003-066X.61.2.99

NICHD Early Child Care Research Network & Duncan, G. J. (2003). Modeling the impacts of child care quality on children's preschool cognitive development. *Child Development, 74*, 1454-1475. doi: 10.1111/1467-8624.00617

Nix, R. L., Bierman, K. L., Domitrovich, C. E., & Gill, S. (2013). Promoting children's social-emotional skills in preschool can enhance academic and behavioral functioning in kindergarten: Findings from Head Start REDI. *Early Education and Development, 24*, 1000-1019. doi: 10.1080/10409289.2013.825565

Nores, M., Belfield, C., & Barnett, W. S. (2005). Updating the economic impacts of High/Scope Perry Preschool Program. *Educational Evaluation and Policy Analysis, 27*, 245-261.

Nap-Kolhoff, E., Schilt-Mol, T. van, Simons, M., Sontag, L., Steensel, R. van, & Vallen, T. (2008). *VVE onder de loep. Een studie naar de uitvoering en effectiviteit van voor- en vroegschoolse educatieve programma's.* Tilburg: IVA.

Onderwijsraad (2008). *Een rijk programma voor ieder kind* [A rich program for every child]. Den Haag, the Netherlands: Onderwijsraad.

Organisation for Economic Cooperation and Development. (2006). *Starting strong II: Early childhood education and care*. Paris: France: Author.

Organisation for Economic Cooperation and Development (2010). Family Database Retrieved from http://www.oecd.org/social/soc/oecdfamilydatabase.htm on 06-20-2014

Organisation for Economic Cooperation and Development (2012). *Starting Strong III: A Quality Toolbox for Early Childhood Education and Care*: OECD Publishing.

Organisation for Economic Cooperation and Development (2013). *Education at a Glance 2013: OECD Indicators. Enrolment rates, by age (2011): Full-time and part-time students in public and private institutions*. doi:10.1787/eag-2013-table130-en

Oudgenoeg-Paz, O., Boom, J., Volman, M.J.M, & Leseman, P.P.M. (submitted manuscript). *Development of exploration of spatial-relational object properties in the second and third year of life*.

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 & Development, 21*, 95-124. doi: 10.1080/10409280902858764

Peisner-Feinberg, E. S., Burchinal, M., R., Clifford, R., M., Culkin, M. L., Howes., C., Kagan, S. L., & Yazejian, N. (2001). The relation of preschool child-care quality to children's cognitive and social developmental trajectories through second grade. *Child Development, 72*, 1534-1553. doi: 10.1111/1467-8624.00364

Pelphrey, K. A., Reznick, J. S., Goldman, B. D., Sasson, N., Morrow, J., Donahoe, A., & Hodgson, K. (2004). Development of visuospatial short-term memory in the second half of the 1st year of life. *Developmental Psychology*, *40*, 836

Perlman, M., Zellman, G. L., & Le, V.-N. (2004). Examining the psychometric properties of the Early Childhood Environmental Rating Scale-Revised (ECERS-R). *Early Childhood Research Quarterly, 19*, 398-412. doi: 10.1016/j.ecresq.07.006

Pessanha, M., Aguiar, C., & Bairrao, J. (2007). Influence of structural features on Portuguese toddler child care quality. *Early Childhood Research Quarterly, 22*, 204-214. doi: 10.1016/j.ecresq.2007.02.003

Philips, D. A., Gormley, W. T., & Lowenstein, A. E. (2009). Inside the pre-kindergarten door: Classroom climate and instructional time allocation in Tulsa's pre-k programs. *Early Childhood Research Quarterly, 24*, 213-228. doi: 10.1016/j.ecresq.2009.05.002

Philips, D., Mekos, D., Scarr, S., McCartney, K., & Abbott-Shim, M. (2000). Within and beyond the classroom door: assessing quality in child care centers. *Early Childhood Research Quarterly, 15*, 475-496.

Philips, D. A., & Lowenstein, A. E. (2011). Early care, education, and child development. *Annual Review of Psychology, 62*, 483-500.

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*, 281-303. doi: 10.1016/S0885-2006(97)90004-1

Pianta, R. C., Barnett, W. S., Burchinal, M., & Thornburg, K. R. (2009). The effects of preschool education: What we know, how public policy is or is not aligned with the evidence base, and what we need to know. *Psychological Science, 10*, 49-88. doi: 10.1177/1529100610381908

Pianta, R. C., & Hamre, B. K. (2009). Conceptualization, measurement, and improvement of classroom processes: Standardized observation can leverage capacity. *Educational Researcher, 38*, 109-119. doi: 10.3102/0013189X09332374

Pianta, R. C., Hamre, B., & Downer, J. (2011). Aligning measures of quality with professional development goals and goals for children's development. In M. Zaslow (Ed.). *Quality measurement in early childhood settings*. (pp. 297-317). Baltimore, MD: Paul H. Brookes

Pianta, R., Howes, C., Burchinal, M., Bryant, D., Clifford, R. Early, D., . . . Barbarin, O. (2005). Features of Pre-Kindergarten programs, classrooms, and teachers: Do they predict observed classroom quality and child-teacher interactions? *Applied Developmental Science, 9*, 144-159. doi: 10.1207/s1532480xads0903_2

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

Pianta, R., Mashburn, A., Downer, J., Hamre, B., & Justice, L. (2008). Effects of web-mediated professional development resources on teacher-child interactions in pre-kindergarten classrooms. *Early Childhood Research Quarterly*, *23*, 431-451 doi: 10.1016/j.ecresq.2008.02.001

Pintrich, P. R. (2002). The role of metacognitive knowledge in learning, teaching, and assessing. *Theory into Practice, 41*, 219-225. doi: 10.1207/s15430421tip4104_3

Preschool Curriculum Evaluation Research Consortium. (2008). *Effects of preschool curriculum programs on school readiness* (NCER 2008-2009). Washington, DC: Government Printing Office.

Ponitz, C. C., McClelland, M. M., Matthews, J. S., & Morrison, F. J. (2009). A structured observation of behavioral self-regulation and its contribution to kindergarten outcomes. *Developmental Psychology, 45*, 605-619. doi: 0.1037/a0015365

Powell, D. R., Burchinal, M. R., File, N., & Kontos, S. (2008). An eco-behavioral analysis of children's engagement in urban public preschool classrooms. *Early Childhood Research Quarterly, 23*, 108-123. doi: 10.1016/j.ecrsq.2007.04.001

Ramey, C.T., & Ramey, S.L. (2004). Early learning and school readiness: Can early intervention make a difference? *Merill-Palmer Quarterly, 50*, 471-491.

Raver, C. C., Smith Carter, J., Charles McCoy, D., Roy, A., Ursache, A., & Friedman, A. (2012). Testing models of children's self-regulation within educational contexts: implications for measurement. *Advances in Child Development and Behavior, 42*, 245-270. doi: 10.1016/B978-0-12-394388-0.00007-1

Reise, S. P., Widaman, K. F., Pugh, R. H. (1993). Confirmatory factor analysis and item response theory: Two approaches for exploring measurement invariance. *Psychological Bulletin, 114*, 552-556.

Rhoades, B. L., Warren, H. K., Domitrovich, C. E., & Greenberg, M. T. (2011). Examining the link between preschool social-emotional competence and first-grade academic achievement: The role of attention. *Early Childhood Research Quarterly 26*, 182-191. doi: 10.1016/j.ecresq.2010.07.003

Riggs, N. R., Jahromi, L. B., Razza, R. P., Dillworth-Bart, J. E., & Mueller, U. (2006). Executive function and the promotion of social-emotional competence. *Journal of Applied Psychology, 27*, 300-309. doi: 10.1016/j.appdev.2006.04.002

Rimm-Kaufman, S. E., Curby, T. W., Grimm, K. J., Nathanson, L. & Brock, L. L. (2009). The contribution of children's self-regulation and classroom quality to children's adaptive behaviors in the kindergarten classroom. *Developmental Psychology, 45*, 985-972. doi:

Rose, S. A., Feldman, J. F., Jankowski, J. J., & van Rossem, R. (2008.) A cognitive cascade in infancy: Pathways from prematurity to later mental development. Intelligence, 36, 367-378. doi:10.1016/j.intell.2007.07.003

Rothbart, M. K., Ahadi, S., & Evans, D. E. (2000). Temperament and personality: Origins and outcomes. *Journal of Personality and Social Psychology, 78*, 122-135. doi: 10.1037//0022-3514,78.1.122

Sabol, T. J., Soliday Hong, S. L., Pianta, R. C., & Burchinal, M. R. (2013). Can rating pre-k programs predict children's learning? *Science, 341*, 845-846. doi: 10.1126/science.1233517

Scerif, G., Cornish, K., Wilding, J., Driver, J., & Karmiloff-Smith, A. (2004). Visual search in typically developing toddlers and toddlers with Fragile X or Williams Syndrome. *Developmental Science, 7*, 116-130.

Schneider, W., Eschman, A., & Zuccolotto, A. (2002) *E-Prime User's Guide*. Pittsburgh: Psychology Software Tools Inc.

Schweinhart, L. J., & Weikart, D. P. (1997). The High/Scope preschool curriculum comparison study through age 23. *Early Childhood Research Quarterly, 12*, 117-143.

Sektnan, M., McClelland, M. M., Acock, A., & Morrison, F. J. (2010). Relations between early family risk, children's behavioral regulation, and academic achievement. *Early Childhood Research Quarterly, 25*, 464-479. doi: 10.1016/j.ecresq.2010.02.005

Shonkoff J. P., & Philips, D. (2000). *From neurons to neighborhoods. Committee on Integrating the Science of Early Childhood Development*. National Research Council and Institute of Medicine. Washington, DC: National Academy Press.

Slot, P.L., Leseman, P. P. M., Mulder, H., & Verhagen, J. (2013). *Handleiding CLASS Toddler* [Manual CLASS Toddler]. Utrecht: Universiteit Utrecht, Onderwijsadvies en Training.

Slot, P. L., Leseman, P. P. M., Verhagen, J., & Mulder, H. (under review). Associations between structural quality aspects and process quality in Dutch early childhood education and care settings. *Submitted to Early Childhood Research Quarterly*

Smilansky, S., & Shefatya, L. (1990). Facilitating play. *A medium for promoting cognitive, socio-emotional and academic development in young children*. Silver Spring, Maryland: PS&E Publications.

Squires, J., Potter, L., & Bricker, D. (1999). The ASQ user's Sylva, K., Melhuish, E., Sammons, P., Siraj-Blatchford, I., & Taggart, B. (2011). Pre-school quality and educational outcomes at age 11: Low quality has little benefit. *Journal of Early Childhood Research, 9,* 109-124. doi: 10.1177/14767X10387900

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

Sylva, K., Taggart, B., Siraj-Blatchford, I., Totsika, V., Ereky-Stevens, K., Gilden, R., . . . Bell, D. (2007). Curricular quality and day-to-day learning activities in pre-school. *International Journal of Early Years Education, 15*, 49-65. doi: 10.1080/09669760601106968

Tabachnik, B. G., & Fidell, L. S. (2007). *Using multivariate statistics* (5th ed.). Boston, MA: Allyn and Bacon.

Thissen, D., & Steinberg, L. 1997. A response model for multiple-choice items. In W. J. van der Linden & R. K. Hambleton (Eds.) *Handbook of modern item response theory* (pp. 51–66). New York, NY: Springer Verlag.

Thomason, A. C., & La Paro, K. M. (2009). Measuring the quality of teacher-child interactions in toddler child care. *Early Education & Development, 20*, 285-304. doi: 10.1080/10409280902773351

Thorell, L.B., Lindquist, S., Bergman-Nutley, S., Bohlin, G ., & Klingberg, T. (2009). Training and transfer effects of executive functions in preschool children. *Developmental Science, 12*, 106-113.

Tomasello, M. (2003). *Constructing a Language: A Usage-Based Theory of Language Acquisition.* Cambridge, Massachusetts: Harvard University Press

Tout, K., Zaslow, M., & Berry, D. (2006). Quality and qualifications: links between professional development and quality on early care and educations settings. In M. Zaslow & I. Martinez-Beck (eds.*), Critical issues in early childhood professional development* (pp. 77-110). Baltimore: Paul H. Brookes Publishing Co.

Ursache, A., Blair, C., & Raver, C. C. (2012). The promotion of self-regulation as a means of enhancing school readiness and early achievement in children at risk for school failure. *Child Development Perspectives, 6*, 122-128. doi:10.1111/j.1750-8606.2011.00209.x

Valloton, C., & Ayoub, C. (2011). Use your words: The role of language in the development of toddlers' self-regulation. *Early Childhood Research Quarterly, 26*, 169-181. doi: 10.1016/j.ecresq.2010.09.002

Vandell, D. L. (2004). Early child care: The known and the unknown. *Merril-Palmer Quarterly, 50*, 387-414.

Van der Ven, S. H. G., Kroesbergen, E. H., Boom, J., & Leseman, P. P. M. (2011). The development of executive functions and early mathematics: A dynamic relationship. *British Journal of Educational Psychology, 82*, 100-119. doi: 10.1111/j.2044-8279.2011.02035.x

Vandell, D. L., Belsky, J., Burchinal, M., Steinberg, L., Vandergrift, N., & NICHD ECCRN (2010). Do Effects of Early Child Care Extend to Age 15 Years? Results From the NICHD Study of Early Child Care and Youth Development, *Child Development, 81*, 737-756. doi : 10.1111/j.1467-8624.2010.01431.x

Van IJzendoorn, M. H., Dijkstra, J., & Bus, A. G. (1995). Attachment, intelligence, and language: A meta-analysis. *Social Development, 4*, 115-128. doi: 10.1111/j.1467-9507.1995.tb00055.x

Van Tuijl, C., & Leseman, P.P.M., (2007). Increases in the verbal and fluid cognitive abilities of disadvantaged children attending preschool in The Netherlands, *Early Childhood Research Quarterly, 22*, 188-203. doi:10.1016/j.ecresq.2007.02.002

Veen, A., J. Roeleveld, & P.P.M. Leseman (2000). *Evaluatie van Kaleidoscoop en Piramide – eindrapportage 2000*. Amsterdam: SCO-Kohnstamm Instituut

Veen, A., Veen, I., Heurter, A., Ledoux, G., Mulder, L., Paas, T., Slot, P. L. 2012). *Pre-COOL cohortonderzoek. Technisch rapport tweejarigencohort, eerste meting 2010-2011*. [Pre- COOL cohort study. Technical report two-year-olds cohort, first measurement 2010-2011]. Amsterdam: Kohnstamm Instituut

Verhagen, J., & Mulder, H. (2011). *Testinstructie voor de testleiders bij het cohortonderzoek pre-COOL: Periode 2011-2012*. [Test instruction for test assistants of the pre-COOL cohort study: Period 2011-2012].

Verhagen, J., De Bree, E., Mulder, H., & Leseman, P. P. M. (under review). *Effects of Vocabulary and Phonotactic Probability on Two-Year-Olds' Nonword Repetition* submitted manuscript.

Vermeer, H. J., van IJzendoorn, M. H., de Kruif, R. E. L., Fukkink, R. G., Tavecchio, L. W. C., Riksen-Walraven, J. M., & van Zeijl, J. (2008). Child Care in the Netherlands: trends in Quality over the years 1995-2005. *The Journal of Genetic Psychology, 169*, 360-385. doi: 10.3200/GNTP.169.4.360-385

Vicari, S., Caravale, B., Carlesimo, G., Casadei, A., & Allemand, F. (2004). Spatial working memory deficits in children at ages 3-4 who were low birth weight, preterm infants. *Neuropsychology, 18*, 673-678. doi: 10.1037/0894-4105.18.4.673

Vieillevoye S., & Nader-Grosbois N. (2008b) Self-regulation during pretend play in children with intellectual disability and in normally developing children. *Research in Developmental Disabilities, 29*, 256–72. doi: 10.1016/j.ridd.2007.05.003

Von Suchodoletz, A., Fäsche, A., Gunzenhauser, C., & Hamre, B. K. (under review). *A typical morning in preschool: Observations of teacher-child interactions in German preschools*. Submitted manuscript

Votruba-Drzal, E., Coley, R. L., & Chase-Lansdale, P. L., (2004). Child care and low-income children's development: Direct and moderated effects. *Child Development, 75*, 296-312. doi: 10.1111/j.1467-8624.2004.00670.x

Vygotsky L. S. (1967) Play and its role in the mental development of the child, *Soviet Psychology, 5*, 6–18.

Walston, J., & West, J. (2004). *Full-day and half-day kindergarten in the United States. Findings from the Early Childhood Longitudinal Study, Kindergarten Class of 1998-99* (NCES 2004-078). Washington, DC: National Center for Education Statistics

Watamura, S. E., Morrisey, T. W., Philips, D. A., McCartney, K., & Bub, K. (2011). Double jeopardy: Poorer social-emotional outcomes for children in the NICHD SECCYD experiencing home and child-care environments that confer risk. *Child Development, 82*, 48-65. doi: 10.1111/j.1467-8624.2010.01540.x.

Weiland, C., Ulvestad, K., Sachs, J., & Yoshikawa, H. (2013). Associations between classroom quality and children's vocabulary and executive function skills in an urban prekindergarten program. *Early Childhood Research Quarterly, 28*, 199-209. doi: 10.1016/j.ecresq.2012.12.002

Weir, R. C., & Gjerde, P. F. (2002). Preschool personality prototypes: Internal coherence, cross-study replicability, and developmental outcomes in adolescence, *Personality and Social Psychology Bulletin, 28*, 1229-1241. doi: 10.1177/01461672022812008

Welsh, J. A., Nix, R. L., Blair, C., Bierman, K. L., & Nelson, K. E. (2010). The development of cognitive skills and gains in academic school readiness for children from low-income families. *Journal of Educational Psychology, 102*, 43-53. doi: 10.1037/a0016738

Wen, X., Leow, C., Hahs-Vaughn, D., Korfmacher, J., & Marcus, S. M. (2012). Are two years better than one year? A propensity score analysis of the impact of Head Start program duration on children's performance in kindergarten. *Early Childhood Research Quarterly, 27*, 684-694. doi: 10.1016/j.ecresq.2011.07.006

Whitebread, D., Bingham, S., Grau, V., Pino Pasternak, D., & Sangster, C. (2007). Development of metacognition and self-regulated learning in young children: Role of collaborative and peer-assisted learning. *Journal of Cognitive Education and Psychology, 6*, 433-455. doi: 10.1891/194589507787382043

Whitebread, D., Coltman, P., Pino Pasternak, D., Sangster, C., Grau, V., Bingham, S., Almeqdad, Q., & Demetriou, D. (2009). The development of two observational tools for assessing metacognition and self-regulated learning in young children. *Metacognition Learning, 4*, 63-85. doi: 10.1007/s11409-9033-1.

Whitebread, D., & Sullivan, L. O. (2012). Preschool children's social pretend play: Supporting the development of meta-communication, metacognition and self-regulation. *International Journal of Play, 1*, 197-213. doi: 10.1080/21594937.2012.693384

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

Williford, A. P., Vick Whittaker, J. E., Vitiello, V. E., & Downer, J. T. (2013). Children's engagement within the preschool classroom and their development of self-regulation. *Early Education & Development, 24*, 162-187. doi: 10.1080/10409289.2011.628270

Willoughby, M., Kupersmidt, J., Voegler-Lee, M., & Bryant, D. (2011). Contributions of hot and cool self-regulation to preschool disruptive behaviour and academic achievement. *Developmental Neuropsychology, 36*, 162-180. doi: 10.1080/87565641.2010.549980

Willoughby, M. T., Wirth, R. J., Blair, C. B., Greenberg, M., & Family Life Project Investigators. (2010). The measurement of executive function at age 3 years: Psychometric properties and criterion validity of a new battery of tasks. *Psychological Assessment, 22*, 306-317. doi:10.1037/a0018708

Winsler, A., Tran, H., Hartman, S. C., Madigan, A. L., Manfra, L., & Bleiker, C. (2008). School readiness gains made by ethnically diverse children in poverty attending center-based childcare and public pre-kindergarten programs. *Early Childhood Research Quarterly, 23*, 314-329. doi: 10.1016/j.ecresq.2008.02.003

Wuensch, K. L. (2012). *An introduction to path analysis*. Retrieved from http://core.ecu.edu/psyc/wuenschk/MV/SEM/Path.pdf on 4-2-2014.

Xue, Y., & Meisels, S. (2004). Early literacy instruction and learning in kindergarten: Evidence from the Early Childhood Longitudinal Study - Kindergarten Class of 1998–1999. *American Educational Research Journal, 41*, 191–229. doi: 10.3102/00028312041001191

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). Retrieved from http://www.acf.hhs.gov/sites/default/files/opre/quality_review_0.pdf

Zaslow, M., Halle, T., Martin, L., Cabrera, N., Calkins, J., Pitzer, L., & Margie, N. G. (2006). Child outcome measures in the study of child care quality. *Evaluation Review, 30*, 577-609. doi: 10.1177/0193841X06291529

Zimmerman, B. J. (2000). Attaining self-regulation: A social cognitive perspective. In M. Boekaerts, M., P. R. Pintrich, & M. Zeidner (Eds). *Handbook of self-regulation*. (pp. 13-39). San Diego, CA, US: Academic Press.

Samenvatting

(Summary in Dutch)

Introductie

In de vroege kindertijd, in de leeftijd van nul tot zes jaar, wordt de basis gelegd voor de verwerving van de taal en voor de ontwikkeling van cognitieve en emotionele controle functies die van groot belang zijn voor de latere schoolcarrière en psychosociale gezondheid. Onderzoek naar de hersenontwikkeling heeft aangetoond dat er in de vroege kindertijd zogenaamde sensitieve perioden zijn die worden gekenmerkt door een hoge mate van breinplasticiteit. Gedurende deze perioden zijn de hersenen maximaal ontvankelijk voor invloeden van buitenaf, maar ook kwetsbaar voor ongunstige omstandigheden. De kwaliteit van de omgeving in de vroege kindertijd is daarom van essentieel belang om kinderen een goede start te geven in hun leven. Kwaliteit verwijst in dit verband naar de emotionele ondersteuning van het kind en de veilige sociale relaties die het kind kan aangaan. Kwaliteit verwijst ook naar de cognitieve stimulering en het taalaanbod, en naar de mogelijkheden om zelfregulatie te ontwikkelen die op een regelmatige en consequente basis aan het kind worden aangeboden in de verschillende contexten waarin een kind opgroeit.

De kwaliteit van de omgeving heeft allereerst betrekking op de thuisomgeving waarin het kind opgroeit. Het gezin biedt kinderen idealiter een veilige thuisbasis met affectieve en ondersteunende sociale relaties en een leeromgeving waarin ze de taal kunnen leren, ervaring kunnen opdoen met geletterdheid en rekenkundige begrippen, en andere vaardigheden kunnen verwerven waar de school op kan voortbouwen. Er ontstaan echter tussen kinderen al vroeg verschillen in ontwikkeling die samenhangen met het opleidingsniveau van ouders, de sociaaleconomische status van het gezin, de culturele achtergrond, opvoedingsstijl en ontwikkelingsstimulering die ouders bieden. Naast het gezin spelen voorschoolse voorzieningen voor opvang en educatie, zoals kinderopvang en peuterspeelzaalwerk, een steeds grotere rol in het leven van kinderen. In Nederland brengt het overgrote deel van de kinderen tot vier jaar, zo'n 80%, gedurende een kortere of langere periode vóór de vierde verjaardag een deel van de week door in zo'n voorziening. Dit roept de vraag op wat de kwaliteit van deze voorzieningen is en hoe de kwaliteit van de voorschoolse opvang en educatie van invloed is op de ontwikkeling van kinderen. Dit is het onderwerp van deze dissertatie.

Kwaliteit van voorzieningen en effecten op ontwikkeling

Internationaal onderzoek heeft positieve effecten aangetoond van voorschoolse voorzieningen voor opvang en educatie op de sociaal-emotionele, cognitieve en

taalontwikkeling van kinderen. Echter, deze positieve effecten blijken in sterke mate af te hangen van de kwaliteit van de voorzieningen. Daarbij wordt vaak onderscheid gemaakt tussen *structurele kwaliteit* en *proceskwaliteit*. Structurele kwaliteit heeft betrekking op van buiten af, door regelgeving reguleerbare, 'distale' kenmerken van de opvang- en educatievoorzieningen, zoals de groepsgrootte, staf-kind ratio en het opleidingsniveau van de pedagogisch medewerkers. Deze structurele kwaliteitskenmerken worden gezien als randvoorwaarden voor de proceskwaliteit. Proceskwaliteit verwijst naar de dagelijkse ervaringen van kinderen en omvat de fysieke, emotionele, sociale en educatieve aspecten van interacties met pedagogisch medewerkers, andere kinderen en materialen in de groep. Proceskwaliteit wordt gezien als 'proximale' determinant van ontwikkelingsuitkomsten aan de kant van het kind. Daarnaast speelt ook het *curriculum,* oftewel het (gerealiseerde) aanbod van activiteiten een rol. Het *gerealiseerde* curriculum kan gezien worden als een aspect van de proceskwaliteit, omdat het betrekking heeft op de aard van de ervaringen van kinderen met materialen en activiteiten. Het verwijst naar de kennisinhouden, aangeboden in activiteiten, die direct van invloed zijn op de kennis en vaardigheden die kinderen kunnen opdoen. In sommige landen is er in het curriculum steeds meer nadruk komen te liggen op activiteiten die gericht zijn op de ontwikkeling van schoolse vaardigheden, zoals taal, geletterdheid en rekenen, om kinderen goed voor te bereiden op het basisonderwijs. Die nadruk gaat ten koste van de ruimte in het curriculum voor typische vroegkinderlijke activiteiten als spel en de vraag is of dat verstandig is. Recent onderzoek wijst op het belang van spel voor de ontwikkeling van zelfregulatie en executieve functies, die sterke voorspellers blijken te zijn voor latere schoolse prestaties, sociale competentie en algemene vaardigheden als werkhouding en taakgerichtheid.

Deze dissertatie

In deze dissertatie worden vier onderzoeken gerapporteerd waarin de kwaliteit van de Nederlandse voorzieningen voor opvang en educatie centraal staat. Er is voor deze onderzoeken gebruik gemaakt van data die verzameld zijn in het kader van het longitudinale cohortonderzoek pre-COOL (onderzoeken 1, 2 en 3) en een daaraan gekoppelde dieptestudie, pre-COOL Groups (onderzoek 4). Pre-COOL wordt in opdracht van het Ministerie van Onderwijs, Cultuur en Wetenschappen en het Nationaal Regieorgaan Onderwijsonderzoek van de Nederlandse Organisatie voor Wetenschappelijk Onderzoek uitgevoerd door een consortium van het Kohnstamm Instituut van de Universiteit van

Amsterdam, het Instituut voor Toegepaste Sociale Wetenschappen uit Nijmegen, en de Afdeling Orthopedagogiek van de Universiteit Utrecht. Het doel van pre-COOL is vast te stellen of gebruik van voorschoolse opvang en educatie voorzieningen bijdraagt aan de brede ontwikkeling van kinderen op zowel cognitief, sociaal als emotioneel vlak. De vraag is vooral in hoeverre deelname aan deze voorzieningen vroege verschillen in ontwikkeling die samenhangen met de gezinsachtergrond ongedaan kan maken. Belangrijke deelvragen van pre-COOL betreffen de kwaliteit van de voorschoolse opvang- en educatievoorzieningen, de relaties tussen structurele kwaliteit, proceskwaliteit en curriculum, en de effecten van al deze kwaliteitsaspecten op de ontwikkeling van kinderen. Deze vragen staan centraal in deze dissertatie. Om deze vragen te kunnen beantwoorden zijn op verschillende manieren gegevens verzameld. Er zijn observaties verricht in kinderdagverblijven en peuterspeelzalen, er is een survey gedaan onder pedagogisch medewerkers en ouders, en er zijn testafnames gedaan bij een grote groep kinderen.

De psychometrische kwaliteit van de CLASS Toddler

In het eerste onderzoek (Hoofdstuk 2) is de psychometrische kwaliteit van het observatie-instrument de Classroom Assessment Scoring System (CLASS Toddler) onderzocht. In pre-COOL is gekozen voor het veelgebruikte instrument CLASS omdat het vergelijking met internationaal onderzoek mogelijk maakt. Aangezien dit instrument ontwikkeld is in de Verenigde Staten (VS), is eerst onderzocht of het instrument betrouwbaar en valide gebruikt kan worden in de Nederlandse situatie. In dit onderzoek zijn twee verschillende methodologische benaderingen toegepast om de psychometrische kwaliteit van de CLASS te onderzoeken. De eerste, tevens de meest gebruikte, is de Klassieke Testtheorie waarin de kwaliteit van het meetinstrument als geheel wordt geëvalueerd, bijvoorbeeld door de factorstructuur te onderzoeken van de afzonderlijke observatieschalen waaruit het instrument is opgebouwd. De tweede is de Item Response Theory (IRT), die veelal gebruikt wordt voor testconstructie omdat deze benadering meer specifieke informatie geeft over de moeilijkheid en het discriminerend vermogen van de afzonderlijke testitems waarmee de primaire informatie over de te testen vaardigheid wordt verzameld.

Binnen de CLASS Toddler worden in Amerikaans onderzoek twee overkoepelende domeinen onderscheiden, namelijk *Emotionele en Gedragsondersteuning* en *Educatieve Ondersteuning*. Voor elk domein geldt dat er een aantal onderliggende dimensies is die verschillende elementen van het betreffende domein vertegenwoordigen. Het domein Emotionele en Gedragsondersteuning kent vijf onderliggende dimensies: *Positieve Sfeer,*

Negatieve Sfeer, Sensitiviteit van de Leidster, Aandacht voor Kindperspectief en *Begeleiding van Gedrag*. Educatieve Ondersteuning omvat drie dimensies: *Faciliteren van Leren en Ontwikkeling, Kwaliteit van Feedback* en *Stimuleren van Taalontwikkeling*. Binnen elke dimensie wordt in de CLASS Toddler voorts een aantal indicatoren (op te vatten als 'items') gedefinieerd aan de hand waarvan tijdens het observeren de concrete gedragingen van de pedagogisch medewerker en de kinderen kunnen worden beoordeeld. Deze indicatoren zijn daarmee, net als de items van een test, te beschouwen als de primaire bronnen van informatie waaruit de dimensiescores en de uiteindelijke totaaloordelen per overkoepelend domein worden opgebouwd.

Met behulp van confirmatieve factoranalyses is allereerst de structuur van domeinen en dimensies van de CLASS Toddler onderzocht. Wat betreft de domeinstructuur blijkt een tweefactormodel goed te passen, in overeenstemming met de Amerikaanse bevindingen, maar blijkt een driefactormodel beter aan te sluiten bij de data. In de Nederlandse data kunnen dus het beste drie domeinen onderscheiden worden: *Emotionele Ondersteuning* (Positieve Sfeer, Sensitiviteit van de Leidster en Aandacht voor Kindperspectief), *Ondersteuning van Gedrag* (Negatieve Sfeer en Begeleiding van Gedrag) en *Educatieve Ondersteuning* (Faciliteren van Leren en Ontwikkeling, Kwaliteit van Feedback en Stimuleren van Taalontwikkeling). Het tweede deel van de analyses richtte zich op de kwaliteit van de primaire informatiebronnen van de CLASS Toddler, de indicatoren. Uitgaande van de drie-domeinen structuur, zijn met behulp van factoranalyse voor categorale data IRT analyses uitgevoerd op de indicatoren. De resultaten tonen dat alle indicatoren binnen aanvaardbare ranges van moeilijkheidsgraad en discriminerend vermogen vallen, dus als indicatoren psychometrisch gezien goed functioneren. Ook de indicatoren met relatief hoge moeilijkheidsgraad (vooral in het domein educatieve ondersteuning), discrimineren naar behoren. Ten slotte, zijn de dimensies en domeinen van de CLASS gerelateerd aan kenmerken van structurele kwaliteit en het geboden curriculum waarmee de op theoretische gronden verwachte verbanden werden gevonden. Alles wegend kan geconcludeerd worden dat de CLASS Toddler een betrouwbaar en valide instrument is en toegepast kan worden in de Nederlandse situatie.

De samenhang van structurele kwaliteit en proceskwaliteit

In het tweede onderzoek (Hoofdstuk 3) zijn de verbanden tussen structurele kwaliteit en proceskwaliteit onderzocht. Structurele kwaliteit heeft betrekking op de randvoorwaarden voor de interactieprocessen in de groep, oftewel de proceskwaliteit. In het huidige onderzoek

zijn de meest gangbare structurele kwaliteitskenmerken, namelijk de groepsgrootte, staf-kind ratio en het opleidingsniveau van de pedagogisch medewerkers meegenomen. Daarnaast zijn twee nieuwe aspecten van structurele kwaliteit aan de analyses toegevoegd, namelijk het gebruik van een voor- en vroegschools educatief [VVE] programma en het aanbod van continue professionele ontwikkeling in het dagverblijf of de peuterspeelzaal. Onder continue professionalisering wordt onder andere verstaan het regelmatig houden van pedagogisch-inhoudelijk teamoverleg, systematisch observeren in de groepen, leren van collega's, individueel begeleiden van medewerkers en het volgen van aanvullende trainingen. Voorts is proceskwaliteit gedefinieerd als een breed construct bestaande uit zowel observaties van de interactieprocessen (gemeten met de CLASS) als uit metingen van het curriculum van aangeboden activiteiten (op basis van zelfrapportages van de pedagogisch medewerkers). De belangrijkste reden van deze aanpak is dat door de combinatie van observaties in *real time* met gerapporteerde gegevens over de groepsprocessen over een *langere tijdspanne*, een betrouwbaarder en omvattender beeld van de proceskwaliteit kan worden verkregen.

Met deze multi-methodische metingen is vervolgens, in lijn met het op Amerikaans CLASS-onderzoek gebaseerde twee-domeinen model, een analysemodel ontworpen met twee latente factoren die respectievelijk de emotionele en educatieve proceskwaliteit representeren, en is nagegaan wat de verbanden zijn van deze factoren met de eerder genoemde structurele kwaliteitskenmerken. De bevindingen tonen dat de emotionele kwaliteit in Nederlandse kinderdagverblijven en peuterspeelzalen *gemiddeld* tot *hoog* is in internationaal perspectief, terwijl de educatieve kwaliteit *laag* tot *gemiddeld* is. De resultaten laten verder zien dat groepsgrootte en staf-kindratio niet gerelateerd zijn aan de emotionele en educatieve proceskwaliteit, zoals gedefinieerd in dit onderzoek. Wat betreft opleidingsniveau van de pedagogisch medewerkers blijkt er een klein positief verband te zijn, wat betekent dat de proceskwaliteit hoger is in groepen met hoger opgeleide medewerkers. Het werken met een VVE-programma hangt ook positief samen met zowel de emotionele als de educatieve kwaliteit. Echter, het sterkste verband met zowel de emotionele als de educatieve kwaliteit wordt gevonden voor het aanbod van continue professionele ontwikkeling. In centra waar veel aandacht is voor professionele ontwikkeling, is de emotionele en educatieve kwaliteit hoger.

Effecten op de ontwikkeling

In het derde onderzoek (Hoofdstuk 4) is nagegaan wat de effecten van kwaliteit en curriculum zijn op de ontwikkeling van twee vroege kernvaardigheden van kinderen: woordenschat

en selectieve aandacht. Met behulp van een multilevel value-added benadering, waarin gecontroleerd is voor achtergrondkenmerken van kinderen en mogelijke selectieve plaatsing in voorzieningen voor opvang en educatie, is onderzocht wat de effecten zijn van de geobserveerde kwaliteit (*Emotionele Ondersteuning, Ondersteuning van Gedrag* en *Educatieve Ondersteuning*) en het gerapporteerde curriculum (aanbod van academische activiteiten, activiteiten om zelfregulatie te bevorderen en aanbod van vrij spel) op de ontwikkeling van de woordenschat en de aandachtsfunctie in de leeftijd van twee naar drie jaar. De resultaten laten zien dat Emotionele Ondersteuning een (kleine) positieve bijdrage levert aan de woordenschatontwikkeling van kinderen terwijl Educatieve Ondersteuning een (groter) positief effect heeft op de ontwikkeling van de aandachtsfunctie van kinderen. Uit het onderzoek blijkt ook dat het aanbod van vrij spel een klein tot middelgroot negatief effect heeft op de ontwikkeling van woordenschat en de aandachtsfunctie heeft. Het aanbod van academische activiteiten of activiteiten gericht op het stimuleren van zelfregulatie heeft geen effect op de taal- en aandachtsontwikkeling van kinderen. Concluderend, kan gesteld worden dat het onderzoek de eerste aanwijzingen heeft opgeleverd voor effecten van voorschoolse opvang en educatie op de ontwikkeling van funderende vaardigheden bij twee- tot driejarige kinderen.

Zelfregulatie in 'doen alsof' spel

In het laatste onderzoek (Hoofdstuk 5) is nader onderzocht hoe de zelfregulatie van kinderen tot uiting komt in een begeleide fantasiespelsituatie. Verder is onderzocht hoe de geobserveerde zelfregulatie verband houdt met de executieve functies van kinderen (gemeten met neuropsychologische tests) en met contextkenmerken, die zowel de globale proceskwaliteit als de kwaliteit van het 'doen alsof' fantasiespel betreffen. Eerder onderzoek heeft laten zien dat met name fantasiespel een belangrijke activiteit kan zijn waarin kinderen zelfregulatie kunnen ontwikkelen. Voor dit onderzoek is gebruik gemaakt van de data van pre-COOL Groups, een dieptestudie verbonden aan het pre-COOL onderzoek. Ten behoeve van dit onderzoek is een nieuw observatie-instrument ontwikkeld om de zelfregulatie van kinderen tijdens spel te kunnen meten waarbij cognitieve (planning, monitoring van doelgericht gedrag, volharding en persistentie) en emotionele (emotieregulatie, omgaan met conflicten en zelfcontrole) aspecten van zelfregulatie zijn onderscheiden. Het observatie instrument blijkt goed te werken en laat zien dat de cognitieve en emotionele aspecten van zelfregulatie inderdaad verschillende vormen van controle betreffen in een spelsituatie waarin ook andere kinderen zijn betrokken. Ook zijn er positieve verbanden

gevonden met de executieve functies van kinderen. Wat betreft de rol van de context, zijn er geen relaties gevonden met de geobserveerde proceskwaliteit gemeten met de CLASS. Echter, er zijn wel middelgrote tot sterke verbanden gevonden tussen zelfregulatie en de complexiteit van het doen alsof spel. Concluderend kan worden gesteld dat de resultaten van deze dieptestudie in overeenstemming zijn met bevindingen in ander onderzoek, deels met sterkere experimentele onderzoeksopzetten, waarin fantasiespel een belangrijke bijdrage aan de ontwikkeling van zelfregulatie blijkt te leveren.

Conclusies

Deze dissertatie rapporteert de eerste bevindingen van de nationale cohortstudie pre-COOL naar de kwaliteit van Nederlandse voorzieningen voor opvang en educatie voor jonge kinderen en de effecten daarvan op de ontwikkeling van kinderen. Het betreft een tussenstand, want er zijn in pre-COOL verschillende vervolgmetingen voorzien die nog nader geanalyseerd zullen worden. De resultaten die in deze dissertatie zijn gerapporteerd, laten zien dat de emotionele proceskwaliteit van Nederlandse voorzieningen gemiddeld tot hoog is, terwijl de educatieve proceskwaliteit laag tot gemiddeld is. Het opleidingsniveau van de pedagogisch medewerkers hangt positief samen met de emotionele en educatieve proceskwaliteit, maar de sterkte van deze samenhang is relatief klein. Ook het gebruik van een educatief programma laat een positief verband zien met de proceskwaliteit, maar het aanbod van continue professionele ontwikkeling op de werkvloer heeft de sterkste positieve samenhang met de emotionele en educatieve proceskwaliteit. De beschreven resultaten laten ook zien dat de proceskwaliteit een aantoonbare bijdrage levert aan de ontwikkeling van funderende vaardigheden van jonge kinderen. Vrij spel blijkt een negatief verband te hebben met zowel de ontwikkeling van woordenschat als de aandachtsfunctie. Echter, begeleid 'doen alsof' spel blijkt een context te bieden waarin kinderen door samen te spelen cognitieve en emotionele zelfregulatie laten zien.

Uit deze dissertatie komen een aantal praktische implicaties naar voren. Gezien het belang van hoge kwaliteit voor de ontwikkeling van kinderen en de bevinding dat de kwaliteit op sommige punten niet optimaal is, is de belangrijkste implicatie dat de proceskwaliteit, met name de educatieve proceskwaliteit, verbeterd moet worden. Continue professionele ontwikkeling op de werkvloer, in het kinderdagverblijf of de peuterspeelzaal, lijkt daarbij een veelbelovende strategie om de wenselijk kwaliteitsverhoging tot stand te brengen (Hoofdstuk 3). Professionele ontwikkeling zou ingebed kunnen worden in een permanent

kwaliteitszorgsysteem en op die wijze kunnen bijdragen aan het verhogen van de kwaliteit. Een andere implicatie die naar voren komt uit deze dissertatie heeft betrekking op het 'spelen versus leren' debat. Enerzijds blijkt uit deze studie (Hoofdstuk 4) dat het aanbieden van veel vrij spel, mogelijk een negatieve invloed heeft op de ontwikkeling van kinderen. Uit de observaties met de CLASS Toddler (Hoofdstuk 3) bleek verder dat vrij spel, vergeleken met bijvoorbeeld educatieve en creatieve activiteiten, zowel in emotioneel als educatief opzicht een minder optimale context biedt. Vrij spel in de Nederlandse kinderdagverblijven en peuterspeelzalen is vooral kind-geïnitieerd met weinig actieve betrokkenheid van de pedagogisch medewerkers. Anderzijds suggereren de bevindingen uit de dieptestudie (Hoofdstuk 5) dat spel onder bepaalde voorwaarden wel zou kunnen bijdragen aan de ontwikkeling van funderende vaardigheden van kinderen. Die voorwaarden betreffen de aard van het spel (een relatief lang volgehouden fantasiespel van een klein groepje kinderen), de kwaliteit van het spel (hoog niveau van symbolisering en rollenspel) en de voorwaardenscheppende en begeleidende rol van de pedagogisch medewerker. Het belangrijkste is dat pedagogisch medewerkers kinderen doelgerichte leermogelijkheden bieden die passen bij hun ontwikkeling en aansluiten op hun interesses, door kennis en ervaringen in te bedden in een betekenisvolle context met voldoende mogelijkheden voor exploratie.

Dankwoord

(Acknowledgements)

It takes a village to raise a child… Ook dit proefschrift had niet tot stand kunnen komen zonder de hulp en steun van veel mensen uit mijn omgeving.

Allereerst wil ik alle kinderen, ouders, leidsters en managers van alle peuterspeelzalen en kinderdagverblijven die hebben meegedaan bedanken voor hun inzet. Daarnaast wil ik de consortium partners en collega's van het Kohnstamm Instituut en het ITS bedanken die voor en achter de schermen hebben meegewerkt aan de dataverzameling en verwerking van pre-COOL. Ook natuurlijk bedankt aan alle studenten en assistenten die hebben meegeholpen met de dataverzameling door het hele land door te reizen om kinderen te testen en observaties te verrichten in de kinderdagverblijven en peuterspeelzalen in alle uithoeken van Nederland. In het bijzonder wil ik Lisette Schonenburg, Marije Meerdink en Lysanne van der Flier bedanken voor alle hulp bij het voorbereiden en trainen van assistenten, naast het verzamelen van de data.

Paul, we kennen elkaar inmiddels al zo'n 10 jaar. Je hebt me zien groeien van bachelor naar master student en daarna als aio. Jouw rotsvaste vertrouwen in combinatie met je scherpe en kritische blik hebben me altijd geïnspireerd om het beste uit mezelf te halen. Dankjewel voor je vertrouwen en je steun de afgelopen jaren. Hanna en Josje, wat fijn dat jullie mijn copromotoren wilden zijn. Jullie waren een welkome aanvulling. Josje, jouw kalmte en pragmatische insteek hebben me erg geholpen op momenten dat ik het even niet meer zag. Hanna, jouw betrokkenheid en luisterend oor waren een grote steun. Samen vormden jullie een goed team, dankjewel hiervoor.

De leden van de leescommissie, Jan van Tartwijk, Mariette de Haan, Janneke Plantenga en Ruben Fukkink, wil ook graag bedanken dat ze de tijd hebben genomen om mijn proefschrift te lezen en beoordelen. A special thanks to Marja-Kristiina for coming all the way from Jyvaskyla (although it's not the end of the world) to act as an opponent. I have warm memories of our first meeting in Montreal, where we had a lovely dinner with Bridget, and of you welcoming me to your lovely house by the lake. I am so happy we are cooperating in the CARE project and hope to continue working with you for a long time.

Ook collega's van de UU wil ik graag bedanken. Allereerst Erica van Doorn die van tijd tot tijd technische ondersteuning heeft gegeven als mijn laptop weer eens niet functioneerde. Joop Hox, Suzanne Jak en Rens van der Schoot wil ik bedanken voor alle statistische adviezen die zeer welkom waren. Jan, fijn dat je mee wilden werken aan mijn artikel over de psychometrische kwaliteit van de CLASS. Nu heb ik ook (een beetje) kennis van IRT analyses. Ook collega's van de gang, speciaal Cathy, Elma en Kim, en vooral ook mijn voormalig kamergenootjes, Sylke en Ilona: jullie waren altijd in voor een praatje.

Bedankt voor de gezellige tijd. Lotte, met jou kon ik lekker sparren over werk en tijdens een congres in Bamberg ontdekten we ook verschillende overeenkomsten en gedeelde interesses op persoonlijk vlak (we dateren allebei van het Popples-tijdperk ontdekten we). Oude en nieuwe collega, Martine, ik heb met veel plezier met je samengewerkt aan pre-COOL Groups en ben heel blij om de samenwerking voort te zetten met het CARE project. We gingen aardig tegelijk op met ons aio-traject en hebben daardoor veel met elkaar kunnen delen (ook de onvermijdelijke frustraties die met zo'n traject gepaard gaan). Ook voormalig collega van O&T, Margriet wil ik bedanken voor de fijne samenwerking aan het 'CLASS' traject. Samen met Jolande hebben we een mooie pilot gedaan. Jolande, jou wil ik ook bedanken voor de diepe gesprekken die we voerden over veel meer dan alleen het CLASS traject. Jouw scherpe visie heeft mij menigmaal aan het denken gezet en we konden heerlijk sparren over van alles en nog wat. Ook Ria wil ik bedanken voor de prikkelende gesprekken en onmisbare ondersteuning.

Saskia en Ora, mijn paranimfen, wat ben ik blij om jullie aan mijn zijde te hebben. Lieve Saskia, we kennen elkaar nu al een aantal jaren, waarbij werk en privé volkomen met elkaar vervlochten zijn. We hebben verschillende congressen bezocht in binnen- en buitenland waar ik hele goede herinneringen aan heb. Montreal met het sjofele hotel, maar met gezellig kamergenootje, samen met Daan en Sophie eten in Seattle, het galadiner en de biergarten in Munchen. Maar ook onze eindeloze telefoongesprekken over van alles en nog wat. Nog even en dan ben jij ook klaar, ik kan niet wachten tot het zover is. Lieve Ora, al zijn we nog maar kort kamergenootjes, het voelt als veel langer. Fijn dat je een vervolg hebt gekregen aan de universiteit, zodat de (soms iets te) gezellige kamer samenstelling ongewijzigd blijft. Vol trots heb ik achter je gestaan tijdens je verdediging en ik ben heel blij dat jij er nu voor mij bent.

Ook alle vrienden, Rodin, Anouk, Patrick, Rachida, Bob, dankjewel voor jullie steun. In het bijzonder dank aan mijn vriendinnen van het eerste uur, Ellen, Marieke en Marieke, met regelmaat constateerden we dat het maar druk is om werk en gezin te combineren. Desondanks hebben we altijd tijd gemaakt voor de gezellige onderonsjes, samen winkelen, lekker eten of gezamenlijke vakanties. Dankjewel, jullie zorgden voor (de broodnodige) ontspanning.

Daarnaast was ook mijn (schoon)familie een grote bron van steun. Mijn schoonouders die vol liefde hebben gezorgd voor Alyza, Luka en Noah en die met zeker regelmaat zorgden voor een heerlijke en gezonde Turkse maaltijd. Zonder hen was dit proefschrift er niet geweest. Ook andere leden van de schoonfamilie, met name Tuba, hebben op veel

verschillende manieren bijgedragen aan dit proefschrift. Niet in de minste plaats door eindeloos vragenlijsten in te vullen om de toepasbaarheid voor pre-COOL te toetsen. Ook mijn moeder en mijn zusje, met wie ik eindeloos kon praten over het leven als werkende vrouw/moeder, wil ik bedanken. Ryan en Yvonne, ik wil ook jullie bedanken voor jullie liefdevolle ondersteuning en rotsvaste vertrouwen in mij en mijn kunnen. Ryan, wat had ik je graag er bij gehad, maar dat heeft helaas niet zo mogen zijn. Weet dat je in gedachten bij me bent. Maar bovenal, wil ik mijn gezin bedanken. Lieve Alyza, Luka en Noah, jullie weten niet beter dan dat mama werkt. Jullie waren erg betrokken bij mijn werk. Zo kwamen jullie graag mee naar de Uithof (al speelde de chocolademelk uit de automaat daarin ook vast een rol) en keken jullie regelmatig mee naar de filmpjes van "de kindjes". Jullie houden mij in balans, dankjewel schatjes. En tot slot mijn lieve man, Ali, wat had ik zonder je gemoeten. Jij bent op alle mogelijke manieren betrokken geweest in dit traject, variërend van technische problemen oplossen met computers, websites, filmpjes etc., tot luisteren en meedenken tijdens de (onontkoombare) moeilijkere momenten. Jij weet inmiddels bijna net zoveel van SEM modellen en Mplus als ik. Dankjewel lieverd, voor wie je bent.

About the author

Curriculum Vitae

Pauline was born on February 25, 1980 in Nijmegen, the Netherlands. She obtained her high school degree (Athneneum) in 1998 from Pax Christi College in Druten. She obtained her bachelor degree in Pedagogical and Educational Sciences in 2007 from Utrecht University. In 2008 she obtained her clinical master degree in special education (Orthopedagogiek) Cum laude also from Utrecht University. She started her PhD in December 2009 at the faculty of Social and Behavioural Sciences at Utrecht University. While working on her PhD she also supervised master students writing their thesis. She became a licensed Train-the-Trainer in the CLASS Toddler and Pre-K and provided several CLASS workshops and trainings on request for national and international researchers and at conferences and a summer school. She also collaborated with the department of Educational Development and Training (Onderwijsadvies en Training) in a project aimed at using the CLASS in a professional development project in a child care organization. In January 2014 she started her work as postdoctoral researcher in the EU CARE project at Utrecht University.

Publications

Scientific publications

Broekhuizen, M. L., Slot, P. L.,Dubas, J. S., & van Aken, M. A. G. (2014). Teacher[1]s Emotional and Behavioral Support and Toddler's Self-Control: Relations With Social and Emotional Well-Being During Play. *Manuscript in preparation*

Leseman, P. P. M., & Slot, P. L. (2014). Breaking the cycle of poverty: challenges for European early childhood education and care. European Early Childhood Education Research Journal, 22, 314-326. doi: 10.1080/13500293X.2014.912894

Slot, P. L., Boom, J.,Verhagen, J & Leseman, P. P. M. (2014). Psychometric properties of the Classroom Assessment Scoring System Toddler. *Manuscript submitted for publication*

Slot, P. L., Leseman, P. P. M., Verhagen, J., & Mulder, H. (2014). Associations between structural quality aspects and process quality in Dutch early childhood education and care settings. *Manuscript under revision*

Slot, P.L., Mulder, H., & Leseman, P. P. M. (2014). Preschoolers' cognitive and emotional self-regulation in pretend play: Relations with executive functions, quality of play, and classroom quality. *Manuscript under revision*

Slot, P.L., Mulder, H., Verhagen, J., Boom, J., & Leseman, P. P. M. (2014). General and specific quality of early childhood education and care predict growth of two-year-olds' vocabulary and attention skills. *Manuscript under revision*

Other publications

Leseman, P.P.M. & Slot, P.L. (2011). De missing link: Een competent team. Kinderen in Europa, 21, 10-11.

Leseman, P.P.M. & Slot, P.L. (2011). Increasing team competence in examining practice critically. *Children in Europe, 21,* 10-11

Slot, P. L., Netjes, J., & Groothuis, M. (2014). Observatie en reflectie: Instrumenten voor professionaliteit. *Beleid Bestuur Management & Pedagogiek in de kinderopvang, 1*, 29-31

Slot, P. L. (2014). Kwaliteit doet ertoe: Aanwijzingen voor effecten op woordenschat en aandachtsontwikkeling. *Beleid Bestuur Management & Pedagogiek in de kinderopvang, 3,* 33-37