Child care quality in the Netherlands: From quality assessment to intervention

Helmerhorst, K.O.W.

Publication date
2014
Document Version
Final published version

Child Care Quality in The Netherlands

From Quality Assessment to Intervention

Katrien Helmerhorst

Cover design: Gijs Helmerhorst & Datawyse
Printed by: Datawyse | Universitaire Pers Maastricht
Photo cover: Joods Historisch Museum

De foto van de *Vereeniging Zuigelingen-Inrichting en Kinderhuis* op de omslag dateert van 1932. Deze crèche lag aan de Plantage Middenlaan te Amsterdam, op een steenworp afstand van waar dit proefschrift is geschreven. Tijdens de Tweede Wereldoorlog werden via deze crèche vele honderden Joodse kinderen gered van deportatie naar de concentratiekampen. Tegenover de crèche ligt de Hollandse Schouwburg, waar vanaf juli 1942 duizenden Joden uit Amsterdam en omstreken zich moesten melden of onder dwang naartoe werden gebracht in afwachting van hun deportatie. De kinderen werden bij de schouwburg gescheiden van hun ouders en ondergebracht in de tegenoverliggende crèche. Directrice van de crèche, Henriëtte Pimentel, bedacht samen met Walter Süskind en Felix Halverstad een list om de kinderen die tijdelijk in de crèche werden opgevangen, onder te laten duiken. Süskind en Halverstad, beiden werkzaam voor de Joodse Raad, verwijderden de namen van de kinderen uit de administratie, waarna de kinderen via de tuin van de crèche werden weg gesmokkeld in manden, zakken en melkbussen naar Limburg en Friesland.

Child Care Quality in The Netherlands: From Quality Assessment to Intervention

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de Universiteit van Amsterdam
op gezag van de Rector Magnificus
prof. dr. D.C. van den Boom
ten overstaan van een door het college voor promoties ingestelde
commissie, in het openbaar te verdedigen in de Aula der Universiteit
op woensdag 12 november 2014, te 11.00 uur

door

Katrien Onny Willemijn Helmerhorst
geboren te Amsterdam

Promotiecommissie

Promotores	Prof. dr. R.G. Fukkink Prof. dr. J.M.A. Riksen-Walraven
Co-promotor	Prof. dr. L.W.C. Tavecchio
Overige leden	Dr. E.M. Albers Dr. H.M.Y. Koomen Prof. dr. P.P.M. Leseman Prof. dr. G.J.J.M. Stams Prof. dr. M.L.L. Volman

Faculteit der Maatschappij- en Gedragswetenschappen

This study was funded by the Dutch Ministry of Social Affairs and Employment (Ministerie van Sociale Zaken en Werkgelegenheid).

Contents

CHAPTER 1
General Introduction

Foto: Joods Historisch Museum

1.1 Background of this thesis

The quality of care provided in Dutch child care centers has been repeatedly measured since 1995 up until now. In 1995 Van IJzendoorn, Tavecchio, Stams, Verhoeven, and Reiling (1998) were the first to assess child care quality in The Netherlands using the Infant/Toddler Environment Rating Scale (ITERS; Harms, Cryer, & Clifford, 1990) and Early Childhood Environment Rating Scale (ECERS; Harms & Clifford, 1980). Results from this study showed that quality of care, on average, was relatively high in international perspective (M = 4.8, SD = 0.61 on a 7-point scale), with none of the groups scoring in the inadequate quality category and 40% of the groups scoring in the high quality category. The second assessment of child care quality in Dutch centers, using the same ITERS and ECERS, was carried out in 2001 by Gevers Deynoot-Schaub and Riksen-Walraven (2005). Compared to the 1995 measurement, quality of care was significantly lower in 2001 (M = 4.3, SD = 0.74). The effect size for the difference between 1995 and 2001 was d = 0.76, corresponding to a large effect according to Cohen (1988). In 2001, inadequate care (score < 3 on a 7-point scale) was observed in 6% of the groups, and only 18% of the groups scored in the high quality category.

Shortly after this second quality assessment in 2001, a study by the Early Child Care Research Network from the National Institute of Child Health and Human Development was published (NICHD ECCRN, 2002). In the comprehensive longitudinal research project conducted by the NICHD ECCRN, children have been followed from infancy up until the 9th Grade (14-15 years old). In 1991, over a thousand children were enrolled in this large-scale longitudinal study, that was explicitly set up to examine the effects of nonparental child care on children's development and functioning (see e.g., Belsky, 2006). Researchers have collected extensive information about the quantity, quality, and type of child care the children have received from birth onwards, as well as about their temperament and socio-emotional and cognitive development and about characteristics of the children's family and other aspects of the context. The NICHD ECCRN has reported both positive and negative impacts of early child care on young children's development. The NICHD ECCRN publication in 2002 demonstrated that high quality of care was positively related to children's cognitive and language development. But it also showed that children who attended child care for more than 30 hours per week, on average, scored higher on behavioral problems at age 4.5 (as rated by their professional caregivers) than children who attended child care for less than 10 hours per week (NICHD ECCRN, 2002). In combination with the decreasing quality scores from the 2001 Dutch quality assessment, the latter negative effect of child care found in the US caused vehement public debates in The Netherlands. A crucial question was whether the findings of the NICHD study were generalizable to the Dutch context at that time, given that the Dutch child care context is considerably different from the US context in a number of respects (see 1.2. below; Tavecchio, 2002; Van IJzendoorn, 2004).

To answer this pressing question, the Dutch Government granted a 12-year subsidy to the Netherlands Consortium for Research in Child Care (Nederlands Consortium Kinderopvang Onderzoek, NCKO) to study the quality of care as provided in child care centers in The Netherlands. The NCKO was established in 2002 by the University of Amsterdam, Radboud University Nijmegen, and Leiden University. The general aim of the NCKO research program was twofold: to assess *and* to improve the quality of center-based child care for 0- to 4-year-olds in The Netherlands (see 1.3 below for a description of the complete NCKO research program). This thesis describes four empirical studies that have been conducted as part of the NCKO research project.

1.2 Child care context in The Netherlands

The number of children attending formal child care in The Netherlands has increased substantially, mainly as a result of the increasing number of mothers with young children who continue to work after bearing a child. Recently, however, the number of children attending child care has been decreasing, which is probably due to the present economic downturn. Today, more than 70% of women with young children are employed, although the majority work part-time, contrary to fathers who generally work fulltime (Merens, Hartgers, & Van den Brakel, 2012). Fifty-six percent of the 0- to 4-year-old children in The Netherlands attend formal child care (30% center care, 26% family day care), with an average of 19 hours per week (OECD, 2014). Higher educated parents choose center care or family day care more often (59%) than middle (36%) and lower educated (25%) parents, who generally choose informal forms of nonparental care (Merens et al., 2012). Child care centers are open to children from 3 months of age (when paid maternity leave ends) up to 4 years, when kindergarten starts.

Dutch child care centers typically distinguish three types of groups with different age compositions: infant groups (0- to 2-year-olds), preschool groups (2- to 4-year-olds), and mixed age groups (0- to 4-year-olds). The mixed-age groups, with the wide age range, are predominant in The Netherlands, while such groups are relatively unknown in other countries. The vast majority of (nearly all female) caregivers have received a 3-year vocational training in general 'social-pedagogic' work, which is not specifically focused on caring for very young children in a child care setting, but rather prepares them for working with a broad variety of age groups in different domains of care (Gevers Deynoot-Schaub & Riksen-Walraven, 2005; Vermeer et al., 2008). Recently, in 2011, caregiver education has changed from social-pedagogic work (SPW) to pedagogic work (PW), which now educates students to work with children from 0 to 12 years old in child care or children up to age 18 in child welfare. This change in caregiver education was implemented after the first two studies of this thesis (Chapter 2 and 3) had been conducted. In other words, this change in education does not apply to studies 1 and 2. During data col-

lection of studies 3 and 4 (Chapter 4 and 5), caregivers were educated in the new system, although the vast majority of caregivers in the sample of these studies had followed the 'old' social-pedagogic course. Both in Dutch caregiver education and in everyday child care practice, the focus typically lies on *care* rather than on *education*, which is also reflected by the fact that most time of the program for young children in child care centers consists of free play and little attention is given to structured developmental activities (Oberhuemer, Schreyer, & Neuman, 2010; OECD, 2006).

In 2005, the introduction of the Child Care Act was a significant policy change that altered the child care funding system and may have affected quality of care afterwards (Donner, 2004). Funding changed from a supply-side system with financial support for providers, to a market-driven system with financial support for parents. Possible implications of this change in the funding system for the quality of child care will be discussed in study 2 (Chapter 3) of this thesis.

This brief description of Dutch child care makes clear that child care in The Netherlands has unique characteristics, which makes it difficult to generalize the findings of the NICHD ECCRN study to the Dutch context. Therefore, it is important to carefully adapt the definition of child care quality and quality assessment instruments to the Dutch situation.

1.3 The NCKO research project

Between 2002 and 2014, the NCKO project addressed the following main issues: 1) the development and validation of a measurement instrument to assess child care quality in the Dutch child care context, 2) the development of a simplified version of this measurement instrument to be used by professionals in child care centers to rate the quality of care in their own centers, 3) national quality assessments in representative samples of child care centers, and 4) the development and evaluation of an intervention program to improve the quality of child care as reflected in the NCKO quality instruments.

The NCKO instrument to assess child care quality

The NCKO started with the development and validation of a measurement instrument of child care quality that is attuned to the Dutch child care context. Point of departure was a theoretical model of child care quality that defined quality from the child's perspective: high quality child care was defined as care that contributes to the child's wellbeing and development. The model (Riksen-Walraven, 2004; see Figure 1) distinguishes the two types of quality that are generally used in child care research, namely process quality and structural quality. Process quality (right hand panel of Figure 1) is defined as the quality of children's everyday experiences in their interactions with the caregivers and with other children and materials in the

care group (see also Lamb, 1998; Vandell & Wolfe, 2000). The caregivers play a pivotal role in determining the quality of children's experiences, because, next to their very important direct interactions with the children, they also play an important role in determining the quality of the children's interactions with their peers and with materials in the center (reflected by the bold arrows in the figure).

Structural quality (left hand panel of Figure 1) is generally defined as the more distal characteristics of the child care environment that do not influence the child directly, but indirectly by affecting the process quality and thereby the children's everyday experiences. Next to the three 'classical' structural quality characteristics, i.e., caregiver education, group size, and caregiver-child ratio (Vandell & Wolfe, 2002), other structural quality characteristics (depicted in the figure) are space, program structure, caregiver work experience, and age composition of the group.

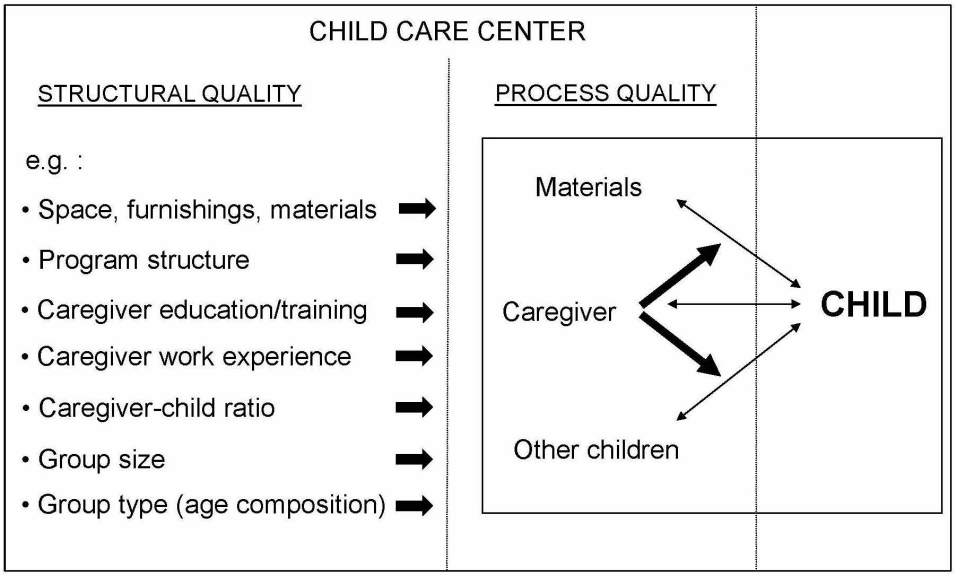

Figure 1 NCKO Quality Model (Riksen-Walraven, 2004)

In the NCKO project, we focused on developing a measure to reliably and validly assess *process* quality in Dutch child care centers, with special attention to the key aspect of process quality, namely the quality of caregiver-child interactions. The NCKO quality model distinguishes six key caregiver interactive skills that are assumed to contribute to 0- to 4-year old children's wellbeing and development, and that should therefore be included in a process quality measure: sensitive responsiveness, respect for autonomy, structuring and limit setting, verbal communication, developmental stimulation, and fostering positive peer interactions (Riksen-Walraven, 2004).

We first searched for available instruments that could be used to assess process quality, including the six caregiver interactive skills, in the Dutch context. The most

widely used instruments designed to assess process quality are the Infant/Toddler Environment Rating Scale-Revised (ITERS-R; Harms, Cryer, & Clifford, 2003) and Early Childhood Environment Rating Scale-Revised (ECERS-R; Harms, Clifford, & Cryer, 1998), which provide a comprehensive and global picture of the quality of the child care environment (Perlman, Zellman, & Le, 2004; Sakai, Whitebook, Wishard, & Howes, 2003). Two previous Dutch quality assessments in 1995 (Van IJzendoorn, et al., 1998) and 2001 (Gevers Deynoot-Schaub & Riksen-Walraven, 2005) used the ITERS/ECERS, because the widespread use of these instruments allowed international comparison. Given that the ITERS-R/ ECERS-R provide a broad picture of child care quality and that they allow international comparisons as well as a comparison with previous Dutch quality assessments, the NCKO decided to hold on to the ITERS-R/ECERS-R as a global measure of child care quality.

Because the six key caregiver interactive skills could not be assessed adequately with the ITERS-R/ECERS-R, nor with any other available instrument, the NCKO also developed a new measure that specifically focuses on the six caregiver interactive skills and that can be applied in combination with the ITERS-R/ECERS-R. The development and validation of this measure, the Caregiver Interaction Profile (CIP) scales, is reported in study 1 (Chapter 2) of this thesis.

The NCKO Quality Monitor

Based on the NCKO instrument (including the ITERS-R/ECERS-R and the CIP scales) described in the former section, the NCKO Quality Monitor was developed (Gevers Deynoot-Schaub, Fukkink, Riksen-Walraven, De Kruif, Helmerhorst, & Tavecchio, 2009). The Monitor is a self-evaluation tool for professionals in child care centers/organizations to get an impression of the child care quality provided in their own centers. Just like the NCKO measurement instrument described above, the Monitor includes a global quality part (based on the ITERS-R/ECERS-R) and a part that specifically focuses on the six caregiver interactive skills (based on the CIP scales).

The part of the Monitor focusing on the caregiver interactive skills includes simplified scales to rate the six interactive skills on a 3-point scale format (high, medium, low) instead of the original 7-point format of the CIP scales. The simplified format allows professionals in the child care field to self-rate quality of caregiver-child interactions.

The global quality part of the Monitor comprises a checklist that is based on the ITERS-R/ECERS-R subscales space and furnishings, activities, language, and program structure, which allows evaluation of the global quality of the child care environment. The checklist includes 26 items on a dichotomous scale with positive anchors (i.e., indicator should be present) and negative anchors (i.e., indicator should not be present). This set-up provides professionals with direct insight in stronger and weaker aspects of the child care environment in a specific child care group, and therefore also with information as to which points need improvement.

The checklist for the global quality of the child care environment was used as a self-evaluation tool for center directors in study 4 (Chapter 5) of this thesis, which describes the evaluation of a consultancy program for center directors to improve global child care quality.

National quality assessments

Using the measurement instrument described above, the NCKO has carried out national quality assessments in 2005, 2008, and 2012 to examine the quality of child care in representative samples of child care centers in The Netherlands. The 2005 quality assessment was performed on request of the Dutch Ministry of Social Affairs and Employment to assess the quality of care just before the introduction of the Child Care Act. This study, which only used the ITERS-R/ECERS-R because the CIP scales were not yet available, has been reported in Vermeer et al. (2008). The second assessment was carried out in 2008, three years after the introduction of the Child Care Act in 2005. This was the first assessment where the newly developed CIP scales were applied together with the ITERS-R/ECERS-R in a nationally representative sample. This study is described in study 2 (Chapter 3) of this thesis. The third and most recent quality measurement was carried out in 2012 (see Fukkink, Gevers Deynoot-Schaub, Helmerhorst, Bollen, & Riksen-Walraven, 2013).

An intervention program to improve child care quality

The fourth and final issue addressed in the NCKO project was to develop and evaluate an intervention program to improve the quality of child care. This part of the NCKO research project constitutes the main topic of the present thesis; the data for the intervention study were collected during the PhD project for the present thesis. The intervention program was aimed at improving the process quality of care as assessed with both components of the NCKO instrument, namely global quality of the child care environment (as measured with the ITERS-R/ECERS-R) and caregiver-child interactions (as measured with the CIP scales). Table 1 provides an overview of the complete intervention program and the design of the effect study. As shown in the table, the intervention program consisted of two parts that ran simultaneously in each care group. One part of the intervention was a 5-week video feedback training for the caregivers to strengthen their interactive skills with the children as defined in the CIP scales (left hand panel of the table). The other part was a consultancy program for the center directors, comprising three consultations in total, to enhance the global quality of the child care environment as assessed with the ITERS-R/ECERS-R (right hand panel of the table). The development and evaluation of the two parallel components of the intervention program is reported in two separate studies in this thesis, namely study 3 in Chapter 4 (CIP training) and study 4 in Chapter 5 (consultancy program).

Table 1 Overview of the complete intervention program and the design of the effect study

Week		
1	PRETEST Measures: CIP scales & ITERS-R/ECERS-R	
	Start intervention program	
	CIP training Directed at: Caregivers Aim: Improve caregiver interactive skills as defined in CIP scales	**Consultancy** Directed at: Center directors Aim: Improve quality of the child care environment as defined in ITERS-R/ECERS-R
2		1) Consultation 1
3	1) Video feedback session 1	*Self-assessment by center director with NCKO Quality Monitor*
4	2) Video feedback session 2	2) Consultation 2
5	3) Video feedback session 3	
6	4) Video feedback session 4	
7	5) Video feedback session 5	3) Consultation 3 by telephone
9	POSTTEST Measures: CIP scales & ITERS-R/ECERS-R	
22	FOLLOW-UP Measures: CIP scales & ITERS-R/ECERS-R	

1.4 Thesis outline

After this introductory chapter, the thesis describes four empirical studies. *Chapter 2* reports on the development and validation of the CIP scales. *Chapter 3* describes the 2008 national quality assessment, in which the CIP scales were used for the first time in a nationally representative sample of child care centers. Chapters 4 and 5 report the development and evaluation of the intervention program (outlined in Table 1) to improve child care quality. *Chapter 4* reports on the effects of the CIP training for caregivers, and *Chapter 5* describes the effects of the consultancy program directed at center directors. Finally, *Chapter 6* presents a summary of the results of the four studies followed by the main conclusion and a general discussion.

References

Belsky, J. (2006). Early child care and early child development: Major findings of the NICHD Study of Early Child Care. *European Journal of Developmental Psychology*, *3*, 95-110. doi: 10.1080/17405620600557755

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

Donner, J. P. H. (2004). *Wet kinderopvang en kwaliteitseisen peuterspeelzalen*. Retrieved from http://wetten.overheid.nl/BWBR0017017/geldigheidsdatum_22-04-2013

Fukkink, R. G., Gevers 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- tot 4-jarigen in Nederlandse kinderdagverblijven in 2012* [Pedagogical quality of child care centers for 0- to 4-year-olds in The Netherlands 2012]. Amsterdam, The Netherlands: Netherlands Consortium of Child Care Research.

Gevers Deynoot-Schaub, M., Fukkink, R., Riksen-Walraven, M., De Kruif, R., Helmerhorst, K., & Tavecchio, L. (2009). *De NCKO-Kwaliteitsmonitor* [The NCKO Quality Monitor]. Amsterdam, The Netherlands: SWP.

Gevers Deynoot-Schaub, M. J. J. M., & Riksen-Walraven, J. M. A. (2005). Child care under pressure: The quality of Dutch center in 1995 and 2001. *Journal of Genetic Psychology, 166*, 280-296. doi: 10.3200/GNTP.166.3.280-296

Gordon, R. A., Fujimoto, K., Kaestner, R., Korenman, S., & Abner, K. (2013). An assessment of the validity of the ECERS-R with implications for measures of child care quality and relations to child development. *Developmental Psychology*, *49*, 146-160. doi: 10.1037/a0027899

Harms, T., & Cilfford, R. M. (1980). *Early Childhood Environment Rating Scale*. New York: Teachers College Press.

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

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

Harms, T., Cryer, D., & Clifford, R. M. (2003). *Infant/Toddler Environment Rating Scale-Revised*. New York, NY: Teachers College Press.

Lamb, M. E. (1998). Nonparental child care: Context, quality, correlates, and consequences. In: W. Damon, I. E. Sigel, & K. A. Renninger (Eds.), *Handbook of Child Psychology, part 4: Child psychology in practice* (pp. 73-133). New York, MA: Wiley.

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

Merens, A., Hartgers, M., & Van den Brakel, M. (2012). *Emancipatiemonitor 2012* [Emancipation Monitor 2012]. The Hague, Sociaal en Cultureel Planbureau/ Centraal Bureau voor de Statistiek.

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

Oberhuemer, P., Schreyer, I., & Neuman, M. J. (2010). *Professionals in early childhood education and care systems. European profiles and perspectives*. Opladen & Farminghton Hills: Barbara Budrich Publishers.

OECD (2006). *Starting Strong II: Early childhood education and care*. Paris, France: Organisation for Economic Co-operation and Development.

OECD (2014). *Family database: Enrolment in childcare and pre-schools*. Retrieved from http://www.oecd.org/els/soc/PF3_2_Enrolment_in_childcare_and_preschools_1May2014.pdf

Perlman, M., Zellman, G. L., & Le, V. (2004). Examining the psychometric properties of the Early Childhood Environment Rating Scale-Revised (ECERS-R). *Early Childhood Research Quarterly*, *19*, 398-412. doi: 10.1016/j.ecresq.2004.07.006

Riksen-Walraven, M. (2004). Pedagogische kwaliteit in de kinderopvang: doelstellingen en kwaliteitscriteria [Pedagogical quality of child care: aims and quality criteria]. In: R. van IJzendoorn, L. Tavecchio, & M. Riksen-Walraven (Eds.). *De kwaliteit van de Nederlandse kinderopvang* (pp. 100-123). Amsterdam, The Netherlands: Boom.

Sakai, L. M., Whitebook, M., Whishard, A., & Howes, C. (2003). Evaluating the Early Childhood Environment Rating Scale (ECERS): Assessing differences between the first and revised edition. *Early Childhood Research Quarterly*, *18*, 427-445. doi: 10.1016/j.ecresq.2003.09.004

Tavecchio, L. W. C. (2002). *Van opvang naar opvoeding* [From care to education]. Amsterdam, The Netherlands: Vossiuspers UvA.

Vandell, D. L. (2004). Early child care: The known and the unknown. *Merill-Palmer Quarterly*, *50*, 387-414. doi: 10.1353/mpq.2004.0027

Vandell, D. L., Belsky, J., Burchinal, M., Steinberg, L., Vandergrift, N., & NICHD Early Child Care Research Network. (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*(3), 737-756. doi:10.1111/j.1467-8624.2010.01431.x

Vandell, D. L., & Wolfe, B. (2000). *Child Care quality: Does it matter and does it need to be improved?* Madison, STATE: Institute for Research on Poverty, University of Wisconsin-Madison.

Van IJzendoorn, R. (2004). Leidt kinderopvang tot agressie? Een Amerikaanse studie naar de kwaliteit en kwantiteit van kinderopvang [Does child care lead to aggression? A US study of quantity and quality of child care]. In: R. van IJzendoorn, L. Tavecchio, & M. Riksen-Walraven (Eds.). *De kwaliteit van de Nederlandse kinderopvang* (pp. 77-99). Amsterdam, The Netherlands: Boom.

Van IJzendoorn, M. H., Tavecchio, L. W. C., Stams, G. J. J. M., Verhoeven, M. J. E., & Reiling, E. J. (1998). Quality of center day care and attunement between parents and caregivers: center day care in cross-national perspective. *The Journal of Genetic Psychology*, *159*, 437-454.

Vermeer, H. J., Van IJzendoorn, M. H., De Kruif, R. E. L., Fukkink, R. G., Tavecchio, L. W. C., Riksen-Walraven, J. M. A., & Van Zeijl, J. (2008). Child care in The Netherlands: Trends in quality over the years 1995-2005. *The Journal of Genetic Psychology, 169*(4), 360-385.

CHAPTER 2

Measuring Interaction Skills of Caregivers in Child Care Centers:

Development and Validation of the Caregiver Interaction Profile Scales

DE ONTVANGST: „ALS HET KINDJE BINNENKOMT"

Foto: Joods Historisch Museum

Helmerhorst, K. O. W., Riksen-Walraven, J. M., Vermeer, H. J., Fukkink, R. G., & Tavecchio, L. W. C. (2014). Measuring interaction skills of caregivers in child care centers: Development and validation of the Caregiver Interaction Profile Scales. *Early Education and Development, 25,* 770-790. doi: 10.1080.10409289.2014.840482

Abstract

High-quality caregiver–child interactions constitute the core of high-quality child care for young children. This article describes the background and development of the Caregiver Interaction Profile (CIP) scales to rate six key skills of caregivers for interacting with 0- to 4-year-old children in child-care centers: sensitive responsiveness, respect for autonomy, structuring and limit setting, verbal communication, developmental stimulation and fostering positive peer interactions. Each interactive skill is rated on a 7-point scale, based on observation of video-recorded caregiver-child interactions. Together, the 6 scale scores constitute an Interaction Profile for individual caregivers that may serve as a starting point for education and training to improve the quality of caregiver-child interactions. This article also presents the results of a first study with the CIP scales, in a sample of 145 caregivers from 75 child-care groups in 47 child-care centers in The Netherlands. Results provide promising preliminary evidence supporting the reliability and validity of the CIP scales.

2.1 Introduction

There is now abundant evidence that the quality of nonparental child care contributes to children's wellbeing and development (for an overview see Belsky et al., 2007; Vandell et al., 2010; Vandell & Wolfe, 2000). Child-care quality can be globally measured by structural quality features such as group size, caregiver-child ratio, and caregiver education and training. A more proximal quality measure is the process quality of care, which is evident through children's interactions with caregivers and other children, and their engagement with the activities and materials provided (e.g., Vandell, 2004). Caregivers play a key role in determining the process quality of care for young children. First and foremost, they affect the children in direct caregiver-child interactions. For example, they have been shown to foster a sense of security in children by sensitively responding to their signals and needs (Ahnert, Pinquart, & Lamb, 2006), and to foster children's cognitive development using sensitive and stimulating interactions (Albers, Riksen-Walraven, & De Weerth, 2010). Furthermore, caregivers may promote children's wellbeing (i.e., the degree to which they feel at ease with the caregivers and how comfortable they are in the physical setting of the center and with the other children in the group; De Schipper, Van IJzendoorn, & Tavecchio, 2004) and development by fostering the children's interactions with peers and play materials in the center. Given the key role of the caregiver in determining the quality of children's experiences in child care, measuring the quality of caregiver-child interactions should be an essential part of regular assessments to monitor and improve child care quality.

In The Netherlands, the process quality of center-based child care has been assessed repeatedly from 1995 onward (see Vermeer et al., 2008). This was done using the Infant/Toddler Environment Rating Scale Revised (ITERS-R; Harms, Cryer, & Clifford, 1990, 2003) and its equivalent for preschool classrooms, the Early Childhood Environment Rating Scale Revised (ECERS-R; Harms, Clifford, & Cryer, 1980, 1998), which have been used extensively worldwide and are therefore interesting for international comparison. The ITERS-R and ECERS-R provide a broad picture of process quality in different domains, such as space, materials, activities, and program structure (for a full description see Method section). In addition to this, there was a need to get a more in-depth picture of what is generally considered the core of process quality, namely the quality of the caregiver-child interactions. The present article reports the development of an instrument to measure relevant aspects of caregiver-child interaction in Dutch child care centers as well as the results of a first study applying this instrument in 75 care groups in Dutch child care centers. Because we developed the measure initially for use in The Netherlands, we first briefly sketch the context of child care in The Netherlands.

Child Care Context in The Netherlands

The percentage of young children attending formal child care has been steeply increasing in The Netherlands in the past decades, reflecting an increase in the number of employed mothers. Nowadays, more than 70% of mothers with young children are employed, although typically part-time (Merens, Hartgers, & Van den Brakel, 2012). Moreover, 56% of 0- to 4-year-old children in The Netherlands attend formal child care (26% family day care, 30% center care) for an average of 19 hours per week (Organisation for Economic Co-operation and Development [OECD], 2011). Because the costs of formal child care are relatively high, this form of nonparental care is mostly used by more highly educated parents; parents with medium-range education or less more often use informal forms of nonparental care for their children (Merens et al., 2012). Dutch child care centers serve children from ages 3 months (when paid maternity leave ends) up to 4 years (when children enter kindergarten). Presently, the vast majority of professional caregivers in Dutch child care centers have received vocational training at an intermediate level, that is, undergoing a broad social-pedagogical 3-year training program (SPW-3) that prepares them to provide various forms of care to various age groups. In earlier years, working with 0- to 4-year-olds in child care was the main focus of caregivers' professional education, and it received much more attention. Both in caregiver education and in everyday practice, the emphasis is typically more on *care* than on *education* of the children. This is reflected in the fact that the program in Dutch child care centers consists mostly of free play for the children, with little structured educational activities (OECD, 2000, 2006). Repeated assessment of the quality of center-based child care in The Netherlands has shown that the quality has been steadily decreasing across the past decade. Between 1995 and 2005, the average quality score on the ITERS-R/ECERS-R decreased from 4.5 (on a scale from 1 to 7) which was high, also in international perspective, to 3.2, which is barely above the "minimal" score of 3 (Vermeer et al, 2008).

Development of the Caregiver Interaction Profile Scales

Our aim was to develop an instrument to measure relevant aspects of the caregiver-child interaction in addition to applying the ITERS-R/ECERS-R. This instrument should meet the following requirements. First, the measure should be relatively time efficient so that it can be completed in combination with the ITERS-R/ECERS-R by a single researcher during a single center visit. Second, the measure should assess individual caregivers' skills in interacting with a *group* of children, thus taking into account caregivers' ability to divide their attention and react consistently across children. Third, the same measure should be applicable to caregivers' interactions with children across the whole age range of 0 to 4 years. Given that most Dutch child care centers have mixed-age groups including 0- to 4-year-olds, the same measure should be applicable to caregivers' interactions with children across

the whole age range. And fourth and finally, the measure should preferably be theory based and supported by empirical evidence underscoring that the relevant aspects of caregiver-child interaction contribute to the wellbeing and development of 0- to 4-year-old children.

Step 1. In our search for such a measure, a first step was to choose a relatively limited set of caregiver interactive skills that can be assumed to essentially contribute to high-quality interactions with 0- to 4-year-olds in group care. To that end, based on a systematic review of the relevant literature, we considered which caregiver behaviors are assumed in developmental theories and models to play an important role in fostering the wellbeing and development of young children from birth onwards, and we looked for empirical evidence underpinning the relation between such caregiver skills and developmental outcomes in children. Our search resulted in the following set of six broad caregiver interactive skills:

(1) *Sensitive responsiveness* refers to the extent to which a caregiver recognizes children's individual emotional and physical needs, and responds appropriately and promptly to their cues and signals (Ainsworth, Blehar, Waters, & Wall, 1978). This broad quality of caregiver behavior, which is also referred to as *sensitivity*, *warmth*, or *supportive presence* (for an overview, see De Wolff & Van IJzendoorn, 1997), is considered the key aspect of caregiving in attachment theory (Ainsworth et al., 1978; Bolwby, 1969) and is generally recognized as the most basic aspect of caregiver behavior in interactions with children from birth onward. Caregiver sensitive responsiveness has been shown to contribute to the development of a secure caregiver-child attachment (Ahnert et al., 2006; De Wolff & Van IJzendoorn, 1997; Howes, Galinsky, & Kontos, 1998) and to foster children's ego resilience in later childhood (Riksen-Walraven & Van Aken, 1997). It is also an essential element in fostering the cognitive development of young children (Albers et al., 2010; National Institute of Child Health and Human Development Early Child Care Research Network [NICHD ECCRN], 2005) and the development of empathy and prosocial behavior (Eisenberg & Fabes, 1998; Lamb & Zakhireh, 1997; Mashburn et al., 2008; NICHD ECCRN, 2001).

(2) *Respect for autonomy* refers to the extent to which a caregiver is nonintrusive but instead recognizes and respects the validity of children's intentions and perspectives. Theoretically speaking, respecting a child's autonomy becomes increasingly important as a caregiver skill in the course of the second year of life, when acquiring a sense of autonomy is considered a central developmental issue in children's development (Erikson, 1950; Sroufe, 1979). However, already in early infancy parental intrusiveness (i.e., lack of respect for child autonomy) has been shown to predict poor development and functioning of children at early school age, even beyond the contribution of a lack of sensitive responsiveness (Egeland, Pianta, & O'Brian, 1993). Caregiver intrusiveness is also associated with lower cognitive performance outcomes in toddlers (Klein & Feldman, 2007). In The Netherlands, parental respect for child autonomy in infancy and toddlerhood has been shown to predict positive developmental outcomes in native Dutch children (Smeekens,

Riksen-Walraven, & Van Bakel, 2007) as well in immigrant children from Surinamese-Dutch families (Riksen-Walraven, Meij, Hubbard, & Zevalkink, 1996).

(3) *Structuring and limit setting* refers to the ability of a caregiver to clearly communicate expectations toward children and structure the situation accordingly, and to set clear and consistent limits on the children's behavior. Just like respect for child autonomy, this caregiver skill also becomes increasingly important in the second year of life, as children acquire a sense of autonomy and quickly expand their locomotor abilities. Adequate and consistent structuring and limit setting contributes to the predictability of the environment and therefore to the development of security and competence (see Thompson, 1998). Moreover, a lack of structuring and limit setting has been associated with the development of noncompliance in children (Arnold, McWilliams, & Arnold, 1998; Feldman & Klein, 2003). In the child-rearing styles model of Diana Baumrind (e.g., Baumrind & Black, 1967) and researchers who have extended her work, the authoritative caregiving style, which has been shown to predict the most favorable developmental outcomes across childhood and adolescence, is characterized by a combination of parental warmth, autonomy granting, and control (see Berk, 2010), which correspond to the three caregiver skills sensitive responsiveness, respect for autonomy, and structuring and limit setting mentioned here.

(4) *Verbal communication* refers to the frequency and quality of verbal interactions between caregiver and children. The regular use of language that is well adjusted to the children's interest and developmental level has an important role in accelerating children's language acquisition (Bloom, 1998; Girolametto & Weitzman, 2002; Girolametto, Weitzman, & Greenberg, 2003; Mashburn et al., 2008) and has been found to contribute to the cognitive and social development of children (Girolametto et al., 2003; Girolametto, Weitzman, & Greenberg, 2004; NICHD ECCRN, 2000). Well-adapted explanations in situations in which children transgress rules (i.e., inductive reasoning) have also been shown to foster moral development, especially if combined with caregiver warmth and control in preschoolers (Turiel, 1998; Zahn-Waxler, Radke-Yarrow, & King, 1979).

(5) *Developmental stimulation* concerns the degree to which a caregiver deliberately attempts to foster children's development (e.g., motor skills, cognitive development and creativity). The caregiver can, for example, draw children's attention to new activities, provide new activities and materials, or suggest new ways to play with materials. Appropriate developmental stimulation means not only providing novel stimuli and learning opportunities but also attuning the stimulation to children's focus of attention, developmental level, and state, thereby challenging the children while at the same time preventing overstimulation. Well-adjusted developmental stimulation in child care centers has been shown to contribute to children's cognitive development from the very first years of life onward (Albers et al., 2010; Belsky, et al., 2007; Vandell, et al., 2010).

(6) *Fostering positive peer interactions* refers to a caregiver's guidance of interactions between children in the child care center. The unique group setting of child

care provides abundant opportunities for children to develop peer relations and social competence, provided that these interactions are scaffolded by caregivers. Positive peer interactions in child care centers predict children's well-being and social-emotional development (Gevers Deynoot-Schaub & Riksen-Walraven, 2006a, 2006b), and children's later social competence with peers (Howes, 2000; Howes & Phillipsen, 1998; NICHD ECCRN, 2001; Williams, Mastergeorge, & Ontai, 2010). The importance of this specific caregiver behavior was emphasized by Dowsett, Huston, Imes, and Gennetian (2008), who emphasize that "one often-neglected dimension of positive caregiving is the ability to help children manage peer interactions effectively", and argued that "in future research measuring quality, more detailed information about ways in which caregivers promote positive interactions with peers is needed" (p. 90).

Step 2. Next, we used a survey to check whether the importance of the six caregiver interaction skills described above was recognized by four groups of stakeholders in child care, that is, parents (n = 241), caregivers (n = 148), center directors (n = 79), and external experts (n = 90), total N = 558 (De Kruif, Riksen-Walraven, Fukkink, Tavecchio, & Van Zeijl, 2009). This was done because we aimed to include the six skills in an instrument to monitor child care quality in The Netherlands on a regular basis, and acceptance of the concepts as indicators of quality is important to ensure acceptance of the outcomes of the monitoring. For this survey, we asked all stakeholders to rate the importance of 15 different process quality topics, including the caregiver interactive skills, on a 4-point scale: 1 = *not at all important (can certainly be left out*), 2 = *of little importance (can possibly be left out)*, 3 = *important (should be included)*, 4 = *very important (should certainly be included)*. Results of the survey indicated that each of the four stakeholder groups acknowledged the importance of the six caregiver interaction skills as indicators of child care quality, with average scores well beyond 3. Mean scores across all stakeholder groups were as follows: sensitive responsiveness, 3.73; fostering positive peer interactions, 3.70; verbal communication, 3.67; respect for autonomy, 3.61; structuring and limit setting, 3.42; and developmental stimulation, 3.33. Agreement between the four stakeholder groups was high; for example, mean scores for fostering positive peer interactions were 3.68 for parents, 3.70 for caregivers, 3.72 for center directors, and 3.69 for the experts. The survey also included the item "stimulating pre-academic skills", to check whether stakeholders valued more academic forms of stimulation. This item was rated much lower than the other caregiver skills, namely 2.84, on average, across all stakeholder groups, with very little variation between the groups. These stakeholder ratings reflect the typical Dutch view of center-based child care, with its stronger emphasis on care than on directed and structured teaching activities.

Step 3. In the next step we examined whether any instruments were available that captured the six caregiver interactive skills and met our other criteria (i.e., assessed individual caregivers' skills in interacting with a group of children, were applicable to the whole age range of 0- to 4-year-olds, were relatively time effi-

cient). The first instrument that might have fit our criteria was the well-known Caregiver Interaction Scale (CIS; Arnett, 1989). The CIS originally included four dimensions – positive interaction, punitiveness, permissiveness, and detachment (Arnett, 1989) –, but these dimensions did not represent the caregiver skills we aimed to assess. A second well-known instrument to rate caregiver-child interactions in child care centers is the Observational Record of the Caregiving Environment (ORCE; NICHD ECCRN, 1996). The ORCE includes several of the caregiver skills in our model, but it focuses on caregiver behavior towards individual children, in one-to-one interactions.

A third instrument that might fit our purpose was the Classroom Assessment Scoring System (CLASS; La Paro, Hamre, & Pianta, 2011; Pianta, La Paro, & Hamre, 2008). Although the CLASS is a valuable instrument that provides a comprehensive description of different dimensions of teacher-child interactions (including dimensions similar to those we aimed to assess), the CLASS was not fit for our purpose for several reasons. For one thing, the CLASS has separate versions for prekindergarten and toddler classrooms, but a version for younger children (under 12 months of age) is not available, so that the CLASS did not fit our requirement of being applicable, in the same form, across the whole range of 0 to 4 years. Furthermore, the CLASS ratings reflect the behavior of *all* caregivers in the classroom, whereas we aimed to obtain individual caregiver ratings that may provide a starting point for individual skills training in the future.

A final instrument we considered was a set of scales devised by De Schipper and Riksen-Walraven (2004) to rate the quality of caregiver behavior toward groups of 0- to 4-year-old children during structured play in child care centers (De Schipper, Riksen-Walraven, & Geurts, 2006). This instrument fits most of our criteria, but did not include scales for all of the six caregiver skills in our model.

Our review made clear that there was no readily available measure to assess the six caregiver skills in interacting with 0- to 4-year-old children in a group setting. Therefore, we developed such a measure ourselves, the CIP rating scales. The scales are partly based on some of the instruments reviewed previously and also inspired by scales to rate comparable interactive skills of parents in one-to-one interactions with their children (Ainsworth, Bell, & Stayton, 1974; Erickson, Sroufe, & Egeland, 1985).

The Present Study

The present study aimed to provide a first step in examining the validity of the CIP scales. We explored three types of validity: convergent, discriminant, and predictive (see, e.g., Campbell & Fiske, 1959). *Convergent validity* is the extent to which the CIP scales correlate with other instruments that aim to measure comparable constructs. For the convergent validity, we examined the correlation between the CIP scales and two established process quality measures that aim to assess comparable, although not identical constructs (see "Measures"). Hence, we expected significant but

moderate correlations. *Discriminant validity* is the extent to which the CIP scales only measure the intended construct and do not at the same time measure different constructs. Discriminant validity was examined by relating the CIP scales to a temperamental characteristic of the caregiver (i.e., caregiver sociability). Sociability can be defined as the general tendency to seek the company of others, which might be reflected in an increased tendency to seek social interactions in general, but does not automatically imply having the specific skills to establish high-quality interactions with young children. Therefore, we expected nonsignificant correlations between these constructs. *Predictive validity* is the extent to which the CIP scales predict theoretically relevant outcomes. Predictive validity was explored by relating the CIP scales to scores reflecting competence and behavior problems of children cared for by the caregivers in question. Given the positive relationship that has been shown between child care quality and developmental outcomes of children (see Belsky et al., 2007; Vandell et al., 2010; Vandell & Wolfe, 2000), we expected higher scores on the CIP scales to predict higher levels of competence and fewer social-emotional problems in the children.

In addition to exploring the validity of the CIP scales, we also examined the *test-retest reliability* of the CIP scales by repeating the measurements three months later. Furthermore, we examined the correlations between the six caregiver interactive skills. Because the six skills are conceptualized as different aspects of a single construct (i.e., caregiver interactive competence), we expected them to be significantly but not very highly correlated.

Summary of study aims. In the introduction of this article we described the background and development of the CIP scales. In the next sections, we report the results of a first study applying the instrument in a sample of 75 care groups in child care centers in The Netherlands. The aim of the study was to get a first impression of the applicability of the scales and to provide preliminary evidence regarding the validity and test-retest reliability of the scales.

2.2 Method

Participants

Originally 171 child care centers from two larger regions in The Netherlands were approached by letter to participate to the study. One week after the mailing, the center directors received a phone call to ask about their willingness to participate in the study. A total of 47 child care centers (27%) agreed to volunteer in the study. The main reason provided for refusal was being too busy. The final sample included 145 caregivers from 75 child care groups in 47 child care centers. In total, 698 children from 0 to 4 years old participated in this study. A total of 55 caregivers worked in infant groups (0- to 2-year-olds; 30 groups), 56 caregivers in preschool groups (2- to 4-year-olds; 29 groups), and 34 caregivers worked in mixed-age

groups (0- to 4-year-olds; 16 groups). All caregivers from a selected classroom were invited to participate in the study using an active consent procedure. Caregivers were mostly female (96%); on average, they were 30.2 years old (*SD* = 9.03, range = 19-57), worked 30.1 hours a week (*SD* = 5.97), and had 5.9 years (*SD* = 4.08) of working experience in child care. The majority of the caregivers (73%) had completed the regular vocational education at intermediate level (SWP-3, a 3-year vocational training in general social-pedagogic work).

Three months later, directors of the centers that participated in the first measurement were approached to cooperate for a second measurement. For this second wave of data collection, 108 caregivers from 69 care groups agreed to participate. Reasons provided for refusal were too busy or caregivers having switched to another care group. Caregivers were sent an active consent form that they could sign for agreement. Agreement meant that they participated in the whole assessment, including being filmed.

Procedure

Each group was visited by two trained researchers. The visit lasted from about 8 a.m. until after lunch (between 12 p.m. and 1 p.m.). The first researcher completed the ITERS-R or ECERS-R and applied the CIS for the individual caregivers (see "Measures"). In accordance with the ITERS-R/ECERS-R manual this researcher also interviewed one of the caregivers at the end of the visit to collect additional information that was not available through observation (i.e., "Is there a possibility for children to do gross motor activities indoors in case of bad weather?"). The second researcher made three to four cycles of video recordings of each caregiver-child interactions for later observations using the CIP scales. In each cycle each caregiver was filmed separately and sequentially for 8-10 minutes in a fixed schedule, during a variety of naturally occurring situations. Time intervals between the recording cycles were approximately an hour. Three months later, the groups participating in the repeat measurement were visited by one researcher, who filmed the caregivers following the same procedure that was also used during the first visit. The retest visits were planned on the same day of the week as the first visit for optimal comparison. After the retest visit, the caregivers also completed a questionnaire to collect individual background information (e.g., education, experience, and the Emotionality, Activity and Sociability Survey [EAS]). In addition, they were asked to complete the Brief Infant-Toddler Social and Emotional Assessment (BITSEA) to rate levels of competence and behavior problems for at least two randomly chosen children ages 12 months or older in the classroom (the BITSEA is not applicable to younger children).

Measures

The CIP scales. The CIP scales measure the six caregiver interactive skills that we more extensively described in the introduction to this article: *sensitive responsiveness*, *respect for autonomy*, *structuring and limit setting*, *verbal communication, developmental stimulation,* and *fostering positive peer interactions* (NCKO, 2009). Each of the six skills is rated on a single 7-point rating scale (7 = very high, 6 = high, 5 = moderate/high, 4 = moderate, 3 = moderate/low, 2 = low, 1 = very low). For each of the six scales, a description is provided that starts with a general definition/explanation of the corresponding caregiver interactive skill. Next a brief description is provided distinguishing scores at the high (6, 7), middle, (3, 4, 5), and low (1, 2) ranges of the scale. For example, for the sensitive responsiveness scale, a caregiver scoring in the high (6, 7) range is described as follows:

> Shows warm and genuine interest in the children and provides emotional support when needed. In general, the caregiver responds promptly and appropriately to the children's signals, thereby functioning as a 'secure base' for the children. If unable to respond, she acknowledges having noticed the signal and provides a more complete response as soon as possible.

A caregiver scoring in the middle range (3, 4, 5)

> provides emotional support to the children, but her support is inconsistent. The emotional support she provides may vary across children and/or across time. She sometimes misses signals and her reactions are not always adequate.

A caregiver scoring in the low (1, 2) range

> hardly provides emotional support to the children. She misses many signals or her reactions are too slow or inadequate. She may show indifferent or detached behavior.

For each of the six scales the description of the high, middle, and low range of scores is followed by more detailed behavioral descriptions for each of the seven scale points.

Six trained observers independently rated the behavior of the caregiver on the six 7-point scales for each 10-min video episode of caregiver-child interactions. Next, for each caregiver a mean score for each of the six skills was calculated by averaging across the three to four episodes. Training consisted of six 4-hr sessions. During each session, two scales were thoroughly discussed and afterwards observers rated example videos. In total each observer scored 36 videos 10 min in length during training; inter-rater reliability was established on at least 80% agreement within 1 scale point with a consensus score provided by an expert. We made sure that observers who rated video episodes from a certain child care center had not visited that particular center for data collection. Also, we made sure that observers who had rated video episodes of a certain caregiver during the second round of data collection, had not rated the video episodes of the same caregiver during the

first round. This was done to keep the observers blind to relevant information that might bias their observations.

Interrater reliability (i.e., intraclass correlations), computed on 10% of the episodes, was .85 for sensitive responsiveness, .81 for respect for autonomy, .76 for verbal communication, .80 for developmental stimulation, and .83 for fostering positive peer interactions. Interrater reliability could not be computed for the structuring and limit setting scale; the scale could not be applied to most of the videotaped episodes, because the relevant caregiver behaviors did not occur during these episodes. Therefore, the structuring and limit setting scale was not included in further analysis.

In addition to the scores for the separate interaction skills, a CIP-total score was also calculated for each caregiver per measurement wave by averaging her scores on the six skills after standardization. For some analyses, group-level scores were calculated by averaging the scores of all caregivers per care group.

The CIS. The CIS (Arnett, 1989) assesses a broad range of caregiver behaviors. The CIS consists of 26 statements that are rated by a trained observer on a 4-point scale with the following anchors: 1 = not at all true, 2 = somewhat true, 3 = quite a bit true, 4 = very much true. In line with other studies (e.g., Burchinal et al., 2000; Peisner-Feinberg et al., 1999), we calculated a total score for each caregiver by averaging ratings across all items after reversing negatively formulated items. Internal consistency was high (α = .87). The CIS was scored by trained observers. Observers were trained to at least 80% agreement, within 1 scale point, on three consecutive observations. The average interrater agreement was 90% (range = 85%-98%). Interrater agreement checks during 10% of the visits resulted in similar agreement scores.

Several studies have demonstrated adequate reliability and validity for the CIS in the Dutch context (De Schipper, Riksen-Walraven, & Geurts, 2007; Fukkink & Tavecchio, 2010; Van IJzendoorn, Tavecchio, Stams, Verhoeven, & Reiling, 1998; Vermeer et al., 2008).

The ITERS-R/ECERS-R. The ITERS-R (Harms et al., 2003) and its equivalent for preschool classrooms, the ECERS-R (Harms et al., 1998), are widely used to measure process quality in child care groups. The ITERS-R was developed for use in groups in which more than 50% of the children are younger than 30 months, whereas the ECERS-R was developed for use in groups in which more than 50% of the children are between the ages of 30 and 48 months. We used these cutoff values to determine which of the two instruments was used in mixed-age groups (0- to 4-year-olds). The ITERS-R consists of 39 items and the ECERS-R consists of 43 items, and both constitute seven subscales: (a) Space and Furnishings, (b) Personal Care Routines, (c) Language, (d) Activities, (e) Interactions, (f) Program Structure, and (g) Provisions for Parents and Staff. The final subscale was not used in the present study. Items are rated on a 7-point scale with descriptors for the scores 1 (inadequate), 3 (minimal), 5 (good), and 7 (excellent). In addition to scores for the sepa-

rate subscales, we also computed a total ITERS/ECERS score by averaging scores across all items.

Prior to data collection, four observers were trained to use the ITERS-R and four observers were trained to use the ECERS-R. Each observer performed at least four on-site visits (range = 4-7) supervised by an expert trainer, and followed by an item-by-item debriefing. Interobserver agreement of 80% (within 1 scale point) on three consecutive visits was required before observers were allowed to collect data independently. The average interobserver agreement was 87% ranging from 84% to 89% for the different scales. Interobserver agreement checks during 10% of the visits resulted in similar agreement scores.

Previous Dutch research in 1995, 2001, and 2005 has reported adequate reliability and validity for the ITERS-R and ECERS-R in The Netherlands (see Vermeer et al., 2008 for an overview).

Most of the ITERS/ECERS subscales do not, like the CIP scales, focus exclusively on caregiver-child interactions, but they primarily reflect physical and organizational aspects of the caregiving environment, such as the provision of materials and organization of activities. Therefore, we expected the CIP scales to be most strongly related to those ITERS/ECERS subscales that focus most on caregiver-child interactions, namely Language and Interactions.

Caregiver sociability. Caregiver sociability was rated by the caregivers themselves using the 4-item sociability subscale of the EAS (Buss & Plomin, 1984). This scale describes the preference to be or work with others instead of being or working alone. Example items are "I like to be with people" and "I find people more stimulating rather than anything else". Caregivers rated each item on a 5-point scale with anchors ranging from 1 = not at all typical for me to 5 = very typical for me. Internal consistency (Cronbach's alpha) for the 4-item subscales was .66.

The BITSEA. The BITSEA (Briggs-Gowan, Carter, Irwin, Wachtel, & Cicchetti, 2004) was used to assess behavior problems and competence for at least two children in each care group. The BITSEA comprises two scales: the 31-item Problem scale to screen for social-emotional/behavioral problems such as aggression, defiance, overactivity, anxiety, and withdrawal, and the 11-item Competence scale to assess social-emotional abilities such as empathy, prosocial behaviors, and compliance. Items are rated on a 3-point scale with the following anchors: 0 = not true/rarely, 1 = somewhat true/sometimes, 2 = true/often. In this study, scores on both subscales were entered into the analyses as continuous variables.

Caregivers in the centers completed the BITSEA for a total of 248 children (between 1 and 4 years old) from 60 care groups. BITSEA scores were averaged across children per care group to obtain an estimate for competence and behavior problems in each of these care groups.

Results

Preliminary Analyses

Table 1 shows the descriptives for the study variables and test-retest correlations for the CIP scales. As can be seen in the table, there were considerable differences between the six skills with regard to the mean scores. Scores for sensitive responsiveness and respect for autonomy were at a moderate level, on average, with large variation between caregivers, covering almost the entire 1-7 range. Mean scores for developmental stimulation and fostering positive peer interactions were at the lowest level and showed a more restricted range, with moderate maximum scores. Test-retest correlations could be computed for 108 caregivers. As shown in Table 1, the correlations between the scores of the caregivers at Time 1 and Time 2 were all significant and moderate, supporting the test-retest reliability of the CIP scales. The correlations also reflect rank-order stability among the caregivers, indicating that, with regard to the six skills, caregivers keep their relative position (high vs. low) compared to other caregivers in the sample over time. Means and standard deviations for the CIP scales at Time 2 (not depicted in the table) were as follows: 4.30 (*SD* = 0.99) for sensitive responsiveness, 4.33 (*SD* = 0.82) for respect for autonomy, 3.27 (*SD* = 0.96) for verbal communication, 2.06 (*SD* = 0.81) for developmental stimulation, and 1.58 (*SD* = 0.63) for fostering positive peer interactions. The differences between the CIP scores at Time 1 and Time 2 were small and nonsignificant for all of the scales, indicating stability in the mean level of the CIP scores across the whole sample over time. Table 2 shows that the correlations between the CIP scales were significant and moderate, as expected. Exploratory factor analysis with oblique rotation on the six CIP scales yielded one factor explaining 68% of the variance.

Because we had data from multiple caregivers per care group and multiple groups per child care center, it was possible to examine the variation in caregiver interactive skills across classrooms and child care centers. Multilevel analysis was performed using MLwiN to study variation in caregiver interaction quality at the three different levels (center, group, caregiver) in a model without predictors (intercept model). This multilevel analysis showed no significant center or group effects for any of the scales. The largest part of the variance resided between caregivers. As is evident from the intraclass correlations in Table 3, the percentage of variance explained at the individual caregiver level was 79% for sensitive responsiveness, 80% for respect for autonomy, 72% for verbal communication, 84% for developmental stimulation, and 96% for fostering positive peer interactions. These findings demonstrate that only a minor part of the variance in caregiver skills is explained by similarity between caregivers working in the same group or in the same child care center. By far the largest part of the total variability lay at the caregiver level.

Table 1 Descriptive Statistics for the CIP Scales and the Validation Measures at Time 1, and Test-Retest Correlations for the CIP Scales

Variable	*N*	*M*	*SD*	Range	Test-retest correlations[a] *r*
CIP					
Sensitive responsiveness	145	4.44	0.90	1.67 – 6.50	.43**
Respect for autonomy	145	4.23	0.83	2.00 – 6.25	.48**
Verbal communication	145	3.38	0.92	1.25 – 6.00	.48**
Developmental stimulation	145	2.20	0.87	1.00 – 4.75	.44**
Fostering pos. peer interactions	145	1.60	0.75	1.00 – 4.75	.36**
Total	145	0.00	0.82	-1.81 – 2.78	.60**
CIS	130	3.12	0.36	1.88 – 3.73	-
Caregiver Sociability	110	3.57	0.61	1.00 – 5.00	-
ITERS-R/ECERS-R					
Space and furnishings	72	3.56	0.69	1.75 – 5.25	-
Personal care routines	72	2.22	0.74	1.00 – 5.00	-
Language	72	3.92	1.21	1.00 – 6.25	-
Interactions	72	4.46	1.40	1.00 – 7.00	-
Activities	72	2.43	0.52	1.33 – 3.89	-
Program structure	72	3.93	0.93	1.33 – 6.33	-
Total	72	3.21	0.56	1.66 – 4.34	-
BITSEA					
Problem behavior	60	0.25	0.08	0.11 – 0.49	-
Competence	60	1.40	0.24	0.73 – 1.76	-

Note. Scores for CIP, CIS, and Caregiver Sociability are at individual caregiver level, scores for ITERS-R/ECERS-R and BITSEA are at care group level.
[a]For test-retest correlations $n = 108$.
** $p < .01$.

Table 2 Pearson Correlations between CIP Scales at Time 1 ($n = 145$, Caregiver Level)

Variable	*1*	*2*	*3*	*4*
1. Sensitive responsiveness	-			
2. Respect for autonomy	.69**	-		
3. Verbal communication	.75**	.66**	-	
4. Developmental stimulation	.57**	.56**	.65**	-
5. Fostering positive peer interactions	.47**	.44**	.59**	.53**

** $p < .01$.

Table 3 Results from Multi Level Analysis on the CIP scales on Time 1: Estimates for Three-Level Model

Variable	*Level*	*Variance component*	SE	*Intraclass Correlation*
Sensitive Responsiveness	Center[a]	0.08	0.10	.10
	Group[b]	0.09	0.12	.11
	Caregiver[c]	0.63*	0.10	.79
Respect for autonomy	Center	0.04	0.08	.05
	Group	0.10	0.11	.14
	Caregiver	0.55*	0.09	.80
Verbal communication	Center	0.02	0.11	.03
	Group	0.22	0.14	.25
	Caregiver	0.61*	0.10	.72
Developmental stimulation	Center	0.12	0.07	.16
	Group	0.00	0.00	.00
	Caregiver	0.63*	0.09	.84
Fostering positive peer interactions	Center	0.02	0.04	.04
	Group	0.00	0.00	.00
	Caregiver	0.53*	0.07	.96

[a]n = 47. [b]n = 75. [c]n =145.
* $p < .05$

Validity of the CIP scales

Convergent validity was examined by correlating the CIP scores with the CIS score and the ITERS-R/ECERS-R scores. As shown in Table 4, the CIP total score and the separate CIP scale scores all correlated significantly and positively with the CIS score and with the ITERS-R/ECERS-R total score. Furthermore, with regard to the ITERS-R/ECERS-R subscales, the CIP scores were most strongly correlated with the subscales Language and Interactions, which, as compared to the other subscales, focus more on caregiver-child interactions and less on physical and organizational features of the classroom environment (see Table 4). Altogether, these findings support the convergent validity of the CIP scales.

Table 4 Validity Results: Correlations Between CIP scales and Criterion Measures

Variable	*N*	*CIP total*	*SR*	*RA*	*VC*	*DS*	*FPI*
Convergent validity							
CIS	145	.50**	.48**	.34**	.46**	.44**	.32**
ITERS-R/ECERS-R							
Total	72	.38**	.38**	.19*	.36**	.34**	.31**
Space and furnishings	72	.23**	.22**	.15	.25**	.14	.20*
Personal care routines	72	.13	.18*	-.04	.09	.12	.17*
Language	72	.34**	.30**	.22*	.32**	.32**	.23**
Interactions	72	.35**	.34**	.16*	.29**	.34**	.28**
Activities	72	.25**	.25**	.14	.27**	.17*	.19**
Program structure	72	.17*	.17*	.02	.17*	.20*	.13
Discriminant validity							
Caregiver Sociability	110	.03	.04	-.02	-.01	.02	.08
Predictive validity							
BITSEA							
Problem behavior	60	-.05	.01	-.07	-.16	.09	-.10
Competence	60	.36**	.25	.34**	.36**	.30*	.26*

Note. SR = Sensitive Responsiveness; RA = Respect for Autonomy; VC = Verbal Communication; DS = Developmental Stimulation; FPI = Fostering Positive Peer Interactions.
For CIS and Caregiver Sociability, correlations are at individual caregiver level. For ITERS-R/ECERS-R and BITSEA, correlations are at care group level.
* $p < .05$. ** $p < .01$.

Discriminant validity was examined by correlating the CIP scales with the EAS sociability score. As expected, the CIP scales were not significantly correlated with caregiver sociability (see Table 4), indicating that the general tendency to seek social interaction does not automatically imply that caregivers have the skills needed for high-quality interactions with young children. This supports the discriminant validity of the CIP scales.

Predictive validity was examined by correlating the total CIP score across caregivers in each group with the average BITSEA scores for problem behavior and competence of randomly selected children in the caregivers' groups. Table 4 shows that, as expected, children from groups with higher mean total scores on the CIP scales had higher competence scores than children from groups with lower total CIP scores. The children's problem scores, however, were not significantly correlated with the total CIP scores in their care group.

Discussion

This article describes the background and development of the CIP scales and presents the results of a first study examining the psychometric properties of the

scales, designed to rate six key skills of caregivers for interacting with 0- to 4-year-old children in child care centers. One of the scales – structuring and limit setting – could not be included in the analyses because the relevant caregiver behaviors occurred too infrequently during the videotaped episodes. For the remaining five CIP scales, interrater agreement and test-retest reliability were satisfying. The results of this study also support the convergent, discriminant and predictive validity of the new measure. The outcomes of our multilevel analysis underscore the fact that the CIP scales provide a unique profile of interaction skills of individual caregivers.

The caregivers in our sample scored higher on the more "basic" interactive skills of sensitive responsiveness and respect for autonomy (*Ms* = 4.4 and 4.2 on a 7-point scale, respectively, indicating moderate to moderate/high average levels) than on the more "educational" skills of verbal communication, developmental stimulation and fostering positive peer interactions (*M* = 2.2 and 1.6, respectively, indicating low to very low average levels). Furthermore, for developmental stimulation and fostering positive peer interactions, we did not observe the entire possible range of scores (1-7). The maximum score (averaged over the three to four episodes per caregiver) observed for these skills was 4.75. A possible explanation for the relatively low scores on these scales could be that the highest scales scores represent an unrealistically high level of skill that is very difficult for caregivers to reach. The description of high scores (6, 7) for developmental stimulation and fostering positive peer interactions, however, show that they are not unrealistically high. A high score for developmental stimulation is described as follows: "The caregiver not only provides much developmental stimulation, but she also attunes the stimulation well to the children's focus of attention, developmental level, and state, thereby encouraging their interest while preventing overstimulation". High scores (6, 7) for fostering positive peer interactions are described as follows: "The caregiver not only reacts consistently positive to spontaneous positive peer interactions, but also actively encourages positive peer interactions, by creating situations to elicit positive peer interactions or encouraging children to initiate positive peer interactions themselves". These descriptions demonstrate that high scores are not overly demanding or impossible to obtain. The weak mean scores for verbal communication, developmental stimulation, and fostering positive peer interactions, and modest maximum scores for developmental stimulation and fostering positive peer interactions are probably best explained by the traditional interpretation of child care in The Netherlands as care rather than education (OECD, 2000, 2006), that was also mentioned in the introduction of this article. The lower scores for instructional support found in this study are in line with previous research, both in The Netherlands (Albers, et al., 2010) and in the United States, the latter in studies using the CLASS (Mashburn et al., 2008;Thomason & La Paro, 2009).

Multilevel analysis showed that the largest part of the variance in caregiver skills resided at caregiver level. This suggests that caregiver behavior is for the most part determined by individual caregiver characteristics, with less influence of

group and center characteristics. The variation between caregivers was also quite large in an absolute sense, especially with regard to sensitive responsiveness, respect for child autonomy, and verbal communication. This implies that children may experience large differences in the quality of interactions with different caregivers in one and the same care group, which makes it worthwhile to measure caregiver behavior at an individual level and take into account the behavior of different caregivers when rating the quality of care experienced by children in a group. Although average classroom caregiver quality can be a valuable measure, it is not unlikely that variation in quality between caregivers is also important in explaining children's well-being and development.

Limitations

Reliability and validity could not be computed for the structuring and limit setting scale, because the relevant caregiver behaviors occurred too infrequently during the videotaped episodes to provide reliable ratings. This was due to the fact that a large proportion of the video fragments consisted of children's free play, during which caregivers mostly do not play an active structuring role in Dutch child care centers. Therefore, the filming procedure has now been adapted to explicitly include activities that call upon caregiver structuring and limit setting skills and make differences between caregivers visible (i.e., structured caregiver-child play and transitions between group activities). A recent study conducted by the NCKO indicated that the structuring and limit setting scale can indeed be reliably applied to such caregiver-child episodes (see De Kruif et al., 2010).

The significant and positive correlations of caregivers' CIP scores with the BITSEA competence scores of children in their group may be considered preliminary evidence in support of the predictive validity of the CIP scales. The expected correlation with the BITSEA problem behavior scores of the children, however, was not found. A probable explanation for this lack of correlation is the very low frequency of occurrence of problem behavior (see Table 1) among these children, who came from a relatively advantaged population; Dutch parents who choose center-based child care for their children are relatively highly educated, probably because of the relatively high costs of center-based child care in The Netherlands (see Albers et al., 2010; Gevers Deynoot-Schaub & Riksen-Walraven, 2005). It should be noted that the conclusions for predictive validity, based on the significant relation between caregiver's interaction skills and BITSEA competence scores, should be considered with caution given the concurrent correlational design of the study, which gives room for alternative explanations. For instance, it is very well possible that caregivers are better at interacting with more socially competent children. Therefore, the use of a longitudinal design is recommended for future studies of the predictive validity of the CIP scales. An alternative explanation may lie in a selection effect – that is, more competent children could have higher educated parents who choose higher quality child care centers for their children. An earlier Dutch study,

however, showed that parental level of education in The Netherlands is unrelated to the quality of caregiver-child interactions (Gevers Deynoot-Schaub & Riksen-Walraven, 2008).

Aside from the aforementioned limitations, there were also some methodological limitations. First, the response rate was relatively low (27%), which might restrict the generalizability of the results. To check for this, we compared characteristics of caregivers in our sample (i.e., age, hours worked per week, and working experience) to those available in two nationally representative samples from Dutch studies with higher response rates (Vermeer et al, 2008). We found no significant differences, but this evidence is, of course, far from conclusive, and further research with more representative samples is recommended.

Second, inter-rater reliability for the CIP scales was established on 10% of the video-recorded episodes. This is a minimal percentage to establish reliability. And third, Cronbach's alpha for the EAS sociability scale was .66, which is a bit low. Comparable and even lower alphas for the EAS sociability scale have been reported in other recent studies (Bould, Joinson, Sterne, & Araya, 2013; Nærde, Røysamb, & Tambs, 2004).

Finally, it should not be forgotten that the CIP scales have been developed for use in child care centers in The Netherlands, and, as a measure of process quality, reflect Dutch values and other characteristics of the Dutch child care context. This should be taken into account when applying the scales in other countries.

Implications for Policy and Practice

As indicated in the introduction, the CIP scales were developed originally as a more detailed measure of caregiver interactive skills to be used in combination with the ITERS-R/ECERS-R, which are regularly applied to monitor more global process quality in representative samples of Dutch child care centers. We opted for applying the CIP scales afterwards based on video recordings to make it possible for a single researcher to collect the relevant ITERS-R/ECERS-R and CIP data during a single visit to the center. The present study showed that this is indeed possible. In interesting and important next question is whether the CIP scales can also be used for live coding in child care centers. We have not yet applied the CIP scales in live observations, but this seems certainly possible. Pilot studies are needed do explore how much observation time in what situations is needed to get a valid and reliable picture of the different caregiver interactive skills using the CIP scales (with special attention to the structure and limit setting scale) during live observations. Furthermore, training for reliable live coding will be more time consuming than training for video observation, because all CIP scales have to be coded in one run, and repeated observation of the same interaction episode is not possible.

To help child care organizations monitor the quality of the care they provide, we developed the NCKO Quality Monitor (Gevers Deynoot-Schaub, Fukkink, Riksen-Walraven, Bollen, & Helmerhorst, 2013; NCKO stands for Netherlands Consortium

of Child Care Research, in which the authors participate). The Quality Monitor is a simplified version of the scientific instrument that was used in the present study and will be regularly applied for national quality assessments in representative samples of Dutch child care centers (i.e., ITERS-R/ECERS-R scales in combination with the CIP scales). The NCKO Quality Monitor has been distributed by the Ministry of Education, Culture, and Science across all child care organizations in The Netherlands and is meant to be used by professionals in child care centers/organizations to globally self-rate the process quality of care in their centers. The simplified version of the CIP scales included in the NCKO Quality Monitor uses a 3-point format (high/medium/low) and can be applied by (trained) professionals. The effects of using the NCKO Quality Monitor are currently being examined.

Furthermore, we developed a 6-week video feedback training program for caregivers in child care centers to improve caregiver interactive skills based on the six CIP scales. The effectiveness of this training program is presently being examined in a randomized controlled trial. Taking this video feedback training a step further, we can imagine a system comparable to MyTeacherPartner of the CLASS (see Allen, Pianta, Gregory, Mikami, & Lun, 2011), which also aims to improve teacher-child interactions. Teachers send their video recordings online to professional trainers. Trainers watch the submitted videos and provide the teachers with feedback on their interactions with the children. A comparable system with the CIP scales could be developed for professional caregivers in child care centers as well.

References

Ahnert, L., Pinquart, M., & Lamb, M. E. (2006). Security of children's relationships with nonparental care providers: A meta-analysis. *Child Development*, *74*, 664-679. doi: 10.1111/j.1467-8624.2006.00896.x

Ainsworth, M. D. S., Bell, S. M., & Stayton, D. J. (1974). Infant-mother attachment and social development: Socialisation as a product of reciprocal responsiveness to signals. In M.P.M. Richards (Eds.). *The integration of a child into a social world*. (pp. 99-137). New York, NY: Cambridge University Press.

Ainsworth, M. D., Blehar, M. C., Waters, E., & Wall, S. (1978). *Patterns of attachment: A psychological study of the strange situation*. Hillsdale, NJ: Erlbaum.

Albers, E. M., Riksen-Walraven, J. M. A., & De Weerth, C. (2010). Developmental stimulation in child care centers contributes to young infants' cognitive development. *Infant Behavior & Development, 33*, 401-408. doi: 10.1016/j.infbeh.2010.04.004

Allen, J. P., Pianta, R. C., Gregory, A., Mikami, A. Y., & Lun, J. (2011). An interaction-based approach to enhancing secondary school instruction and achievement. *Science*, *333*, 1034-1037. doi: 10.1126/science.1207998

Arnett, J. (1989). Caregivers in day-care centers: Does training matter? *Journal of Applied Developmental Psychology*, *10*, 541-552. doi: 10.1016/0193-3973(89)90026-9

Arnold, D. H., McWilliams, L., & Arnold, E. H. (1998). Teacher discipline and child misbehavior in day care: Untangling causality with correlational data. *Developmental Psychology*, *34*(2), 276-287.

Baumrind, D. & Black, A. E. (1967). Socialization practices associated with dimensions of competence in preschool boys and girls. *Child Development, 38*(2), 291-327.

Belsky, J., Vandell, D. L., Burchinal, M., Clarke-Stewart, K. A., McCartney, K., Owen, M. T., & the NICHD Early Child Care Research Network (2007). Are there long-term effects of early child care? *Child Development, 78*, 681-701. doi: 10.1111/j.1467-8624.2007.01021.x

Berk, L. (2010). *Child Development*. Boston: Allyn & Bacon.

Bloom, L. (1998). Language acquisition in its developmental context. In: W. Damon, D. Kuhn, & S. Siegler (Eds.), *Handbook of child psychology, part 2: Cognition, perception, and language* (pp. 309-371). New York, NY: Wiley.

Bould, H., Joinson, C., Sterne, J., & Araya, R. (2013). The emotionality activity sociability temperament survey: Factor analysis and temporal stability in a longitudinal cohort. *Personality and Individual Differences*, *54*, 628-633. doi: org/10.1016/j.paid.2012.11.010

Bowlby, J. (1969). *Attachment and Loss, Vol. 1. Attachment.* New York: Basic Books.

Briggs-Gowan, M. J., Carter, A. S., Irwin, J. R., Wachtel, K., & Cicchetti, D. V. (2004). The Brief-Toddler Social and Emotional Assessment: Screening for social-emotional problems and delays in competence. *Journal of Pediatric Psychology*, *29*, 143-155. doi: 10.1093/jpepsy/jsh017

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

Buss, A., & Plomin, R. (1984). *Temperament: Early developing personality traits*. Hillsdale, NJ: Lawrence Erlbaum.

Campbell, D.T., & Fiske, D.W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin, 56(2),* 81-105.

De Kruif, R. E. L., Riksen-Walraven, J. M. A, Fukkink, R. G., Tavecchio, L. W. C., & Van Zeijl, J. (2009, April). *What constitutes high-quality center child care? Stakeholders' views*. Poster presented at the meeting of Society for Research in Child Development, Denver, CO.

De Kruif, R. E. L., Riksen-Walraven, J. M. A., Gevers Deynoot-Schaub, M. J. J. M., Helmerhorst, K. O. W., Tavecchio, L. W. C., & Fukkink, R. G. (2010). *Pedagogische kwaliteit van de opvang van 0- tot -4 jarigen in Nederlandse kinderdagverblijven 2008* [Pedagogical quality of child care centers for 0- to 4-year-olds in The Netherlands 2008]. Amsterdam: NCKO.

De Schipper, E. J., & Riksen-Walraven, J. M. A. (2004). *Rating scales for the quality of interactions during structured play between caregivers and groups of children in child care centers*. Unpublished Manuscript, Radboud University Nijmegen, The Netherlands.

De Schipper, E. J., Riksen-Walraven, J. M. A., & 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

De Schipper, E. J., Riksen-Walraven, J. M. A., & Geurts, S. A. E. (2007). Multiple determinants of caregiver behavior in child-care centers. *Early Childhood Reserarch Quarterly*, *22*, 312-326. doi:10.1016/j.ecresq.2007.04.004

De Schipper, J. C., Van IJzendoorn, M. H., & Tavecchio, L. W. C. (2004). Stability in center day care: Relations with children's well-being and problem behavior in day care. *Social Development*, *13*, 531-550. doi: 10.1111/j.1467-9507.2004.00282.x

De Wolff, M., & Van IJzendoorn, M. H. (1997). Sensitivity and attachment: A meta-analysis on parental antecedents of infant attachment. *Child Development, 68* (4), 571-591.

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

Egeland, B., Pianta, R., & O'Brian, M. A. (1993). Maternal intrusiveness in infancy and child maladaptation in early school years. *Developmental and Psychopathology, 5* (3), 359-370.

Eisenberg, N., & Fabes, R. A. (1998). Prosocial development. In: W. Damon & N. Eisenberg (Eds.), *Handbook of child psychology, part 3: Social, emotional, and personality development* (pp. 701-778). New York, NY: Wiley.

Erickson, M., Sroufe, L. A., & Egeland, B. (1985). The relationship between quality of attachment and behavior problems in preschool in a high risk sample. In I. Bretherton and E. Waters (Eds.), Growing points of attachment theory and research. *Monographs of the Society for Research in Child Development, 50* (1/2), 147-166.

Erikson, E. H. (1950). *Childhood and society.* New York: Norton.

Feldman, R. A., & Klein, P. S. (2003). Toddlers' Self-Regulated Compliance to Mothers, Caregivers, and Fathers: Implications for Theories of Socialization. *Developmental Psychology*, *39* (4), 680-692.

Fukkink, R. G., & Tavecchio, L. W. C. (2010). Effects of video interaction guidance on early childhood teachers. *Teacher and Teaching Education, 26*, 1652-1659. doi: 10.1016/j.tate.2010.06.016

Gevers Deynoot-Schaub, M., Fukkink, R., Riksen-Walraven, M., Bollen, I., & Helmerhorst, K. (2013, April). *Applying the quality monitor to compare the quality of caregiver-child interactions in child care centers during lunch and free play*. Poster presented at the meeting of Society for Research in Child Development, Seattle, WA.

Gevers Deynoot-Schaub, M. J. J. M., & Riksen-Walraven, J. M. A. (2005). Child care under pressure: The quality of Dutch centers in 1995 and 2001. *Journal of Genetic Psychology, 166*, 280-296. doi: 10.3200/GNTP.166.3.280-296

Gevers Deynoot-Schaub, M. J. J. M., & Riksen-Walraven, J. M. A. (2006a). Peer contacts of 15-month-olds in child care: Links with child temperament, parent-child interaction and quality of child care. *Social Development*, *15*, 709-729. doi: 10.1111/j.1467-9507.2006.00366.x

Gevers Deynoot-Schaub, M. J. J. M., & Riksen-Walraven, J. M. A. (2006b). Peer interaction in child care centres at 15 and 23 months: Stability and links with children's socio-emotional adjustment. *Infant Behavior & Development, 29*, 276-288. doi: 10.1016/j.infbeh.2005.12.005

Gevers Deynoot-Schaub, M. J. J. M., & Riksen-Walraven, J. M. A. (2008). Infants in group care: Their interactions with professional caregivers and parents across the second year of life. *Infant Behavior & Development, 31*, 181-189. doi: 10.1016/j.infbeh.2007.10.010

Girolametto, L., & Weitzman, E. (2002). Responsiveness of child care providers in interactions with toddlers and preschoolers. *Language, Speech and Hearing Services in Schools, 33* (4), 268-281.

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

Girolametto, L., Weitzman, E., & Greenberg, J. (2004). The effects of verbal support strategies on small-group peer interactions. *Language, Speech, and Hearing Services in Schools, 35* (3), 254-268.

Harms, T., Clifford, R., & Cryer, D. (1980). *Early Childhood Environment Rating Scale*. New York, NY: Teachers College Press.

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

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

Harms, T., Cryer, D., & Clifford, R. M. (2003). *Infant/Toddler Environment Rating Scale-Revised*. New York, NY: Teachers College Press.

Howes, C. (2000). Social-emotional classroom climate in child care, child-teacher relationships and children's second grade peer relations. *Social Development*, *9*, 191-204. doi: 10.1111/1467-9507.00119

Howes, C., Galinsky, E., & Kontos, S. (1998). Child care caregiver sensitivity and attachment. *Social Development, 7*, 25-36. doi: 10.1111/1467-9507.00048

Howes, C., & Phillipsen, L. (1998). Continuity in children's relations with peers. *Social Development*, *7*, 340-349. doi: 10.1111/1467-9507.00071

Klein, P. S., & Feldman, R. (2007). Mothers' and caregivers' interactive and teaching behavior with toddlers. *Early Child Development and Care*, *177*, 383-402. doi: 10.1080/03004430600551682

Lamb, S., & Zakhireh, B. (1997). Toddlers' attention to the distress of peers in a daycare setting. *Early Education and Development*, *8*, 105-118.

La Paro, K. M., Hamre, B. K., & Pianta, R. C. (2011). *Classroom assessment scoring system, Toddler manual*. Baltimore, Brookes.

Mashburn, A. J., Pianta, R. C., Hamre, B. K., Downer, J. T., Barbarin, O. A., Bryant, D., Burchinal, M., Early, D. M., & 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

Merens, A., Hartgers, M., & Van den Brakel, M. (2012). Emancipatiemonitor 2012 [Emancipation Monitor 2012]. The Hague, Sociaal en Cultureel Planbureau / Centraal Bureau voor de Statistiek.

Nærde, A., Røysamb, E., & Tambs, K. (2004). Temperament in adults – reliability, stability, and factor structure of the EAS temperament survey. *Journal of Personality Assessment*, *82* (1), 71-79.

NCKO (2009). *Caregiver Interaction Profile* (2nd edition). Amsterdam/Nijmegen, The Netherlands: Netherlands Consortium for Child Care Research.

NICHD Early Child Care Research Network (1996). Characteristics of infant care: Factors contributing to positive caregiving. *Early Childhood Research Quarterly*, *11*, 269-306. doi: 10.1016/S0885-2006(96)90009-5

NICHD Early Child Care Research Network (2000). The relation of child care to cognitive and language development. *Child Development*, *71* (4), 960-980.

NICHD Early Child Care Research Network (2001). Child care and children's peer interaction at 24 and 36 months: The NICHD Study of Early Child Care. *Child Development*, *72* (5), 1478-1500.

NICHD Early Child Care Research Network (2005). Early child care and children's development in the primary grades: Results from the NICHD Study of Early Child Care. *American Educational Research Journal, 42* (3), 537-570.

OECD (2000). *Early childhood education and care policy in The Netherlands. Background report to the OECD-project.* The Hague: Ministry of Health, Welfare & Sport, Ministry of Education, Culture and Science.

OECD (2006). Starting Strong II: Early childhood education and care. Paris: Organisation for Economic Co-operation and Development.

OECD (2011). *PF3.2: Enrolment in childcare and pre-schools*. Retrieved from OECD website: http://www.oecd.org/els/soc/37864698.pdf

Peisner-Feinberg, E. S., Burchinal, M. R., Clifford, R. M., Yazejijan, N., Culkin, M. L., Zelazo, J., Howes, C., Byler, P., Kagan, S.L., & Rustici, J. (1999). *The children of the Cost, Quality and Outcome Study go to school: Technical Report.* Chapel Hill, NC: University of North Carolina.

Pianta, R. C., La Paro, K. M., & Hamre, B. K. (2008). *Classroom assessment scoring system (CLASS). PRE-K manual.* Baltimore, Brookes.

Riksen-Walraven, J. M. A., Meij, J. Th., Hubbard, F. O., & Zevalkink, J. (1996). Intervention in lower-class Surinam-Dutch families: Effects on mothers and infants*. International Journal of Behavioral Development, 19 (4)*, 739-756. doi: 10.1080/016502596385550

Riksen-Walraven, J. M. A., & Van Aken, M. A. G. (1997). Effects of two mother-infant intervention programs upon children's development at 7, 10, and 12 years. In: W. Koops, J. Hoeksma & D. van den Boom (Ed.), *Development of interaction and attachment: Traditional and non-traditional approaches* (pp.79-93). Amsterdam: North-Holland.

Smeekens, S., Riksen-Walraven, J. M. A., & Van Bakel, H. J. A. (2007). Multiple determinants of externalizing behavior in 5-year-olds: A longitudinal model. *Journal of Abnormal Child Psychology, 35(3), 347-361.* doi: 10.1007/s10802-006-9095-y

Sroufe, L. A. (1979). The coherence of individual development: Early care, attachment, and subsequent developmental issues. *American Psychologist, 34(10),* 843-841.

Thomason, A. C., & La Paro, K. M. (2009). Measuring the quality of teacher-child interactions in toddler child care. *Early Education and Development, 20*(2), 285-304. doi:10.1080/10409280902773351

Thompson, R. (1998). Early sociopersonality development. In W. Damon & N. Eisenberg (Eds.), *Handbook of child psychology, part 3: Social, emotional, and personality development* (pp. 25-104). New York, NY: Wiley.

Turiel, E. (1998). The development of morality. In: W. Damon & N. Eisenberg (Ed*.), Handbook of child psychology, part 3: Social, emotional, and personality development* (p. 863-932). New York, NY: Wiley.

Vandell, D. L. (2004). Early child care: The known and the unknown. *Merill-Palmer Quarterly*, 50, 387-414. doi: 10.1353/mpq.2004.0027

Vandell, D., Belsky, J., Burchinal, M., Steinberg, L., Vandergrift, N., & the NICHD Early Child Care Research Network (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

Vandell, D. L., & Wolfe, B. (2000) *Child Care quality: does it matter and does it need to be improved?* Madison: Institute for Research on Poverty, University of Wisconsin-Madison.

Van IJzendoorn, M. H., Tavecchio, L. W. C., Stams, G. J. J. M., Verhoeven, M. J. E., & Reiling, E. J. (1998). Quality of center day care and attunement between parents and caregivers: Center day care in cross-national perspective. *The Journal of Genetic Psychology*, *159*, 437-454. doi: 10.1080/-00221329809596163

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: 10.1080/00221329809596163

Williams, S. T., Mastergeorge, A. M., & Ontai, L. L. (2010). Caregiver involvement in infant peer interactions: Scaffolding in a social context. *Early Childhood Research Quarterly*, *25*, 251-266. doi: 10.1016/j.ecresq.2009.11.004

Zahn-Waxler, C., Radke-Yarrow, M., & King, R. M. (1979). Childrearing and children's prosocial initiations toward victims of distress. *Child Development, 50,* 319-330.

CHAPTER 3
Child Care Quality in The Netherlands over the Years: A Closer Look

Foto: Joods Historisch Museum

Helmerhorst, K. O. W., Riksen-Walraven, J. M. A., Gevers Deynoot-Schaub, M. J. J. M., Tavecchio, L. W. C., & Fukkink, R. G. (2014). Child care quality in The Netherlands over the years: A closer look. *Early Education and Development*. doi: 10.1080/10409289.2014.948784

Abstract

We assessed the quality of child care in a nationally representative sample of 200 Dutch child care centers using the ITERS-R/ECERS-R and compared it with the previous assessment in 2005. The Caregiver Interaction Profile (CIP) scales were used to rate the quality of caregiver-child interactions. Results show a significant and substantial decline in quality as compared to 2005, with 49% of the groups now scoring below the minimal level. The CIP scales showed relatively high scores for the basic caregiver interactive skills sensitive responsiveness, respect for autonomy, and structuring and limit setting, but much lower scores for the more educational skills verbal communication, developmental stimulation, and fostering positive peer interactions. Caregiver sensitive responsiveness was significantly lower in infant groups (0-2 years) than in preschool groups (2-4 years); caregiver respect for autonomy, verbal communication, developmental stimulation, and fostering positive peer interactions were significantly lower in infant groups than in preschool groups and mixed-age groups (0-4 years). Quality of child care is not stable across the years and regular quality assessments are therefore needed to monitor child care quality. The low scores on the more educational versus the more basic caregiver interactive skills indicate that these skills deserve more attention in caregiver education and training. Training programs should be attuned to the individual "interaction skill profile" of caregivers.

3.1 Introduction

Quality of early child care can be broadly defined as the extent to which the child care environment offers experiences to the children that enhance their wellbeing and development (see, e.g., Layzer & Goodson, 2006). There is now ample empirical evidence that high quality child care may have considerable and enduring effects on children's cognitive and socio-emotional development (for an overview see Belsky et al., 2007; Vandell & Wolfe, 2000; Vandell et al., 2010). It is therefore important and in the interest of children to closely monitor the quality of care and improve it when needed.

In the Netherlands, repeated quality assessments in nationally representative samples of child care centers have shown a steady decrease in average quality from 1995 to 2001 and 2005 (Gevers Deynoot-Schaub & Riksen-Walraven, 2005; Van IJzendoorn, Tavecchio, Stams, Verhoeven, & Reiling, 1998; Vermeer et al., 2008). In these Dutch studies, child care quality was assessed using the Infant/Toddler Environment Rating Scale (ITERS-R; Harms, Cryer, & Clifford, 2003) and its equivalent for preschool classrooms, the Early Childhood Environment Rating Scale (ECERS-R; Harms, Clifford, & Cryer, 1998). These scales are used to rate the global quality of the child care environment in different domains, such as space and furnishings, personal care routines, language, activities, interactions, and program structure. In 1995, this instrument was used for the first Dutch child care quality assessment (Van IJzendoorn et al., 1998), because it was used in many countries and therefore allowed international comparison (e.g., Tietze, Cryer, Bairrão, Palacios, & Wetzel, 1996). In later national quality assessments, we retained the ITERS-R and ECERS-R as measures to be able to compare global quality ratings over time. The present study reports the results of the fourth Dutch quality assessment, in 2008, with the ITERS-R and ECERS-R in a nationally representative sample of child care centers. In addition to the ITERS-R and ECERS-R, which globally assess a broad range of aspects of child care environment, we developed the Caregiver Interaction Profile (CIP) scales (see Helmerhorst, Riksen-Walraven, Vermeer, Fukkink, & Tavecchio, 2014) to gain a more detailed picture of what is generally considered the core aspect of child care quality (i.e. the quality of interactions between professional caregivers and the children in their care; Layzer & Goodson, 2006; Vandell & Wolfe, 2000). This is the first time the CIP scales have been applied in a nationally representative sample in combination with the ITERS-R and ECERS-R, yielding both a broad and detailed picture of the quality of care and education provided in Dutch centers.

Before specifying the aims of the study, we provide more information on the Dutch child care context. Today, the vast majority of Dutch mothers with young children (> 70%) are employed, although mostly part time. Parents of children attending formal child care are generally highly educated, which is explained by the relatively high costs of formal care (Merens, Hartgens, & Van den Brakel, 2012). Children attend child care from age 3 months onwards (when paid maternity leave

ends), to age 4 years (when kindergarten starts). The vast majority of caregivers have characteristically received vocational training at intermediate level (SPW-3), which is not specifically directed at very young children and focuses on care rather than education. This focus is also reflected in the typical daily program of child care centers in the Netherlands, which consists mostly of free play with little time allotted to structured educational activities (OECD, 2000, 2006).

The first aim of the present study was to compare the quality of center-based child care in 2008 as assessed with the ITERS-R and ECERS-R to the quality in earlier assessments, especially the previous assessment in 2005. The decrease in quality assessed with the ITERS-R and ECERS-R observed in the Netherlands between 1995 and 2005 was significant and substantial. In 1995, the average overall quality score on the ITERS/ECERS was 4.8 on a 7-point scale, which was relatively high in international perspective (Van IJzendoorn et al., 1998). In 2001, the average overall quality had decreased to 4.3 (Gevers Deynoot-Schaub & Riksen-Walraven, 2005) and in 2005 it had further decreased to 3.2, which is low, both in absolute terms and in international perspective (Vermeer et al., 2008). Two main factors have been proposed to explain the observed decrease in child care quality across the years 1995-2005. First, the explosive growth in the number of young children attending child care, together with a shortage of qualified staff and increased workload, may have put quality under pressure. A second possible explanation lies in changes in the vocational training of child care professionals in the Netherlands in the past decade. In earlier years, caregivers received specialized vocational training focused on working with very young children in a child care setting (*LKC*; Caregiver Child Care Centers), while from1996 onwards most caregivers receive a more general social pedagogical 3-year training (*SPW-3*; Social Pedagogical Work) that prepares them for working with a broader variety of age groups in different domains of care (Gevers Deynoot-Schaub & Riksen-Walraven, 2005; Vermeer et al., 2008).

It is difficult to predict whether child care quality has further declined since the last measurement in 2005 or whether the decline has stopped or even turned into an increase. The number of children attending child care centers has continued to increase after 2005 (in 2008, still more than 31,000 children were on waiting lists of child centers, see Van Beem & Wever, 2008), so this probably unfavorable influence on quality remained. An important policy change that may have affected child care quality in the Netherlands in the past years is the introduction of the Child Care Act in 2005. This Act redirected financial support from providers to parents to give parents more freedom to choose child care that fits their needs. This may lead to quality improvement if parents choose high-quality child care for their children, but it is questionable whether such an effect – if any (see also Plantenga, 2012; Vandell & Wolfe, 2000) – could already be visible three years after introduction of the Act. In addition, the introduction of the Child Care Act did not change regulations for quality monitoring, which could have led to a change in quality. In the Netherlands, quality monitoring is regulated by the Association of Community Health Services (GGD), which is responsible for (unannounced) inspection and

enforcement of quality in child care centers. Based on this, there was no ground for formulating hypotheses regarding possible changes in the quality of child care since the former national quality assessment in 2005.

The second aim of the present study was to apply the newly developed CIP scales as an extension of the ITERS-R/ECERS-R in this large and nationally representative sample. The ITERS-R and ECERS-R are very useful to apply in regular national quality assessments, because they provide a comprehensive picture of the quality of the child care environment in a broad set of domains, including caregiver-child interactions. In addition to this more global picture, we aimed to get a more in-depth picture of the interactions between the professional caregivers and children by applying the CIP scales to video recordings made in the centers. The CIP scales were especially developed to be used in combination with the ITERS-R/ECERS-R in national quality assessments in Dutch child care centers. The CIP scales rate six important caregiver skills in interacting with 0- to 4- year-old children in a group setting (see the Method section). The results of a previous study supported the validity of the scales, suggesting that the CIP scales are indeed a valuable extension of the ITERS-R and ECERS-R (Helmerhorst et al., 2014).

The third aim of the study was to relate the process quality of child care, as measured by the ITERS-R/ECERS-R and the CIP scales, to three structural features of child care groups, namely group size, caregiver-child ratio, and group type (age composition). Research has demonstrated the relation between process quality (especially caregiver-child interactions) and the structural features group size and caregiver-child ratio; in general, higher process quality is observed in smaller groups and with fewer children per caregiver (De Schipper, Riksen-Walraven, & Geurts, 2006; National Institute of Child Health and Human Development Early Child Care Research Network [NICHD ECCRN], 1996, 2002; Vandell & Wolfe, 2000). Therefore, we expected a smaller group size and fewer children per caregiver to be associated with higher process quality, and especially with higher quality of caregiver-child interactions as measured with the CIP scales and, more globally, with the Interactions subscale of the ITERS-R/ECERS-R. With regard to age composition, we distinguished three types of groups: infant groups (0- to 2-year-olds), preschool groups (2- to 4-year-olds) and mixed-age groups (0- to 4-year-olds). The mixed-age group, with its relatively wide age range, is predominant in Dutch centers, while it is relatively unknown in other countries. Although there has been much discussion among professionals and parents in the Netherlands about the pros and cons of especially mixed-age versus infant groups, both for children and for caregivers, relevant empirical evidence is still lacking. Therefore, we also explored the relation between group type and the process quality of care provided in the groups in the present representative sample.

Internationally, caregiver education is also considered an important structural quality feature, and research in various countries has shown it to be associated with child care quality (e.g., NICHD ECCRN, 2002; for a review see Vandell & Wolfe, 2000). But in the Netherlands, earlier studies have consistently failed to show an

association between caregiver education and child care quality. This is probably due to the lack of variation in education among caregivers in Dutch child care centers; in the Netherlands, the vast majority of caregivers have received the same vocational training at intermediate level (see e.g., De Schipper, Riksen-Walraven, & Geurts, 2007; Helmerhorst et al., 2014; Vermeer et al, 2008). Given this lack of variation in caregiver education we did not include it as a structural feature in the present study.

To summarize, this study had three aims. First, we examined the global process quality of child care in a nationally representative sample of child care centers in the Netherlands with the ITERS-R/ECERS-R and compared the results of the present assessment to the previous assessment in 2005. Second, we used the newly developed CIP scales to observe the quality of caregiver-child interactions in more detail and related this quality to child care quality as assessed with the ITERS-R/ECERS-R. And third, we examined group size, caregiver-child ratio, and group type in relation to process quality, and especially to the quality of caregiver-child interactions.

3.2 Method

Participants

A random sample of 200 child care centers was taken from a nationally representative pool of approximately 3000 child care centers located across The Netherlands. The sample was stratified by region. Four backup samples per child care center were selected that could be approached in case the center did not participate. Participation rate was 33%. The main reason for refusal was being "too busy". Child care centers were approached by letter with an explanation of the purpose and procedures of the study and an invitation to cooperate. One week after the mailing, the center directors received a phone call to ask about their willingness to participate in the study. The final sample of 200 child care groups included 55 infant groups (0- to 2-year olds), 52 preschool groups (2- to 4-year olds), and 93 mixed-age groups (0- to 4-year olds). We randomly selected one care group per child care center. All caregivers of a selected care group were invited to participate in the study. The final sample include 425 caregivers, who were mostly female (98%), were born in The Netherlands (90%), had an average age of 32.7 years (*SD* = 9.7, range 19-64), worked 27.9 hours a week (*SD* = 6.5), on average, and had an average of 8.1 years (*SD* = 6.1) of working experience in childcare. The vast majority of the caregivers (81%) had completed the regular vocational education at intermediate level. Only 3% had not (yet) completed this regular vocational education, and 12% had completed vocational education at a higher level. Just like the education at the intermediate level, the higher education was also more general and not specifically directed at early child care. This distribution shows there is hardly any variation in

caregiver education, which is in line with previous research in child care centers in The Netherlands De Schipper, Riksen-Walraven, & Geurts, 2007; Helmerhorst et al., 2014; Vermeer et al, 2008). All participating caregivers and all parents of the children from the selected groups gave their active consent for filming procedures.

Procedure

Observations took place between September 2007 and January 2009. Each group was visited by a trained researcher from about 8 a.m. until approximately 3 p.m. . The researcher completed the ITERS-R or ECERS-R and also made video recordings of caregivers in interaction with the children throughout the day (for applying the CIP scales afterwards). Each individual caregiver was filmed for 8 to 10 minutes in four different situations: diapering, lunch/snack, free play, and transition between group activities. Caregivers were not given any specific instructions in advance and were asked to follow the regular program, in order to capture a day as usual. In accordance with the ITERS-R/ECERS-R manual, the researcher had a short interview with one of the caregivers in the afternoon to obtain information on ITERS-R/ECERS-R items that could not be observed during the visit. Prior to the visit, caregivers had been sent questionnaires in order to gather individual background information (e.g. age, education, work experience, and working hours).

Measures

ITERS-R/ECERS-R. The Infant/Toddler Environment Rating Scale-Revised (ITERS-R; Harms et al., 2003) and the Early Childhood Environment Rating Scale-Revised (ECERS-R; Harms et al., 1998), were used to measure global process quality in the child care groups. The ITERS-R was developed for use in groups in which more than 50% of the children is under the age of 30 months, whereas the ECERS-R was developed for use in groups in which more than 50% of the children is between the ages of 30 and 48 months. We used these cut-off values to determine which of the two instruments was used in mixed-age groups (0- to 4-year-olds). The ITERS-R consists of 39 items and the ECERS-R consists of 43 items, and both constitute seven subscales: (a) Space and Furnishings, (b) Personal Care Routines, (c) Language, (d) Activities, (e) Interactions, (f) Program Structure, and (g) Provisions for Parents and Staff. The subscales Personal Care Routines and Provisions for Parents and Staff were not used in the present study. Items are rated on a 7-point scale with descriptors for score 1 (inadequate), 3 (minimal), 5 (good), and 7 (excellent). In addition to mean scores for the separate subscales, we also computed a total ITERS-R/ECERS-R mean score by averaging scores across all items. Based on the mean scores, groups were classified according to the quality levels *low* ($M < 3$), *moderate* ($3 \leq M < 5$) and *high* ($M \geq 5$) scores.

Prior to data collection, eight observers were trained to use the ITERS-R, six observers were trained to use the ECERS-R, and three observers were trained to

use both instruments. Each observer performed at least 3 on-site visits (range 3-6) supervised by an expert trainer, and followed by an item-by-item debriefing. Interobserver agreement of 80% (within 1 scale point) on three consecutive visits was required before observers were allowed to collect data independently. The average interobserver agreement was 88%, ranging from 81% to 95% for the different scales. Interobserver agreement checks during 10% of the visits was 89%.

The Caregiver Interaction Profile (CIP) scales (for a more detailed description of development and first validation results see Helmerhorst et al., 2014) measure six caregiver interactive skills: (1) *sensitive responsiveness* or the extent to which a caregiver recognizes children's individual emotional and physical needs, and responds appropriately and promptly to their cues and signals; (2) *respect for autonomy* or the extent to which a caregiver is non-intrusive but instead recognizes and respects the validity of children's intentions and perspectives; (3) *structuring and limit setting* or the ability of a caregiver to clearly communicate expectations towards children and structure the situation accordingly and to set clear and consistent limits to the children's behavior; (4) *verbal communication* or the frequency and quality of verbal interactions between caregiver and children; (5) *developmental stimulation* or the degree to which a caregiver deliberately attempts to foster children's development (e.g., motor development, cognitive development and creativity), while attuning the stimulation to the children's focus of attention, developmental level, and state; (6) *fostering positive peer interactions* or a caregiver's guidance of interactions between children in the child care group. Each of the six CIP scales is rated on a single 7-point Likert-type scale (7 = *very high*, 6 = *high*, 5 = *moderate/high*, 4 = *moderate*, 3 = *moderate/low*, 2 = *low*, 1 = *very low*). In the extensive description of each of the six scales, first a general definition of the corresponding caregiver interactive skill is given, followed by a brief description distinguishing scores at the high (6, 7), middle (3, 4, 5), and low (1, 2) ranges of the scale. Finally, for each of the scales, a detailed behavioral description for each of the seven scale points is provided. For a more comprehensive description of the CIP scales see Helmerhorst et al. (2014).

Fifteen observers independently rated the videotaped caregiver-child interactions episodes. Observers attended six training sessions before applying the CIP scales. The training was followed by a test, and intra-class correlations were computed. Observers had to meet a criterion of .70 for each scale before they could start to observe. Observers independently rated the behavior of the caregiver on the six 7-point scales for each of the four videotaped episodes. Per caregiver a mean score for each of the six skills was calculated by averaging across the four episodes. For analyses at the level of the child care groups, scores of all caregivers per group were averaged. Analogous to the ITERS-R/ECERS-R, mean scores on the CIP scales were classified in three quality levels. The levels are labeled *inadequate* ($M < 3.5$), *moderate* ($3.5 \leq M < 4.5$) and *adequate to good* ($M \geq 4.5$).

Inter-rater reliability (i.e., intraclass correlations computed for 10% of the tapes) was .83, on average.

Group size and caregiver-child ratio. Group size was defined as the total number of children present in the group during the visit. Caregiver-child ratio was defined as the total number of children divided by the total number of caregivers in the group during the visit.

3.3 Results

Global Process Quality as Compared to Earlier Assessments

Table 1 presents the means and standard deviations for the ITERS-R/ECERS-R scores for the present assessment (2008) and the earlier assessments (1995, 2001, and 2005). The table shows that the declining trend both for the subscales and the total score from 1995 to 2001 and 2005 (see also Vermeer et al., 2008) has continued to 2008. The t-tests show that the mean total score and the mean scores for the space and furnishings subscale and the interactions subscale in the present assessment are substantially and significantly lower than the scores for the preceding assessment in 2005. Effect sizes are d = 0.52, d = 0.50, and d = 1.01, respectively. The latter effect size, for the decrease of the total score, is strong in terms of Cohen's (1988) definition (d = 0.20 is a weak effect, d = 0.50 a modest effect, and d = 0.80 a strong effect). In terms of the ITERS-R/ECERS-R scale point definitions, the present ITERS-R/ECERS-R total score (M = 3.0, SD = 0.60) is just at the minimal level. For the subscales, the lowest score was obtained for activities (M = 2.2, SD = 0.58), followed by space and furnishings (M = 3.0, SD = 0.78), language (M = 3.4, SD = 1.08), program structure (M = 3.7, SD = 1.03), and interactions (M = 3.8, SD = 1.17).

Figure 1 shows the percentages of child care groups with high, moderate, and low scores in the present assessment in terms of the predefined ITERS-R/ECERS-R quality categories. With regard to the total score (see right-hand bar in the figure), none of the groups scored in the high category, 51% scored moderate, and 49% scored low. In the preceding assessment in 2005 there were no high total scores either; 64% of the groups scored moderate and 36% scored low. This indicates that between 2005 and 2008, 13% of the groups shifted from the moderate to the low category.

As can be seen in Figure 1, results from the present assessment demonstrate that for the space and furnishings subscale and for the activities subscale none of the groups scored high, and the vast majority of the groups scored low. For the subscales language, interactions and program structure, a minor percentage of groups (between 10% and 16 %) scored high, the majority of the groups scored moderate, and between 18 and 33% of the groups scored low.

Table 1 ITERS-R/ECERS-R Scores, Group size, and Child-Caregiver Ratio in 1995, 2001, 2005, and 2008, with t-Values and Effect Sizes for Differences Between 2005 and 2008

	1995 (N = 43)		*2001 (N = 51)*		*2005 (N = 42)*		*2008 (N = 200)*		*2008 vs 2005*	
Variable	M	SD	M	SD	M	SD	M	SD	t	*Cohen's* d
ITERS-R/ECERS-R										
Space and Furnishings	5.1	0.98	4.9	1.07	3.4	.69	3.0	.78	-3.08**	.52
Personal Care Routines	4.3	1.06	4.4	1.11	-	-	-	-	-	-
Language	4.7	1.01	3.7	1.30	3.5	1.13	3.4	1.08	-0.54	.09
Activities	4.3	0.82	3.6	.71	2.4	.69	2.2	.58	-1.96	.33
Interactions	5.3	0.97	5.1	1.46	4.4	1.37	3.8	1.17	-2.93**	.50
Program structure	5.2	1.24	-	-	3.9	1.00	3.7	1.03	-1.15	.20
Total[a]	4.8	0.61	4.3	.74	3.6	.57	3.0	.60	-5.94**	1.01
Group size	11.1	4.0	9.1	2.6	9.7	2.6	10.9	3.2	2.28*	.39
Caregiver-child ratio	.27	0.12	.26	0.08	.22	0.05	.21	0.05	-1.18	.20

Note. [a]Total scores for 1995 and 2001 are based on all items, including Personal Care Routines, the total scores for 2005 and 2008 are based on all items, excluding Personal Care Routines.
* $p < .05$, ** $p < .01$

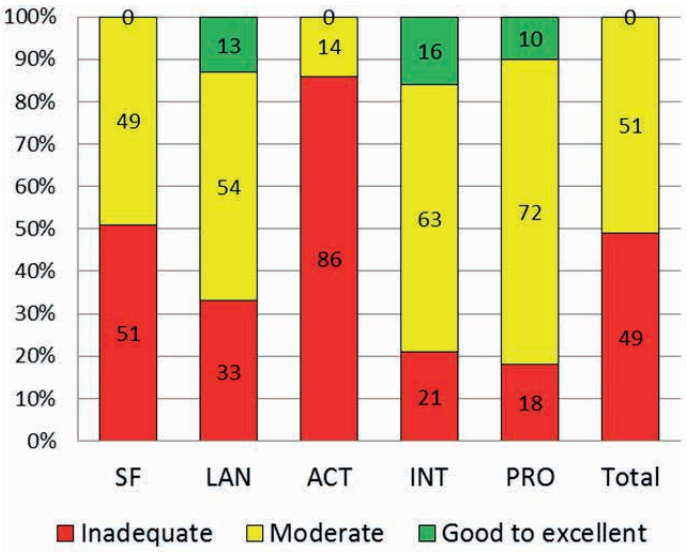

Figure 1 Percentages of care groups (N = 200) with Mean Scores "Inadequate", "Moderate", and "Good to Excellent" on the ITERS-R/ECERS-R. SF = Space and furnishings; LAN = Language; ACT = Activities; INT = Interactions; PRO = Program structure

Caregiver Interactive Skills

The correlations among the CIP scales were all significant, ranging between .34 and .75. The means, standard deviations, and ranges for the CIP scores are presented in Table 2 (left-hand side). The table shows substantial variation between the means of the different CIP scales. The means for the most "basic" caregiver skills (sensitive responsiveness, respect for autonomy, and structuring and limit setting) are in the adequate to good category (score ≥ 4.5), while the means for the more "educational" skills are moderate (3.5 ≤ score < 4.5) for verbal communication, and even inadequate (score < 3.5) for developmental stimulation and fostering positive peer interactions.

Figure 2 presents the percentages of groups that scored adequate to good, moderate, and inadequate, respectively, for the six CIP caregiver skills. The vast majority of the caregivers scored adequate to good on sensitive responsiveness, respect for autonomy, and structuring and limit setting. Less than 10% of the caregivers scored inadequate on these scales. For the remaining skills we found a quite different pattern. Here only a small minority of the caregivers had a high score (13%, 3%, and 1%, respectively, for verbal communication, developmental stimulation, and fostering positive peer interactions), while almost half to almost all caregivers had inadequate scores (42%, 89%, and 99%, respectively, for verbal communication, developmental stimulation, and fostering positive peer interactions. To summarize, the vast majority of caregivers scored adequate to good on the more basic caregiving skills sensitive responsiveness, respect for autonomy, and structuring and limit setting, while the vast majority scored inadequate for the more educational skills developmental stimulation and fostering positive peer interactions.

The right side of Table 2 shows the correlations between the CIP scores and the ITERS-R/ECERS-R scores. In general, the correlations were significant and moderate (ranging from .11.to .49), indicating that the CIP scales and the ITERS-R/ECERS-R, as intended, measure distinct but related aspects of the same construct (i.e., process quality).

Process Quality as Related to Group Size, Caregiver-child Ratio and Group Type

Table 1 shows the means and standard deviations for group size and caregiver-child ratio. Compared to 2005, the average group size has significantly increased in 2008 by more than one child per group to $M = 10.9$ ($SD = 3.2$). With regard to caregiver-child ratio, the difference with 2005 was not significant. On average, there was one caregiver per five children ($M = 0.21$, $SD = .05$)

Table 2 Descriptives for the CIP Scales and Pearson correlations between the CIP Scales and ITERS-R/ECERS-R Subscales

			ITERS-R/ECERS-R (N = 200)						
Variable	M	SD	*Range*	*SF*	*LAN*	*ACT*	*INT*	*PRO*	*Total*
CIP (*N* = 425)									
Sensitive responsiveness	4.81	0.93	1.50 – 7.00	.24**	.34**	.29**	.40**	.36**	.45**
Respect autonomy	4.51	0.84	1.25 – 6.75	.28**	.38**	.38**	.36**	.34**	.49**
Structuring & liming setting	4.90	1.17	2.00 – 7.00	.11	.18*	.22**	.18*	.17*	.24**
Verbal communication	3.60	0.90	1.25 – 6.00	.28**	.28**	.28**	.33**	.26**	.42**
Developmental stimulation	2.16	0.93	1.00 – 5.50	.22**	.25**	.23**	.24**	.24**	.33**
Fostering positive peer interactions	1.72	0.83	1.00 – 7.00	.13	.39**	.27**	.24**	.23**	.35**

Note. SF = Space and furnishings; LAN = Language; ACT = Activities; INT = Interactions; PRO = Program structure. Correlations between all CIP scales and ITERS-R/ECERS-R subscales are at care group level (*N* = 200).

* $p < .05$, ** $p < .01$ (two-tailed)

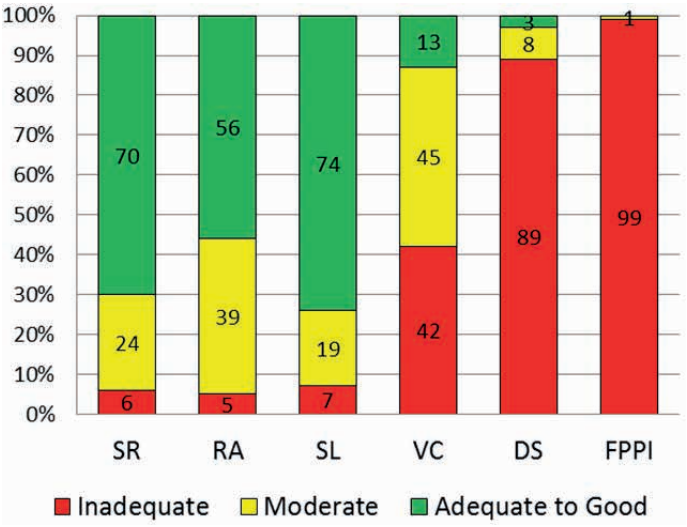

Figure 2 Percentages of care groups (*N* = 199) with Mean Scores "Inadequate", "Moderate", and "Adequate to Good" on the CIP Scales. SR = sensitive responsiveness; RA = respect for autonomy; SL = structuring and limit setting; VC = verbal communication; DS = developmental stimulation; FPPI = fostering positive peer interactions.

Correlations between group size and caregiver-child ratio on the one hand and the ITERS-R/ECERS-R subscales on the other revealed significant correlations with the subscale interactions (-.18, $p < .05$ and .18, $p < .01$, respectively), indicating that quality of interactions was lower with larger group sizes and more children per caregiver. The other subscales were not significantly related to group size and caregiver-child ratio. Group size was not significantly related to any of the CIP scales. Caregiver-child ratio, however, was significantly related to two of the scales, i.e., verbal communication (-.22, $p < .01$) and developmental stimulation (-.15, $p < .05$) indicating, surprisingly, that caregivers scored higher on verbal communication and developmental stimulation in groups with more children per caregiver.

Finally, we examined whether process quality, as measured with the ITERS-R/ ECERS-R and the CIP scales, was related to group type (infant, preschool, and mixed-age). Table 3 shows means and standard deviations for the ITERS-R/ ECERS-R subscales and the CIP scales separately for the three group types. For the ITERS-R/ECERS-R, we conducted a MANOVA with type of group as between-subjects factor and the ITERS-R/ECERS-R subscales as dependent variables. We found a significant overall effect of group type on the ITERS-R/ECERS-R (Wilks' Lambda = 0.792, $F(12, 384) = 3.95$, $p < .00$). As can be seen in Table 3, subsequent univariate analyses revealed a significant effect of group type only for the subscale space and furnishings. Post hoc tests showed that scores for space and furnishings were significantly lower in infant groups compared to the preschool groups. For the CIP scales, we conducted a MANOVA with type of group as between-subjects factor and the six CIP scales as dependent variables. Results demonstrated a significant overall effect of group type on the CIP scales (Wilks' Lambda = 2.394, $F(12, 381) = 1.87$, $p < .01$). Subsequent analyses of variance showed significant effects of group type for all CIP scales except for structuring and limit setting (see Table 3). Post hoc tests showed that caregivers in infant groups scored significantly lower on sensitive responsiveness than caregivers in preschool groups. For respect for autonomy, verbal communication, developmental stimulation, and fostering positive peer interactions, caregivers in infant groups also scored significantly lower than caregivers in the mixed-age groups.

Table 3 Means and Standard Deviations for Three Different Group Types

	Infant (I) (n = 55)		*Preschool (P)* (n = *52)*		*Mixed-age (M)* (n= *92)*[a]			
Variable	M	SD	M	SD	M	SD	F	*Significant Contrasts*[b]
CIP								
Sensitive responsiveness	4.57	0.90	5.05	0.65	4.85	0.75	5.21**	I < P
Respect for autonomy	4.26	0.77	4.73	0.60	4.55	0.68	6.53**	I < P, M
Structuring and limit setting	4.98	1.16	5.12	0.66	4.81	1.07	1.78	
Verbal communication	3.23	0.66	3.87	0.64	3.72	0.73	13.21**	I < P, M
Developmental stimulation	1.87	0.68	2.34	0.80	2.30	0.86	6.38**	I < P, M
Fostering positive peer interactions	1.49	0.53	1.82	0.65	1.79	0.76	4.15*	I < P, M
ITERS-R/ECERS-R								
Space and furnishings	2.70	0.63	3.29	0.68	3.00	0.62	7.58**	I < P
Language	3.45	1.09	3.56	1.08	3.28	1.08	2.29	
Activities	2.31	0.56	2.13	0.50	2.27	0.63	0.58	
Interactions	3.81	1.25	3.70	0.97	3.75	1.25	1.65	
Program structure	3.47	1.18	3.77	0.93	3.71	0.98	2.32	
Total	2.93	0.60	3.08	0.60	3.01	0.61	-	

Note. [a]For one mixed aged group the CIP scores were not available due to a technical problem with the camera. [b]Scheffé post-hoc test.
* $p < .05$, ** $p < .0$

3.4 Discussion

This study shows, first of all, that the decreasing trend in the quality of child care in Dutch centers as measured with the ITERS-R/ECERS-R across the years 1995-2005 has continued in the three years thereafter. The ITERS-R/ECERS-R total score in 2008 was significantly and substantially lower than in 2005, with almost half of the groups now scoring below the minimal level. The CIP scales, that were used for the first time in a nationally representative sample of child care centers, provided a more detailed picture of the core aspect of process quality, namely the caregiver-child interaction. The CIP-scores revealed the typical strengths and weaknesses of caregivers' interactive skills in Dutch child care centers, with relatively high scores for the more basic caregiving skills sensitive responsiveness, respect for autonomy, and structure and limit setting, and clearly lower scores for the more educational skills verbal communication, developmental stimulation, and fostering positive peer interactions.

The research design of this study does not allow drawing definite conclusions about possible causes for the further decline in global process quality. As outlined in the introduction, the number of children entering child care has continued to increase after the former quality assessment, which may partly explain the de-

crease in quality. A second possible explanation for the further decline might lie in changes in caregiver vocational training. As also mentioned in the introduction, in 1996 caregiver vocational training with a specific focus on working with young children (LKC) changed to a more general social pedagogical training (SPW-3). Between 2005 and 2008 however, caregiver education has remained unchanged. Therefore caregiver's vocational training is not a plausible explanation for the decline in quality across those years.

Whether or not the introduction of the Child Care Act in 2005 – a major policy change since the former quality assessment – contributed to the further decline in child care quality between 2005 and 2008 also cannot be stated with certainty. As noted in the introduction, the Child Care Act introduced a market-driven system which could lead to quality improvement if it stimulates parents to choose more high-quality child care for their children. The continuing decrease in quality after the introduction of the Act suggests that the child care market did not (yet) work as intended. This apparent market failure may have different causes, the most important of which is lack of information related to pedagogical quality for parents (Vandell & Wolfe, 2000). In the Netherlands, information regarding the comparative quality of care provided in child care centers is not publicly available. In addition, it appears to be difficult for parents to reliably judge the quality of child care themselves based on their own observations; they rate the quality of Dutch child care centers as more than sufficient (Berden & Kok, 2009; Kok, Groot, Mulder, Sadiraj, & Ham, 2005), whereas the results of the present study point in a different direction. In choosing a child care center for their children, Dutch parents are primarily led by convenience factors like availability (length of waiting lists) and distance to home and/or work (Berden & Kok, 2009; Kok, et al., 2005). And parents who are not satisfied with the quality of the care provided in a center hardly ever switch to another center, even when alternative options are available (Berden & Kok, 2009), because of the emotional burden for the children (Plantenga, 2012). The above findings help explain why the child care market that was introduced by the Child Care act in 2005 evidently did not lead to the intended improvement in the quality of child care in the subsequent years.

This study is the first to apply the CIP scales in a nationally representative sample of child care centers together with the ITERS-R/ECERS-R to get more detailed information about the quality of caregiver-child interactions. The results of the present study underscore those of a previous study with a smaller and non-representative sample (Helmerhorst et al, 2014), indicating that the CIP scales are indeed a valuable extension to the ITERS/ECERS. The average score on the ITERS-R/ECERS-R subscale interactions was 3.8 on a 7-point scale (moderate), which is in line with the moderate average score across the six CIP scales (M = 3.6 on a 7-point scale). But the results from the CIP scales also demonstrate substantial differences between the different caregiver interactive skills, with adequate to good scores for the more basic caregiving skills sensitive responsiveness (M = 4.8), respect for autonomy (M = 4.5) and structuring and limit setting (M = 4.9), to inadequate scores

for the more educational skills verbal communication (M = 3.6), developmental stimulation (M = 2.2), and fostering positive peer interactions (M = 1.7). The CIP scores also revealed interesting differences in caregiver-child interactions in different group types (Table 3) that were not visible in the ITERS-R/ECERS-R scores. Overall, these results underscore that the CIP scales provide valuable additional information on caregiver-child interactions next to the ITERS-R/ECERS-R subscale interactions.

Our present finding that caregivers scored higher on the more basic interactive skills than on the more educational skills of the CIP is in line with findings in earlier studies with smaller and nonrepresentative samples of caregivers in Dutch centers (Albers et al, 2010; Helmerhorst, et al., 2014), and with findings with the CLASS in the US (Gua, Piasta, Justice, & Kaderavek, 2010; Mashburn et al., 2008; La Paro, Pianta, & Stuhlman, 2004; LoCasale-Crouch et al., 2007; Thomason & La Paro, 2009). The lower scores for the more educational skills can also be explained by the focus on "care" rather than "education" in Dutch child care centers (Helmerhorst et al., 2014; OECD, 2006). The positive influence of caregivers' educational or instructional support on children's cognitive and language development, however, has been demonstrated in numerous studies (e.g., Albers et al., 2010; Burchinal, Vandergrift, Pianta, & Mashburn, 2010; Howes, et al., 2008; Mashburn et al., 2008). These results endorse the importance of the more educational interactive skills, and encourage policies and practices to improve quality of caregivers' educational skills by training and professional development (LoCasale-Crouch et al., 2007; Mashburn, et al, 2008).

In a previous study, in which we introduced the CIP scales and examined their validity, we were unable to obtain reliable ratings for structuring and limit setting, because the relevant caregiver behaviors occurred too infrequently on the videotaped interaction episodes. We assumed this to be due to the filming procedure, which was based on a fixed pre-set time schedule across the day. This yielded a large proportion of video fragments in free play situations, during which caregivers mostly did not have an active structuring role towards the children. Therefore, we adapted the filming procedure that now implies filming caregivers in four situations: diapering, lunch/snack, free play, and a transition between group activities. The transition moment was specifically chosen to rate the skill structuring and limit setting. With the adapted filming procedure used in the present study the structure and limit setting scale could be reliably applied. The scores for the other CIP scales were comparable to those in the former study (Helmerhorst et al., 2014), suggesting that the adaptation of the filming procedure did not affect the CIP-results.

Our expectation that a smaller group size and fewer children per caregiver would be associated with higher quality caregiver-child interactions was supported for the ITERS-R/ECERS-R subscale Interactions, but not for the CIP scales. Group size was unrelated to any of the CIP scales, while the association between the caregiver-child ratio and two of the CIP scales (i.e., verbal stimulation and developmental stimulation) was even in contrast to our expectation: Caregivers scored signifi-

cantly *higher* instead of lower on the CIP scales verbal communication and developmental stimulation with more children per caregiver. With regard to the lack of association between group size and CIP-scores: this lack of association could be characteristic for the Netherlands, given that this result is in line with findings of earlier Dutch child care studies. Van IJzendoorn et al. (1998) and Vermeer et al. (2008) found group size to be unrelated to caregivers' scores on the Caregiver Interaction Scales (CIS; Arnett, 1989), and De Schipper, Riksen-Walraven, and Geurts (2007) found it to be unrelated to caregiver-child interactions as observed with the CIS as well as with rating scales that are comparable to the CIP-scales. A methodological explanation could be that the lack of correlation is due to a restricted range of group size in Dutch centers, but the score distribution for group size – just as that for caregiver-child ratio – is not substantially different from the distribution reported in other countries (see Vandell & Wolfe, 2000). This raises the question if there are any other specific characteristics of the Dutch situation that might explain the lack of association between group size and caregiver-child interactions. What comes to mind is that the parents of children attending child care are relatively highly educated, which may contribute to a smaller number of children with problem behaviors per group. Groups with fewer highly demanding children could be easier to manage for caregivers, which could explain why a few more children per group do not go at the cost of the quality of caregiver-child interactions. Of course this is merely an assumption and more research is warranted to examine this further.

The unexpected finding that more children per caregiver was related to higher caregiver scores on the CIP-scales verbal communication and developmental stimulation was quite puzzling. A possible explanation for the higher verbal communication scores for caregivers who have to interact with more children could be that more children per caregiver increases the number of children asking for attention in the group, which may lead to more verbal communication between caregiver and children – not only in caregivers' interactions with individual children, but also in group-directed verbal interactions. The significantly higher levels of developmental stimulation provided by caregivers who have to care for more children remains puzzling. It cannot be excluded that more highly skilled caregivers are placed in groups with more children per caregiver – more research is needed to examine this. Moreover, given that the unexpected negative correlations between caregiver-child ratio and the two CIP-scales were quite low and applied to only two of the six CIP scales, replication is needed before drawing definitive conclusions about the relation between CIP-scores and caregiver-child ratio.

When comparing the quality of caregiver-child interactions in groups with different age-compositions, we found that scores on the CIP scale sensitive responsiveness were significantly lower in infant groups than in preschool groups. Furthermore, caregiver scores for respect for autonomy, verbal communication, developmental stimulation, and fostering positive peer interactions were significantly lower in infant groups than in preschool groups and mixed-age groups. This is in

line with the results of an earlier study in Dutch child care centers that also showed that more children under the age of two in a care group significantly predicted lower quality caregiver-child interactions (De Schipper et al., 2007). Given that sensitive responsiveness is generally considered to be the most basic aspect of caregiver behavior in interaction with infants, the significant lower score for sensitive responsiveness in infant groups is reason for concern and needs attention.

Limitations

The question of whether individual infants are better off in mixed-age groups than in infant groups cannot be answered in this study, because the CIP scales were designed to reflect caregivers' behavior toward children in the care group in general and not their interactions with individual children. Comparing the quality of care for infants in infant groups versus mixed-age groups requires observations and rating scales that focus on the children instead of the caregivers during caregiver-child interactions (e.g, the Observational Record of the Caregiving Environment; see NICHD ECCRN, 1996).

The response rate of the child care centers in the present study (33%) was comparable to that in our former study using the CIP scales (27%; Helmerhorst et al., 2014), but somewhat lower than in the former national quality assessment (49%; Vermeer et al., 2008). It was unclear why the response rate of the present study was lower. The procedures for the recruitment in the present study was exactly the same as in 2005, and the main reason for refusal to participate in the present study (i.e., 'too busy') was also the same as in the 2005 assessment. Unfortunately, it is not possible to draw any conclusions about causes of this lower response rate based on the present data.

Although the present study sample seems to be representative for the Netherlands, the results of this study cannot be automatically generalized to other countries given the typical characteristics of the Dutch child care context, outlined in the introduction section of this paper. Furthermore, it should be noted that the CIP scales were developed for use in child care centers in the Netherlands and may therefore reflect Dutch values and other characteristics of the Dutch child care context (see Helmerhorst et al., 2014). This should be taken into account when applying the scales in other countries.

However, although the results may be typical for the Dutch context, they may also be interesting for an international audience, because repeated child care quality assessments are quite rare in Europe, and may contribute to understanding how quality is affected by changing circumstances and policies in a different context. The study may increase awareness that quality of child care should always be considered in its own national and cultural context. In fact, the Dutch research program of which the present study is part (including repeated quality assessments and the development of culturally valid assessment instruments like the CIP scales) started

because there were doubts whether results of the NICHD study in the US (NICHD ECCRN, 2002) could be generalized to the Dutch context.

Implications for Policy and Practice

The continuing and substantial decrease in the quality of child care in The Netherlands as measured with the ITERS-R/ECERS-R in the three years since the former national quality assessment shows that the quality of child care is far from stable and that regular assessments are therefore needed to monitor child care quality.

The present study is the first to have used the CIP scales together with the ITERS-R/ECERS-R in assessing the quality of child care in a national representative sample of child care centers. The results underscore that the CIP scales provide valuable additional information beyond the ITERS-R/ECERS-R, and it is therefore recommended that the CIP scales be included in future quality assessments.

The low average scores on the more educational versus the more basic CIP scales indicate that these skills (i.e., verbal communication, developmental stimulation, and fostering positive peer interactions) should get more attention in caregiver education and training in general. The large variability in CIP scores between caregivers, however, suggests that an individual approach in training may also be useful. This is underscored by the moderate intercorrelations between the CIP scales, which demonstrate that a high score for one skill does not necessarily imply a high score for other skills. This suggests that training programs should be well attuned to the "interaction skill profile" of individual caregivers. We recently developed a video-feedback training program that aims to improve professional caregivers' interactive skills based on their individual profile of scores on the CIP scales. The effects of this training are presently being examined and will be reported in the near future.

References

Albers, E. M., Riksen-Walraven, J. M. A., & de Weerth, C. (2010). Developmental stimulation in child care centers contributes to young infants' cognitive development. *Infant Behavior & Development, 33*, 401-408. doi: 10.1016/j.infbeh.2010.04.004

Arnett, J. (1989). Caregivers in day-care centers: Does training matter? *Journal of Applied Developmental Psychology*, *10*, 541-552. doi: 10.1016/0193-3979(89)90026-9

Belsky, J., Vandell, D. L., Burchinal, M., Clarke-Stewart, K. A., McCartney, K., Owen, M. T., & the NICHD Early Child Care Research Network (2007). Are there long-term effects of early child care? *Child Development, 78*, 681-701. doi: 10.1111/j.1467-8624.2007.01021.x

Berden, C., & Kok, L. (2009). Ontwikkelingen op de markt voor kinderopvang 2004-2008 [Developments for the market in child care 2004-2008]. Amsterdam: Stichting Economisch Onderzoek.

Burchinal, M., Vandergrift, N., Pianta, R., & Mashburn, A. (2010). Threshold analysis of association between child care quality and child care outcomes for low-income children in pre-kindergarten programs. *Early Childhood Research Quarterly*, *25*, 166-167. doi: 10.1016/j.ecreq.2009.10.004

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

De Schipper, E. J., Riksen-Walraven, J. M. A., & 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

De Schipper, E. J., Riksen-Walraven, J. M. A., & Geurts, S. A. E. (2007). Multiple determinants of caregiver behavior in child-care centers. *Early Childhood Research Quarterly*, *22*, 312-326. doi:10.1016/j.ecresq.2007.04.004

Gevers Deynoot-Schaub, M. J. J. M., & Riksen-Walraven, J. M. A. (2005). Child care under pressure: The quality of Dutch center in 1995 and 2001. *Journal of Genetic Psychology, 166*, 280-296. doi: 10.3200/GNTP.166.3.280-296

Guo, Y., Piasta, S. B., Justice, L. M., & Kaderavek, J. N. (2010). Relations among preschool teachers' self-efficacy, classroom quality, and children's language and literacy gains. *Teaching and Teacher Education*, *26*, 1094-1103. doi: 10.1016/j.tate.2009.11.005

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

Harms, T., Cryer, D., & Clifford, R. M. (2003). *Infant/Toddler Environment Rating Scale-Revised*. New York, NY: Teachers College Press.

Helmerhorst, K. O. W., Riksen-Walraven, J. M. A, Vermeer, H. J., Fukkink, R. G., & Tavecchio, L. W. C. (2014). Measuring interactive skills of caregivers in child care centers: Development and validation of the caregiver interaction profile scales. *Early Education & Development, 25,* 770-790. doi: 10.1080/10409289.2014.840482

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.

Kok, L. Groot, I., Mulder, J., Sadirai, K, & Van Ham, M. (2005). *De markt voor kinderopvang 2004* [The child care market in 2004]. Amsterdam: Stichting Economisch Onderzoek.

La Paro, K. M., Pianta, R., & Stuhlman, M. (2004). The classroom assessment scoring system; Findings from the prekindergarten year. *The Elementary School Journal*, *104*, 409-426.

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

LoCasale-Crouch, J., Konold, T., Pianta, R., Howes, C., Burchinal, M., Bryant, D., . . . Barbarbarin, O. (2007). Observed classroom quality profiles in state-funded pre-kindergarten programs and associations with teacher, program, and classroom characteristics. *Early Childhood Research Quarterly*, *22*, 3-17. doi: 10.1016/j.ecresq.2006.05.001.

Mashburn, A. J., Pianta, R. C., Hamre, B. K., Downer, J. T., Barbarin, O. A., Bryant, D., Burchinal, M., Early, D. M., & 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

Merens, A., Hartgers, M., & Van den Brakel, M. (2012). *Emancipatiemonitor 2012* [Emancipation Monitor 2012]. The Hague, The Netherlands: Sociaal en Cultureel Planbureau/Centraal Bureau voor de Statistiek.

NICHD Early Child Care Research Network (1996). Characteristics of infant child care: Factors contributing to positive caregiving. *Early Childhood Research Quarterly*, *11*, 269-306. doi: 10.1016/S0885-2006(96)90009-5

NICHD Early Child Care Research Network. (2002). CHILD-CARE STRUCTURE → PROCESS → OUTCOME: Direct and indirect effects of child-care quality on young children's development. *Psychological Science, 13*(3), 199-206.

OECD (2006). *Starting Strong II: Early childhood education and care*. Paris: Organisation for Economic Co-operation and Development.

Plantenga, J. (2012). Local providers and loyal parents: competition and consumer choice in the Dutch childcare market. In: L. Lloyd & H. Penn (Ed.), *Childcare markets: Can they deliver an equitable service* (p. 63-78). Bristol, UK: The Policy Press. doi:10.1332/policypress/9781847429339.003.0004

Thomason, A. C., & La Paro, K. M. (2009). Measuring the quality of teacher-child interactions in toddler child care. *Early Education and Development, 20*(2), 285-304. doi:10.1080/10409280902773351

Tietze, W., Cryer, J., Bairrão, D., Palacios, J., & Wetzel, G. (1996). Comparisons of observed process quality in early child care and education programs in five countries. *Early Childhood Research Quarterly, 11*, 447-475. doi: 10.1016/S0885-2006(96)90017-4

Van Beem, M., & Wever, Y. (2008). *Wachtlijsten en wachttijden buitenschoolse opvang en dagopvang (2e meting)* [Waiting lists and waiting periods in after school care and child daycare centers (2nd measurement). Den Haag: B&A Consulting bv.

Vandell, D. L., Belsky, J., Burchinal, M., Steinberg, L., Vandergrift, N., & NICHD Early Child Care Research Network. (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*(3), 737-756. doi:10.1111/j.1467-8624.2010.01431.x

Vandell, D. L., & Wolfe, B. (2000). *Child Care quality: Does it matter and does it need to be improved?* Madison: Institute for Research on Poverty, University of Wisconsin-Madison.

Van IJzendoorn, M. H., Tavecchio, L. W. C., Stams, G. J. J. M., Verhoeven, M. J. E., & Reiling, E. J. (1998). Quality of center day care and attunement between parents and caregivers: center day care in cross-national perspective. *The Journal of Genetic Psychology*, *159*, 437-454.

Vermeer, H. J., Van IJzendoorn, M. H., De Kruif, R. E. L., Fukkink, R. G., Tavecchio, L. W. C., Riksen-Walraven, J. M. A., & Van Zeijl, J. (2008). Child care in The Netherlands trends in quality over the years 1995-2005. *The Journal of Genetic Psychology, 169*(4), 360-385.

CHAPTER 4
Effects of the Caregiver Interaction Profile Training on Caregiver-Child Interactions in Child Care Centers

Foto: Joods Historisch Museum

Helmerhorst, K. O. W., Riksen-Walraven, J. M. A., Fukkink, R. G., Tavecchio, L. W. C., & Gevers Deynoot-Schaub, M. J. J. M. (2014). *Effects of the Caregiver Interaction Profile Training on caregiver-child interactions in child care centers*. Manuscript submitted for publication.

Abstract

In this study we used a randomized controlled trial to examine the effects of a five-week video feedback training for caregivers in early child care centers aimed at improving six key interactive caregiver skills: Sensitive responsiveness, respect for autonomy, structuring and limit setting, verbal communication, developmental stimulation, and fostering positive peer interactions. The sample included 68 mixed-age groups from 33 child care centers (35 intervention groups and 33 control groups) with a total of 139 caregivers. Results at posttest indicate a significant positive training effect on all six caregiver interactive skills. Three months after the posttest, caregivers in the experimental group still scored significantly higher on sensitive responsiveness, respect for autonomy, verbal communication, and fostering positive peer interactions than caregivers in the control group. Possible ways to further improve the training and to implement it in practice and education are discussed.

4.1 Introduction

Numerous studies have demonstrated the importance of early child care quality for children's socio-emotional and academic development (for an overview see Belsky et al., 2007; Vandell & Wolfe, 2000; Vandell et al., 2010). The quality of caregiver-child interactions is generally acknowledged as a key aspect of quality (e.g., Vandell & Wolfe, 2000); high-quality interactions with adults have been characterized as proximal processes through which young children develop (Bronfenbrenner & Ceci, 1994). Therefore, it is important to monitor child care quality and improve it when needed. In The Netherlands, the quality of child care has been assessed repeatedly with the internationally widely used Infant/Toddler Environment Rating Scale-Revised (ITERS-R; Harms, Cryer, & Clifford, 2003) and Early Childhood Environment Rating Scale-Revised (ECERS-R; Harms, Clifford, & Cryer, 1998), which broadly and globally assess the quality of the child care environment. To get a more detailed picture of the caregiver-child interaction, we developed the Caregiver Interaction Profile (CIP) scales, to be applied in combination with the ITERS-R and ECERS-R. The CIP scales rate six key skills of caregivers for interacting with 0- to 4-year-old children: *Sensitive responsiveness*, *respect for autonomy*, *structuring and limit setting*, *verbal communication*, *developmental stimulation*, and *fostering positive peer interactions* (for background, development, and validation results, see Helmerhorst, Riksen-Walraven, Vermeer, Fukkink, & Tavecchio, 2014b).

In a next study, we applied the CIP scales in a large nationally representative sample of caregivers in Dutch child care centers (Helmerhorst, Riksen-Walraven, Gevers Deynoot-Schaub, Tavecchio, & Fukkink, 2014a). The results showed caregivers to score highest, on average, on the three more basic caregiving skills, i.e., sensitive responsiveness, respect for autonomy, and structuring and limit setting, although about one third of the caregivers scored inadequate or moderate on these skills, indicating that there is still room for improvement. Scores were clearly lower for the more "educational" skills, i.e., verbal communication, developmental stimulation, and fostering positive peer interactions, with the majority of the caregivers scoring in the moderate to inadequate range. These results are in line with findings in the United States (Guo, Piasta, Justice, & Kaderavek, 2010; La Paro et al., 2009; Mashburn et al., 2008; Thomason & La Paro, 2009), Australia (Tayler, Ishimine, Cloney, Cleveland, & Thorpe, 2013) and Spain (Sandstrom, 2012), and underscore the need to improve caregivers' skills by training and professional development (La Paro et al., 2009; LoCasale-Crouch et al., 2007; Mashburn et al., 2008). Based on the above results we therefore developed a training program for caregivers to improve the six interactive skills included in the Caregiver Interaction Profile. The present study examined the effects of this training using a randomized controlled trial.

Earlier Studies of the Effects of Training Programs to Improve Caregiver-Child Interactions

In designing the training program, we first reviewed earlier effect studies to identify effective features of training programs for improving caregiver-child interactions. In a meta-analysis of effect studies published between 1998 and 2005, Fukkink and Lont (2007) examined the effects of a broad set of training programs directed at caregivers in child care. They demonstrated that, on average, training significantly improved caregiver skills. However, not all training programs were effective. The largest effects were found for programs with a fixed curriculum and programs including fewer trainees. Large-scale programs designed for a variety of training formats and for a wide variety of learners were less effective. Furthermore, larger effects were reported in studies using outcome measures that were closer aligned with the training content. The meta-analysis included only four studies with a randomized controlled trial (RCT) design – the design that is now generally considered the "gold standard" for assessing intervention effects (Fantuzzo et al., 1996; Fantuzzo et al., 1997; Girolametto, Weitzman, & Greenberg, 2003; Girolametto, Weitzman, & Greenberg, 2004). All four studies used individualized video feedback and showed significant effects of training on caregivers' interactions with children. Trained caregivers were more emotionally supportive (Fantuzzo et al., 1996; Fantuzzo et al., 1997) as well as more verbally supportive and better able to facilitate peer interactions (Girolametto et al., 2003; Girolametto et al., 2004).

Since the meta-analysis by Fukkink and Lont (2007), other RCT studies have been published which examined the effects of an intervention to improve caregiver interactive skills. We summarize the results of the five most relevant studies. First, Raver et al. (2008) demonstrated that preschool teachers who received five 6-hour training sessions in behavior management and weekly coaching by mental health consultants who provided individual on-site feedback to the teachers throughout the year, had higher scores for positive climate, teacher sensitivity, and behavior management as measured with the Classroom Assessment Scoring System (CLASS; La Paro, Pianta, & Stuhlman, 2004), than caregivers in a control group.

In a second study, Pianta, Mashburn, Downer, Hamre, and Justice (2008) evaluated the web-mediated MyTeachingPartner program (MTP), which includes video exemplars and individualized feedback targeted at caregivers' own interactions with children. The MTP program is based on the conceptual framework of the CLASS (Pianta, LaParo, & Hamre, 2008) for defining classroom interactions. Pianta et al. (2008) demonstrated that teachers who received the individualized MTP consultation for an entire year showed significantly greater increases in the quality of their interactions than teachers in the web-only control condition who only had access to video clip examples of high-quality interactions.

In a third randomized controlled study, Domitrovich et al. (2009) examined the effects of a program for Head Start classroom teachers to improve teaching quality in both the emotional-behavioral and cognitive-linguistic domain. The one-year

program involved professional development workshops and weekly mentoring in the classroom by trained educational consultants. Domitrovich et al. (2009) showed that the program improved the teachers' language/literacy stimulation and emotional support to the children in their classroom.

Fourth, Piasta and colleagues (2012) evaluated the effects of a personal development program for preschool teachers to improve their conversational responsivity. The program included workshops and, every two weeks, written individual feedback on the teachers' own videotaped classroom interactions. Piasta et al. (2012) demonstrated that the teachers who received the program showed significantly larger improvements in communication-facilitating strategies than teachers in a comparison group who received a comparable program that was not specifically focused on their conversational responsivity.

Finally, a video feedback intervention in Dutch child care centers, aimed at improving caregiver sensitivity and – to a lesser extent - verbal stimulation, showed positive effects on the relevant caregiver interactive skills (Fukkink & Tavecchio, 2010). These effects remained visible at follow-up test, three months after the video feedback intervention. The program usually included four weekly or two-weekly individual sessions, during which trainer and caregiver together discussed videos of the caregivers' own interactions with the children in their group.

The Present Study

Altogether, results from earlier effect studies show that it is possible to improve caregiver's interactive skills through training, although there is relatively little evidence for caregivers of very young children (0- to 2-year-olds). Most studies pertained to preschool teachers of 2- to 5-year-old children. Furthermore, although the studies are difficult to compare, their results allow several conclusions that are relevant for decisions regarding the content, method, and duration of the intervention that we aimed to develop.

First, with regard to the intervention method, individual video feedback on caregivers' own everyday interactions with the children in their group appeared to be an effective method. We decided to adopt this method, in combination with an introductory meeting to introduce and discuss the conceptual framework.

Second, with regard to the content of the training program, it is important to note that the largest effects were found in effect studies that used outcome measures that were closely aligned with the content of the program. Given that we aimed to improve the six caregiver interactive skills included in the CIP scales, which we therefore used as outcome measures in the present effect study, we based the training on the same conceptual framework that also underlies the CIP scales.

Third, with regard to the duration of the training, the programs reviewed above showed a broad variation in duration, from four sessions to an entire year. The effect studies demonstrated that even the relatively short program evaluated by Fukkink and Tavecchio (2010) was effective. This is in line with the conclusion by

Bakermans-Kranenburg, Van IJzendoorn, and Juffer (2003) from a meta-analysis of intervention programs aimed at improving parent-child interactions. They demonstrated that programs with a moderate number of sessions (less than 16) were more effective than programs with more sessions, and in addition that effectiveness of programs with less than 5 sessions was comparable to programs with 5 to 16 sessions. In line with this, we opted for a training with 5 weekly sessions.

In sum, based on our own earlier studies with the CIP scales in Dutch child care centers and on the results of prior effect studies of other intervention programs for caregivers and teachers, we developed a 5-week video feedback training for caregivers, based on the conceptual framework underlying the CIP scales and aiming to improve the 6 CIP skills. The present RCT study examined the effects of this CIP training program.

4.2 Method

Participants and Randomization

Child care groups in this study were recruited from child care centers in and around the city of Amsterdam, The Netherlands. Directors of the child care centers responded to appeals in (digital) newsletters and announcements on child care websites.

Eligibility criteria were: (1) child care centers had to have mixed-age groups and (2) had to participate with an even number of groups to assign to the study. In The Netherlands, three types of child care groups are distinguished: infant groups (0- to 2-year-olds), preschool groups (2- to 4-year-olds), and mixed-age groups (0- to 4-year-olds). For this study we selected mixed-age groups to test the effects of the intervention in groups with children across the whole possible age range, and because mixed-age groups are most prevalent in The Netherlands (see Helmerhorst et al., 2014a). Per center, half of the participating groups were randomly assigned to the experimental condition and half to the control condition. When child care organizations participated with multiple centers, we randomly assigned half of the centers (with all participating groups) either to the experimental condition or to the control condition. All caregivers from the groups selected for the study were invited to participate using an informed consent procedure. Out of the invited caregivers, two caregivers – both in the control condition – did not want to be filmed and were therefore not included in the sample. The parents of the children in the selected care groups were also asked to give their informed consent for the filming procedures. Whenever parents did not give permission to film their children, we made sure that these children were not filmed. As a result, groups in which not all parents gave their consent were not excluded from the study.

The final study sample included 68 mixed-age groups from 33 child care centers, with 35 intervention groups and 33 control groups. Originally, the sample

consisted of 35 intervention groups and 35 control groups, but two control groups dropped out after randomization and before the pretest. A total of 139 caregivers participated in the study. In 64 groups two caregivers participated and in 4 groups only one caregiver participated. Caregivers were all female and were on average 32.4 years old (*SD* = 9.65, range 18-56), worked 28.3 hours a week (*SD* = 6.44, range 16-40), and had 8.2 years (*SD* = 6.48, range 0-25) of working experience in child care. The majority of the caregivers (89.5%) had completed the regular 3-year vocational training in general social-pedagogic work, 7% had a bachelor degree, and 3.5% worked in practice as part of their education. There were no significant differences between caregivers in the experimental and control group at pretest for caregivers' age, education, years of experience, and working hours per week. The average group size at pretest was 10.7 children per group (*SD* = 2.17), with an average child-caregiver ratio of 4.8 children per caregiver (*SD* = 0.83). The average child-caregiver ratio in the experimental group (*M* = 5.0, *SD* = 0.78) was significantly higher than in the control group (*M* = 4.7, *SD* = 0.87), $t = 2.14$, $p < .05$, indicating that the ratios were more favorable in the control condition.

Figure 1 shows the flow chart of the participating caregivers per phase. As can be seen, 27 (19%) of the 139 participants dropped out in the course of the intervention program. Main reasons for discontinuing in the control group were maternity leave and caregivers being sick at the day of the observation.

Independent *t* tests show no differences between caregivers who completed the intervention and those who dropped out in terms of age, education, working hours per week, and experience at pretest.

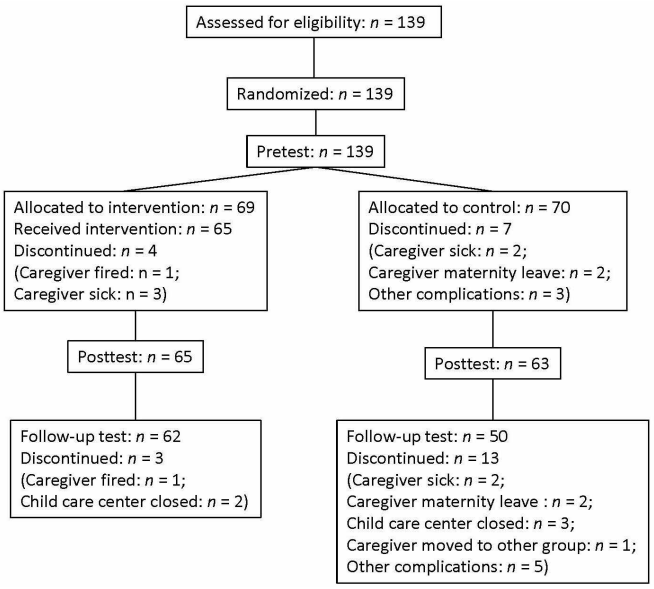

Figure 1 Flow Chart of the Study's Progress in Terms of Numbers of Caregivers

Design and General Procedure

This study is a RCT with random assignment of the child care groups to two conditions (experimental or control group). Caregivers' performance on the six CIP scales was the dependent variable in this study. Their performance was rated using the CIP scales at pretest, at a posttest immediately after the end of the intervention, and at a follow-up test three months after the posttest.

Parallel with the present training for caregivers to improve the quality of caregiver-child interactions, a separate consultancy program for center directors was conducted. This consultancy program had a different aim, i.e., improving the global quality of the child care environment as measured with the ITERS-R/ECERS-R; during the consultancy, no attention was given to caregiver-child interactions. When examining the effects of the training for caregivers on their interactive skills as measured with the CIP scales in the present study, we controlled for possible effects of the consultancy program (see below under Measures: ITERS-R/ECERS-R).

For the pretest, posttest, and follow-up assessments, the groups were visited by a trained experimenter from 8 a.m. until approximately 3 p.m. . During the visit, each individual caregiver was filmed for 8 to 10 minutes in each of four different situations: diapering, lunch/snack, structured play, and transition between group activities. The video episodes were rated afterwards by observers who had not visited the care group in question. Further, the trained experimenter applied the ITERS-R and ECERS-R to assess the global quality of the child care environment.

Two weeks after the pretest, the five-week training program started at the same day of the week as the pretest. One week after the last training session, or six weeks after the pretest for the control groups, each group was visited for the posttest by a second experimenter. Three months after the posttest, each group was visited by a third experimenter for the follow-up test. For an optimal comparison, the posttest and follow-up visits were planned on the same day of the week as the pretest. At the pretest, caregivers also completed a questionnaire to collect individual background information (e.g., age, education, work experience).

The Intervention

Experimental group. The CIP training was developed by the Netherlands Consortium for Child Care Research, based on the conceptual framework underlying the CIP scales (for details on the framework see Helmerhorst et al., 2014b). The CIP scales rate six caregiver skills in interacting with 0- to 4-year-old children in a group setting: *sensitive responsiveness*, *respect for autonomy*, *structuring and limit setting*, *verbal communication, developmental stimulation,* and *fostering positive peer interactions* (see Method section). Five CIP trainers (n = 5) with a bachelor (n = 2) or master degree (n = 3) in psychology trained the caregivers at the childcare centers during day/work time. All trainers had been trained by the first author using a

manual and video examples. The training for trainers consisted of five sessions in total, including home video assignments.

Table 1 shows an overview of the CIP training. All CIP trainers used a manual with a standardized protocol for delivering the training. The training comprises five onsite visits in total, each lasting two hours per caregiver. The first four visits were individual training sessions with the caregiver receiving feedback on her own videotaped interactions. One caregiver was trained during the morning and the second caregiver in the afternoon. The fifth and final training session was provided to both caregivers of the care group together; the session was cancelled if only one caregiver per group participated in the training. Before each training session, the trainer had analyzed video recordings to be used during that session. For the first session, the trainer used the pretest recording, and for the second, third and fourth session the trainer collected new video material of the caregiver during daily routines in the care group, after the training session. As can be seen in Table 1, the first session was used to inform the caregiver about the upcoming training sessions and procedures, which were described in an intervention booklet that all caregivers received at the start of the first session. Directly after this general introduction, the skills *sensitive responsiveness* and *respect for autonomy* were discussed, because those are considered to be the most basic aspects of caregiver behavior in interactions with children. The set-up of the first four sessions was the same: (1) first the trainer and caregiver read a description and discussed the relevant CIP skill, followed by (2) three exemplary videos of a model caregiver interacting with children. The caregiver was requested to rate the video examples as high, medium, or low in terms of the relevant CIP scale. A description for high, medium and low performance was given in the intervention booklet. After the caregiver had rated the examples, (3) the trainer showed the caregiver short fragments (between 1 and 3 minutes) of the earlier selected video recordings of the caregiver's own interactions. After watching the video fragment together, the trainer asked the caregiver to comment on her own video. Based on the caregiver's reaction, the trainer and caregiver discussed the caregiver's behavior, and when needed reviewed the episode again. Next, (4) the caregiver was asked to indicate goals for the upcoming week by means of a checklist (also in the booklet), which listed concrete behaviors related to the specific CIP skill (e.g., make eye contact with the children, use a warm and calm voice when talking to the children for sensitive responsiveness).

The second training session started with a short review of sensitive responsiveness and respect for autonomy, followed by the same training for two new behaviors: *structuring and limit setting* and *verbal communication*. The third session again started with a review of structuring and limit setting and verbal communication, and proceeded with *developmental stimulation* and *fostering positive peer interactions*. After the third session, the caregiver determined which two out of the six CIP skills she wanted to repeat during the fourth session. The vast majority of caregivers chose to repeat developmental stimulation and fostering positive peer interactions. Whenever a caregiver chose one of the other four skills, developmen-

tal stimulation and fostering positive peer interactions were repeated during the start of the fourth session to ensure that all six skills were discussed two times with all caregivers during the intervention. After the fourth session, the caregiver was asked to choose three to five video episodes that she wanted to show to her colleague during the last session. The fifth and final session was used to share experiences and video episodes between both caregivers. This way, caregivers learned from each other by seeing what the fellow-caregiver had been working on during the past four weeks.

Control group. Caregivers in the control group received no training at all and were only contacted for pretest, posttest, and follow-up test observations.

Table 1 Overview of Sessions and Content of the CIP Training

Session	Content	Individual/ with colleague	Video episodes collected
1	Information about procedures and general introduction Video feedback *sensitive responsiveness* & *respect for autonomy* Setup video feedback: Read description CIP skill Caregiver rates video examples model in high, medium, low Watch caregiver's own video episodes Caregiver fills in behavior checklist	Individual	Pretest
2	Review *sensitive responsiveness* & *respect for autonomy* Video feedback *structuring and limit setting* & *verbal communication* Setup video feedback: see session 1	Individual	After session1
3	Review *structuring and limit setting* & *verbal communication* Video feedback *developmental stimulation* & *fostering positive peer interactions* Setup video feedback: see session 1	Individual	After session 2
4	Review *developmental stimulation* & *fostering positive peer interactions* Video feedback booster - two skills of caregiver's own choice Setup video feedback: see session 1	Individual	After session 3
5	Learn from each other - caregivers choose three to five of their own video episodes to show their fellow-caregiver	With colleague	Pretest; after session 1, 2, 3

Measures

CIP scales. These scales measure six caregiver interactive skills: (1) *sensitive responsiveness* refers to the extent to which a caregiver recognizes children's individual emotional and physical needs, and responds appropriately and promptly to their cues and signals; (2) *respect for autonomy* refers to the extent to which a caregiver

is non-intrusive but instead recognizes and respects the validity of children's intentions and perspectives; (3) *structuring and limit setting* refers to the ability of a caregiver to clearly communicate expectations towards children and structure the situation accordingly and to set clear and consistent limits to the children's behavior; (4) *verbal communication* refers to the frequency and quality of verbal interactions between caregiver and children; (5) *developmental stimulation* concerns the degree to which a caregiver deliberately attempts to foster children's development, e.g., motor skills, cognitive development, and creativity; (6) *fostering positive peer interactions* refers to the extent to which the caregiver guides or facilitates positive interactions between children in the child care group. For a more extensive description of the CIP scales see Helmerhorst et al. (2014b). Each of the six CIP skills is rated on a single 7-point scale (7 = *very high*, 6 = *high*, 5 = *moderate/high*, 4 = *moderate*, 3 = *moderate/low*, 2 = *low*, 1 = *very low*), with detailed behavioral descriptions for each of the seven scale points. Scores on the CIP scales can be classified in three quality levels: *inadequate* (score < 3.5), *moderate* (3.5 ≤ score < 4.5) and *adequate to good* (≥ 4.5).

Thirteen trained observers independently rated the behavior of the caregiver on the six 7-point scales for each of the four videotaped episodes. Observers that rated the video episodes had not visited the care group for data collection. Per caregiver a mean score for each of the six skills was calculated by averaging across the four episodes. Inter-rater reliability (i.e., intraclass correlations computed for 10% of the tapes) was .87 on average (range .71 – 1.0).

ITERS-R/ECERS-R. Both the ITERS-R (Harms et al., 2003, for children younger than 30 months) and the ECERS-R (Harms et al., 1998, for children between the ages of 30 and 48 months) were used in each child care group to assess global process quality. This was done to control for possible effects of the parallel consultancy program for center directors, that was aimed at increasing global process quality of child care as measured with these instruments. We applied only those ECERS-R/ITERS-R subscales and items that were addressed during the consultancy: Space and furnishings, Language, Activities, Interactions and Program structure. Subscale items are rated on a 7-point scale with descriptors for the scores 1 (inadequate), 3 (minimal), 5 (good), and 7 (excellent). Total scores for the ITERS-R (26 items) and for the ECERS-R (31 items) at pretest, posttest, and follow-up were computed by averaging item scores across subscales. Given their high correlation (r = .87), we aggregated the ITERS-R and ECERS-R total scores. Per care group, two ITERS-R/ECERS-R gain scores were subsequently computed. First, by subtracting the group's combined total ITERS-R/ECERS-R score on the pretest from the total score on the posttest and second, by subtracting the group's combined total ITERS-R/ECERS-R score on the pretest from the total score on the follow-up test. These gain scores were included as a control variable in our analyses testing the effects of the CIP training in the present study.

Missing Data

We examined the effects of the CIP training using an *intent-to-treat* analysis, in which all data for all caregivers were analyzed, regardless of attrition. In the experimental group a total of 7 caregivers (10%) discontinued between pretest and follow-up. In the control group a total of 20 caregivers (29%) discontinued. We used Expectation Maximization (EM; SPSS 20.0) to estimate missing parameters for the experimental group and control group separately. Results demonstrated that the data were missing completely at random (Little's MCAR test, $p = .27$ for the experimental group and $p = .34$ for the control group). We imputed scores for the missing values of caregivers at posttest and follow-up and ran all analysis with both the imputed data and the original collected data. We found no differences between the analysis with the imputed data and the original collected data.

4.3 Results

Effects of the CIP training at Posttest

Table 2 shows the means and standard deviations for the six CIP scales in the experimental and control group at pretest, posttest, and follow-up. Because caregivers (level 1) were nested within groups (level 2), and groups were nested within centers (level 3), we conducted multilevel analysis, using MLwiN. First, to examine the effects of the training at posttest for each of the six scales, we tested whether the CIP scores at posttest differed between the experimental and control group, controlling for the mean pretest CIP scores and controlling for the ITERS-R/ECERS-R gain score (to control for possible effects of the consultancy training for child care directors). The dependent variable in each model was the posttest score for the six CIP scales. Three variables were entered as predictors: a group (dummy) variable which indicated whether the group was allocated to the experimental or the control condition (0 = control, 1 = experimental), the mean pretest score of the CIP skill, and the ITERS-R/ECERS-R pretest-posttest gain score to control for possible effects of the parallel consultancy intervention. Table 3 shows the outcomes of the six multilevel analyses for the separate CIP skills at posttest. We found a significant difference between the experimental and control group at posttest for all of the CIP skills, indicating that the training had a positive effect on sensitive responsiveness, respect for autonomy, structuring and limit setting, developmental stimulation, and fostering positive peer interactions. Table 2 shows that caregivers in the experimental group scored on average about half a scale point (on a 7-point scale) higher for sensitive responsiveness, respect for autonomy, and verbal communication than caregivers in the control group at the posttest (0.55, 0.44, and 0.53, respectively). For structuring and limit setting caregivers in the experimental group scored on average 0.33 higher than caregivers in the control group. For developmental stimu-

lation and fostering positive peer interactions, differences between caregivers' mean scores in the experimental group and control group were more than three-quarters of a scale point (0.77 and 0.84, respectively) in favor of the experimental group.

Effects of the CIP training at Follow-up

Next, we analyzed whether the effects of the training were retained at follow-up, three months after completion of the training. Multilevel models for the analysis of the follow-up results were comparable to the posttest models described before, with only the dependent variable being replaced by the follow-up scores on the CIP scales and the ITERS-R/ECERS-R pretest-posttest gain score by the pretest-follow-up gain score. Table 4 shows the outcomes of the six multilevel analyses. The results demonstrate a significant difference between the experimental and control group for four of the six scales at the follow-up: sensitive responsiveness, respect for autonomy, verbal communication, and fostering positive peer interactions. No significant differences were found for structuring and limit setting, and developmental stimulation. As can be seen in Table 2, caregivers in the experimental group still scored about half a scale point higher, on average, than caregivers in the control group (0.45 for sensitive responsiveness, 0.52 for respect for autonomy, 0.57 for verbal communication, and 0.42 for fostering positive peer interactions) at follow-up. This suggests that the positive effects of the training were retained over time for sensitive responsiveness, respect for autonomy, verbal communication, and fostering positive peer interactions.

Table 2 Mean Scores and Standard Deviations for the Experimental and Control Group at Pretest, Posttest, and Follow-up Test

	Experimental Group			*Control group*		
Variable	N	M	(SD)	N	M	(SD)
Pretest						
Sensitive responsiveness	69	5.04	(0.86)	70	5.00	(0.64)
Respect for autonomy	69	4.58	(0.77)	70	4.69	(0.77)
Structuring and limit setting	69	5.09	(0.82)	70	4.91	(0.85)
Verbal communication	69	3.94	(0.76)	70	3.80	(0.62)
Developmental stimulation	69	2.61	(0.82)	70	2.43	(0.75)
Fostering peer interactions	69	2.17	(0.85)	70	1.96	(0.77)
Posttest						
Sensitive responsiveness	65	5.33	(0.66)	63	4.78	(0.78)
Respect for autonomy	65	4.92	(0.85)	63	4.48	(0.73)
Structuring and limit setting	65	5.24	(0.84)	63	4.98	(0.92)
Verbal communication	65	4.25	(0.86)	63	3.72	(0.73)
Developmental stimulation	65	3.02	(0.94)	63	2.25	(0.65)
Fostering peer interactions	65	2.63	(1.06)	63	1.79	(0.76)
Follow-up						
Sensitive responsiveness	62	5.27	(0.80)	50	4.82	(0.82)
Respect for autonomy	62	4.91	(0.77)	50	4.39	(0.76)
Structuring and limit setting	62	5.16	(0.81)	50	4.93	(0.86)
Verbal communication	62	4.03	(0.76)	50	3.46	(0.69)
Developmental stimulation	62	2.53	(0.82)	50	2.33	(0.77)
Fostering peer interactions	62	2.22	(0.85)	50	1.80	(0.71)

Meaningful Change

Although the abovementioned training effects were significant, the absolute differences were relatively small. Nevertheless, the increase in scores can be meaningful if it marks the difference between inadequate care and moderate or adequate to good care. To examine this, we compared the percentages of caregivers who scored in the categories *inadequate* (score < 3.5), *moderate* (3.5 ≤ score < 4.5) and *adequate to good* (≥ 4.5) at the pretest, posttest, and follow-up test. This classification is based on the detailed behavioral description of the seven scale points of each of the six CIP scales, where a score 3 is defined as inadequate and a score 5 as adequate (see Helmerhorst et al., 2014b). Figure 2 shows the percentages of caregivers scoring in the three quality categories on the more basic CIP skills (i.e., sensitive responsiveness, respect for autonomy, and structuring and limit setting). With re-

gard to changes in the experimental group between pretest and posttest, the following picture emerges. At pretest, 3% of the caregivers in the experimental group scored inadequate on sensitive responsiveness, 3% on respect for autonomy, and 2% on structuring and limit setting. After completing the training, none of the caregivers in this group still scored inadequate on sensitive responsiveness and structuring and limit setting. The percentage of caregivers scoring in the inadequate range for respect for autonomy in the experimental group remained stable. Furthermore, the percentage of caregivers scoring in the adequate to good range improved by 9% for sensitive responsiveness (from 81% to 90%) and by 22% for respect for autonomy (from 52% to 74%), but remained stable for structuring and limit setting.

Figure 3 shows the percentages of caregivers scoring in the three quality categories on the educational CIP skills (i.e., verbal communication, developmental stimulation and fostering positive peer interactions). Between pretest and posttest, the percentage of caregivers that scored inadequate decreased by 11% for verbal communication and by 15% for developmental stimulation and fostering positive peer interactions. Furthermore, the percentage of caregivers scoring adequate to good increased by 21% for verbal communication, by 11% for developmental communication, and by 10% for fostering positive peer interactions. Altogether, these results suggest that the significant positive effects of the training on caregivers' interactive skills were relatively small in an absolute sense, but may yet be meaningful for children, as it decreases the percentage of caregivers providing inadequate care and increases the percentage of caregivers providing care of adequate to good quality.

Table 3 Effects of CIP Training on Caregiver Interactive Skills at Posttest (Multilevel Analysis, *N* = 139)

	Sensitive responsiveness Model		*Respect for autonomy Model*		*Structuring and limit setting Model*		*Verbal communication Model*		*Developmental stimulation Model*		*Fostering peer interactions Model*	
	Estimate	*SE*	Estimate	*SE*	Estimate	*SE*	Estimate	*SE*	Estimate	*SE*	Estimate	*SE*
Fixed parameters												
Intercept	3.339*	0.386	3.294*	0.407	4.517*	0.395	2.690*	0.346	1.973*	0.237	1.339*	0.206
CIP training	0.561*	0.131	0.544*	0.132	0.287*	0.146	0.483*	0.155	0.659*	0.134	0.767*	0.173
ITERS-R/ECERS-R gain score	0.045	0.127	-0.088	0.128	-0.113	0.147	0.044	0.151	0.100	0.142	-0.049	0.166
SR pretest	0.283*	0.076	-		-		-		-		-	
RA pretest	-		0.250*	0.087	-		-		-		-	
SL pretest	-		-		0.093	0.076	-		-		-	
VC pretest	-		-		-		0.267*	0.085	-		-	
DS pretest	-		-		-		-		0.134	0.085	-	
FPPI pretest	-		-		-		-		-		0.237*	0.085
Random parameters												
Center level	0.042	0.067	0.064	0.071	0.136	0.099	0.000	0.000	0.244	0.085	0.000	0.000
Group level	0.094	0.080	0.009	0.084	0.080	0.099	0.287	0.079	0.000	0.000	0.320	0.096
Caregiver level	0.338	0.056	0.482	0.080	0.428	0.071	0.300	0.050	0.411	0.061	0.412	0.068
Deviance	283.486		310.445		321.957		300.082		317.436		335.307	

Note. SR = sensitive responsiveness; RA = respect for autonomy; SL = structuring and limit setting; VC = verbal communication; DS = developmental stimulation; FPPI = fostering positive peer interactions.

*$p < .05$

Table 4 Effects of CIP Training on Caregiver Interactive Skills at Follow-up (Multilevel Analysis, $N = 139$)

	Sensitive responsiveness Model		*Respect for autonomy Model*		*Structuring and limit setting Model*		*Verbal communication Model*		*Developmental stimulation Model*		*Fostering peer interactions Model*	
	Estimate	*SE*	Estimate	*SE*	Estimate	*SE*	Estimate	*SE*	Estimate	*SE*	Estimate	*SE*
Fixed parameters												
Intercept	3.400*	0.403	2.425*	0.348	4.383*	0.406	2.040*	0.306	1.854*	0.247	1.840*	0.186
CIP training	0.366*	0.134	0.539*	0.122	0.173	0.144	0.488*	0.121	0.238	0.148	0.521*	0.135
ITERS-R/ECERS-R gain score	0.216*	0.100	0.202*	0.091	0.018	0.106	0.122	0.090	-0.014	0.115	-0.005	0.102
SR pretest	0.270*	0.079	-	-	-	-	-	-	-	-	-	-
RA pretest	-	-	0.404*	0.073	-	-	-	-	-	-	-	-
SL pretest	-	-	-	-	0.112	0.079	-	-	-	-	-	-
VC pretest	-	-	-	-	-	-	0.373*	0.076	-	-	-	-
DS pretest	-	-	-	-	-	-	-	-	0.179*	0.088	-	-
FPPI pretest	-	-	-	-	-	-	-	-	-	-	-0.058	0.080
Random parameters												
Center level	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.028	0.082	0.017	0.065
Group level	0.126	0.061	0.119	0.050	0.069	0.072	0.111	0.049	0.166	0.105	0.072	0.088
Caregiver level	0.375	0.063	0.258	0.048	0.516	0.086	0.290	0.048	0.402	0.067	0.452	0.075
Deviance	293.375		261.604		324.372		261.343		313.904		306.882	

Note. SR = sensitive responsiveness; RA = respect for autonomy; SL = structuring and limit setting; VC = verbal communication; DS = developmental stimulation; FPPI = fostering positive peer interactions.

*$p < .05$

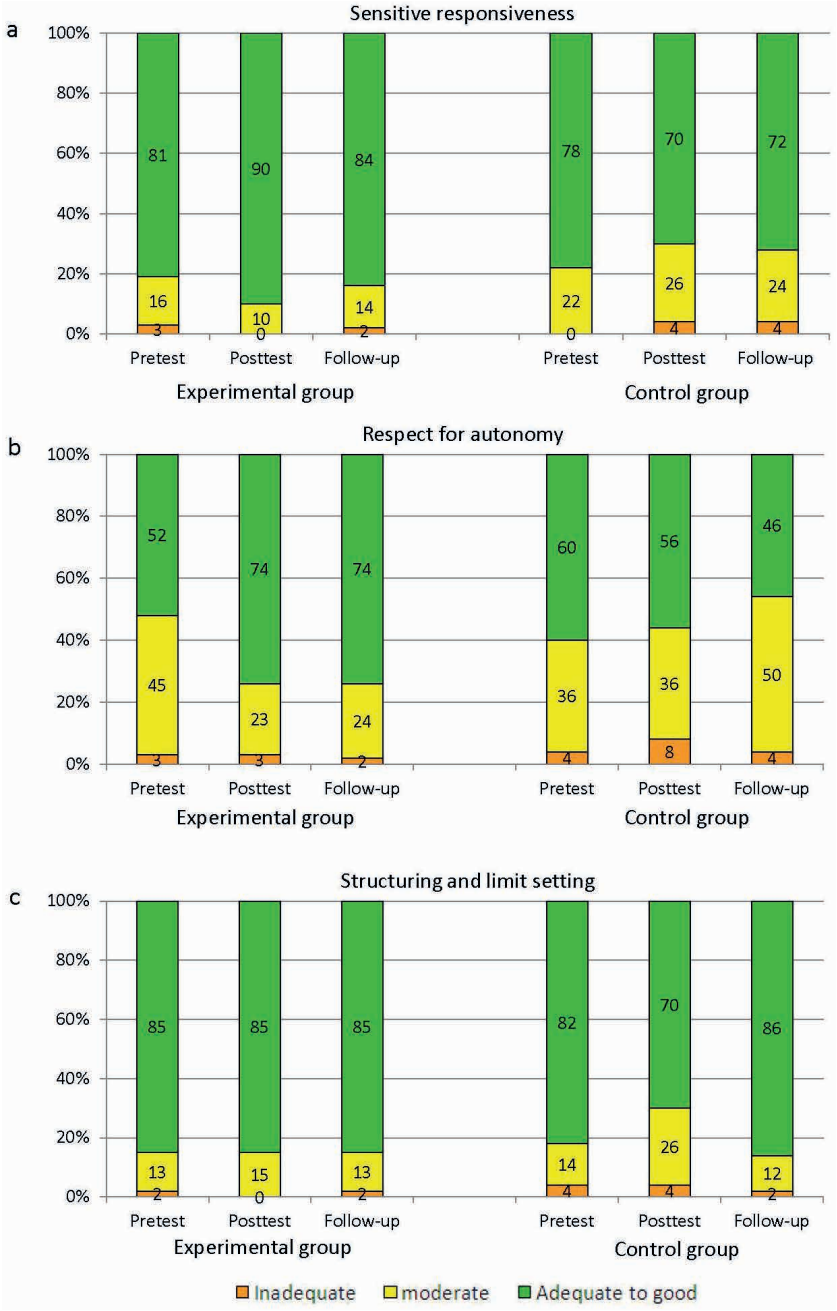

Figure 2a-c Percentages of Caregivers in Experimental Group (n = 62) and Control Group (n = 50) with Mean Scores "Inadequate", "Moderate", and "Adequate to Good" on Sensitive Responsiveness, Respect for Autonomy and Structuring and Limit Setting at Pretest, Posttest and Follow-up.

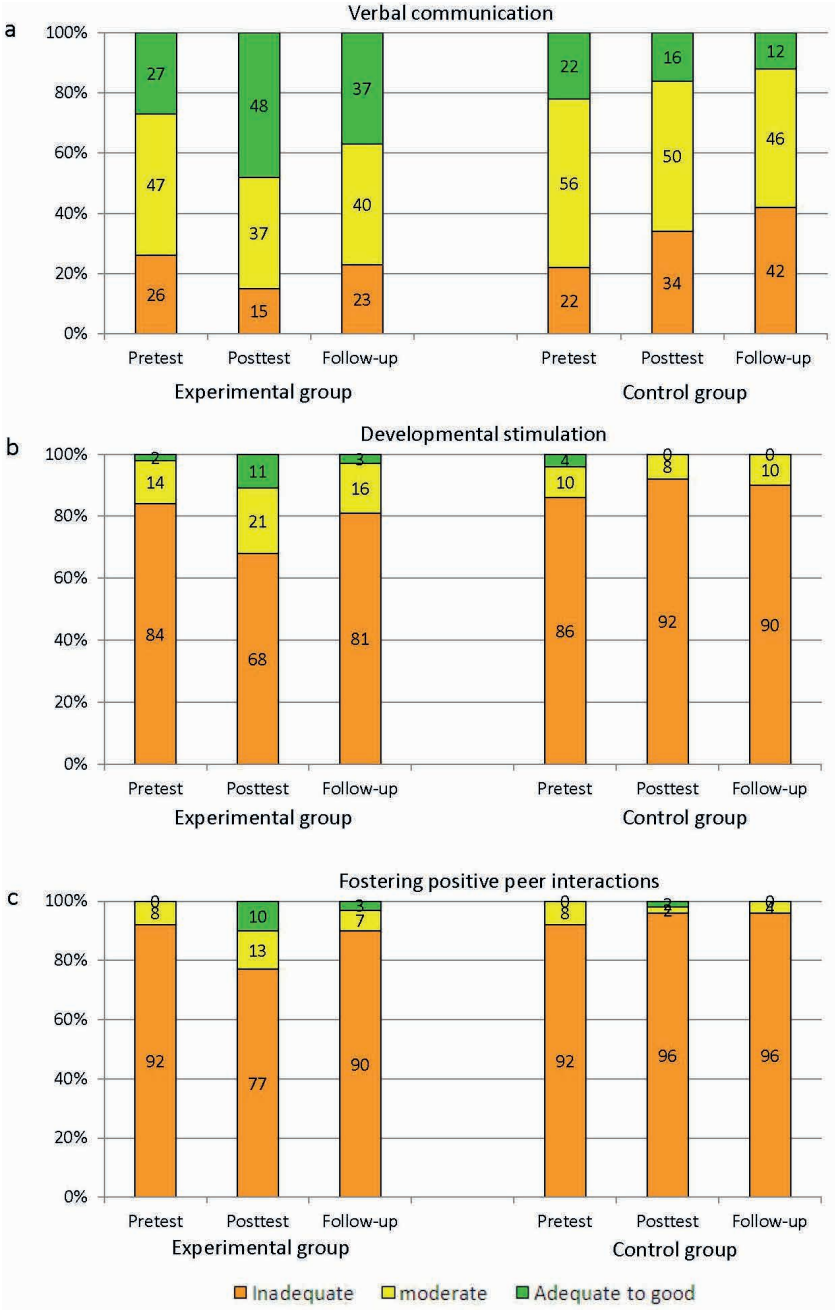

Figure 3a-c Percentages of Caregivers in Experimental Group (n = 62) and Control Group (n = 50) with Mean Scores "Inadequate", "Moderate", and "Adequate to Good" on Verbal Communication, Developmental Stimulation and Fostering Positive Peer Interactions at Pretest, Posttest and Follow-up.

4.4 Discussion

This RCT-study examined the effects of the CIP training, a 5-week video feedback training for caregivers in child care centers aimed at strengthening their interactive skills as measured with the CIP scales. Results at posttest indicate that the training had a positive effect on all six caregiver interactive skills. Three months after the posttest, at follow-up, caregivers in the experimental group still scored significantly higher on sensitive responsiveness, respect for autonomy, verbal communication, and fostering positive peer interactions than caregivers in the control group. Although training effects were relatively small in absolute terms, they appear meaningful in terms of decreasing percentages of caregivers with inadequate scores and increasing percentages of caregivers scoring in the adequate to good range.

The impact of the CIP training was most apparent for caregivers' educational skills (i.e., verbal communication, developmental stimulation and fostering positive peer interactions). This finding is highly relevant because these educational skills are of particular concern in The Netherlands. Results from the pretest scores of the present study (Figure 3) and large Dutch representative samples (Helmerhorst et al., 2014a) show that the majority of caregivers scored in the moderate quality category for verbal stimulation and in the inadequate category for developmental stimulation and fostering positive peer interactions. Given these low scores on the educational skills, it is promising that the results of the present study indicate that it is possible to strengthen these educational skills through video feedback training. But despite the significant improvement on the educational skills at the posttest, most caregivers still scored in the inadequate or moderate quality range after the training (see Figure 3). Furthermore, the increase in educational skills was only temporary; at follow-up, three months after the training, caregivers' scores for the educational skills were back at pretest level. A possible explanation may lie in the very low pretest level for the educational skills, especially for developmental stimulation and fostering positive peer interactions. This low level of educational skills is probably due to the nature of the regular vocational education in general social-pedagogic work that was completed by the vast majority of caregivers in our sample. This education prepares them to provide care for a broad age group (from children to elderly) and has a strong emphasis on care rather than education. So, especially with regard to the educational skills, the CIP training had no foundation to build on, which may explain why the posttest-levels were still low and training effects quickly disappeared. Evidently, a five-week training with only limited attention to the educational skills is not enough to produce sufficient and enduring improvement.

To increase the effect of the CIP training for the educational skills, we would first suggest adding extra sessions for developmental stimulation and fostering positive peer interactions because the initial level for these skills is, generally speaking, low. In the present training, these two skills are focused upon only by the end of the training course, namely during the third and fourth session (see Table 1).

It is very well possible that this did not give the caregivers enough time to strengthen these skills and to integrate them in their everyday behavioral repertoire. Therefore, the duration of the CIP program could be prolonged with additional booster sessions.

To prevent the drop in scores after the posttest we would, in addition, recommend a monitoring system after the training. This could be realized by using a system comparable with MyTeachingPartner of the CLASS (see Pianta et al., 2008). In this training, caregivers upload their own videos online and receive online feedback on their interactive skills complemented with concrete actions points by a trainer. Another and possibly more efficient long-term solution to ensure monitoring would be to provide an in-company training for staff members of the child care centers/organizations to assess and monitor their caregivers' interactive skills. This would require a *train the trainer* program, which we are currently developing. The first phase of the train the trainer program is devoted to training the staff members in reliably observing caregivers' interactive skills with a simplified version of the CIP scales, with high/medium/low ratings instead of a 7-point scale. In a next step, we train the staff members in providing video feedback to the caregivers. Future research should make clear whether the abovementioned adaptations contribute to the effectiveness of the CIP training, also in the long run.

Limitations and Future Directions

Despite the strengths of this study, including the randomized controlled design, relatively large sample size, and follow-up observations, we also acknowledge some important limitations. First of all, child care centers volunteered to the study, which implies that our sample was selective, which limits generalizability of the findings to the general population of Dutch child care centers. To check for possible selectivity, we compared caregivers' pretest scores in our sample with scores on the CIP scales obtained in a large representative sample of Dutch caregivers in the same period (see Fukkink, Gevers Deynoot-Schaub, Helmerhorst, Bollen, & Riksen-Walraven, 2013). Results indicated no significant differences for any of the CIP scales, except for fostering positive peer interactions. On this scale, caregivers in the present study scored significantly lower than those in the concurrent representative sample ($M_{present}$ = 2.03, $SD_{present}$ = 0.8, $M_{representative}$ = 2.50, $SD_{representative}$ = 1.0, t = 4.01, p < .01), indicating that the present sample is indeed not representative in this respect for caregivers in The Netherlands.

A second limitation lies in the fact that the present CIP training for caregivers was conducted parallel to a consultancy program for directors of the same centers. Although the consultancy program was not directed at improving caregiver interactive skills and although we controlled for possible effects of the consultancy program, possible confounding cannot be completely ruled out as an explanation for the effects of the CIP training. Further research examining the effects of an independently conducted CIP training may clarify this question.

Third, the positive results do not automatically imply that this training is also effective in other countries. The effectiveness of the training may depend, for example, on caregiver education, which shows considerable variation across countries (Oberhuemer, Schreyer, & Neuman, 2010). It should also be kept in mind that the CIP scales were initially developed for use in child care centers in The Netherlands and therefore reflect Dutch childcare values. As described in a former paper (Helmerhorst et al., 2014b) our choice of the six caregiver skills included in the CIP was also based on the results of a survey among different groups of stakeholders in child care, which showed that Dutch parents, caregivers, center directors, and external experts recognized these skills as important child care quality indicators. Although the caregiver skills included in the CIP scales are generally recognized as important by researchers and are also included in measures developed in other countries (e.g., the ORCE, see NICHD ECCRN, 1996; and the CLASS, see La Paro et al., 2004), it cannot be automatically assumed that these skills are also seen as key aspects of pedagogical quality in other countries.

A fourth limitation is that not all caregivers per care group were trained. For practical reasons, we only trained the caregivers that were present during the day of the pretest observations. Because most caregivers work part-time, it is very common in The Netherlands that children are cared for by more than three caregivers during the week. Prior research has shown that caregiver behavior is for the most part determined by individual caregiver characteristics, which suggests that children may experience large differences in the quality of interactions with all the different caregivers in the care group (Helmerhorst et al., 2014b). Therefore, to get a complete picture of children's everyday experiences in the care group, it would be better to observe all caregivers of the care group. To improve the quality of the children's everyday experiences in the care group, it would be beneficial to train all caregivers to strengthen their interactive skills. For future research, we therefore recommend to include all caregivers of the care group in a CIP training. Taking this a step further, it would also be interesting to not only examine the effects of the CIP training on the caregivers' interactive skills, but to also examine its effects on the development and functioning of the children in their care group. This would also make it possible to examine whether including more caregivers per group in the CIP training will indeed increase the effect of the training on the children.

Implications for Policy and Practice

The results of the present study show that it is possible to strengthen important caregiver interactive skills through an extensive training, which underscores that it is worthwhile to invest in this type of training. The present training could be made available nationally to booster the interactive skills of caregivers in child care centers. The present set-up of the training may be too financially challenging in case of large-scale implementation. Therefore, we are currently working on an adjustment of the training to make implementation in practice more feasible. Although prior

research and the present study suggest that training caregivers individually is an important element in effective programs, it seems worthwhile to examine the possibility of training all caregivers of the care group together in focus groups. The advantage could be that caregivers learn from each other from the beginning instead of only during the last session, which may in turn lead to better retention of the gains made during training, because caregivers could monitor each other afterwards. Pilot studies are needed to examine whether the program could also be effectively used when all caregivers of the care group are trained together.

An even more effective way of improving caregiver interactive skills and thereby the quality of child care, would be to incorporate the present training in the regular education of caregivers. The current regular vocational education for caregivers in child care does not educate students adequately in how to interact with young children in child care centers. Therefore, a straightforward next step would be to adapt the CIP training for implementation in the regular caregiver education, which requires the development of a *train the trainer* program for teacher educators in caregiver education. We are currently developing an adapted version of the training and will examine the effectiveness of its implementation in caregiver education in future research.

The most powerful way to improve the quality of caregiver-child interactions in child care centers would be to combine the two measures for strengthening caregiver skills that we recommended in the former paragraphs. This means that the future approach for enhancing the quality of care and education for young children should be twofold: training caregivers who already work in the field *and* investing in the interactive skills of future caregivers by implementing the CIP training in the curriculum of caregiver education.

References

Bakermans-Kranenburg, M. J., Van IJzendoorn, M. H., & Juffer, F. (2003). Less is more: Meta-analyses of sensitivity and attachment interventions in early childhood. *Psychological Bulletin*, *129*, 195-215. doi: 10.1037/0033-2909-129.2.195

Belsky, J., Vandell, D. L., Burchinal, M., Clarke-Stewart, K. A., McCartney, K., Owen, M. T., & the NICHD Early Child Care Research Network (2007). Are there long-term effects of early child care? *Child Development, 78*, 681-701. doi: 10.1111/j.1467-8624.2007.01021.x

Bronfenbrenner, U., & Ceci, S. J. (1994). Nature-nurture reconceptualized: A bioecological model. *Psychological Review, 101,* 568–586.

Domitrovich, C. E., Gest, S. D., Gill, S., Bierman, K. L., Welsh, J. A., Jones, D. (2009). Fostering high-quality teaching with enriched curriculum and professional development support: The Head Start REDI program. *American Educational Research Journal*, *46*, 567-597. doi: 10.3102/0002831208328089

Fantuzzo, J., Childs, S., Stevenson, H., Coolahan, K. C., Ginsburg, M., Gay, K., . . . & Watson, C. (1996). The Head Start Teaching Center: An evaluation of an experimental, collaborative training model for Head Start teachers and parent volunteers. *Early Childhood Research Quarterly*, *11*, 79-99.

Fantuzzo, J., Childs, S., Hampton, V., Ginsburg-Block, M., Coolahan, K. C., & Debnam, D. (1997). Enhancing the quality of early childhood education: A follow-up of an experimental, collaborative training model for Head Start. *Early Childhood Research Quarterly*, *12*, 425-437.

Fukkink, R. G., Gevers 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- tot 4-jarigen in Nederlandse kinderdagverblijven in 2012* [Pedagogical quality of child care centers for 0- to 4-year-olds in The Netherlands 2012]. Amsterdam, The Netherlands: Netherlands Consortium of Child Care Research.

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

Fukkink, R. G., & Tavecchio, L. W. C. (2010). Effects of video interaction guidance on early childhood teachers. *Teaching and Teacher Education*, *26*, 1652-1659. doi: 10.1016/j.tate.2010.06.016

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

Girolametto, L., Weitzman, E., & Greenberg, J. (2004). The effects of verbal support strategies on small-group peer interactions. *Language, Speech, and Hearing Services in Schools, 35* (3), 254-268.

Guo, Y., Piasta, S. B., Justice, L. M., & Kaderavek, J. N. (2010). Relations among preschool teachers' self-efficacy, classroom quality, and children's language and literacy gains. *Teaching and Teacher Education*, *26*, 1094-1103. doi: 10.1016/j.tate.2009.11.005

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

Harms, T., Cryer, D., & Clifford, R. M. (2003). *Infant/Toddler Environment Rating Scale-Revised*. New York, NY: Teachers College Press.

Helmerhorst, K. O. W., Riksen-Walraven, J. M. A., Gevers Deynoot-Schaub, M. J. J. M., Tavecchio, L. W. C., & Fukkink, R. G. (2014a). Child care quality in The Netherlands over the years: A closer look. *Early Education and Development*. doi: 10.1080/10409289.2014.948784.

Helmerhorst, K. O. W., Riksen-Walraven, J. M. A, Vermeer, H. J., Fukkink, R. G., & Tavecchio, L. W. C. (2014b). Measuring interactive skills of caregivers in child care centers: development and validation of the caregiver interaction profile scales. *Early Education & Development, 25,* 770-790. doi: 10.1080/10409289.2014.840482

La Paro, Hamre, B. K., Locasale-Crouch, J., Pianta, R. C., Bryant, D., Early, D., . . . Burchinal, M. (2009). Quality in kindergarten classrooms: observational evidence for the need to increase children's learning opportunities in early education classrooms. *Early Education and Development*, *20*, 657-692. doi: 10.1080/10409280802541965

La Paro, k., Pianta, R., & Stuhlman, M. (2004). Classroom assessment scoring system (CLASS). Findings from the pre-k year. *The Elementary School Journal*, *104*, 409-426.

LoCasale-Crouch, J., Konold, T., Pianta, R., Howes, C., Burchinal, M., Bryant, D., . . . Barbarbarin, O. (2007). Observed classroom quality profiles in state-funded pre-kindergarten programs and associations with teacher, program, and classroom characteristics. *Early Childhood Research Quarterly*, *22*, 3-17. doi: 10.1016/j.ecresq.2006.05.001.

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

NICHD Early Child Care Research Network (1996). Characteristics of infant child care: Factors contributing to positive caregiving. *Early Childhood Research Quarterly*, *11*, 269-306. doi: 10.1016/S0885-2006(96)90009-5

Oberhuemer, P., Schreyer, I., & Neuman, M. J. (2010). *Professionals in early childhood education and care systems. European profiles and perspectives*. Opladen & Farminghton hills: Barbara Budrich Publishers.

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

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

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

Raver, C. C., Jones, S. M., Li-Grining, C. P., Metzger, M., Champion, K. M., & Sardin, L. (2008). Improving preschool classroom processes: Preliminary findings from a randomized trial implemented in Head Start settings. *Early Childhood Research Quarterly*, *23*, 10-26. doi: 10.1016/j.ecresq.2007.09.001

Sandstrom, H. (2012). The characteristics and quality of pre-school education in Spain. *International Journal of Early Years Education*, *20,* 130-158. doi: 10.1080/09669760.2012.715012

Tayler, C., Ishimine, K., Cloney, D., Cleveland, G., & Thorpe, K. (2013). The quality of early childhood education and care services in Australia. *Australasian Journal of Early Childhood, 38*(2), 13-21.

Thomason, A. C., & La Paro, K. M. (2009). Measuring the quality of teacher-child interactions in toddler child care. *Early Education and Development, 20*(2), 285-304. doi:10.1080/10409280902773351

Vandell, D., Belsky, J., Burchinal, M., Steinberg, L., Vandergrift, N., & the NICHD Early Child Care Research Network (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

Vandell, D. L., & Wolfe, B. (2000). *Child Care quality: Does it matter and does it need to be improved?* Madison, STATE: Institute for Research on Poverty, University of Wisconsin-Madison.

CHAPTER 5
Improving Quality of the Child Care Environment through a Consultancy Program for Center Directors

Foto: Joods Historisch Museum

Helmerhorst, K. O. W., Fukkink, R. G., Riksen-Walraven, J. M. A., Gevers Deynoot-Schaub, M. J. J. M., & Tavecchio, L. W. C. (2014). *Improving quality of the child care environment through a consultancy program for center directors*. Manuscript submitted for publication.

Abstract

Repeated quality assessments in The Netherlands in 1995, 2001, 2005, and 2008 in representative samples of child care centers have demonstrated that quality of center based care has been steadily decreasing to a worrisome level over the last decades. Therefore, we developed a consultancy program for center directors to increase quality. This study examined the effects of a newly developed on-site consultancy program to improve global quality of the child care environment as measured with the ITERS-R and ECERS-R. Using a randomized controlled trial study design with a pretest, posttest, and follow-up test after 3 months, we compared an experimental group (n = 35) with a control group (n = 33). The consultancy program comprises three consultations in total, with two on-site consultations, lasting about two hours, and a third consultation given per telephone. The consultancy program had no significant overall effect on the ITERS-R and ECERS-R total scores. But when the analysis was restricted to include only those items that were specifically targeted during the consultancy, results demonstrated a significant increase between pretest and posttest and between posttest and follow-up for both the ITERS-R and the ECERS-R items targeted during the consultancy. The results were also meaningful in terms of changes from inadequate to adequate quality categories as defined for the ITERS-R and ECERS-R. Quality of the child care environment improved during the consultancy and this effect was retained over time.

5.1 Introduction

Ample research has shown that the quality of child care contributes to children's social emotional and cognitive development (for an overview see Belsky et al., 2007; Vandell & Wolfe, 2000; Vandell et al., 2010). However, repeated quality assessments in 1995, 2001, 2005 and 2008 in nationally representative samples of child care centers indicate that the quality of center based care in The Netherlands has been steadily decreasing over the last decades (Helmerhorst, Riksen-Walraven, Gevers Deynoot-Schaub, Tavecchio, & Fukkink, 2014b; Vermeer et al., 2008). In these studies, the quality of the child care environment was measured using the Infant/Toddler Environment Rating scale (ITERS-R; Harm, Cryer, & Clifford, 2003) and the Early Childhood and Environment Rating Scale (ECERS-R; Harms, Clifford, & Cryer, 1998). The ITERS-R/ECERS-R total score decreased from 4.8 on a 7-point scale in 1995 to 4.3 in 2001, 3.6 in 2005, and 3.0 in 2008. These scores are low both in absolute terms and in international perspective (Helmerhorst et al., 2014b; Vermeer et al., 2008). In 2008, none of the groups scored in the category good to excellent, 49% scored moderate, and more than half (51%) of the groups scored inadequate. With regard to the subscales, mean scores for the space and furnishings subscale and the interactions subscale in the last assessment were significantly lower than the scores from the 2005 assessment (Helmerhorst et al., 2014b). These results clearly underscore the need for improving quality of care in Dutch child care centers. Therefore, we developed a comprehensive intervention program to improve quality of care in Dutch child care centers. The complete intervention program consists of two parts with a different focus and directed at different staff members of the child care center. One part was a video-feedback training for caregivers, which aimed to strengthen their skills in interacting with the children. The other part, which is described in this study, is a consultancy program for center directors aiming to improve global quality of the child care environment (i.e., space, furnishings, materials, program) as measured with the ITERS-R/ECERS-R. Henceforth, quality of the child care environment as measured with the ITERS-R/ECERS-R will be referred to as "global quality" in this paper. We chose to direct the intervention for improving global quality at center directors because, according to Bloom and Sheerer (1992), it is the center director who generally manages financial resources and decides about provisions, materials, and the program. They can therefore be seen as the "gatekeeper to quality". The present study is an effect study with randomized controlled design of an on-site consultancy program directed at center directors that aims to improve global quality of the child care environment of child care groups.

Results of Previous Studies on Effectiveness of Consultancy in Child Care Settings

Up till now, research focusing on improving global quality through on-site consultancy has come up with mixed results. Because our intervention was directed at

center directors, we started by searching for prior interventions aimed at improving global quality and directed at center directors. We found only one study, by Bloom and Sheerer (1992), who evaluated a 16-month training that focused on a broad set of domains (i.e., personal and professional self-knowledge, child development, organizational theory, leadership, parent relations), with quality of the classroom being one of the outcome variables. Overall classroom quality, as measured with the Early Childhood Classroom Observation Scale (ECCOS; Bredekamp, 1986), was significantly higher in target groups than in comparison groups. Next, we broadened our search and found several intervention studies directed at caregivers of a care group. In a review study Zaslow, Tout, Halle, Vick Whittaker, and Lavelle (2010) discuss five studies that are specifically focused on strengthening global quality of the child care environment (Campbell & Milbourne, 2005; Fiene, 2002; Kontos, Howes, & Galinksy, 1996; Palsha & Wesley, 1998; Wesley, 1994). First, a study by Fiene (2002) describes a 4-month intensive mentoring program for infant caregivers given by an early childhood professional. Unfortunately, little information is given about the content of the mentoring program. No effects were found on global quality of the classroom environment as measured with the ITERS (Harms, Clifford, & Cryer, 1990). A second study describes an on-site training for family child care providers (including workshops and home-visits) that modestly improved global quality as measured with FDCRS (Harms & Clifford, 1989): 19% of the providers made "observable" change (improving one scale point or more on the FDCRS), 73% made no change, and 8% got observable worse (Kontos et al., 1996). Modest effects were also found in a third study, by Campbell and Milbourne (2005). In this study, infant-toddler caregivers received both a course on issues related to infant-toddler care and three 1-hour consultation visits. Caregivers used a self-assessment instrument to indicate targets for improvement that were related to the ITERS subscales space and furnishings, personal-care routines, activities, interactions, program structure, and adult needs. Results showed no significant improvement on the ITERS subscales, although 21% of the caregivers in the consultation group showed observable change (e.g., a change from a rating of *inadequate* to *adequate* quality), as compared to 8% in the non-consultation group. Finally, two studies describing on-site interventions demonstrated significant improvement on global quality in center-based infant-toddler and preschool classrooms (Palsha & Wesley, 1998; Wesley, 1994). Sample sizes in these studies were small and a control group was lacking. Both studies used the ITERS and Early Childhood Environment Rating Scale (ECERS; Harms & Clifford, 1980) rating scales as the basis for the intervention; the intervener trained the child care provider to use the rating scales themselves, and the intervener and child care provider together drew up an action plan for improvement. The study of Palsha and Wesley (1998) found a significant improvement by more than a half scale point on the ITERS (n = 6 groups) and ECERS (n = 14 groups) total score.

Taken together, results from previous studies on effectiveness of on-site interventions aimed at improving global quality are promising but far from conclusive.

None of the above mentioned intervention studies examined the effects of the intervention using a randomized controlled trial (RCT) design, which is generally acknowledged the "gold standard" for evidence of intervention effects. There thus seemed to be a need for research using a randomized controlled trial to draw conclusions about the effects of this type of intervention.

Consultation Model

The abovementioned studies of Palsha and Wesley (1998) and Wesley (1994), in which they used their on-site consultation model, showed the most favorable results. The model describes a framework with several phases in which the consultant and consultee work in productive collaboration on overall quality improvement with a strong focus on tailoring the consultancy to the individual needs of the consultee or specific center. The model was used as a theoretical framework underlying the current consultancy intervention and served as a foundation for the set-up (Palsha & Wesley, 1998; Wesley & Buysse, 2004). The first phase in the consultation process according to this model is *establishing the consulting relationship*. The consultant starts with gaining the trust of the consultee by creating a collaborative basis and together they develop an action plan for improvement. During the next phase, *assessment or problem identification*, the consultant focuses on determining the factors needed for improvement. Self-assessment with a broad quality measure as the ECERS-R/ITERS-R can be useful for this. This self-assessment is important because it shows staff that their input in the consultation process is vital; staff is actively involved in setting goals and standards for their future professional development; and it allows staff to get insight in where they stand (Wesley, 1994). Altogether, this provides better chances for maintaining quality improvement over time when consultancy is completed. Subsequently, consultant and consultee collaboratively *identify and define the goal for change*. This phase highlights the often unique and group-specific nature of the intended quality improvement. In the next phase of *strategy selection*, consultant and consultee together work on strategies for improvement. The consultee plays an active role in this phase. During the *implementation* phase the consultant gives advice on how to implement the plan of action. In the *evaluation* phase *the match between the desired and actual outcomes identified in the original consultation plan* is assessed. During this assessment underlying mediating factors, that may influence the intervention outcomes, have to be identified. Finally, during the *termination* phase, consultants should gradually assign their tasks over to the consultee. That way consultees can continue to work on quality improvement when the consultant has left.

Present Study

The main aim of this study was to examine the effects of a newly developed on-site consultancy program to improve the global quality of the child care environment as

measured with the ITERS-R and ECERS-R. The consultancy program, which included three sessions, was based on the consultancy model described above and focused on four quality domains, represented by the ITERS-R/ECERS-R subscales *Space and furnishings, Language, Activities,* and *Program structure.* Consultancy specifically targeted items that consultant and consultee together identified as "weak areas" for a given care group. We examined the effects of the consultation program both on the overall quality of the child care environment as measured with the relevant ITERS-R/ECERS-R subscales and, more specifically, on the specific ITERS-R/ECERS-R items targeted during consultation.

5.2 Method

Participants and Randomization

Child care groups in this study were recruited from child care centers in and around Amsterdam, The Netherlands. Center directors responded to appeals in (digital) newsletters and announcements on child care websites. Centers could participate only with an even number of care groups, because we aimed to randomly assign half of the participating groups per center to the experimental condition and half to the control condition. Furthermore, participating groups had to be mixed-age groups (0- to 4-year-olds), because we wanted to focus on groups with children across the whole possible age range and because mixed-age groups are more prevalent in The Netherlands than the other two possible types of groups in Dutch centers, i.e., infant groups (0- to 2-year-olds) and preschool groups (2- to 4-year-olds) (see Helmerhorst et al., 2014b).

A total of 68 mixed-age groups from 33 child care centers participated in the study: 35 intervention groups and 33 control groups, which were randomly assigned to the two conditions per center. Across the 35 experimental groups, a total of fourteen center directors participated in the consultancy program with one group, six center directors with 2 groups, and three center directors with 3 groups.

Figure 1 shows the flow chart of the participating groups per phase. Originally, the sample consisted of 35 intervention groups and 35 control groups, but two control groups dropped out after randomization and before the pretest. Three groups (one in the experimental condition and two in the control condition) discontinued between posttest and follow-up because the child care center or the group had been closed. None of the participating center directors dropped out in the course of the intervention program.

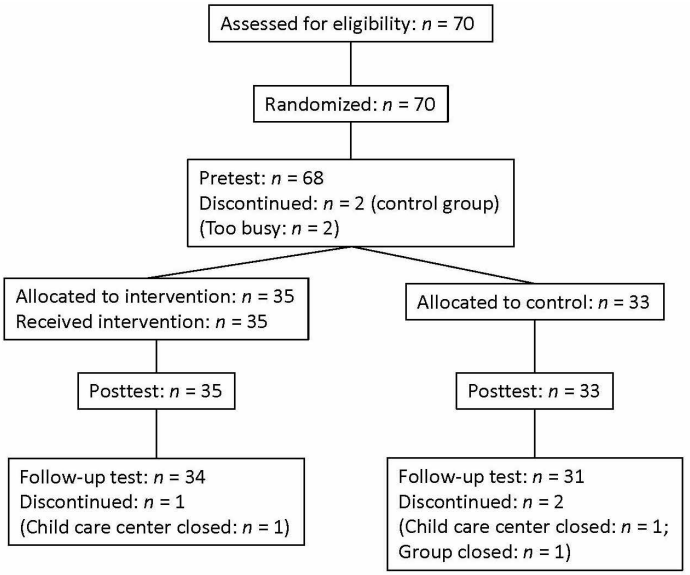

Figure 1 Flow Chart of the Study's Progress in Terms of Groups.

On average, child care centers existed 12.4 years (*SD* = 9.45, range 0.5-30) and had 3.7 groups per center (*SD* = 1.84, range 1-8). Center directors were all female; on average, they were 42.8 years old (*SD* = 9.77, range 27-61), worked 28.3 hours a week (*SD* = 7.7), and had 13.4 years (*SD* = 8.5) of working experience in child care. About half (52%) of the center directors had a bachelor degree, 28% completed regular vocational training, and 10% had a master degree. The remaining 10% only completed secondary school.

We found no differences at the pretest between the experimental and the control group for either the ITERS-R or ECERS-R subscales (Wilks' Lambda = .947, $F(4, 63) = 0.87$, $p = .49$, $\eta^2_{partial} = .053$ and Wilks' Lambda = .955, $F(4, 63) = 0.74$, $p = .57$, $\eta^2_{partial} = .045$, respectively).

Design and General Procedure

We used a randomized controlled study design with an experimental and control condition. The dependent variables, ITERS-R and ECERS-R scores, were measured at pretest, at posttest directly after the intervention, and at 3-month follow-up.

As noted in the Introduction, this study was part of a more comprehensive intervention program; simultaneously to the present consultancy program for center directors aimed at improving global quality, we also conducted a video-feedback training for professional caregivers that specifically aimed at improving the quality of caregiver-child interactions as measured with the Caregiver Interaction Profile scales (CIP scales; Helmerhorst, Riksen-Walraven, Vermeer, Fukkink, & Tavecchio, 2014c). In the present study, we focused on the effects of the consultancy aimed at

global quality of the child care environment as measured with the ITERS-R and ECERS-R and controlled for possible effects of the CIP intervention (see below under Measures: CIP scales).

All groups in experimental and control condition were visited by a trained observer for pretest, posttest, and follow-up test. The visit lasted from about 8 a.m. until about 3 p.m. . In order to obtain independent ratings, observers visited a care group only once. At the pretest, posttest, and follow-up, the observer scored global process quality for the group with the ITERS-R and ECERS-R, and interviewed one of the caregivers to collect additional information with regard to the ITERS-R and ECERS-R (see *Measures*) that was not available through observation at the end of the observation. In addition, the observer made video recordings of each caregiver, which were rated afterwards for a separate study examining the effects of the parallel video-feedback training to improve caregivers' interactive skills (see Helmerhorst, Riksen-Walraven, Fukkink, Tavecchio, & Gevers Deynoot-Schaub, 2014a). Two weeks after the pretest, the center director was visited by the consultant for the first time. Two weeks after the last consultation, or six weeks after the pretest for the control groups, each group was visited for the posttest. Three months after the posttest, each group was visited for the follow-up observation. The posttest and follow-up visits were planned on the same day of the week as the pretest for an optimal comparison. At the pretest, center directors also completed a questionnaire to collect individual background information (i.e., education and work experience).

Intervention program

Experimental group. The present consultancy program was conducted by two consultants; one consultant had a master degree in developmental psychology, the other had a PhD in psychology. Both consultants had experience in working with the quality of the child care environment as evaluated with the NCKO Quality Monitor (Gevers Deynoot-Schaub et al., 2009; see below) and used a standardized manual.

The consultancy program comprised three consultations in total, with two onsite consultations, each lasting about two hours, and a third consultation given per telephone. During the first consultation, two weeks after the pretest, the consultant explained the general consultancy procedure and provided the center director with background information of the research. As a first step in the consultancy process, the center director and the consultant jointly chose targets for the consultancy program by means of a self-assessment tool. For this self-assessment we used the NCKO Quality Monitor, which is a tool developed for professionals in child care centers to broadly self-evaluate the quality of the child care environment of the care groups with a checklist. This quality of the child care environment (QCCE) checklist is rated with 26 quality indicators on a dichotomous scale with a positive and a negative anchor and is based on the ITERS-R and ECERS-R items from the subscales space and furnishings, activities, language and program structure. The positive

anchor represents items that should be present (i.e., represented by a "thumbs up" symbol) in the care group and the negative anchor represents items that should not be present (i.e., represented by a "thumbs down" symbol). By specifying the negative items and positive items, the QCCE checklist provides the center director with direct insight into a center's stronger and weaker points and which items need improvement. During the first consultation, the consultant explained the center directors how to administer the QCCE checklist. After the first consultation, center directors were asked to fill in the QCCE checklist by themselves, and once completed, to send it back to the intervener by mail.

To prepare for the second session, the consultant analyzed both the QCCE checklist filled in by the center director and the ITERS-R and ECERS-R scores from the pretest, and drew up an action plan based on both forms. The consultancy protocol states that ITERS-R/ECERS-R items with a score 3 or lower would be marked with priority on the action plan. The number of action items was 13.4, on average (SD = 2.91), with a minimum of 8 and a maximum of 20 action points. Table 1 shows all items that were targeted during the consultancy and the number of care groups for which each item was targeted. As shown in the table, in all groups consultancy concerned activities for blocks and math/numbers (n = 35), in most groups it concerned child related display of visual materials (i.e., posters, collages, children's artwork, pictures) for children, and music/movement (n = 34). In general, items of the subscale program structure were targeted least.

During the second consultation, about two weeks after the first, the consultant brought back the QCCE checklist filled in by the center director together with the action plan. After careful consideration, the center director and the consultant jointly determined the definitive action points and developed a plan for quality improvement for the group in the given areas. This way the consultancy was tailored to the individual and unique needs of the group. The consultant advised the center director on how to implement the improvement plan. After the second visit, the center director worked on the action items list.

The third consultation was a final phone call about two weeks later to check with the center director whether the action items were feasible, and to answer questions when needed.

Control group. For the groups that were assigned to the control condition, center directors received no intervention at all and were only contacted to schedule the pretest, posttest, and follow-up test.

Table 1 Items Targeted during Consultancy and Numbers of Care Groups for Which the Items Were Targeted (Experimental Group, *n* = 35)

Item Targets	*n*
Space and Furnishings	
Indoor space	2
Furniture for care, play and learning	17
Furnishings for relaxation and comfort	28
Room arrangement for play	24
Child-related display	34
Space for gross motor play	14
Gross motor equipment	20
Activities	
Fine motor	30
Art	16
Music/movement	34
Blocks	35
Dramatic play	27
Sand/water	18
Nature/science	27
Math/number	35
Use of TV, video, and/or computers	3
Promoting acceptance of diversity	33
Language	
Encouraging children to communicate	1
Using language to develop reasoning skills	4
Books and pictures	30
Program structure	
Schedule	3
Free play	7
Group time	1

Measures

ITERS-R/ECERS-R. The ITERS-R (Harms et al., 2003) and its equivalent for preschool classrooms, the ECERS-R (Harms et al., 1998), are widely used to measure process quality in child care groups. The ITERS-R was developed for use in groups in which more than 50% of the children is under the age of 30 months, whereas the ECERS-R was developed for use in groups in which more than 50% of the children are between the ages of 30 and 48 months. Because only mixed-aged groups (0- to 4-year olds) participated in this study, we used both ITERS-R and ECERS-R in each group at the pre-, posttest, and follow-up to capture quality for both infants and

toddlers. Both the ITERS-R and the ECERS-R include 7 subscales: (a) Space and Furnishings, (b) Personal Care Routines, (c) Language, (d) Activities, (e) Interactions, (f) Program Structure, and (g) Provisions for Parents and Staff. Items are rated on a 7-point scale with descriptors for the scores 1 (inadequate), 3 (minimal), 5 (good), and 7 (excellent). We only used the subscales that were targeted in the consultancy program, i.e., Space and furnishings, Language, Activities, and Program structure. A total score (20 items for the ITERS-R and 24 items for the ECERS-R) was computed by averaging item scores across the four subscales. In addition, groups were classified according to the quality levels *low* ($M < 3$), *moderate* ($3 \leq M < 5$), and *high* ($M \geq 5$). Internal consistency for the ITERS-R (Cronbach's alpha) was .77 at the pretest, .63 at the posttest, and .77 at the follow-up test. For the ECERS-R Cronbach's alpha was .70 at the pretest, .69 at the posttest, and .78 at the follow-up test.

Prior to data collection, eight observers were trained to use both the ITERS-R and the ECERS-R. Each observer performed at least 4 on-site visits (range 4 - 6) supervised by an expert trainer, followed by an item-by-item debriefing. The average interobserver agreement (i.e., intraclass correlations) during training was .84, on average, ranging from .80 to .88.

CIP scales. Quality of caregiver-child interactions was rated from the videotaped episodes with the CIP scales (Helmerhorst et al., 2014b); the scores were used here to control for possible effects of the caregiver training (CIP training). The CIP scales reflect six caregiver interactive skills: *sensitive responsiveness, respect for autonomy, structuring and limit setting, verbal communication, developmental stimulation*, and *fostering positive peer interactions*. Each of the CIP scales is rated on a single 7-point scale (7 = *very high*, 6 = *high*, 5 = *moderate/high*, 4 = *moderate*, 3 = *moderate/low*, 2 = *low*, 1 = *very low*). For a more comprehensive description of the CIP scales see Helmerhorst et al. (2014c). Per care group a CIP-total score was calculated by first averaging the scores of the six skills per caregiver and then averaging the scores of all caregivers per group. Next, a gain score was computed by calculating the difference between the pretest and follow-up CIP-total scores. This gain score was included in the analyses to control for possible effects of the parallel video-feedback training for caregivers that aimed to improve their interactive skills as measured with the CIP scales.

5.3 Results

Effects of the Consultancy Program on the Total ITERS-R and ECERS-R Scores

Table 2 shows the descriptives for the ITERS-R and ECERS-R pretest, posttest, and follow-up scores across the four relevant subscales and the total score in both the experimental and control groups. Table 2 shows that the total scores and subscale scores in the experimental groups increased over time. The ITERS-R and ECERS-R

total scores in the control group decreased between pretest and posttest, but increased again between posttest and follow-up.

Because we had measurements at three times (pretest, posttest, follow-up; level 1) per group, and groups nested within center directors (level 2), we conducted a multilevel analysis. Two models were tested, one with the ITERS-R total score and one with the ECERS-R total score as the dependent variable, with 'group' as the between-subjects variable (a dummy variable: 0 = control group, 1 = experimental group) and time (pretest, posttest, follow-up) as the within-subjects variable. The gain score of the parallel intervention was included as a covariate in the model. We found no significant Group × Time interaction effect for the consultancy program, indicating that the effect of the consultancy on the ITERS-R and ECERS-R total scores was not significant.

Effects of the Consultancy Program on Targeted Items

Next, we examined the effect of the consultancy more specifically on the ITERS-R and ECERS-R items that had been targeted during consultancy. For that purpose we computed an aggregated variable for each experimental group for the ITERS-R and ECERS-R by averaging scores on only those items that had been targeted for the care group in question. We hypothesized that this variable, which is referred to as "targeted items", was more sensitive to intervention effects, because the exact focus of the consultancy varied considerably between groups and was, therefore, related to different ITERS-R and ECERS-R items. We examined the effect of the consultancy on the targeted items. This analysis only applied to the experimental group (there were no targeted items in the control group). We used multilevel repeated measures analysis separately for the ITERS-R and ECERS-R. The dependent variable in both models was the aggregated targeted items variable, time was the repeated measures factor with a pretest, posttest and follow-up, and again the gain score of the parallel CIP training was included as a control variable.

Table 2 Mean Subscale and Total Scores for the ITERS-R and ECERS-R in the Experimental and Control Group at Pretest, Posttest and Follow-up

Variable	*Experimental group*			*Control group*		
	Pretest (n = *35*)	*Posttest* (n = *35*)	*Follow-up* (n = *34*)	*Pretest* (*n = 33*)	*Posttest* (n = *33*)	*Follow-up* (n = *31*)
	M (SD)	M (SD)	M (SD)	M (SD)	M (SD)	M (SD)
ITERS						
Space and furnishings	2.84 (0.51)	3.25 (0.61)	3.30 (0.76)	3.01 (0.72)	3.12 (0.51)	3.16 (0.59)
Language	3.44 (0.80)	3.80 (1.07)	4.21 (1.07)	3.26 (0.93)	3.15 (0.95)	3.49 (1.22)
Activities	2.18 (0.51)	2.34 (0.67)	2.49 (0.70)	2.21 (0.65)	2.45 (0.49)	2.49 (0.62)
Program structure	3.99 (0.88)	4.25 (0.98)	4.58 (1.09)	4.09 (1.11)	3.62 (0.78)	4.32 (1.08)
Total	2.91 (0.34)	3.14 (0.47)	3.31 (0.56)	3.14 (0.65)	2.98 (0.42)	3.30 (0.60)
ECERS						
Space and furnishings	3.09 (0.46)	3.34 (0.53)	3.59 (0.73)	3.26 (0.59)	3.23 (0.41)	3.52 (0.56)
Language	3.19 (0.69)	3.62 (0.92)	3.88 (0.97)	3.13 (0.79)	3.11 (0.75)	3.41 (0.85)
Activities	1.90 (0.36)	2.04 (0.50)	2.27 (0.57)	1.87 (0.36)	1.99 (0.33)	2.13 (0.50)
Program structure	3.45 (0.72)	3.71 (0.66)	3.74 (0.88)	3.52 (0.83)	3.13 (0.60)	3.67 (0.74)
Total	3.04 (0.39)	3.27 (0.49)	3.44 (0.59)	3.06 (0.47)	2.93 (0.34)	3.23 (0.48)

Table 3 shows the results for the two multilevel models. We found a significant increase in the targeted items score for both the ITERS-R and ECERS-R. Results showed a significant increase between the pretest and the posttest and between posttest and follow-up for both the ITERS-R and ECERS-R items targeted during the consultancy. As can be seen from the estimated means in Table 3, ITERS-R items targeted during the consultancy improved, on average, by 0.36 point between pretest and posttest and an additional 0.20 point between posttest and follow-up test. ECERS-R item scores improved, on average, by 0.30 point between pretest and posttest and an additional 0.30 point between posttest and follow-up test. The positive effects of the consultancy on the targeted items thus remained after three months, and showed a small, but significant additional increase between posttest and follow-up test.

Table 3 Consultancy Effects on ITERS-R and ECERS-R items Targeted in Experimental Groups (Multilevel Analysis, *n* = 35)

	ITERS-R items targeted		*ECERS-R items targeted*	
	Estimate	SE	Estimate	SE
Fixed parameters				
Intercept (pretest)	2.005**	0.103	2.075**	0.087
Effects at posttest	0.364*	0.121	0.298**	0.104
Effects at follow-up	0.557**	0.126	0.601**	0.117
Gain score parallel CIP-Training	-.108	0.118	-.084	0.095
Deviance	175.381		145.205	

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

Meaningful Change

The effects of the consultancy for the targeted items were significant, but relatively small. The effect on the targeted items can nevertheless be meaningful whenever they mark the difference between inadequate and adequate quality. We therefore examined whether there was *meaningful change* for the targeted items and for the total scores in the experimental group by comparing the percentages of groups scoring in the categories *inadequate* (score < 3), *moderate* (3 ≤ score < 5), and *adequate to good* (score ≥ 5) at the pretest, posttest, and follow-up test. With regard to the targeted items, Figure 2a shows improvement over time for the ITERS-R items. Whereas 97% of the groups scored in the inadequate category at pretest, 13% shifted to the moderate category at posttest, and at the follow-up another 13% shifted to the moderate category, totaling an improvement of 26% for the experimental group between pretest and follow-up. For the ECERS-R we found comparable results for the targeted items. As can be seen in Figure 2b, at the pretest all groups scored low, on average, on the targeted items. After the consultancy, 9% improved from low to moderate, and at the follow-up test, even 25% scored moderate, indicating that a quarter of the groups improved from low to moderate between pretest and follow-up test.

We also examined whether there was meaningful change over time for the ITERS-R and ECERS-R total scores in both the experimental and the control group. As shown in Figure 3a-b, in the experimental group 20% of the groups improved from inadequate care to moderate care between pretest and posttest on the total ITERS-R score and 16% of the groups on the ECERS-R total score. Between posttest and follow-up, an additional 5% of the groups shifted from inadequate to moderate care for the ITERS-R and 7% for the ECERS-R. Furthermore, we examined whether there was a difference in the percentage of groups that shifted from inadequate to moderate care for the ITERS-R and ECERS-R total scores between the experimental and control group, using a chi-square test. With regard to the ITERS-R total score, we found no significant differences between the experimental group and the control group at the posttest (χ^2 (1) = 0.51, p = .48, N = 68) or the follow-up (χ^2 (1) = 0.17, p = .68, N = 68). With regard to the ECERS-R total score, there was significantly more positive meaningful change in the experimental group than in the control group (χ^2 (1) = 5.88, p < .05, N = 68) at the posttest. However, no significant differences between the experimental group and the control group at the follow-up were found (χ^2 (1) = 1.53, p = .22, N = 68).

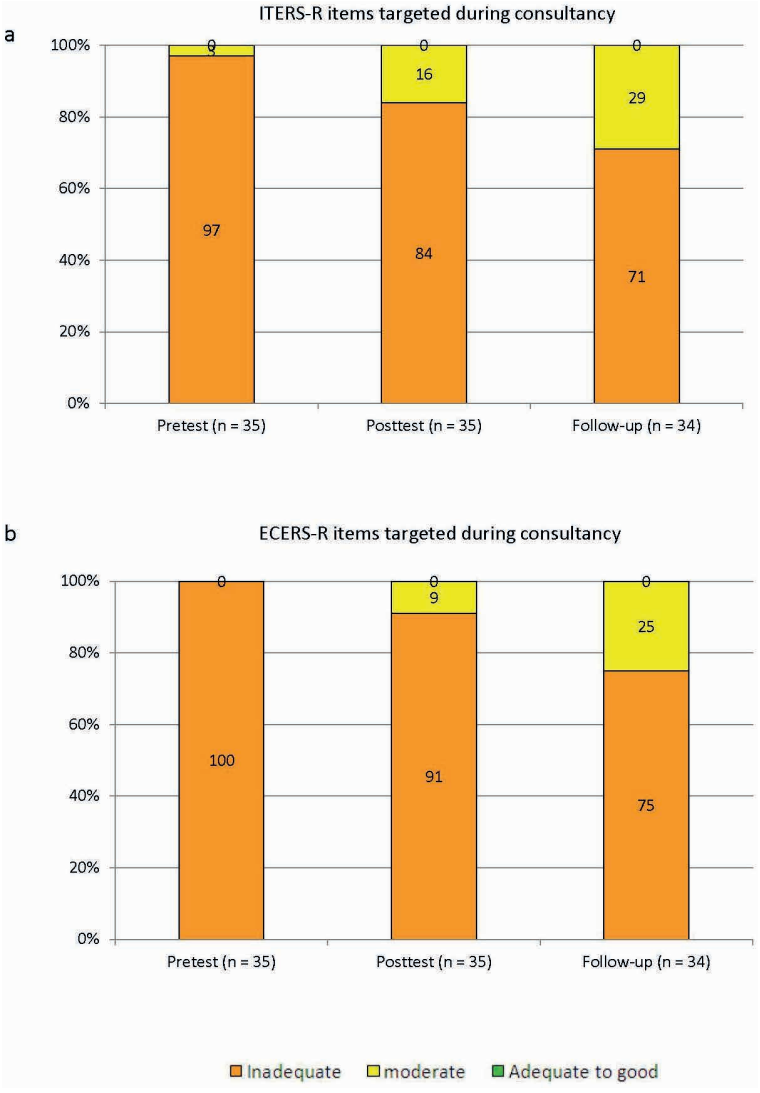

Figure 2 Percentages of Groups scoring Low, Moderate, and High on the ITERS-R and ECERS-R Items Targeted During Consultancy in the Experimental Group at Pretest, Posttest, and Follow-up test.

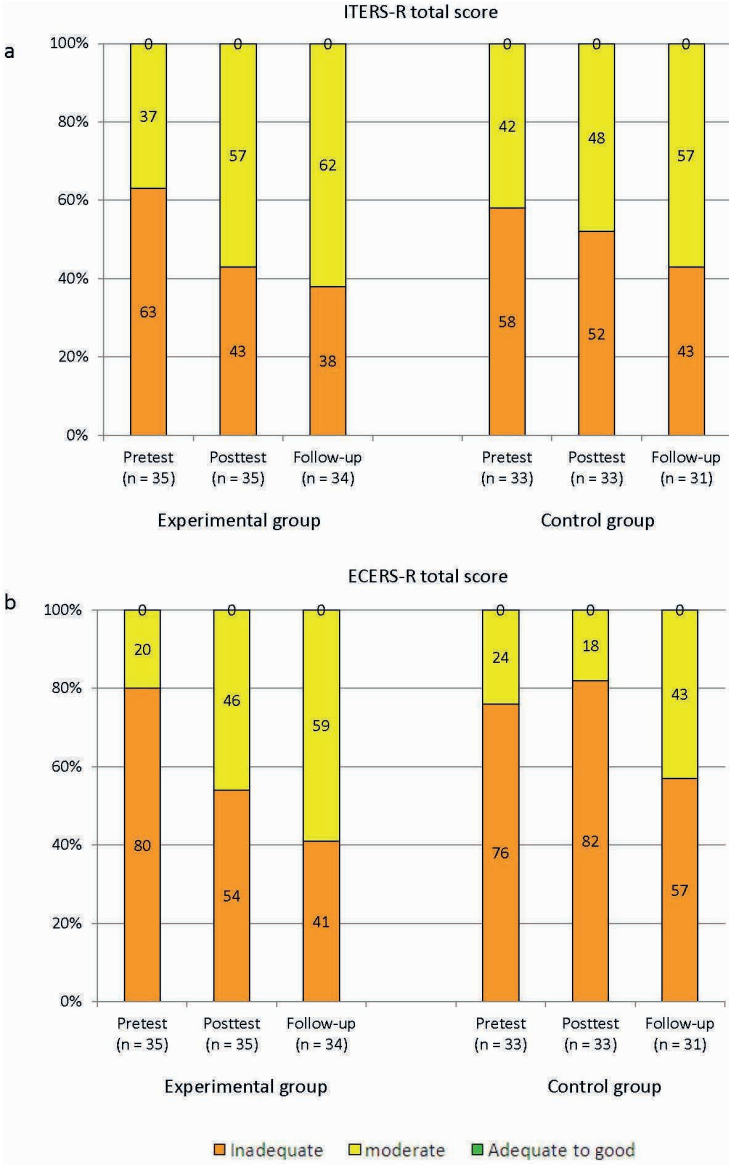

Figure 3 Percentages of Groups scoring Low, Moderate, and High on ITERS-R and ECERS-R Total Score in the Experimental and Control Group at Pretest, Posttest, and Follow-up test.

5.4 Discussion

The controlled experimental evaluation of our on-site consultancy program for center directors that aimed to improve quality of the environment in child care groups with a broad age range (0- to 4-year-olds) demonstrated a significant posi-

tive effect of the intervention for the items targeted during the consultancy. This indicates that quality of the child care environment improved during the intervention and this effect remained at follow-up (three months after the posttest). The positive results were also meaningful in terms of changes form the lower to higher quality categories as defined for the ITERS-R and ECERS-R. It should be noted that the improvement was specific and was directly linked to the exact focus (targeted items) of the consultation, which varied significantly between the different groups. Hence, results of our consultancy intervention were not found using the ITERS-R and ECERS-R total scores. But a carefully constructed measure that matched the specific consultancy focus with the targeted items from these holistic measures, was able to capture the subtle effects of the QCCE consultancy.

We did not find a significant improvement in the ITERS-R/ECERS-R total scores. A possible explanation for this could be that, next to the increase in total scores over time in the experimental groups, we also found a remarkable increase over time for the total score in the control group (see Table 2). In the control group, ITERS-R and ECERS-R total scores decreased between pretest and posttest, but the total scores increased between posttest and follow-up. It is not clear what the specific underlying mechanism for this increase between posttest and follow-up was. The fact is that most center directors participated with a child care location in which both experimental and control groups were enrolled in the study. As a consequence of the design of the present study it cannot be excluded that center directors also influenced global quality in control groups. We instructed center directors not to make any changes in the control groups before follow-up observations were conducted. But, unfortunately, we do not know if center directors followed our instructions or not and perhaps did improve global quality in control groups. For future effect studies, we would therefore recommend to randomly assign center directors with all their groups participating in the study to the experimental or control condition.

Although we did not find an overall effect on the subscales of the ITERS-R and ECERS-R, we did, however, find a significant and positive effect of the QCCE consultancy for the specific items targeted during the consultancy. These findings highlight the importance of a sensitive measure to detect specific improvement, particularly in the context of a consultancy program with individual adaptation and a wide variation in goals. The fact is that quality of the child care environment comprises many facets, which are aggregated in global quality measures, whereas the strength of many consultancy programs is that they are tailored to the needs, aims and preferences of the child care provider and are thus highly flexible (Wesley & Buysse, 2004). However, global assessment instruments have a wide scope and may therefore not capture the effect that was the specific goal of the intervention (Zaslow et al., 2010; Zaslow, Tout, & Halle, 2011). This emphasizes the importance of taking into account the specific purpose of measurement in choosing an assessment instrument (see also Snow & Van Hemel, 2008; Zaslow et al., 2011), because whenever the measurement instrument does not align closely with the content of the inter-

vention, effects do not present themselves (Fukkink & Lont, 2007). In short, there seems to be a discrepancy between the highly individualized intervention programs and the assessment instruments with a very broad scope that have been used to measure effectiveness of interventions.

The results of the present study underscore the fact that improving quality of the child care environment is a challenging task. Quality of the child care environment improved, on average, with 0.65 on a 7-point scale for the ITERS-R items targeted during the consultancy and with 0.58 for the ECERS-R items. These moderate outcomes are comparable to the study of Palsha and Wesley (1998), who also found a significant improvement between half and one scale point. A first possible explanation for the modest improvement may be the intensity of the consultancy. A more intensive consultancy program with additional support might lead to more improvement. In the present set-up, center directors received three consultations in total, of which the first consultation was mostly introductory and instructive, while the consultancy for improvement was only discussed during the second on-site consultation and the telephone call, which constituted the third consultancy contact. Improving quality of the environment through consultation may not be effective enough with the current intensity. As mentioned by Zaslow and colleagues (2010), research into the effect of program intensity is difficult to find, and is highly dependent on the aims of the intervention. In developing the current QCCE consultancy, we aimed for a time-efficient approach considering the costs benefits, and because we reasoned that the QCCE list gave direct insight into which points to improve, because it was filled in by the center directors themselves, who generally manage financial resources and decide about the program. Although we have no reason to believe that the points of improvement were insufficiently evident for center directors, effects might have been larger if we would have kept a closer finger on the pulse by visiting the center director for a third on-site consultation to monitor the status of the improvements.

A second explanation for the only modest improvement may lie in the nature of the scoring system used with the ITERS-R and ECERS-R. In the present study we used the stop-scoring procedure, in which indicators of higher quality are rated only when requirements of indicators of lower quality are met. A recent study by Gordon, Fujimoto, Kaestner, Korenman, and Abner (2013) recommends scoring all indicators of items instead of the usual stop-scoring procedure, because scoring all indicators provides centers with more information about improvement points and also about what they are doing well. It may be possible that center directors indeed improved on indicators of the items, but that we were unable to detect this improvement on item level due to the scoring procedure. In other words, the stop-scoring procedure may have underestimated the actual improvement made in the care groups. In fact, a study comparing the traditional stop-scoring procedure with scoring all indicators, demonstrated that out of 268 care groups, about half (n = 135) of the groups improved by one ECERS-R quality category when all items were scored (Hofer, 2010).

Although the effects of the QCCE consultancy in the present study are not large in absolute sense, analysis of meaningful change demonstrated that taking part in the QCCE consultancy program marked the difference between low and moderate quality for 25% of the experimental groups. This suggests that we were able to eliminate one fourth of the lowest quality through the consultancy, which can be considered meaningful and substantial.

Limitations and Future Directions

Our study was not without limitations. First, center directors volunteered to the study, which may have resulted in a relatively highly motivated, thus selective group. Therefore the results of this study may not be generalized to the general population of child care centers. Because the ITERS-R and ECERS-R were used in previous Dutch quality assessments, we could compare the subscale scores of the present study with those found in the latest quality assessment (Fukkink, Gevers Deynoot-Schaub, Helmerhorst, Bollen, & Riksen-Walraven, 2013). Comparison yielded only one significant difference, namely for the ITERS-R subscale Activities (M_{2008} = 2.46, SD_{2008} = 0.60, N_{2008} = 105; M_{2012} = 2.18, SD_{2012} = 0.51, N_{2012} = 35; t = 2.48, p = < .05), indicating that scores were significantly lower in the present study. No significant differences were found for the remaining ITERS-R subscales or the ECERS-R subscales. Nevertheless, we acknowledge that further research with representative samples is needed to draw definite conclusions.

A second limitation is that we assigned groups and not center directors to the experimental or control condition and therefore we were unable to compare center directors in experimental and control groups. This limitation could have led to diffusion of treatment and we acknowledge this as an important threat to the internal validity of our study. Therefore, we would strongly recommend to randomly assign center directors and not groups to experimental and control condition for future effect studies.

Future research should address several topics. First, future research could establish whether differential effects of the QCCE consultancy on center directors may occur when possible moderators are analyzed. In addition, future research could focus on adding new elements to the consultancy (such as focusing on how the center director can convey the improvement points to the caregivers of the care group and how the director can monitor these action points) and examine whether these elements enhance the consultancy intervention.

The current study introduces a consultancy program to enhance quality of the child care environment. Results of this first RCT study are promising and comparable to previous consultancy programs (Palsha & Wesley, 1998; Wesley, 1994). The outcomes suggest that with self-assessment and support from a consultant, center directors are able to make improvements in the quality of the child care environment. Nevertheless, the effects are modest and therefore scrutiny is certainly warranted. Future research should address ways to improve the effects of the consul-

tancy program. Furthermore, a broader implementation (i.e., in different group types (infant and preschool groups), or in different countries) of the consultancy program is needed to examine whether the favorable outcomes of this study can be confirmed.

References

Belsky, J., Vandell, D. L., Burchinal, M., Clarke-Stewart, K. A., McCartney, K., Owen, M. T., & the NICHD Early Child Care Research Network (2007). Are there long-term effects of early child care? *Child Development, 78*, 681-701. doi: 10.1111/j.1467-8624.2007.01021.x

Bloom, P. J., & Sheerer, M. (1992). The effect of leadership training in child care program quality. *Early Childhood Research Quarterly*, *7*, 579-594.

Bredekamp, S. (1996). The reliability and validity of the Early Childhood Classroom Observation Scale for accrediting early childhood programs. *Early Childhood Research Quarterly*, *1*, 103-118.

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

Fiene, R. (2002). Improving child care quality through an infant caregiver mentoring project. *Child & Youth Care Forum*, *31*, 79-87.

Fukkink, R. G., Gevers Deynoot-Schaub, M. J. J. M., Helmerhorst, K. O. W., Bollen, I., & Riksen-Walraven, J. M. A. (2013). *Pedagogische kwaliteit van de kinderopvang van 0- tot 4-jarigen in Nederlandse kinderdagverblijven in 2012* [Pedagogical quality of child care centers for 0- to 4-year-olds in The Netherlands 2012]. Amsterdam, The Netherlands: Netherlands Consortium of Child Care Research.

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

Gevers Deynoot-Schaub, M., Fukkink, R., Riksen-Walraven, M., De Kruif, R., Helmerhorst, K., & Tavecchio, L. (2009). *De NCKO-Kwaliteitsmonitor* [The NCKO Quality Monitor]. Amsterdam, The Netherlands: SWP.

Gordon, R. A., Fujimoto, K., Kaestner, R., Korenman, S., & Abner, K. (2013). An assessment of the validity of the ECERS-R with implications for measures of child care quality and relations to child development. *Developmental Psychology*, *49*, 146-160. doi: 10.1037/a0027899

Harms, T., & Clifford, R. M. (1980). *Early Childhood Environment Rating Scale*. New York, NY: Teachers College Press.

Harms, T., & Clifford, R. (1989). *The Family Day Care Rating Scale*. New York, NY: Teachers College Press.

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

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

Harms, T., Cryer, D., & Clifford, R. M. (2003). *Infant/Toddler Environment Rating Scale-Revised*. New York, NY: Teachers College Press.

Helmerhorst, K. O. W., Riksen-Walraven, J. M. A., Fukkink, R. G., Tavecchio, L. W. C., & Gevers Deynoot-Schaub, M. J. J. M. (2014a). *Effects of the Caregiver Interaction Profile Training on caregiver-child interactions in child care centers*. Manuscript submitted for publication.

Helmerhorst, K. O. W., Riksen-Walraven, J. M. A., Gevers Deynoot-Schaub, M. J. J. M., Tavecchio, L. W. C., & Fukkink, R. G. (2014b). Child care quality in The Netherlands over the years: A closer look. *Early Education and Development*. doi: 10.1080/10409289.2014.948784.

Helmerhorst, K. O. W., Riksen-Walraven, J. M.A., Vermeer, H. J., Fukkink, R. G., & Tavecchio, L. W. C. (2014c). Measuring interactive skills of caregivers in child care centers: development and validation of the caregiver interaction profile scales. *Early Education & Development, 25,* 770-790. doi: 10.1080/10409289.2014.840482

Hofer, K. G. (2010). How measurement characteristics can affect ECERS-R scores and program funding. *Contemporary Issues in Early Childhood*, *11*, 175-191. doi: 10.2304/ciec.2010.11.2.175

Kontos, S., Howes, C., & Galinsky, E. (1996). Does training make a difference to quality in family child care? *Early Childhood Research Quarterly*, *11*, 427-445. doi: 10.1016/S0885-2006(96)90016-2

La Paro, Hamre, B. K., Locasale-Crouch, J., Pianta, R. C., Bryant, D., Early, D., . . . Burchinal, M. (2009). Quality in kindergarten classrooms: observational evidence for the need to increase children's learning opportunities in early education classrooms. *Early Education and Development*, *20*, 657-692. doi: 10.1080/10409280802541965

Palsha, S. A., & Wesley, P. W. (1998). Improving quality in early childhood environments through on-site consultation. *Topics in Early Childhood Special Education*, *18*, 243-253.

Plantenga, J. (2012). Local providers and loyal parents: competition and consumer choice in the Dutch childcare market. In: L. Lloyd & H. Penn (Ed.), *Childcare markets: Can they deliver an equitable service* (p. 63-78). Bristol, UK: The Policy Press. doi:10.1332/policypress/9781847429339.003.0004

Smith, S., Schneider, W., & Lee Kreader, J. (2010). *Features of professional development and on-site assistance in child care quality rating improvement systems: A survey of Stat-wide systems*. New York, NY: National Center for Children in Poverty. Retrieved from: http://www.nccp.org/

Snow, C., & Van Hemel, S. (2008). *Early childhood assessment: Why, what and how? Report of the Committee on Developmental Outcomes and Assessment for Young Children*. Washington, DC: National Academies Press.

Vandell, D. L., Belsky, J., Burchinal, M., Steinberg, L., Vandergrift, N., & NICHD Early Child Care Research Network. (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*(3), 737-756. doi:10.1111/j.1467-8624.2010.01431.x

Vandell, D. L., & Wolfe, B. (2000). *Child Care quality: Does it matter and does it need to be improved?* Madison: Institute for Research on Poverty, University of Wisconsin-Madison.

Vermeer, H. J., Van IJzendoorn, M. H., De Kruif, R. E. L., Fukkink, R. G., Tavecchio, L. W. C., Riksen-Walraven, J. M. A., & Van Zeijl, J. (2008). Child care in The Netherlands trends in quality over the years 1995-2005. *The Journal of Genetic Psychology, 169*(4), 360-385.

Wesley, P. W. (1994). Providing on-site consultation to promote quality in integrated childcare programs. *Journal of Early Intervention*, *18*, 391-402. doi: 10.1177/105381519401800408

Wesley, P. W., & Buysse, V. (2004). Consultation as a framework for productive collaboration in early intervention. *Journal of Education and Psychological Consultation*, *15*, 127-150. doi: 10.1207/-s1532768xjepc1502_2

Zaslow, M., Tout, K., Halle, T., Vick Whittaker, J. V., & Lavelle, B. (2010). *Towards the identification of features of effective professional development for early childhood educators*. Washington, DC: U.S. Department of Education.

Zaslow, M., Halle, T., & Tout, K. (2011). Differing purposes for measuring quality in early childhood settings: aligning purpose with procedures. In: M. Zaslow, I. Martinez-Beck, K. Tout, & T. Halle (Ed.), *Quality Measurement in Early Childhood Settings* (p. 389-410). Baltimore, MD: Paul H. Brookes Publishing Co.

CHAPTER 6
Summary, Conclusions and General Discussion

Foto: Joods Historisch Museum

6.1 Summary of the research project

The studies presented in this thesis have been conducted as part of the Netherlands Consortium for Research in Child Care (Nederlands Consortium Kinderopvang Onderzoek, NCKO) research program, which has received funding from the Dutch Ministry of Social Affairs and Employment since 2002. The general purpose of the NCKO project was to assess *and* to improve child care quality for 0- to 4-year-olds in The Netherlands. The main aim of this thesis has been the development and evaluation of an intervention program to improve the quality of child care (study 3 and 4).

The first study describes the validation of the Caregiver Interaction Profile (CIP) scales. Next, this thesis reports on the application of the CIP scales in combination with the widely used Infant/Toddler Environment Rating Scale-Revised (ITERS-R; Harms, Cryer, & Clifford, 2003) and Early Childhood Environment Rating Scale-Revised (ECERS-R; Harms, Clifford, & Cryer, 1998) in a nationally representative sample of Dutch child care centers. Outcomes of this second study were the starting point for quality improvement. To this purpose, we developed a broad intervention program that aimed to improve both the quality of caregiver-child interactions as assessed with the CIP scales (described in study 3) and the more global quality of the child care environment as measured with the ITERS-R/ECERS-R (described in study 4) in the same care group.

Study 1. The first study, presented in Chapter 2, describes the development and validation of the CIP scales, which rate six caregiver behaviors that are assumed to play an important role in fostering the wellbeing and development of young children from birth onwards: *sensitive responsiveness*, *respect for autonomy*, *structuring and limit setting*, *verbal communication*, *developmental stimulation* and *fostering positive peer interactions*.

We examined *convergent validity* with another, somewhat comparable observational rating scale for caregiver interactive behavior and with a measure for global process quality. *Discriminant validity* was examined using ratings of caregivers' temperamental sociability, and *predictive validity* was examined by relating the CIP scales with children's competence and behavior problems as rated by their caregivers. Child care groups were visited by two trained researchers. The first researcher observed overall quality of the child care environment with the ITERS-R (Harms et al., 2003) and ECERS-R (Harms et al., 1998), while the second researcher recorded caregiver-child interactions on videotape in 8-10 minute cycles based on a fixed time schedule for later observation with the CIP scales. During a second visit three months later, caregiver-child interactions were recorded once more to determine test-retest reliability for the CIP scales and caregivers completed a questionnaire to rate children's competence and behavior problems.

Results showed promising preliminary evidence supporting the reliability and convergent, discriminant, and predictive validity of the CIP scales. The structuring

and limit setting scale could not be applied in this study because the relevant caregiver behaviors occurred too infrequently during the videotaped episodes (therefore, we adapted the video recording procedure for future studies – see study 2 below). Results of a multilevel analysis demonstrated that the CIP scales provide a unique interaction profile of individual caregiver skills; variation of caregiver behavior was for the most part situated at individual caregiver level, with relatively little variance at group and center level.

Study 2. The second study (Chapter 3) reports on the results of the second NCKO quality assessment, in 2008, in a nationally representative sample of 200 child care centers for 0- to 4-year-olds in The Netherlands. Global quality of the child care environment as measured with the ITERS-R and ECERS-R was compared with global quality at the prior assessment in a nationally representative sample in 2005. The newly developed CIP scales were applied for the first time to assess caregiver interactive skills in a nationally representative sample. Furthermore, child care quality was examined in relation to three structural features of child care groups, namely, group size, caregiver-child ratio, and group type (age composition).

Care groups were visited by a researcher who completed the ITERS-R/ECERS-R and made video recordings of each caregiver in interaction with children in four different situations, i.e., diapering, lunch/snack, free play, and transitions (8-10 minutes per situation). Video recordings were rated afterwards with the CIP scales by an independent and trained coder.

Results showed a significant decline in global quality as assessed with ITERS-R/ECERS-R scores in 2008 as compared to 2005, with almost half of the groups scoring below the minimal level. Outcomes on the CIP scales revealed relatively high scores for the more basic caregiver skills sensitive responsiveness, respect for autonomy, and structuring and limit setting, and clearly lower scores for the more educational skills verbal communication, developmental stimulation, and fostering positive peer interactions. Caregiver sensitive responsiveness was significantly lower in infant groups (0-2 years) than in preschool groups (2-4 years); caregiver respect for autonomy, verbal communication, developmental stimulation and fostering positive peer interactions were significantly lower in infant groups than in preschool groups and mixed age groups (0-4 years). Group size was not related to any of the CIP scales. In contrast to our expectations, the significant relation between caregiver-child ratio and two of the CIP scales, i.e., verbal communication and developmental stimulation, indicated that caregivers scored higher on verbal communication and developmental stimulation in groups with more children per caregiver.

The low CIP scores on the more educational versus the more basic caregiver interactive skills indicate that these educational skills deserve extra attention in caregiver education and training.

Study 3. This study, presented in Chapter 4, examined the effects of a video feedback training for caregivers that aimed to strengthen quality of the caregiver-child interactions as assessed with the CIP scales. The training was based on the conceptual framework underlying the CIP scales and aimed to improve the six CIP skills.

The study sample included 68 mixed-age groups from 33 child care centers, with 35 groups randomly assigned to the experimental condition and 33 groups to the control condition. A total of 139 caregivers participated in the study. To assess the effects of the training, the quality of caregiver-child interactions was assessed with the CIP scales at pretest, posttest, and follow-up three months after the posttest. Following the procedure described in study 2, each caregiver was filmed in four situations (diapering, lunch/snack, structured play, and transition) for 8-10 minutes per situation. The ITERS-R/ECERS-R were also administered in the groups; these measures were used to control for possible effects of the parallel part of the intervention program (described in study 4).

Two weeks after the pretest, the video-feedback trainer visited the caregivers at the child care center for the first time for the video-feedback training, which consisted of five weekly meetings in total. Per group, all participating caregivers were trained individually with feedback on their own videotaped interaction episodes in the first four sessions. The fifth and final session was used to share experiences and video observations with the other trained caregiver from the same group.

Results showed a significant positive training effect on all six CIP skills at possttest. Three months after the posttest, at follow-up, caregivers who had received training still scored significantly higher on four of the six CIP skills: sensitive responsiveness, respect for autonomy, verbal communication, and fostering positive peer interactions. Although training effects were relatively small in absolute terms, they appear meaningful in terms of decreasing percentages of caregivers with inadequate scores and increasing percentages of caregivers scoring in the adequate to good range.

The results demonstrate that it is possible to enhance the quality of caregiver-child interactions through this video feedback training based on the CIP scales. Extending the CIP training with additional booster sessions is suggested to further improve the effectiveness of the training. Possible ways to implement the training in practice and education are discussed.

Study 4. The fourth study, presented in Chapter 5, examined the effects of the second part of the broader intervention program, namely an on-site consultancy program directed at center directors that aimed to improve the quality of the child care environment as reflected in the ITERS-R/ECERS-R (i.e., space, furnishings, materials, and program structure). The consultancy program ran parallel to the video-feedback training for caregivers in the same care group.

Participating groups were the same as in study 3, i.e., 68 mixed-age groups from 33 child care centers, with 35 groups randomly assigned to the experimental condi-

tion and 33 groups to the control condition. A total of 14 center directors participated in the consultancy program with one group, 6 center directors with two groups, and 3 center directors with three groups. All groups were visited for pretest, posttest, and follow-up three months after the posttest to complete the ITERS-R and ECERS-R that were used to examine the effects of the consultancy program. CIP scores for the caregivers in the groups, obtained at pretest, posttest, and follow-up to examine the effects of the video-feedback training in study 3, were used to control for possible effects of that training in the present study.

The consultancy program consisted of three consultations in total, with two onsite consultations and a third consultation given per telephone. During the consultancy program, center directors used a self-assessment tool (the NCKO Quality Monitor; Gevers Deynoot-Schaub et al., 2009) to self-evaluate the quality of the child care environment in care groups. Based on the self-assessment and the ITERS-R/ECERS-R pretest scores, the consultant and center director together drew up an action plan for improvement which was tailored to the individual needs of the group: the consultancy specifically targeted items that were identified as 'weak areas' by the consultant and center director.

Results demonstrated no significant effects for the ITERS-R and ECERS-R total scores. The results for the total scores were nonetheless meaningful in terms of percentages of groups switching from inadequate to adequate care categories as defined for the ITERS-R and ECERS-R: 25% of the experimental groups improved from inadequate care to moderate care between pretest and follow-up on the ITERS-R and 23% of the groups on the ECERS-R. Furthermore, we found a significant improvement for the specific ITERS-R and ECERS-R items that had been targeted during the consultancy.

These promising, although moderate, outcomes demonstrate that the quality of the child care environment can be improved using the consultancy program for center directors, and that this effect remains over time. The results also show that the effects of the program do not generalize to quality topics beyond those addressed during the consultancy, so that care should be taken to include all topics that deserve improvement in the intervention aims.

6.2 Conclusions

The results of this thesis can be summarized in the following main conclusions:

- Reliability and validity of the CIP scales in the Dutch child care context is supported, which means that the scales can be applied for assessment of child care quality in The Netherlands to get a detailed picture of caregivers' interactive skills, and provide important information on child care quality in addition to the ITERS-R and ECERS-R that measure more global quality of the child care environment.

- Caregivers in The Netherlands score, on average, relatively high on the more basic CIP skills (i.e., sensitive responsiveness, respect for autonomy, and structuring and limit setting), but the range of scores is considerable and also includes inadequate scores. Caregivers' scores on the more educational CIP skills (i.e., verbal communication, developmental stimulation, and fostering positive peer interactions) are clearly lower and mostly in the inadequate range.
- It is important to measure and train caregiver interactive skills at an individual level, because caregivers have a personal and individually determined CIP interactive skills profile.
- The CIP training effectively strengthens caregivers' interactive skills. Additional booster sessions may be needed to increase the effectiveness of the training, by further improving the caregivers' educational skills.
- Global process quality in The Netherlands as assessed with the ITERS-R and ECERS-R is at a low to medium level and needs improvement. Therefore, it is important to continue monitoring child care quality.
- The consultancy program for center directors to improve global child care quality is effective, but only for items that are specifically targeted during the consultancy. Effects of the program do not generalize to topics beyond those addressed during the consultancy.

6.3 General discussion

Several issues regarding this thesis deserve further discussion. First, it is an important implication of the studies examining the effects of the intervention program that the current set-up of the intervention program with both the CIP training and the consultancy program for care groups should be retained, because the two components of the intervention program evidently affect different aspects of child care quality. The complete intervention program has a double focus, with one part aiming at improving the quality of the caregiver-child interaction and the other part aiming at improving the broader quality of child care environment (i.e., space, furnishings, materials, and program structure). Effects of the two intervention components were measured with the CIP scales and the ITERS-R/ECERS-R. Although both aspects of child care quality are related – validation of the CIP scales (see study 1) yielded a positive and significant correlation between the CIP scales and the ITERS-R/ECERS-R (r = .38 for the CIP total score and the ITERS-R/ECERS-R total score) – the experimental studies for the two parallel parts of the intervention program nevertheless demonstrated that correlations between the gain scores obtained for the CIP training and the ITERS-R/ECERS-R gain scores obtained for the consultancy program were low and non-significant (not mentioned in study 3 and 4). Evaluation of both parts of the intervention program thus demonstrated that improvement of caregiver interactive skills did not necessarily lead to improvement of the global

quality of the child care environment, or vice versa. This is a clear indication that it is important to retain the double focus approach of the intervention program. Merely choosing one of the two intervention components is only justified when either one of the quality measures (caregiver interactive skills or the global quality of the child care environment) is notably lower than the other quality measure.

A second point that deserves attention is the finding that the effects of the CIP training were more apparent than the effects of the consultancy program. Obviously, the two complementary programs have a different focus and method. The most evident strength of the CIP training is the video feedback method. Previous studies have already endorsed the use of video feedback for training programs targeting parent or caregiver behavior (Bakermans-Kranenburg, Van IJzendoorn, & Juffer, 2003; Fukkink & Lont, 2007; Fukkink & Tavecchio, 2010; Groeneveld, Vermeer, Van IJzendoorn, & Linting, 2011). Given that the consultancy program does not focus on the center director's interactive behavior, but rather on bringing about changes in environmental characteristics such as space, furnishings, and materials, video feedback did not directly apply for the consultancy program. Nevertheless, it would be interesting to incorporate some visual tools in the consultancy. For instance, photographs of good practice examples for space, furniture, and materials could be very useful to visualize the intended purpose of the program.

Another important difference between the two parts of the intervention program is that the intensity of the consultancy program was lower than the intensity of the CIP training. The CIP training consisted of 5 sessions, while the consultancy program merely consisted of two onsite consultations and a third more 'distant' session by telephone. Although prior research is not clear about the effects of program intensity in the case of consultancy programs (Zaslow et al., 2010), the lower intensity of the consultancy program may have contributed to the less favorable outcomes as compared to the CIP training.

A final difference between the two intervention program parts that may explain their difference in effectiveness is that the trainees in the CIP training (i.e., the caregivers) could directly realize their intervention aims in their own behavior in interaction with the children. In the consultancy program, however, center directors had to attain their aims in a more indirect way. They were asked to improve the child care environment in the care group of the caregivers, whose cooperation was required to realize many of the intervention targets. Hence, instructing the center directors on how to improve environment characteristics was only a first step in the actual improvement process; the second step was for center directors to realize these changes in close collaboration with the caregivers in the care group. For example, center directors could purchase new materials, but implementing these materials eventually boiled down to what the caregivers in the care group did with them. Obtaining adequate to good scores for materials, according to the ITERS-R/ECERS-R, requires not only availability of the materials but also involvement of the caregiver when children play with the materials. The main reason that the consultancy was nevertheless directed at center directors and not caregivers was be-

cause caregivers' influence on the broader quality of the child care environment is limited, given that it is the center director who decides about provisions, materials, and the program. For future research, however, we would suggest to also involve caregivers in the consultancy process, provided that this does not lay a heavy burden on the caregivers. To prevent an overload for caregivers, the two training programs could be run sequentially instead of simultaneously. The order in which the programs are offered does not seem very important. An advantage of starting the intervention with the CIP training might be that there is room for the suggested additional booster sessions of the CIP training (see study 3, chapter 4) during the subsequent consultancy program. Altogether, the abovementioned differences between the CIP training and the consultancy program may explain why the effects of the CIP training appear more favorable than the effects of the consultancy program.

A third point of discussion concerns initial vocational education for caregivers. Caregiver education in The Netherlands evidently does not pay enough attention to the educational CIP skills. Study 1 and 2 (Chapter 2 and 3) of this thesis demonstrated that although caregivers in The Netherlands score, on average, relatively high on the more basic CIP skills (i.e., sensitive responsiveness, respect for autonomy, and structuring and limit setting), scores on the more educational CIP skills (i.e., verbal communication, developmental stimulation, and fostering positive peer interactions) are mostly in the inadequate range. Thus, from the perspective of the child, caregiver education falls short; students are not well enough instructed on how to adequately stimulate children's language, cognitive, creative, motor, and social development. In recent years some adaptions in Dutch caregiver education have been made in response to the acknowledged shortcomings of the previous vocational training. As from September 2014, caregiver education in The Netherlands will get a new profile with a specific focus on all six CIP skills (Calibris, 2014). Of course, this requires specific expertise and new skills of teacher educators to implement this profile and therefore a straightforward next step would be to implement the CIP training in caregiver education. The NCKO is currently developing a *train the trainer* program of the CIP training for teacher educators in caregiver education. Effects of the implementation of this adapted CIP training (including the train the trainer) in caregiver education will be examined in future research.

A final point of discussion pertains to the international relevance of the studies described in this thesis, considering that they are based on the NCKO Quality Model (see Chapter 1 of this thesis), which may reflect Dutch values and characteristics of the Dutch child care context. Although there are evidently culturally specific values for defining, measuring, and improving child care quality, recent interest in the NCKO research project from Norway suggests that there is broader support for the Dutch view on child care quality and the corresponding line of research. The complete NCKO research plan has been adopted by Norwegian colleagues and their proposal (which also includes some possible adaptions to the Norwegian context) has been granted by the Norwegian government for a period 2012-2020 (Bjørnestad, Gulbrandsen, Johansson, & Os, 2013; Johansson, 2012). As a result, the CIP

scales have now been translated into Norwegian and the first promising observation results suggest that the CIP scales can be satisfactorily applied in the Norwegian child care context.

Interest for the CIP scales has also been shown by researchers in the United States, and plans are being made for a collaborative study to examine how the CIP scales are related to the Classroom Assessment Scoring System (CLASS; Pianta, La Paro, & Hamre, 2008; Roorbach Jamison, Cabell, LoCasale-Crouch, Hamre, & Pianta, 2014) – which assesses teacher-child interactions on several dimensions, including some that are similar to the CIP scales. Both examples of international collaboration demonstrate that there is international interest in the NCKO research and future research should explore whether the NCKO Quality Model is also applicable in other countries with different child care contexts.

6.4 Limitations and future directions

Beyond the limitations that are already mentioned above and those mentioned in the chapters reporting on the separate studies, two limitations and future directions deserve further attention. First, the NCKO Quality Monitor was nationally distributed with government funding at the level of child care organizations in 2009. Thus, it is possible that center directors with groups allocated to the control condition had already worked with the monitor before the intervention program had started off, which may have acted as a confounder in the effect studies. To check for this, center directors in the experimental and control condition were asked whether they had heard about the Quality Monitor, and, if so, whether they had been working with (parts of) the Quality Monitor. About 80 percent of the center directors indicated that they had heard about the monitor. Eight center directors answered that their organization had received the monitor, yet only two directors indicated that they had actually used the monitor: one center director with groups that were allocated to the control condition, and one center director with groups that were allocated to the experimental condition. We compared pretest scores of the groups of which the center directors had indicated that they had worked with the monitor (9 groups with 14 caregivers), with the groups that had not previously worked with the monitor (59 groups with 124 caregivers). Results indicate that there were no significant differences in ITERS-R and ECERS-R or CIP pretest scores between the groups for which the center directors had or had not worked with the monitor previously.

Second, the intervention was carried out in mixed-age groups only. Although mixed-age groups cover the child care age range from 0- to 4-year-olds, it cannot be automatically assumed that the effects of the intervention program can be generalized to infant and preschool groups. Therefore, future research in samples with homogeneous infant and preschool groups in child care centers is recommended to draw definite conclusions about the effects of the intervention program in other

group types. This issue is also interesting from an international perspective, given that mixed-age groups are typical for the Dutch situation but are less common in other countries.

References

Bakermans-Kranenburg, M. J., Van IJzendoorn, M. H., & Juffer, F. (2003). Less is more: Meta-analyses of sensitivity and attachment interventions in early childhood. *Psychological Bulletin*, *129*, 195-215. doi: 10.1037/0033-2909-129.2.195

Bjørnestad, E., Gulbrandsen, L., Johansson, J. E., & Os, E. (2013). *Metodiske idealkrav og nødvendige tilpasninger: Foreløpig tilstandrapport fra prosjektet* [Better provisions for Norway's children in ECEC: A study of children's well-being and development in ECEC]. Oslo, Norway: Høgskolen I Oslo og Akershus (HiOA).

Calibris (2014). *Kwaliteitsdossier MBO, Pedagogisch Werk*. [Quality file Intermediate Vocational Training, Pedagogical Work]. Bunnik, The Netherlands, Calibris. Retrieved from http://www.calibris.nl/-Onderwijsdocumenten/Kwalificatiedossiers/KD-s-2014-2015-(concept)/Kwalificatiedossier-Pedagogisch-Werk

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

Fukkink, R. G., & Tavecchio, L. W. C. (2010). Effects of video interaction guidance on early childhood teachers. *Teaching and Teacher Education*, *26*, 1652-1659. doi: 10.1016/j.tate.2010.06.016

Gevers Deynoot-Schaub, M., Fukkink, R., Riksen-Walraven, M., De Kruif, R., Helmerhorst, K., & Tavecchio, L. (2009). *De NCKO-Kwaliteitsmonitor* [The NCKO Quality Monitor]. Amsterdam, The Netherlands: SWP.

Groeneveld, M. G., Vermeer, H. J., Van IJzendoorn, M. H., & Linting, M. (2011). Enhancing home-based child care quality through video-feedback intervention. *Journal of Family Psychology*, *25*, 86-96. doi: 10.1037/a0022451

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

Harms, T., Cryer, D., & Clifford, R. M. (2003). *Infant/Toddler Environment Rating Scale-Revised*. New York, NY: Teachers College Press.

Johansson, J. (2012). *Better provision for Norway's children in ECEC. A study of children's well-being and development in ECEC, and new tools for quality evaluation*. Application to The Research Council of Norway. Oslo, Norway: Oslo and Akershus University College for Applied Sciences.

Oberhuemer, P., Schreyer, I., & Neuman, M. J. (2010). *Professionals in early childhood education and care systems. European profiles and perspectives*. Opladen & Farminghton Hills: Barbara Budrich Publishers.

Palsha, S. A., & Wesley, P. W. (1998). Improving quality in early childhood environments through on-site consultation. *Topics in Early Childhood Special Education*, *18*, 243-253.

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

Roorbach Jamison, K., Cabell, S. Q., LoCasale-Crouch, J., Hamre, B. K., & Pianta, R. (2014). CLASS-Infant: An observational measure for assessing teacher-infant interactions in center based care. *Early Education and Development, 25*, 553-572. doi: 10.1080/10409289.2013.822239

Zaslow, M., Halle, T., & Tout, K. (2011). Differing purposes for measuring quality in early childhood settings: aligning purpose with procedures. In: M. Zaslow, I. Martinez-Beck, K. Tout, & T. Halle (Ed.), *Quality Measurement in Early Childhood Settings* (p. 389-410). Baltimore, MD: Paul H. Brookes Publishing Co.

Samenvatting en Conclusies

Samenvatting van het onderzoeksproject

De studies die in dit proefschrift worden beschreven, zijn onderdeel van het onderzoeksprogramma van het Nederlands Consortium Kinderopvang Onderzoek (NCKO), dat in 2002 is gestart en wordt gesubsidieerd door het Ministerie van Sociale Zaken en Werkgelegenheid. Het NCKO heeft als doel het meten *en* verbeteren van de kwaliteit van de Nederlandse kinderdagopvang (0-4 jaar). Het hoofddoel van dit proefschrift is de ontwikkeling en evaluatie van een interventieprogramma ter verbetering van de kwaliteit van de kinderopvang (studie 3 en 4).

De eerste studie in dit proefschrift betreft de ontwikkeling en validatie van de Caregiver Interaction Profile (CIP) schalen. In de tweede studie zijn de CIP schalen, in combinatie met de veelvuldig en wijd gebruikte Infant/Toddler Environment Rating Scale (ITERS; Harms, Cryer, & Clifford, 1990) and Early Childhood Environment Rating Scale (ECERS; Harms & Clifford, 1980), toegepast in een landelijk, representatieve steekproef van Nederlandse kinderdagverblijven om de pedagogische kwaliteit in kaart te brengen. De resultaten van deze tweede studie waren het uitgangspunt voor kwaliteitsverbetering. Daartoe is een breed interventieprogramma ontwikkeld dat erop gericht is om zowel de kwaliteit van de interactievaardigheden van de pedagogisch medewerkers zoals gemeten met de CIP schalen (studie 3), als de meer globale kwaliteit van de leefomgeving zoals gemeten met de ITERS-R/ECERS-R (studie 4) in dezelfde groep te verbeteren.

Studie 1. De eerste studie, in Hoofdstuk 2, betreft de ontwikkeling en validatie van de CIP schalen, die zes interactievaardigheden voor pedagogisch medewerkers meten, waarvan verondersteld wordt dat deze belangrijk zijn voor het welbevinden en de ontwikkeling van jonge kinderen (0-4 jaar): *sensitieve responsiviteit*, *respect voor autonomie*, *structureren en grenzen stellen*, *praten en uitleggen*, *ontwikkelingsstimulering* en *begeleiden van interacties*.

Convergente validiteit werd onderzocht door vergelijking met een ander, vergelijkbaar observatie-instrument dat de kwaliteit van interactievaardigheden van pedagogisch medewerkers meet en een instrument dat de bredere, meer globale proceskwaliteit meet. Discriminante validiteit werd onderzocht door vergelijking met de schaal sociabiliteit uit een instrument dat het temperament van pedagogisch medewerkers meet. Predictieve validiteit werd onderzocht door de CIP schalen te relateren aan competentie en probleemgedrag van kinderen, zoals beoordeeld door hun pedagogisch medewerkers. De groepen werden bezocht door twee getrainde observatoren; de eerste observator scoorde de globale proceskwaliteit van de

groepen met behulp van de ITERS-R of ECERS-R en de tweede observator maakte video-opnames van 8 tot 10 minuten (in vaste filmrondes) van de pedagogisch medewerker-kind interacties om die op een later moment te kunnen beoordelen met de CIP schalen. Tijdens een tweede bezoek, drie maanden later, werden de interactievaardigheden van de medewerkers opnieuw gefilmd om de test-hertest betrouwbaarheid vast te kunnen stellen. Pedagogisch medewerkers vulden tevens een vragenlijst in om competentie en probleemgedrag van de kinderen te beoordelen.

De resultaten gaven veelbelovende, eerste aanwijzingen voor de betrouwbaarheid en voor de convergente, discriminante en predictieve validiteit van de schalen. De schaal structuren en grenzen stellen kon niet worden gebruikt in deze studie, omdat dit gedrag te weinig voorkwam op de filmbeelden om goed gescoord te kunnen worden. Om die reden is de filmprocedure voor latere studies aangepast (zie studie 2 hieronder). De uitkomsten van de multilevel analyses toonden aan dat de CIP schalen een uniek interactieprofiel voor individuele pedagogisch medewerkers opleveren; de variantie in de interactievaardigheden van pedagogisch medewerkers lag voor het overgrote deel op individueel niveau, en minder op groeps- of vestigingsniveau.

Studie 2. De tweede studie (Hoofdstuk 3) rapporteert de resultaten van de tweede landelijke kwaliteitsmeting van het NCKO in 2008, in een landelijk representatieve steekproef van 200 kinderdagverblijven (0-4 jaar) in Nederland. Deze studie had drie doelen: ten eerste het meten van de algemene proceskwaliteit en deze vergelijken met de eerdere peiling uit 2005; ten tweede het toepassen van de nieuw ontwikkelde CIP schalen voor het meten van de kwaliteit van interactievaardigheden in een landelijk representatieve steekproef; en ten derde het onderzoeken van de relatie tussen de pedagogische kwaliteit en drie structurele kwaliteitskenmerken, namelijk groepsgrootte, staf-kind ratio en type groep (leeftijdssamenstelling).

De groepen werden bezocht door een observator die de ITERS-R of ECERS-R afnam en die daarnaast per pedagogisch medewerker video-opnames van 8 tot 10 minuten in vier verschillende situaties (lunch/snack, verschonen, spel en transitie) maakte van de pedagogisch medewerker-kind interacties. De filmbeelden werden later door een andere getrainde observator beoordeeld met de CIP schalen.

Uit de resultaten blijkt dat de globale proceskwaliteit significant is gedaald tussen 2005 en 2008. In 2008 scoorde bijna de helft van de groepen onder het minimum niveau. Voor de interactievaardigheden geldt dat pedagogisch medewerkers relatief hoog scoorden op de basale vaardigheden sensitieve responsiviteit, respect voor autonomie en structureren en grenzen stellen. Op de meer educatieve vaardigheden praten en uitleggen, ontwikkelingsstimulering en begeleiden van interacties scoorden de pedagogisch medewerkers duidelijk lager dan op de basale vaardigheden. Daarnaast werd gevonden dat pedagogisch medewerkers in de babygroepen (0-2 jaar) significant lager scoorden op sensitieve responsiviteit dan de pedagogisch medewerkers in de peutergroepen (2-4 jaar). De scores voor respect

voor autonomie, praten en uitleggen, ontwikkelingsstimulering en begeleiden van interacties waren in de babygroepen significant lager dan in de peutergroepen en verticale groepen (0-4 jaar). De interactievaardigheden bleken niet gerelateerd aan de groepsgrootte, maar in tegenstelling tot de verwachting wel aan de staf-kind ratio; pedagogisch medewerkers scoorden hoger op praten en uitleggen en ontwikkelingsstimulering in groepen met meer kinderen per medewerker.

Concluderend geven de lagere scores op de educatieve vaardigheden ten opzichte van de basale vaardigheden aan dat praten en uitleggen, ontwikkelingsstimulering en begeleiden van interacties extra aandacht verdienen in de opleiding en training van pedagogisch medewerkers.

Studie 3. De studie die in Hoofdstuk 4 wordt beschreven, onderzocht de effectiviteit van een video-feedback training voor pedagogisch medewerkers ter verbetering van hun interactievaardigheden met de kinderen zoals gemeten met de CIP schalen. De training is gebaseerd op het conceptuele model dat ten grondslag ligt aan de CIP schalen.

De steekproef bestond uit 68 verticale groepen in 33 kinderdagverblijven, waarbij 35 groepen random werden toegewezen aan de experimentele conditie en 33 groepen aan de controle-conditie. In totaal deden 139 pedagogisch medewerkers mee aan dit onderzoek. Om de effecten van de training te meten werden de interactievaardigheden van pedagogisch medewerkers driemaal gemeten met de CIP schalen: op een voormeting, een nameting en een follow-up meting drie maanden na de nameting. Net als bij studie 2 werd iedere medewerker 8 tot 10 minuten gefilmd in vier situaties (lunch/snack, verschonen, gestructureerd spel en transitie). Daarnaast werden de ITERS-R en de ECERS-R afgenomen om te controleren voor mogelijke effecten van het consultancy programma dat parallel in dezelfde groepen liep (beschreven in studie 4).

Twee weken na de voormeting bezocht de video-feedback trainer de pedagogisch medewerkers voor de eerste bijeenkomst van de video-feedback training, die in totaal uit 5 bijeenkomsten bestond. Per groep werden alle medewerkers die op de dag van de voormeting aanwezig waren, tijdens de eerste vier sessies individueel getraind middels feedback op hun eigen filmbeelden. Tijdens de laatste en vijfde sessie wisselden de medewerkers ervaringen uit en toonden elkaar hun beelden.

De resultaten op de nameting lieten een significant positief trainingseffect zien op alle zes CIP schalen. Drie maanden later, op de follow-up meting, scoorden de pedagogisch medewerkers nog steeds significant hoger dan bij de voormeting op vier van de zes CIP schalen: sensitieve responsiviteit, respect voor autonomie, praten en uitleggen en begeleiden van interacties. De trainingseffecten zijn in absolute zin relatief klein, maar desalniettemin betekenisvol als we kijken naar het afnemende percentage pedagogisch medewerkers met scores in de categorie onvoldoende en het toenemende percentage pedagogisch medewerkers met scores in de categorie voldoende tot goed scoort.

Deze studie toont aan dat het mogelijk is om de interactievaardigheden van

pedagogisch medewerkers te verbeteren met behulp van video feedback gebaseerd op de CIP schalen. Om de effectiviteit nog verder te verbeteren wordt voorgesteld om de training te verlengen met extra 'booster' sessies. Mogelijke opties voor de implementatie van de CIP training in praktijk en opleiding worden besproken.

Studie 4. In de vierde studie (Hoofdstuk 5) werd de effectiviteit van het tweede onderdeel van de brede interventie onderzocht, namelijk een consultancy programma voor leidinggevenden ter verbetering van de kwaliteit van de leefomgeving zoals gemeten met de ITERS-R/ECERS-R (ruimte, meubilering, materialen en programmastructuur). Het consultancy programma liep parallel met de videofeedback training voor pedagogisch medewerkers in dezelfde groep. In totaal participeerden 14 leidinggevenden met één groep, 6 leidinggevenden met twee groepen en 3 leidinggevenden met drie groepen aan het onderzoek. Op alle groepen werden tijdens een voormeting, een nameting en een follow-up meting de ITERS-R en ECERS-R afgenomen om de effectiviteit van het consultancy programma te evalueren. Daarnaast werden ook de CIP scores van de pedagogisch medewerkers op voor-, na- en follow-up meting gebruikt om te controleren voor mogelijke effecten van de parallelle video-feedback training.

Het consultancy programma bestond uit drie consulten in totaal, met twee consulten op de groep van het kinderdagverblijf en een laatste consult per telefoon. Leidinggevenden gebruikten tijdens de consultancy de NCKO-Kwaliteitsmonitor als zelf-evaluatie instrument om zo zelf de kwaliteit van de leefomgeving op de groep in kaart te brengen. Op basis van de scores van de zelf-evaluatie en de ITERS-R/ECERS-R scores van de voormeting stelden de leidinggevenden en consultant in overleg per groep een plan van aanpak op, dat specifiek was toegesneden op de behoefte van de groep. Op deze manier richtte de consultancy zich specifiek op die punten die door de leidinggevende en consultant werden gekenmerkt als 'zwakker'.

Er werd geen significant overall-effect gevonden op de ITERS-R en ECERS-R totaalscores. De resultaten voor de totaalscores waren desalniettemin betekenisvol als we kijken naar het percentage groepen dat van onvoldoende naar voldoende kwaliteit verschoof: in 25% van de groepen in de experimentele conditie verbeterde de score op de ITERS-R tussen de voormeting en de follow-up van onvoldoende naar matig, en op de ECERS-R was dit in 23% van de groepen het geval. Bovendien werd ook een significante verbetering gevonden voor de punten waar specifiek aan gewerkt was tijdens de consultancy.

Deze uitkomsten tonen aan dat de kwaliteit van de leefomgeving verbeterd kan worden met behulp van een consultancy programma voor leidinggevenden en dat deze effecten bovendien drie maanden na het programma nog zichtbaar zijn. De resultaten laten echter ook ook zien dat de effecten van het programma niet te generaliseren zijn naar onderwerpen die niet behandeld zijn tijdens de consultancy. Daarom is het belangrijk om bij de consultancy alle specifieke punten die aandacht verdienen op te nemen in het plan van aanpak.

Conclusies

De resultaten van dit proefschrift kunnen worden samengevat in de volgende voornaamste conclusies:

- De CIP schalen blijken betrouwbaar en valide voor de Nederlandse context. Dit betekent dat de schalen kunnen worden gebruikt voor landelijke kwaliteitsmetingen om een gedetailleerd beeld te krijgen van de interactievaardigheden van pedagogisch medewerkers. De CIP schalen geven belangrijke informatie over de pedagogische kwaliteit van kinderdagverblijven naast de ITERS-R en ECERS-R, die een meer globaal beeld van de kwaliteit van de leefomgeving geven.
- Pedagogisch medewerkers in Nederland scoren gemiddeld relatief hoog op de basale CIP interactievaardigheden (sensitieve responsiviteit, respect voor autonomie en structureren en grenzen stellen), maar de spreiding van de scores is aanzienlijk en er komen ook onvoldoende scores voor op deze schalen. Op de educatieve vaardigheden (praten en uitleggen, ontwikkelingsstimulering en begeleiden van interacties) scoren pedagogisch medewerkers beduidend lager en vaker in de categorie onvoldoende.
- Het is belangrijk om de interactievaardigheden van pedagogisch medewerkers op individueel niveau te meten en verbeteren, omdat medewerkers een persoonlijk en individueel bepaald profiel hebben op de CIP schalen.
- De CIP training versterkt de interactievaardigheden van pedagogisch medewerkers. Toevoeging van extra booster sessies zou met name de educatieve vaardigheden van medewerkers verder kunnen verbeteren en daarmee de effectiviteit van de training nog verder kunnen vergroten.
- De globale proces kwaliteit zoals gemeten met de ITERS-R/ECERS-R is van laag tot matig niveau in Nederland en vraagt om verbetering. Daarom is het belangrijk om regelmatig landelijke kwaliteitspeilingen uit te voeren.
- Het consultancy programma voor leidinggevenden ter verbetering van de kwaliteit van de leefomgeving is effectief, maar alleen voor de punten waar gericht aan gewerkt is tijdens de consultancy. De effecten van het programma zijn niet te generaliseren naar punten die niet behandeld zijn tijdens de consultancy.

Dankwoord

Op de eerste plaats wil ik mijn promotores en co-promotor bedanken. Ik ben hen zeer dankbaar voor de inspanningen die dit promotieonderzoek hebben mogelijk gemaakt. Prof. dr. Fukkink, beste Ruben, ik wil je graag bedanken voor je rol als dagelijks begeleider met de daarbij behorende snelle beschikbaarheid in voor jou zeer drukke tijden. Het was prettig om een begeleider te hebben die statistisch sterk onderlegd is en die bovendien de tijd nam om op dat gebied geduldig uit te leggen. Prof dr. Riksen-Walraven, beste Marianne, ik wil je specifiek bedanken voor het vertrouwen dat je me de afgelopen vier jaar gaf. De afstand Nijmegen-Amsterdam heb ik geenszins gevoeld, je was altijd heel dichtbij. Na een intensieve en altijd productieve dag in Nijmegen zat ik immer met een strak omlijnd plan en vol inspiratie weer in de trein terug naar huis. Je bent zeer belangrijk geweest voor mijn wetenschappelijke vorming! Aan het SRCD congres in Seattle met jou en Mirjam bewaar ik goede herinneringen. Prof. dr. Tavecchio, beste Louis, ik wil je graag bedanken voor het lezen van mijn manuscript. Je stond bovendien samen met Marianne aan de wieg van het NCKO onderzoek waar je je tot en met je afscheidsrede hard voor hebt gemaakt. Daar ben ik je zeer erkentelijk voor.

Ik wil hier ook graag mijn NCKO collega Mirjam Gevers Deynoot-Schaub bedanken. Niet alleen voor de belangrijk rol die je vervulde in mijn onderzoek, maar vooral voor je steun in de rug en prettige kijk op het leven. Ik hoop dat je weet hoeveel dat voor mij betekent… Ik ben blij dat we al zo lang collega's zijn en dat we dat in de toekomst, ook buiten de UvA, zullen blijven. Met jou zaken doen, voelt niet als werk.

Dit onderzoek had niet plaats kunnen vinden zonder de medewerking en inzet van leidinggevenden, pedagogisch medewerkers en ouders van de kinderdagverblijven. Dank dat jullie de kinderdagverblijven openstelden voor de studentes, trainers en mij.

De afgelopen drie jaar hebben verschillende studentes van de UvA mij enorm geholpen bij de dataverzameling: Brenda Servaas, Daria Henning, Irene ten Hagen, Rosanne Spikker, Vera Copini, Whitney de Haan, Charlotte Fokkinga, Lisette Vermeulen en Lisanne Jillink. Jullie waren voor mij absoluut onmisbaar en ik ben jullie zeer dankbaar voor de inzet: voor dag en dauw gingen jullie op pad naar kinderdagverblijven en waren aldaar het visitekaartje van mijn onderzoek. Bovendien scoorden jullie talloze videobeelden. Een speciaal woord van dank gaat uit naar Brenda en Daria, jullie waren mij gedurende de volle drie jaar van het onderzoek trouw.

Bij de uitvoering van de interventie werd ik geholpen door een team van zeer ervaren trainers en voormalig NCKO medewerkers. Leontien Simis, Monique Lesser, Laura Kaats, Mahalia van der Spek en Christel Adriaens; alleen had ik dit nooit voor elkaar gekregen. Jullie hebben veel werk verzet in korte tijd, zorgden voor een klankbord bij de ontwikkeling van de interventie en zijn bovenal van doorslaggevend belang geweest voor de effectiviteit ervan. Dank daarvoor.

Mijn fijne collega's van OOS (en later OOI) wil ik laten weten dat het ontzettend prettig is om deel uit te maken van een hechte groep en om met jullie samen te werken. Henny, Loes, Floor, Anne, Eva en Elly, ons gezamenlijke symposium in Lausanne was een knap staaltje teamwork. Daar ben ik trots op. Prof. dr. Overbeek, beste Geertjan, je komst naar de UvA bracht een frisse wind, maar vooral ook aanstekelijke inspiratie. Ik kijk met vertrouwen uit naar de toekomstige samenwerking.

Een promotie gaat gepaard met bloed, zweet en tranen. Dan is het fijn als je aio's om je heen hebt die in hetzelfde schuitje zitten. Dominik, Bonny, Cathy, Eva, Imane, Loes en Loes, jullie aanwezigheid op de UvA heeft ervoor gezorgd dat ik vooral veel plezier heb beleefd aan de afgelopen vier jaar. Loesies en Dominik, jullie wil ik graag nog speciaal genoemd hebben, omdat jullie inmiddels veel meer zijn dan collega's.

Gelukkig was er de afgelopen jaren meer dan werk. Ik prijs mij dan ook zeer gelukkig met mijn lieve vriendinnen, die veel belangstelling toonden maar ook mijn aandacht wisten af te leiden in drukke tijden. Hannah, speciale dank voor al je advies op gebied van de Engelse taal. En Eef, omdat we na 30 jaar vriendschap aan een half woord genoeg hebben, dat is heel bijzonder.

Als laatste dank ik hen die mij het meest nabij zijn: mijn lieve familie en ons mooie gezin. Floor, samen zoevend op de fiets door Rome was zo heerlijk ongecompliceerd na een intense en spannende periode. Onze dinsdagavond samen is mij zeer dierbaar. Frans, wat fijn dat je een wetenschappelijk klankbord was al die tijd. Lieve ouders, Theo en Moniek, jullie hebben ons drieën een zeer liefdevolle en rijke opvoeding gegeven die ons stimuleerde om gretig en op eigen wijze de wereld te ontdekken ("*is niet een muis, lijke muis*"). Want zo klom ik dus werkelijk op jullie schouders en kijk ik inderdaad vol vertrouwen naar de toekomst. Het uitzicht is mooi en het vangnet stevig. Tenslotte mijn lieve broers Gijs en Hendrik, beiden ook onderweg naar een promotie: dat jullie links en rechts van me staan op de dag van de verdediging is voor mij vanzelfsprekend, want ik ben allang niet meer de grote zus…

Curriculum Vitae

Katrien Helmerhorst was born on August 23, 1982 in Amsterdam The Netherlands. She obtained her gymnasium degree at the Ignatius Gymnasium in Amsterdam. In 2006, she graduated from the study Pedagogical Sciences at the University of Amsterdam. During her master's program she wrote her thesis with the Netherlands Consortium for Research in Child Care (NCKO). After receiving her master's degree, she worked as a junior-researcher for the NCKO at the University of Amsterdam and was responsible for the daily co-ordination of the data collection during the 2008 quality measurement. During this project she was trained to observe with the ITERS-R/ECERS-R and the CIP scales. In 2009, she started as a PhD candidate at the Research Institute of Child Development and Education at the University of Amsterdam. Between 2010 and 2012 she was co-chair of the PhD candidates counsel. Her research focusses on quality of the care in daycare centers (0- to 4-year-olds), playschools (2- to 4-year-olds), and family/home care (0- to 12-year-ols). Her dissertation resulted in the development of the CIP training, a video feedback training that shows caregivers how to improve their interactive skills with young children. Together with the NCKO research team, she works in close collaboration with researchers from Oslo and Akershus University College of Applies Sciences in Norway, where the NCKO research program (in a slightly adapted form) is performed in the Norwegian context. Since 2010, she teaches a bachelor course at the University of Amsterdam and is a guest lecturer at the minor Child Care at Amsterdam University of Applied Sciences (Hogeschool van Amsterdam). In March 2014, the NCKO research team was granted a subsidy from the Ministry of Social Sciences and Employment, which allows Katrien to currently work as a postdoctoral researcher at the Research Institute of Child Development and Education at the University of Amsterdam. This project will focus on the implementation of the CIP training in caregiver vocational training (MBO – PW3 & PW4).

SOPHIE ZAAL