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

Helmerhorst, K.O.W.

Publication date
2014
Document Version
Final published version

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.