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

Helmerhorst, K.O.W.

Publication date
2014
Document Version
Final published version

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.

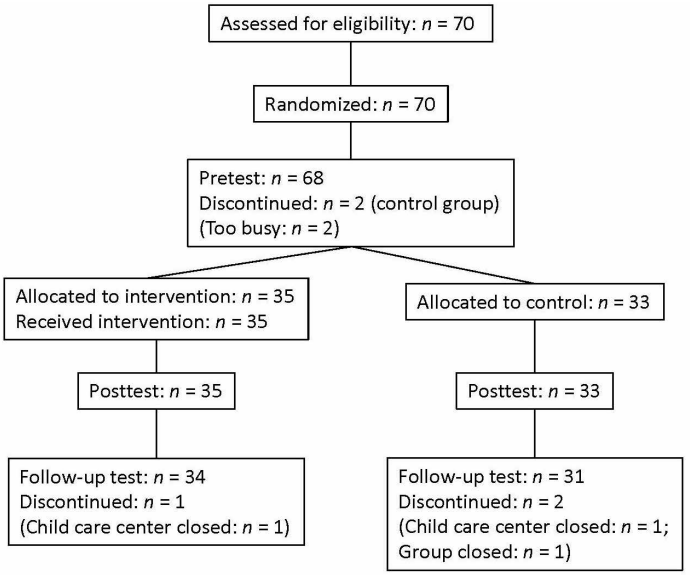

Figure 1 Flow Chart of the Study's Progress in Terms of Groups.

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).

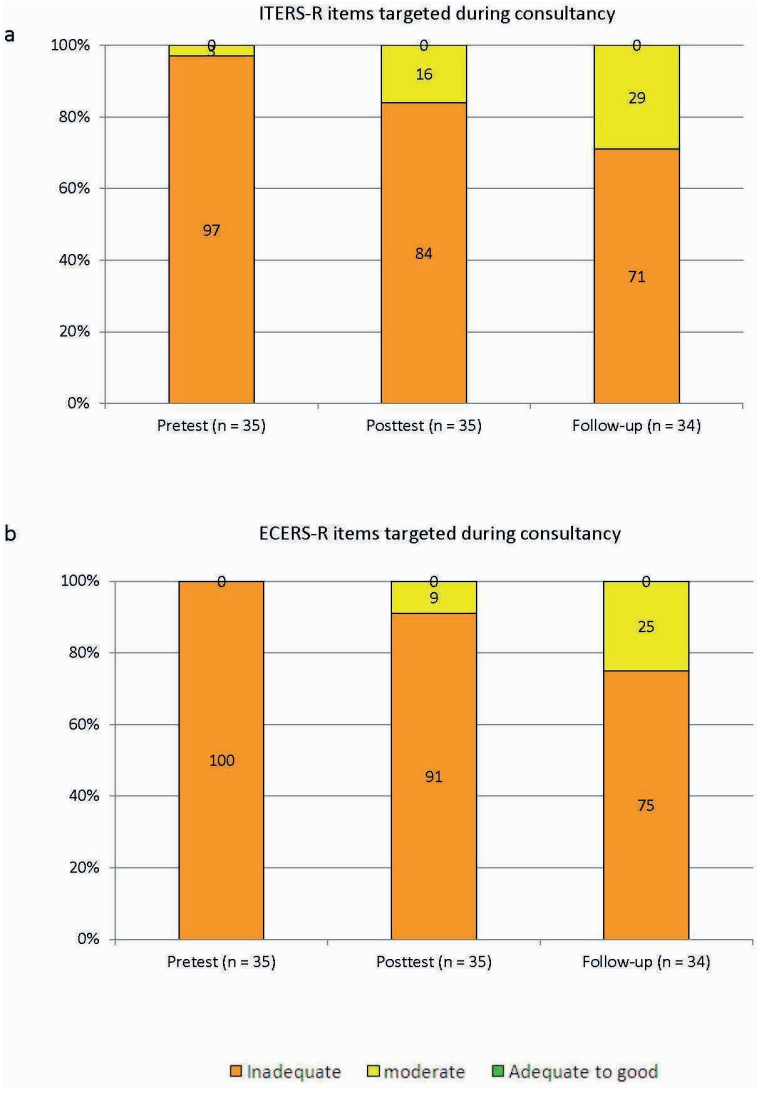

Figure 2 Percentages of Groups scoring Low, Moderate, and High on the ITERS-R and ECERS-R Items Targeted During Consultancy in the Experimental Group at Pretest, Posttest, and Follow-up test.

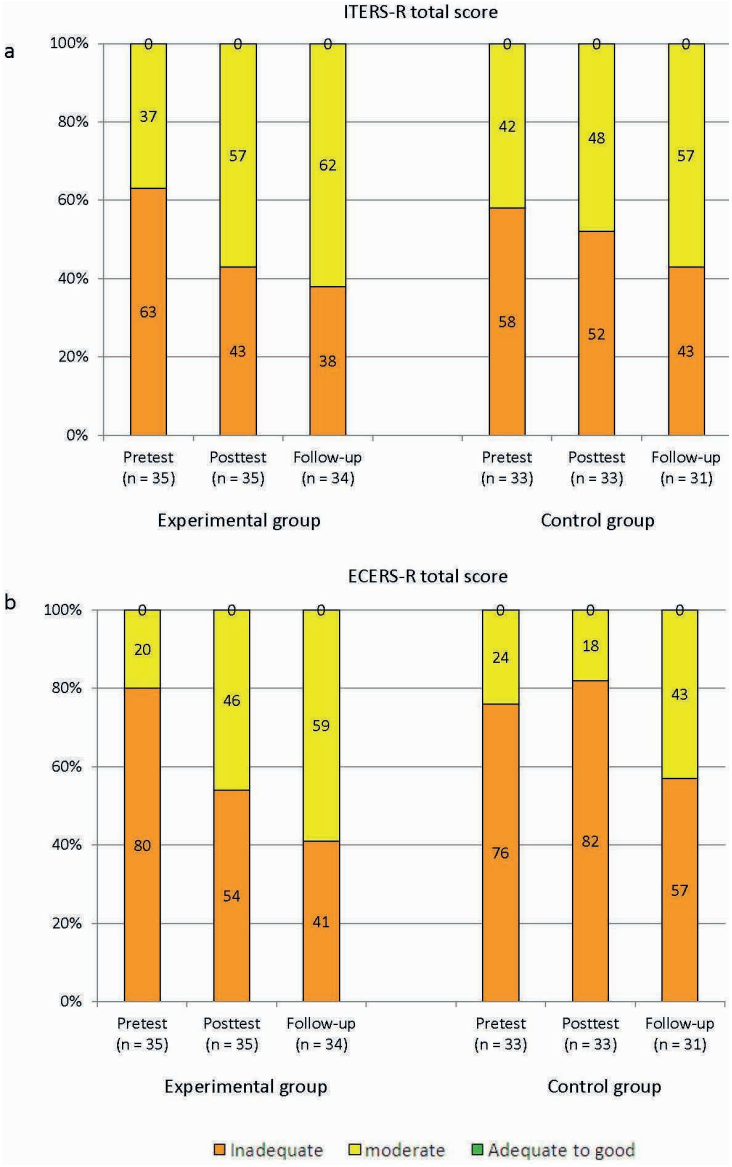

Figure 3 Percentages of Groups scoring Low, Moderate, and High on ITERS-R and ECERS-R Total Score in the Experimental and Control Group at Pretest, Posttest, and Follow-up test.

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.