

Does training matter? A meta-analysis and review of caregiver training studies

Ruben G. Fukkink*, Anna Lont

SCO-Kohnstamm Instituut, Universiteit van Amsterdam, Wibautstraat 2-4, 1091 GM Amsterdam, The Netherlands

Received 4 May 2006; received in revised form 10 April 2007; accepted 20 April 2007

Abstract

A review of studies published between 1980 and 2005 shows a significant positive effect of specialized training on the competency of caregivers in childcare ($d=0.45$, $S.E. = 0.10$). Experimental results from the meta-analysis were significantly smaller for settings with no fixed curriculum content, delivery of the training at multiple sites and large-scale programs. Results were also smaller when tests were used that did not align closely with the content of the training. Furthermore, experimental results were smaller for the skills domain, compared to the knowledge and attitude domain. A subset of experiments with both caregiver and child data also showed a positive effect, supporting the causal link between caregiver training, caregiver competencies and child behavior in childcare, although this effect was not significant due to the small number of studies ($d=0.55$, $S.E. = 0.30$). Based on these findings, we advocate the inclusion of instruction related to teacher–child interaction in the curriculum of vocational training for caregivers. © 2007 Elsevier Inc. All rights reserved.

Keywords: Caregiver training; Teacher–child interaction; Caregiver competency; Process quality; Childcare; Meta-analysis

1. Introduction

Correlational research has suggested that the training of caregivers is a cornerstone for quality in early care. Caregivers with higher educational levels provide better personal care (Davis, Thornburg, & Ispa, 1996; Honig & Hirallal, 1998), are more sensitive (Burchinal, Cryer, & Clifford, 2002; Clarke-Stewart, Lowe Vandell, Burchinal, O'Brien, & McCartney, 2002; Ghazvini & Mullis, 2002; Honig & Hirallal, 1998; Howes, 1997; Howes, Whitebook, & Phillips, 1992; Whitebook, Howes, & Phillips, 1990), are more involved with children (Blau, 1997; Burchinal, Howes, & Kontos, 2002; Ghazvini & Mullis, 2002; Whitebook et al., 1990), and have more knowledge of developmentally appropriate practice (Snider & Fu, 1990) than caregivers with lower educational levels. Furthermore, more educated early educators offer richer learning experiences (Berk, 1985; Clarke-Stewart et al., 2002; Davis et al., 1996), provide more language stimulation (Berk, 1985; Honig & Hirallal, 1998; Howes, James, & Ritchie, 2003; Ruopp, Travers, Glantz, & Coelen, 1979), and stimulate the social and physical skills of children more often (Honig & Hirallal, 1998) than other educators. Formal education has also been linked empirically with such global childcare quality measures as the Infant and Toddler Environment Rating Scale (ITERS) and Early Childhood Environment Rating Scale (ECERS) instruments for center-based care (Burchinal et al., 2002a; Ghazvini & Mullis, 2002; Phillips, Mekos, Scarr, McCartney, & Abbott-Shim, 2000; Phillipsen, Burchinal, Howes, & Cryer, 1997; Scarr, Eisenberg, & Deater-Deckard, 1994; Whitebook et al., 1990), the

* Corresponding author. Tel.: +31 20 525 1351; fax: +31 20 525 1200.
E-mail address: R.G.Fukkink@uva.nl (R.G. Fukkink).

Family Day Care Rating Scale (FDCRS) instrument for family daycare (Burchinal et al., 2002b), and the childcare version of the HOME instrument (Clarke-Stewart et al., 2002). Interestingly, recent studies have shown that both the general educational level (e.g., the number of years of education) and specialized caregiver training are both significant and strong predictors of sensitive and stimulating caregiver–child interactions and global quality ratings (Ghazvini & Mullis, 2002). Empirical evidence has extended to professional development after initial training. In contrast to earlier studies, Burchinal et al. (2002a) were able to show that both formal and informal training (e.g., attending specialized workshops) contribute to childcare quality. Norris (2001) found that family child care providers who continuously participated in training offered higher quality care than providers who attended training intermittently or never participated in training. Supervision while working in childcare is another example of informal training that has appeared as a predictor of effective teaching (Howes et al., 2003).

Seen from this perspective, it is not surprising that both education and training appeared better predictors of childcare quality than did caregiver age (Clarke-Stewart et al., 2002), work experience (Burchinal et al., 2002a; Clarke-Stewart et al., 2002; Honig & Hirallal, 1998; Howes et al., 1992; Phillippsen et al., 1997; Snider & Fu, 1990), professionalism (Clarke-Stewart et al., 2002), stability of the childcare position (Honig & Hirallal, 1998), mental health (Clarke-Stewart et al., 2002), or other caregiver-related characteristics in correlational studies. In some studies, formal education and training were also stronger and more robust predictors of childcare quality than was the adult–child ratio and group size (Burchinal et al., 2002b; NICHD ECCRN, 2002a).

Education and training of caregivers have also been associated with positive developmental outcomes for children. Research from within the “structure → process → outcomes” paradigm (see NICHD ECCRN, 2002a) suggests that the effects of education and training transfer from caregiver practice to children’s outcomes in several developmental domains. Children who attend centers that have more educated and trained caregivers have higher levels of language skills (Burchinal et al., 2002a; Clarke-Stewart et al., 2002; NICHD ECCRN, 1999, 2002b), social competence (NICHD ECCRN, 2002b), and school readiness (NICHD ECCRN, 1999) than children at other centers. They also have fewer behavioral problems at 3 years of age (NICHD ECCRN, 1999), and they have been rated as more cooperative (Clarke-Stewart et al., 2002). In a study by Burchinal et al. (2002a), the positive relationship between education and training of the caregiver and developmental outcomes for children remained after correcting for group size and adult–child ratio.

Despite the conclusions reported above, empirical evidence on the value of caregiver schooling is not uniformly positive. Several studies have found that the positive relationship of education and training with process quality varied across child age groups. Phillippsen et al. (1997) found that process quality was higher in classrooms with caregivers with higher education, whereas education was not related to process quality in infant/toddler classrooms. Phillips et al. (2000) found that process quality was related to training in infant and toddler center classrooms and to education in toddler center classrooms only, but was not related to either training or education in preschool center classrooms. In contrast, the NICHD ECCRN (1996) reported process quality in relative, child care home and center-based settings was reliably related to both caregiver education and specialized training for toddlers and preschoolers, but not for infants. The NICHD ECCRN (2000) study, however, reported significant relationships between caregiver education and process quality for toddlers and preschoolers. No relationship between specialized training and positive caregiving was found for this population, however. Similarly, not all studies have supported the positive influence of caregiver education or training on children’s development. For example, Burchinal et al. (2000) found no clear association between the developmental outcomes of children and the educational level of caregivers.

Equally important, a number of key scientific and practical questions remain regarding the design and efficacy of caregiver training. First, correlational evidence does not allow causal claims about the effects of caregiver education or training, as has been stressed by many authors, nor does it allow the direct estimation of their effects on childcare quality. In fact, several authors have highlighted the empirical difficulty of disentangling the effects of caregiver education from those of the socio-economic status of caregivers and overall quality of childcare centers (Berk, 1985; Blau, 1997; Kontos & Wilcox-Herzog, 2001). An additional complication is inherent in the fact that years of formal education and specialized caregiver education are intertwined (Phillipsen et al., 1997; Kontos & Wilcox-Herzog, 2001).

Correlational research involves an additional limitation, in that it provides no direct insight into which concrete interventions may be effective for fostering quality in childcare (see Whitebook, 2003). Although it is possible to draw distinctions between the quantity of education (e.g., number of years), general level (e.g. baccalaureate level or lower), specificity (e.g., general versus specialized education), and the recency of training experiences (e.g., having taken a course in the past year or not), correlational research does not reveal the specific content of the key variable of “caregiver training.” Because caregiver training remains a “black box” in the analysis, this line of research cannot be

used to guide the development of effective training, which is a key challenge facing the discipline of childcare (Taylor, Dunster, & Pollard, 1999; Whitebook, 2003; Zaslow & Martinez Beck, 2006). Experimental and quasi-experimental research offers more and direct insight into the content of training methods and their efficacy. In addition, only (quasi-) experimental research is able to offer empirical support for the assumed causal link between caregiver training and quality in childcare.

1.1. *Research design*

As many authors have noted, research on the extent to which specialized training enhances caregiving have been inconclusive, and enhances child outcomes has been even less conclusive. Explorations of the effects of caregiver training on on-the-job performance, including caregiver interaction behavior with children are particularly limited (see Berk, 1985; Daniels & Shumow, 2003; Kontos & Wilcox-Herzog, 2001; Taylor et al., 1999; Whitebook, 2003). Problems include the use of self-appraisals in the early evaluations (Peters & Kostelnik, 1981; Snow, 1982) and failure to pursue or publish evaluations of training (Peters & Kostelnik, 1981; Saracho, 1993), perhaps due a belief that work with young children is not something that requires training (Genishi, Ryan, Ochsner, & Yarnall, 2001). Nevertheless, there have been a number of experimental or quasi-experimental studies published in recent years.

As many authors have noted, (quasi-)experimental research into the benefits of specialized caregiver training is infrequent and inconclusive in many areas. Explorations of the effects of caregiver training on on-the-job performance, including caregiver interaction behavior with children are particularly limited (see Berk, 1985; Daniels & Shumow, 2003; Kontos & Wilcox-Herzog, 2001; Taylor et al., 1999; Whitebook, 2003). Peters and Kostelnik (1981) and Snow (1982) have criticized the relatively poor methodological standards of evaluation, including the popular practice of evaluation with self-appraisals in their narrative reviews of studies that were published before 1980.¹ A different sort of problem is that many evaluations of caregiver training are not reported, and, hence, many ideas are lost, and few are shared, according to Peters and Kostelnik (1981). Saracho (1993) suggested that this state of affairs might result from a relative lack of interest in scientific evaluation within this domain. Also Powell and Stremmel (1989: p. 340) have mentioned “the perception that work with young children neither possesses nor needs a technical database derived from scientific theory and research,” which is related to the conviction, or even “folk belief” (see Genishi et al., 2001: p. 1183) that is held by some people that activities involved in the education and care of young children are familiar to everyone. The fact remains, however, that a number of interesting studies have been published in the past years. Some have reported favorable outcomes, while others have not. The question that was posed by Arnett (1989), “Does training matter?”, has yet to be answered conclusively.

1.2. *Goal of this study*

A paradoxical situation emerges from the findings of correlational and (quasi-)experimental research with regard to the training hypothesis. Although correlational research provides abundant indirect evidence, many questions remain concerning support from (quasi-)experimental studies, which test the hypothesis directly. Similarly, little is known about factors that influence the effectiveness of caregiver training. The goal of this study is to integrate findings from (quasi-)experimental studies into the effects of specialized caregiver training on caregiver competencies. In this study, caregiver competencies have been delineated as the professional knowledge, attitudes, and skills that are related to teacher-child interaction. Two other research questions are addressed in an explorative fashion, taking into account the limited empirical evidence. The first explorative question concerns the study characteristics that are associated with experimental results. The second question involves the transfer effects of specialized caregiver training on children's behavior and development. A few studies have explored the effects of training both at the level of the caregiver and at the level of the children, thereby providing further insight into the assumed causal link between caregiver training and quality of childcare.

Following the basic “structure → process → outcomes” paradigm (see NICHD ECCRN, 2002a), we assume that instruction influences the professional competencies of caregivers, which underlie their professional performance, or,

¹ This comment still applies to a part of the literature (see, for example, Aguirre and Marshall, 1988; Griffin et al., 1998; Rellon, 1998; Restaino-Kelly and Handler, 1996).

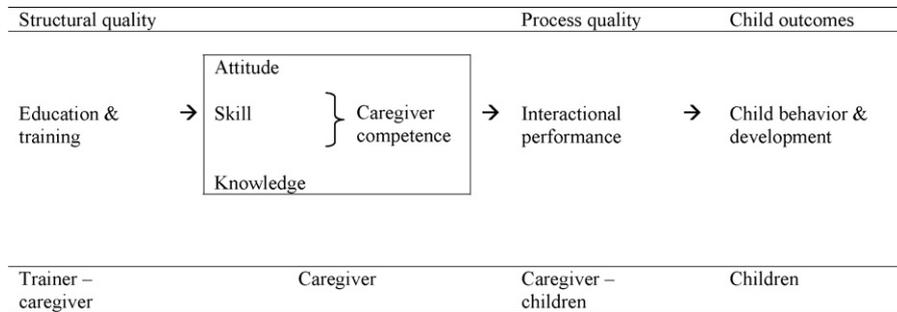


Fig. 1. Basic model of this study.

more specifically, their interaction with children. Subsequently, this interaction influences the behavior of children in the childcare center and their development. Although skills are the most “visible” part of caregiver competencies, and, hence, are most directly related to process quality of childcare, we distinguish between skills, attitude and knowledge as separate and complementary learning domains. This fits in with current educational theories, which emphasize that skills and their application in practice are connected to professional knowledge and attitudes (see [Ellström, 1997](#); [Spencer & Spencer, 1993](#); [Weinert, 2001](#)); for example, a caregiver’s beliefs and knowledge, or a lack thereof, may promote or hinder the transfer of a trained skill to practice. Our framework (see [Fig. 1](#)), which is not considered an exhaustive theory of the complex processes involved, distinguishes the key elements at a basic level in order to categorize the findings from training studies.

We seek to answer these questions by reviewing (quasi-)experimental studies that were published between 1980 and 2005 (see [Peters & Kostelnik, 1981](#) and [Snow, 1982](#) for a qualitative overview of several training programs for the period before 1980).

2. Method

2.1. Selection of studies

We conducted a search in the ERIC, PsychINFO, and SSCI databases, combining descriptors for the independent variable (“train*, instruct*, educat*, prepar*, or experiment**”) and setting (“childcare, early childhood, preschool, caregiver*, or teacher**”) to locate relevant studies that were published between 1980 and 2005. Studies were also located by searching for additional references in the references of collected reports. In addition, the databases were searched in a second round, using the names of the authors of relevant reports that had been identified in the original search. At the suggestion of [Peters and Kostelnik \(1981\)](#) and [Saracho \(1993; see the Introduction\)](#) and meta-analytical guidelines (e.g., [Rosenthal, 1994](#)), unofficial publications (“fugitive literature”) were also included in the search. The broad-ranging search was followed by a selection of studies, based on a reading of their full content. Selection was based on the following criteria: First, we included only reports that provided information about studies involving specialized caregiver training with a focus on interaction skills with children (see [Kontos & Wilcox-Herzog, 1997](#)) in a regular childcare setting. Second, we included only those studies in which the caregiver was the primary focus of the evaluation, possibly supplemented with a focus on children. We did not include studies that focused specifically on children with disabilities, early childhood special education or residential childcare. Studies involving childcare counseling and consultation were also not included. Furthermore, we did not include studies that involved teacher competencies that are not directly related to regular caregiver–child interaction.² Both two-group designs (i.e., with a treatment and comparison group) and one-group designs with a pretest and posttest were included. Like in many other applied research areas, the pool of relevant childcare studies comprises both designs ([Shadish, Cook, & Campbell, 2002](#); [Morris & DeShon, 2002](#)). Following meta-analytic guidelines ([Gibbons, Hedeker, & Davis, 1993](#); [Morris & DeShon, 2002](#); [Wampler, Reifman, & Serovich, 2005](#)), we selected all studies to include all empirical evidence available

² For example, the enhanced frequency of elaborated use of target words during interactive reading (see [Wasik and Bond, 2001](#)), specific caregiver behaviors during music activities (see [De l’Etoile, 2001](#)) or caregiver attitude towards math activities (see [Arnold et al., 2002](#)).

pertaining to the central hypothesis of this study. Finally, we included only studies that reported the statistics that are necessary to deduce effect sizes. A number of studies did not contain this information. These studies, which included both group designs and single-subject designs, are briefly discussed in a narrative review, but could not be included in the meta-analysis.

Finally, a decision was made with regard to the inclusion of treatments. Arnett (1989) reports data for two treatments (Levels 2 and 3 from the Bermuda College Training Program) and a Level 4 group with 4-year college degree in early childhood education. We did not consider the latter reference group as an experimental group, as the formal education of the caregivers was not a focus of the experiment. Campbell and Milbourne (2005) report pretest and posttest data for two treatments (the First Beginnings program with and without consultation). Only the latter group meets our eligibility criteria and has been included (i.e., as a pretest-posttest design). Because Cassidy, Hicks, Hall, Farran, and Gray (1998) describe and evaluate the effects of both training and service placement separately, they have been addressed as two separate treatments. For one treatment, caregiver data and child data were reported in separate studies (Kontos, Howes, & Galinsky, 1996; Howes, Galinsky, & Kontos, 1998: Study Three); the findings from both studies have been combined. Finally, Horm-Wingerd, Caruso, Gomes-Atwood, and Golas (1997) reported an initial evaluation of a training, the effects of which have been studied in an extended sample by Horm, Caruso, and Golas (2003). The results from the latter, more comprehensive, study were used to determine effect sizes.

2.2. Coding

The training studies that were included in the meta-analysis were coded for instructional and methodological characteristics. We coded whether the training was integrated into childcare practice (i.e., in the case of in-service training at the center or in case classroom activities were combined with practice at the center) or not (i.e., delivery outside the center, without follow-up activities in the center). It was also coded whether supervision was part of the training, i.e. the provision of individual feedback and support by a personal trainer or coach. Scope of the training was coded as either narrow or broad, to differentiate between courses that cover various topics and courses that have a specific focus. Location of the training was scored as one-site or multi-site. In addition, we coded whether the intervention was characterized by a fixed curriculum (i.e., the curriculum content of the training was identical for each trainee) or not (i.e., the curriculum differed between training sites or individuals). Other instructional characteristics included the explicit focus of the training on knowledge, skills, or attitude (or possible combinations); the use of video feedback; the setting (center-based care or family daycare), the number of training sessions and their duration. Finally, the trainees' age and working experience were coded (see Table 1). The coded methodological characteristics included publication type (scientific journal or other publications); type of design (within design, like the one-group pretest posttest design, or between design, like the untreated control group design); random assignment or matching (at the level of the individual caregiver or childcare center); design with pretest and posttest (yes/no); the number of participants at the start; attrition (percentage); and the definitive number of participants (see Table 2).

Effect sizes were determined for the direct post-test data. The effect sizes (i.e., dependent variables) were coded for the following characteristics: self-report measures of caregiver competencies versus 'objective' test or observation by independent observer; focus of the test on knowledge, skill or attitude (or possible combinations); standard measure or newly developed measure; and close alignment of the test with the content of the training (yes/no). These characteristics were judged by the coders for each dependent variable. Dependent variables related to participants' satisfaction with the course, work commitment or job satisfaction were not the scope of this study, and, hence, they were not included.

Two independent coders coded all the studies. Inter-coder reliability was estimated by determining the Cohen's kappa (κ) in case of nominal variables and the intra-class correlation (ICC) for interval variables, using the value .70 as a cut-off score for inclusion. Reliability proved satisfactory, with κ ranging from .77 to 1 and ICC ranging from .99 to 1 for the characteristics of the training studies. The derivation of effect sizes from the report was also reliable (ICC was 1 for caregiver data and .95 for child data). In the case of divergent codes, final codes were established by discussion and used in subsequent analysis. A number of coded variables were not included in the meta-analysis, however, due to lack of information in the description of the actual training in the reports (e.g., caregiver-child ratio, group size, informal or formal status of training, length of training in weeks, class size, previous education of trainee).

The following hypotheses were studied. First, an association was expected between publication in a scientific journal and larger effect sizes (Hypothesis 1), derived from the literature, as discussed in the Introduction (see Peters & Kostelnik, 1981), and based on the meta-analytical literature (see Begg, 1994). The second and third hypothesis

Table 1
Overview of experiments: instructional characteristics

Study	Program name	Integrated with practice	Supervision	Scope	Location	Fixed Curriculum	Focus	Video feedback	Setting	Nr. sessions	Hrs	Age	Exp	Educ
Arnett (1989)	Bermuda, two course program	No	No	B	One	Yes	K/S/A	No	C	–	–	27	2.5	12
Arnett (1989)	Bermuda, four course program	Yes	Yes	B	One	Yes	K/S/A	No	C	–	–	36	6.5	12
Bloom and Sheerer (1992)	Early Childhood Leadership	Yes	Yes	B	One	No	K/S/A	No	C	77	4	38	12	15
Campbell and Milbourne (2005)	First Beginnings, no consultation	Yes	No	B	Multi	No	K/S/A	No	C	5	3	38	7.4	13
Caruso et al. (1998)	NEHSTC	No	No	B	One	No	K/S/A	No	C	4	6	35	4	13
Cassidy, Buell, Pugh-Hoese, and Russel (1995)	TEACH	No	No	B	Multi	No	K/S/–	No	C	4	–	–	2.5	12
Cassidy et al. (1998)—I	NCCCC—training	Yes	Yes	B	One	Yes	K/S/A	No	C	22	3	22	0	13
Cassidy et al. (1998)—II	NCCCC—service-only	Yes	Yes	B	Multi	No	K/S/A	No	C	–	–	22	0	13
Fantuzzo et al. (1996)	Collaborative training	Yes	Yes	B	Multi	Yes	–/S/–	Yes	C	6	4.7	–	8	14
Fantuzzo et al. (1997)	Collaborative training revised	Yes	Yes	B	Multi	Yes	–/S/–	Yes	C	5	5.6	–	8	14
Girolametto et al. (2003)	Learning Language and Loving it	Yes	Yes	N	One	Yes	–/S/–	Yes	C	14	1.7	38	15.5	14
Girolametto et al. (2004)	Learning Language and Loving it	Yes	Yes	N	One	Yes	K//S/–	Yes	C	6	1.5	38	10	14
Horm-Wingerd et al. (1997)	NEHSTC	Yes	Yes	B	One	Yes	K/S/–	No	C	4	6	35	4	15
Johnson (1994)	Project Success	No	No	B	Multi	No	K/S/A	No	C	–	–	38	10	12
Kaplan and Conn (1984)	MDCPTP	No	No	B	Multi	No	K/S/–	No	C	7	2.9	26	5	13
Kontos et al. (1996)	Family-to-family	No	No	B	Multi	No	K/S/–	No	F	5	4.3	38	5.4	14
Rhodes and Hennessy (2000)	Foundation course	No	No	B	One	Yes	K/S/A	No	C	48	2	–	6	14
Total/mean score		Yes: 11, No: 6	Yes: 9, No: 8	B: 15, N: 2	One: 9, Multi: 8	Yes: 9, No: 8	14/17/9	Yes: 4, No: 13	C: 16, F: 1	16(22)	3.7 (1.6)	33 (6.4)	6.3 (4.1)	13.4 (1.0)

Note—Scope: B = broad, N = narrow; focus: K = knowledge, S = skill, A = attitude; setting: C = center-based care, F = family care; Nr. sessions = number of sessions of the program; Hrs: average duration of one session in hours; Age: Average age of trainees; Exp (years) = average working experience of trainees; Educ = years of formal education (in years).

Table 2
Overview of experiments: methodological characteristics

Study	Journal	Design	Assign	Pre-test– post-test	Nr. Exp + Con Start	Attrition Exp + Con	Nr Exp + Con Def	Nr. ES (caregiver)	Nr. ES (children)
Arnett (1989)—I	Yes	B	NR	No	12 + 14	0 + 0%	12 + 14	5	–
Arnett (1989)—II	Yes	B	NR	No	16 + 14	0% + 0%	16 + 14	5	–
Bloom and Sheerer (1992)	Yes	B	NR	Yes	22 + 22	0% + 0%	22 + 22	4	–
Campbell and Milbourne (2005)	Yes	W	–	Yes	37	41%	22	4	–
Caruso et al. (1998)	No	B	NR	Yes	231 + 130	62% + 68%	91 + 41	2	–
Cassidy et al. (1995)	Yes	B	MA	Yes	21 + 20	10% + 25%	19 + 15	4	–
Cassidy et al. (1998)—I	Yes	W	–	Yes	35	31%	24	3	–
Cassidy et al. (1998)—II	Yes	W	–	Yes	24	0%	24	9	–
Fantuzzo et al. (1996)	Yes	B	RA	No	24 + 24	17% + 17%	20 + 20	6	–
Fantuzzo et al. (1997)	Yes	B	RA	No	70 + 70	31% + 29%	48 + 50	5	–
Girolametto et al. (2003)	Yes	B	RA	Yes	8 + 8	0% + 0%	8 + 8	10	8
Girolametto et al. (2004)	Yes	B	RA	Yes	8 + 9	0% + 0%	8 + 9	1	4
Horm-Wingerd et al. (1997)	Yes	B	NR	Yes	295 + 153	64% + 67%	105 + 51	2	–
Johnson (1994)	No	W	–	Yes	188	50%	94	2	–
Kaplan and Conn (1984)	Yes	W	–	Yes	26	35%	17	7	–
Kontos et al. (1996)	Yes	B	NR	Yes	130 + 112	27% + 0%	95 + 112	5	1
Rhodes and Hennessy (2000)	Yes	B	MA	Yes	16 + 17	0% + 29%	16 + 12	4	2
Total/mean score	Yes: 15, No: 2	W: 5, B: 12	RA/MA: 6, NR: 6	Yes: 13, No: 4	68 (88) + 49 (53)	22% (23) + 20% (25)	38 (35) + 31 (30)	78	15

Note—Design: W = within-design, B = between design; assign = assignment to conditions is not random (NR), random (RA), or a matching procedure was used (MA); pre-test–post-test = experimental design with pre-test and post-test; Nr Exp + Con Start = number of participants in the experimental and control group at the pretest; attrition Exp + Con = attrition (%) for experimental and control group; Nr Exp + Con def. = definitive number of participants (at the post-test); Nr ES = number of effect sizes (caregiver or child level) for the treatment.

pertain to the educational and methodological relationship between the treatment and the test. We hypothesize that programs with fixed curriculum content (i.e., the independent variable) are more effective than are those with no fixed curriculum content (Hypothesis 2). Close alignment of measures (i.e., the dependent variable) with the training content is expected to be associated with larger study outcomes (Hypothesis 3).

2.3. Analysis

Most of the studies identified in the search included multiple outcomes reflecting different aspects of caregiver competencies. Effect sizes were determined as either pre–post or treatment-comparison group comparisons for each outcome variable in each study. This procedure yielded a database containing 78 effect sizes at the caregiver level, derived from 17 treatments and reported in 15 studies. At the child level, the database was much smaller, with 15 effect sizes from 4 treatments, reported in 4 studies.

The research literature consists of a mixture of within designs and between-designs. Effect sizes were derived directly from reported means and standard deviations or test statistics for both type of designs (Morris & DeShon, 2002; formula 28). Hedges' *g*, which corrects for bias resulting from small samples, was used as the effect size estimate (Hedges & Olkin, 1985). For inherently negative variables, effect sizes were recoded by changing the sign (e.g., a decrease in the detachment of a caregiver after training has a positive effect size). The large scale approximation formula of Becker (1988; see also Morris, 2000) has been used to estimate the variance for effect sizes from within-designs. Correlation values for the pre- and post-test were never reported and conservative estimates of .5 were therefore used to estimate the variance of the effect sizes of these designs.

Effect sizes were subsequently integrated into an overall effect size according to a random effects model (Hedges & Olkin, 1985; Raudenbush, 1994), using a multilevel approach. This model acknowledges the hierarchical nature of the data, with effect sizes (i.e., the lowest level) nested under treatments (i.e., the highest level) (Bryk & Raudenbush, 2002). The multi-level approach also allows the explanation of heterogeneous outcomes through moderator analysis, including explanatory variables at either level. Moderator analysis also allows the empirical examination of any variations in effect sizes from between and within designs, or from other characteristics of coded studies (Morris & DeShon, 2002). The residual variance was tested for significance using the Chi-square test proposed by Bryk and Raudenbush (2002). The specification and testing of models was conducted with MLwiN, using restricted maximum-likelihood estimation (Hox, 2002).

3. Results

3.1. Description of studies: instructional characteristics

The treatments that were included in the meta-analysis were typically classroom-based courses, which were often integrated with childcare practice (see Table 1). Some form of supervision (broadly defined as coaching, mentoring, guided practice, or home visits for home-care providers) constituted a supplementary part of many courses (9 out of 16 treatments). Four treatments included videotaped practice sessions, which allowed individualized and practice-oriented feedback. As expected, the treatments showed a focus on skills and their application in childcare practice. Words from the training descriptions like “workshop sessions,” “participatory approach,” “experiential learning,” “in-service training,” “hands-on activities,” or “guided practice” all refer to this recurring theme. The teaching of theoretical knowledge was a significant accompanying component of most courses. An emphasis on attitudinal aspects was less common (at least, it was not emphasized in the description of the program). Most courses were characterized by a broad scope and the inclusion of various childcare topics. Interaction skills were always part of the training, often including general learning activities or language development (in, respectively, 14 and 5 out of 17 treatments). In some courses, cultural diversity (6) was a specific focus. In most courses, the interaction component was complemented with various topics, including health and hygiene (7), food and nutrition (6), parents and staff communication (7), children with special needs (5), management skills (4), safety (3), and program structure (2 out of 16 treatments). An exception was found in both studies of Girolametto, Weitzman, and Greenberg (2003, 2004), which involved training that was tailored to the teaching of five concrete interaction strategies within the domain of language development. The predominantly hybrid curriculum of the courses partially reflected their introductory nature. Even at the more advanced level, however, courses for experienced caregivers covered a relatively broad array of topics.

As stated before, it was not possible to deduce instructional time, training period, class size, child–adult ratio and the previous level of caregiver education from all of the reports. For the studies that provided the relevant information, the programs lasted, on the average, 55 h (S.D. = 83, minimum–maximum = 9–308 h), distributed over 16 sessions (S.D. = 22, minimum–maximum = 4–77; see also Table 2). The average training period lasted approximately 6 months, and varied from an intensive 4-day course to a long period of 80 weeks. These statistics clearly show considerable variation; this is partly due to one extreme score (i.e., the treatment in Bloom & Sheerer, 1992). Class size varied between 8 and 24 trainees, with an average of 19 (S.D. = 6.4). The average child/adult ratio was 5.9, varying from 3.6 to 8 (S.D. = 1.9).

Further analysis revealed additional didactic similarities and differences between the interventions. The (implicit or explicit) deductive subject matter sequence in most courses was associated with the skill-oriented nature of the courses, beginning with the presentation of theory, followed by practice and, finally, feedback. Despite the claimed practice-oriented perspective of all courses, differences in the emphasis on skills remained. For example, the programs of Arnett (1989; Level 2) and of Kaplan and Conn (1984) apparently relied heavily on theoretical knowledge. The presentation of the subject matter differed as well. While some courses included a textual presentation of the subject matter (e.g., using books and lectures), others applied a combination of textual and visual presentation. For example, in studies by Fantuzzo et al. (1997) and by Girolametto et al. (2004), trainees observed teachers on video during the instructional phase. In some courses, separate activities (e.g., simulation, role-play, or group discussions) were part of the teaching strategy, often preceding the step from theory to practice (e.g., Horm-Wingerd et al., 1997; Girolametto et al., 2003, 2004).

3.2. Description of studies: methodological characteristics

The meta-analysis included four experimental studies (i.e., employed random assignment), two matched quasi-experimental studies, five studies with convenient comparison groups, and four one-group pre–post-test studies. Of these, 13 collected pre and post assessments and 4 (2 experiments and 2 studies with comparison groups) did not. Attrition is relatively large in three of the included treatments (see Table 2). Consistent with the perspective of the instruction, most of the measures from the studies (86%) involved caregiver skills. Examples included the Caregiver Interaction Scale (Arnett, 1989), the Adult Involvement Scale (Howes & Stewart, 1987), the Caregiver Behavior Checklist (Kaplan & Conn, 1984), and the Appropriate Care-giving scale from the ITERS/ECERS instrument. Knowledge was involved in 10 percent of the (quasi-)experimental comparisons (e.g., McPhee's Knowledge of Infant Development Inventory, 1983), and 12% pertained to attitudinal measures (e.g., the Teacher Belief Scale of Charlesworth, Hart, Burts, & Hernandez, 1990, or the Parental Modernity Scale, see Arnett, 1989). A small number of the measures tapped a combination of knowledge, skill, and attitudinal aspects (thus explaining why the percentages do not add up to 100%). A small proportion of these measures were based on self-report (13%). In three quarters (74%) of all cases, the tests were considered to align closely with the content of the training. Finally, individual effect sizes revealed significant variation (see Fig. 2). Roughly one quarter of the effect sizes (28%) from the sample fell within the negative to zero range.

The four studies with evaluation at both the caregiver and child level comprises 15 effect sizes for child-related measures (see Table 2). On average, the number of children is 40 (S.D. = 21) and 36 (S.D. = 24) for the experimental and control group, respectively. The children were randomly assigned at center level in two experiments, whereas assignment was not random in one study; one study adopted a within-design. The average age of the children is 37.4 months (S.D. = 7.8).

3.3. Aggregation of results

As shown in Table 3, the aggregation of results revealed a statistically significant effect size of $d = 0.45$ (S.E. = 0.10), corresponding to a medium effect size, according to the guidelines of Cohen (1988). This effect can be interpreted as the aggregated effect of specialized training on general caregivers' competency. Learning gains were found in the knowledge, attitude and skills domain with aggregated effect sizes that amounted to 0.43, 0.65 and 0.40, respectively.

The homogeneity test indicated significant variation. The variation in effect sizes was not significantly related to publication type (Hypothesis 1). Reports in scientific journals did generally report larger effect sizes (i.e., a difference of $d = 0.46$), but this association did not reach statistical significance due to the small sample but also the large variation in articles from scientific journals and other publications. The fact is that negative and positive effects have been reported

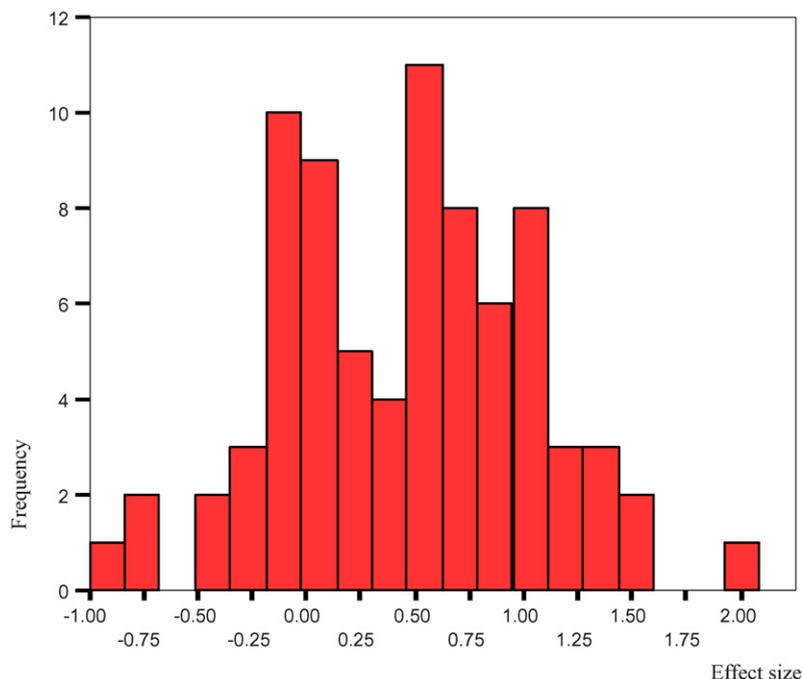


Fig. 2. Distribution of effect sizes (caregiver level).

in both scientific journals and other publications. As expected, studies that involved fixed-curriculum courses appeared to be more effective than were courses that did not have this curricular structure (Hypothesis 2). The hypothesized positive effect of the alignment of tests was also supported in this meta-analysis (Hypothesis 3). The treatment effect sizes were, on average, not statistically significant in studies that involved courses without fixed content or a choice of measures that did not align closely with the subject matter of the instruction.

In an explorative analysis, three other characteristics appeared also significantly associated with study outcomes. Delivery of the program at multiple sites and the number of the trainees at the onset of the study (regression coefficient: -0.002 , S.E. = 0.001) were both negatively associated with study outcome, indicating that large-scale training programs are not very effective. Finally, the analysis showed that caregiver skills are amenable to instruction, although learning gains are slightly smaller for this domain. The aggregated effect size of 0.40 for skill-related measures is statistically significant, but was smaller than for the other two learning domains combined ($d=0.68$). None of the other coded characteristics showed any significant relationships with the study outcomes. Furthermore, significant residual variation remained after the inclusion of the predictors in the moderator analysis (Table 4).

Further analysis revealed that the significant predictors from the moderator analysis are correlated only modestly with each other, with the exception of “fixed curriculum” and “number of training sites” (contingency coefficient = 0.47 , $p=0.03$). This multicollinearity complicates the interpretation of these two predictors. More specifically, it is not clear

Table 3
Parameter estimates for caregiver level ($k=78$) and child level ($k=15$)

Parameters	Caregiver level		Child level	
	g	S.E.	g	S.E.
Fixed effect				
Integrated effect size	0.45*	0.10	0.55	0.30
Random effects				
Variance between experimental comparisons	0.15*	0.06	0.34	0.26
Test of homogeneity	$\chi^2(77)=321^*$		$\chi^2(14)=206^*$	

Note: statistically significant effect sizes ($p<0.05$) are marked with an asterisk.

Table 4
Parameter estimates for the moderator analysis of caregiver effects

Characteristic	<i>g</i>	% variance explained	$\Delta\chi^2$
Curriculum		30	82
Fixed	0.66		
Not fixed	0.22		
Alignment test		38	99
Yes	0.54		
No	0.07		
Location		32	76
One site	0.68		
Multi-site	0.21		

whether offering a course at multiple sites (i.e., with possible degrees of freedom in the implementation due to different settings and trainers) or lack of a fixed curriculum (i.e., with possible degrees of freedom in the implementation due to different instructional content) affects the potential of training.

3.4. Effects on children

This section provides an explorative discussion of findings from the few studies that have included data for both caregivers and the children, in order to study the causal link from training to caregiver competencies and child characteristics. A preliminary analysis revealed an aggregated positive effect of training at the caregiver level for the subset of relevant studies with both caregiver and child data. This sample meets therefore the necessary criterion of a positive effect at the caregiver level ($d=0.75$, S.E. = 0.39). The aggregated effect size for the child data is also positive, although not statistically significant (see Table 3). In conclusion, the limited empirical evidence only tentatively supports the two-step causal link from caregiver training to improved caregiver competencies (a) and its positive influence on children's behavior (b).

The few studies that have been conducted have evaluated the transfer hypothesis by examining various dimensions of teacher–child interaction that affect the social, emotional, language, and cognitive development of children, as distinguished in the review by Kontos and Wilcox-Herzog (1997). Positive effects have been found for the domains of roles (e.g., socializing, encouraging play, managing misbehavior), sensitivity (e.g., being responsive, not harsh or detached), and teacher talk (e.g., frequency and quality of verbal support and verbal stimulation). The findings from Rhodes and Hennessy (2000) show convincingly that the caregivers in their study were able to intensify children's play after training. The effects on two scales, the Peer Play Scale and Play with Objects Scale, were large ($d=0.95$ and 0.85 , respectively). Howes et al. (1998) reported that the security of children's attachments increased after the training ($d=0.24$). Due to the many "fine-grained" measures in the studies of Girolametto and colleagues, different effect sizes from the sample are related to the language domain. The results from Girolametto et al. (2003) show a dramatic increase in children's language production. The effect sizes for the different measures (e.g., number of different words, multi-word utterances, and peer-directed utterances during reading and play) are all positive and average at more than one standard deviation ($d=1.25$, S.D. = 0.56). The only study with mixed results at the child level (Girolametto et al., 2004) highlights an interesting aspect of caregiver training. This study also reveals large effect sizes for the interaction measures, but some effects are positive (i.e., "acknowledge" and "uptake"), and others are negative (i.e., "override" and "no response"). One possible explanation is that the caregivers may have provided too much language stimulation after their training, which led to a simultaneous increase in the contingent and non-contingent responses of the children. This finding makes clear that there is no straightforward relation between caregiver-level and child-level effects.

3.5. Findings from other studies

As stated before, a number of studies could not be included in the meta-analysis because no effect sizes could be deduced from their reports. The following narrative review discusses the main findings of these studies, linking them with the outcomes from the meta-analysis.

Most of the group-design studies that could not be included in the meta-analysis report favorable outcomes. Mathews, Thornburg, Espinosa, and Ispa (2000) report that four repeated measurements with standardized measures (ECERS, FDCRS and the Caregiver Interaction Scale of Arnett, 1989) showed meaningful improvements on both the global quality of care and the quality of interactions. Caregivers became more sensitive and less detached in their interactions with children, although their level of harshness remained unchanged. Their attitude toward developmentally appropriate practice also improved. Based on a comparison between a training and coaching group and a comparison group, Landry (2002) reported positive changes in a teacher behavior checklist, which included, among others, responsive teaching practices and oral language use. The study of Aguirre and Marshall (1988), which focused primarily on the knowledge domain, also reported learning gains. A study by Naber (1995), which focused on the attitude domain, reported a more positive attitude on the part of caregivers toward developmentally appropriate practice.

Two studies with group-designs report disappointing findings, however. The Philadelphia Early Childhood Collaborative program, which focused on increasing coordination among organizations and providing various types of training (Jaeger, Shlay, & Weinraub, 2000), produced no significant effects on caregiver knowledge, attitude, or behavior. Also Fuqua and Greenman (1982) found negative results in their evaluation of a statewide training approach that included various courses. These two studies thus suggest a link with the findings of the meta-analysis, in which large-scale interventions, delivery at multiple sites, and lack of a clear curricular focus were associated with “null” results. Unfortunately, no conclusion regarding program efficacy can be drawn from Griffin et al. (1998), who also describe a large-scale collaborative approach, because no statistical data are reported. Finally, DeBord and Sawyers (1996) report mixed findings. Based on the finding that home providers showed significant gains on the FDCRS (Harms & Clifford, 1989) only if they were not affiliated with professional organizations, they conclude that training may not always be effective.

Single-subject studies, which typically study clearly focused interventions in highly structured settings, have generally reported encouraging results. For example, Tennant, McNaughton, and Glynn (1988) report that caregivers demonstrated both quantitative and qualitative changes in their language interactions after training. Gowen (1987) also reports favorable outcomes, showing that caregivers became more involved in children’s play after training. The multiple-probe design of Venn and Wolery (1992) also yielded positive outcomes, although no transfer to new care situations was observed. The frequency of trained interactive behaviors increased during the trained diapering situations, but this pattern did not generalize to other care situations.

The limited validity of these findings is an important issue. Although the single-subject studies are internally valid, their external validity is limited, and the internal validity (e.g., statistical conclusion validity) of some group studies is debatable. The image that emerges from this narrative review, however, seems consistent with the conclusions of the meta-analysis. Caregiver training produces learning gains in the domain of knowledge, attitude, and skills. Furthermore, large-scale interventions appeared ineffective in both the qualitative and quantitative reviews.

4. Discussion

Training seems to matter. Taken together, the current empirical evidence demonstrates that specialized training improves the pedagogical competencies of caregivers in childcare, including their professional attitude, knowledge, and skills. Further study is still needed to reach firmer conclusions with regard to the effects of caregiver training at the child level. Interestingly, the exploration of effects of caregiver training at the child level also showed positive, although non-significant, results. Unlike the evidence from correlational studies, this line of study provides direct support for the direct causal link from caregiver training to improvement of caregiver competencies; methodological limitations of the quasi-experimental research in this domain should be acknowledged, however (see below). The limited empirical data provide only tentative support yet for the assumed link from caregiver training to caregiver competencies and, subsequently, to its positive effect on children’s behavior (see also Fig. 1). The evaluations at the caregiver level or both the caregiver and child level have advanced the field since the 1980s, when Peters and Kostelnik (1981) indicated that the almost no programs had “established a relationship between program-induced personnel changes and subsequent benefits in the development of children” (p. 30).

The result of this review generalizes to caregivers with different educational levels, varying from high school (e.g., Arnett, 1989; Cassidy et al., 1998) to bachelor and graduate level (e.g., Bloom & Sheerer, 1992; Girolametto et al., 2003) and to caregivers in different settings, including center-based care (e.g., Girolametto et al., 2004), Head Start centers (e.g., Fantuzzo et al., 1996; Horm-Wingerd et al., 1997), and other, less frequently studied settings, like

family care (Kontos et al., 1996) and play groups (Rhodes & Hennessy, 2000). The result of this review, which is based on evaluations of specialized training with a focus on teacher–child interaction, cannot be generalized to other types of instruction (like, for example, initial vocational training, service placement or courses with different content). Furthermore, a broad scope and often an introductory nature appeared a dominant characteristic of the curriculum of many evaluated training programs, and, hence, it is less clear which results can be expected from programs at a more advanced level or with a narrow focus (see Girolametto et al., 2003, 2004).

Despite the positive general outcome of caregiver training at the caregiver level, it should be stressed that not all interventions are equally effective. In fact, some programs yielded null or even negative results, which contrast sharply with the encouraging outcomes of other treatments. This may explain the cautious conclusions with regard to caregiver training efficacy that have previously appeared in the literature. As hypothesized, lack of alignment between testing with course content was associated with smaller study outcomes. The latter finding suggests that possible learning gains do not seem to be given a fair chance in some evaluations, because the testing instruments were too widely remote from the actual program. Programs that have no fixed curriculum are not a recipe for success, as expected. It should be noted, however, that this factor was related to delivery of the training at multiple sites in large-scale programs, which also appeared to be associated with negative outcomes. Hence, the effect of these factors could not be disentangled empirically. The findings from this study suggest, however, that large-scale programs that are designed to a variety of training formats and to a wide variety of learners are not highly effective. The influence of the identified factors is significant. The effects of caregiver training do not simply decrease with each identified factor; they disappear.

Additionally, the interventions are not equally effective in all domains. Learning gains appeared somewhat larger for the attitude domain, compared to the skills and knowledge domain. This result is difficult to explain. A possible explanation is that attitudinal changes may precede behavioral changes, which may require a longer period of training. The fact is that some programs were relatively brief. A different, more critical explanation is that the posttest scores for attitudinal measures may reflect, at least partially, a tendency of trainees to “match” their responses with the content of the training they have just completed. In this case, attitude measures may be biased to some extent for the intervention group and may overestimate their attitudinal change after the training.

Training appeared to be less effective when it reflected a distinction between “actual” and “ideal” curriculum. For example, the choice to offer a course without a fixed curriculum is a problem from the initial developmental phase. Offering a course for many trainees and at different sites is typically a problem from the subsequent implementation phase. The lack of structure and delineation in the curriculum, delivery, and target population apparently affect the learning potential of caregiver training. Even the final phase of course evaluation is not free of problems, as demonstrated by the significant effect of using tests that are not closely aligned with course content. It should be noted that the road from the initial design of a course to its final evaluation is long, and this review has certainly not charted it in adequate detail.

4.1. Implications for future research

Many questions regarding caregiver training remain unanswered. As stated before, the current database is relatively small, and more studies are seriously needed. The small number of studies included in this meta-analysis precluded a strong test of the relationship between study characteristics and outcomes.

In future studies, the general research question, “Does training matter?” can still serve as null hypothesis. However, the educational rationale of the intervention should be expressed in greater detail. Developers of future programs should therefore explain their educational choices clearly, paying special attention to the assumed learning mechanisms. Furthermore, further experimental investigation is needed with regard of the “transfer” effects of training on children, as stated before. These two suggestions imply that pedagogic knowledge about teacher–child interaction (e.g., which caregiver competencies foster a secure and stimulating environment for children?) and educational theories about vocational training (e.g., how can we improve key caregiving competencies of current and future staff?) should be linked in future development and research to discover which interventions improve childcare quality. Obviously, this is not a simple task. A step in the right direction in this line of study is to focus on smaller learning areas and, accordingly, specific training modules in the teacher–child interactions domain. These focused studies could provide insight into which caregiver competencies can be enhanced and how this affects children in childcare. Possibly, this line of study could profit from insights from other disciplines. Other meta-analyses of related interventions with other populations, such as parents (Bakermans-Kranenburg, van IJzendoorn, & Juffer, 2003) and the family caregivers of older

adults (Sörensen, Piquart, Habil, & Duberstein, 2002), have also reported positive outcomes, showing that caregiver competencies are amenable to instruction. Acknowledging the differences between these different populations, the findings from these fields may provide guidelines for future program development and evaluation.

Finally, the empirical foundation could be strengthened by careful attention to methodological considerations in the design and report of studies in applied settings. Future studies should match caregivers or centers, or, preferably, should randomly assign them to an experimental and a control condition. The combined usage of newly developed tests that are tailored to curriculum content and standard measures for process quality, if possible, is another methodological aspect that merits attention. Some studies used standard testing instruments that were not tailored to the training. Other studies, however, relied upon newly developed tailored tests that lacked supplementary standard and validated measures. Both types of study leave therefore questions unanswered. Further attention should be paid to the issue of measurement development in this field to evaluate programs with validated measures that are directly related to their educational content. Alternatively, the use of newly developed instruments that measure specific aspects of the training could be supplemented with validated instruments that measure a related, but more general construct. In summary, future researchers should use methodologically sound experiments to provide more insight into the question of what kind of training matters for caregivers and children.

Finally, a number of reports could not be included in the meta-analysis due to missing statistical information. Other reports provided relatively little information pertaining to the intervention. A methodological aspect that could not be deduced from most reports, was the selection of participants. Specifically, it was often not clear whether the caregivers pursued training or whether the training was required. It was not clear either whether the training was connected to credit. This precluded a test of the hypothesis that training is more effective for motivated caregivers who volunteer for a specialized caregiver course. To summarize, reports of future studies should provide a complete picture of the experiment by addressing in detail both the intervention part and the evaluation part.

4.2. *Implications for childcare practice*

Licensing standards for childcare have traditionally been rather low. Related to this situation is the unclear role of training of caregivers. For example, Early and Winton (2001: p. 286) note that “being 18 years old, having a driver’s license, and having no criminal record is all that is required” in some states of the USA, while other states require a master’s degree (see Ackerman, 2004; Whitebook, 2003). This situation has generated highly critical remarks regarding current staff. For example, Kellogg (1999) claims that the workforce in the United States is “largely ill-prepared” and “randomly trained” (p. 57). The situation in Europe is similar, with some countries that have rigorous and uniform training requirements (e.g., Germany, Finland, nursery classes in the United Kingdom) and other countries that require no formal training (e.g., playgroups in the United Kingdom or Ireland) (see Oberhümer & Ulich, 1997; OECD, 2001; Tietze & Cryer, 1999 for an overview of several European countries).

Unfortunately, many providers, caregivers, and parents consider experience as a parent and other earlier life experiences sufficient preparation for work in the field of childcare (Gable & Hansen, 2001; see also Howes et al., 1992; Hübner, Walker, & McFarland, 2003), whether in family childcare settings (Taylor et al., 1999) or center-based care (Larkin & Newman, 2001). This conviction seriously undervalues the importance of specialized caregiver training, and has suppressed its development and implementation in practice. Although the belief that being a parent or having experienced other life events is relevant – or that it is a qualification, as some assert – for becoming a caregiver may not seem entirely irrational at first sight, it ignores the fact that workers become better caregivers through specialized training. The related “training versus practice” issue (see Berk, 1985) is a conviction that does not appear to support constructive thinking about the caregiver curriculum either. The question of whether caregivers should be prepared in schools or whether they should learn in practice is easily recognized as a false dilemma. At the same time, the statement that training and practice should be combined is true, but requires further specification, as illustrated by the study of Cassidy et al. (1998) in this review. The evaluation of the program in this study showed, perhaps surprisingly, that trainees made no further progress in practice after their first effective training.

Many authors emphasize the importance of training as a means of improving and maintaining process quality (e.g., Clarke-Stewart et al., 2002; Gormley, 1999; Howes et al., 2003; NICHD ECCRN, 2002b; Zaslow & Martinez Beck, 2006). It is hoped that the positive findings from the training studies that are considered in this review will promote understanding in the field that specialized training helps caregivers to improve their performance in their important jobs. The early childhood curriculum should devote a significant portion of the instruction to caregiver behavior in interacting

with children (see also Honig & Hirallal, 1998; Howes et al., 2003; Kellogg, 1999; Whitebook, 2003). Training should include information on child development (Daniels & Shumow, 2003; Stott & Bowman, 1996) and focus on responsive and stimulating interaction skills with babies and young children (Burchinal et al., 2000; Girolametto & Weitzman, 2002; Snider & Fu, 1990; Whitebook, 1999). The relatively modest role of specialized caregiver training in early childhood education in different countries and the positive effects of specialized training suggest that significant gains in the quality of teacher–child interaction can be realized in childcare practice.

References³

- Ackerman, D. J. (2004). States' efforts in improving the qualifications of early care and education teachers. *Educational Policy*, 18(2), 311–337.
- # Aguirre, B. E., & Marshall, M. G. (1988). Training family day care providers using self-study written and video materials. *Child & Youth Care Quarterly*, 17(2), 115–127.
- * Arnett, J. (1989). Caregivers in day-care centers: Does training matter? *Journal of Applied Developmental Psychology*, 10, 541–552.
- Arnold, D. H., Fisher, P. H., Doctoroff, G. L., & Dobbs, J. (2002). Accelerating math development in Head Start classrooms. *Journal of Educational Psychology*, 94(4), 762–770.
- Bakermans-Kranenburg, M. J., van IJzendoorn, M. H., & Juffer, F. (2003). Less is more: Meta-analyses of sensitivity and attachment interventions in early childhood. *Psychological Bulletin*, 129(2), 195–215.
- Becker, B. J. (1988). Synthesizing standardized mean-change measures. *British Journal of Mathematical and Statistical Psychology*, 41, 257–278.
- Begg, C. B. (1994). Publication bias. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 399–409). New York: Russell Sage Foundation.
- Berk, L. E. (1985). Relationship of caregiver education to child-oriented attitudes, job satisfaction, and behaviors toward children. *Child Care Quarterly*, 14(2), 103–129.
- Blau, D. M. (1997). The production of quality in child care centers. *The Journal of Human Resources*, 32(2), 354–387.
- * Bloom, P. J., & Sheerer, M. (1992). The effect of leadership training on child care program quality. *Early Childhood Research Quarterly*, 7, 579–594.
- Bryk, A. S., & Raudenbush. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage Publications.
- Burchinal, M. R., Cryer, D., & Clifford, R. M. (2002). Caregiver training and classroom quality in child care centers. *Applied Developmental Science*, 6(1), 2–11.
- Burchinal, M., Howes, C., & Kontos, S. (2002). Structural predictors of child care quality in child care homes. *Early Childhood Research Quarterly*, 17, 87–105.
- Burchinal, M. R., Roberts, J. E., Riggins, R., Zeisel, S. A., Neebe, E., & Brant, D. (2000). Relating quality of center-based child care to early cognitive and language development longitudinally. *Child Development*, 71(2), 339–357.
- * Campbell, P. H., & Milbourne, S. A. (2005). Improving quality of infant-toddler care through professional development. *Topics in Early Childhood Special Education*, 25(1), 3–14.
- * Caruso, D. A., Horm-Wingerd, D. M., & Golas, J. C. (1998). *Head Start Teaching Center: Outcome evaluation of 3 years of participatory training* (ERIC Document Reproduction Service No. 420427).
- * Cassidy, D. J., Buell, M. J., Pugh-Hoese, S., & Russel, S. (1995). The effect of education on child care teachers' belief and classroom quality: Year one evaluation of the TEACH Early Childhood Associate Degree Scholarship Program. *Early Childhood Research Quarterly*, 10, 171–183.
- * Cassidy, D. J., Hicks, S. A., Hall, A. H., Farran, D. C., & Gray, J. (1998). The North Carolina Child Care Corps: The role of national service in child care. *Early Childhood Research Quarterly*, 13(4), 589–602.
- Charlesworth, R., Hart, C. H., Burts, D. C., & Hernandez, S. (1990). *Kindergarten teachers beliefs and practices*. Boston, MA: American Educational Research Association (ERIC Document Reproduction Service No. 318 571).
- Clarke-Stewart, K. A., Lowe Vandell, D., Burchinal, M., O'Brien, M., & McCartney, K. (2002). Do regulable features of child-care homes affect children's development? *Early Childhood Research Quarterly*, 17, 52–86.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.
- Daniels, D. H., & Shumow, L. (2003). Child development and classroom teaching: A review of the literature and implications for educating teachers. *Applied Developmental Psychology*, 23, 495–526.
- Davis, N., Thornburg, K. R., & Ispa, J. (1996). Training determinants for quality infant child care. *Early Child Development and Care*, 124, 25–32.
- De l'Etoile, S. K. (2001). An in-service training program in music for child-care personnel working with infant and toddlers. *Journal of Research in Music Education*, 49(1), 6–20.
- # DeBord, K., & Sawyers, J. (1996). The effects of training on the quality of family child care for those associated with and not associated with professional child care organizations. *Child and Youth Care Forum*, 25(1), 7–15.
- Early, D. M., & Winton, P. J. (2001). Preparing the workforce: Early childhood teacher preparation at 2- and 4-year institutions of higher education. *Early Childhood Research Quarterly*, 16, 285–306.

³ * Studies included in the meta-analysis are marked with an asterisk; # Studies included in the narrative review are marked with this symbol.

- Ellström, P. E. (1997). The many meanings of occupational competence and qualification. In W. J. Nijhof & J. N. Streumer (Eds.), *Key qualifications in work and education* (pp. 39–50). Dordrecht, Netherlands: Kluwer Academic Publishers.
- * Fantuzzo, J., Childs, S., Stevenson, H., Coolahan, K. C., Ginsburg, M., Gay, K., et al. (1996). The Head Start Teaching Center: An evaluation of an experiential, collaborative training model for Head Start teachers and parent volunteers. *Early Childhood Research Quarterly*, *11*, 79–99.
- * Fantuzzo, J., Childs, S., Hampton, V., Ginsburg-Block, M., Coolahan, K. C., & Debnam, D. (1997). Enhancing the quality of early childhood education: A follow-up of an experiential, collaborative training model for Head Start. *Early Childhood Research Quarterly*, *12*, 425–437.
- # Fuqua, R. W., & Greenman, J. T. (1982). Training of caregivers and change in day care center environments. *Child Care Quarterly*, *11*(4), 321–324.
- Gable, S., & Hansen, J. (2001). Child care provider perspectives on the role of education and training for quality caregiving. *Early Child Development and Care*, *166*, 39–52.
- Genishi, C., Ryan, S., Ochsner, M., & Yarnall, M. M. (2001). Teaching in early childhood education: Understanding practice through research and theory. In V. Richardson (Ed.), *Handbook of research on teaching* (4th ed., pp. 1175–1210). Washington, DC: American Educational Research Association.
- Ghazvini, A., & Mullis, R. L. (2002). Center-based care for young children: Examining predictors of quality. *The Journal of Genetic Psychology*, *163*(1), 112–125.
- Gibbons, R. D., Hedeker, D. R., & Davis, J. M. (1993). Estimation of effect size from a series of experiments involving paired comparisons. *Journal of Educational Statistics*, *18*(3), 271–279.
- Girolametto, L., & Weitzman, E. (2002). Responsiveness of child care providers in interactions with toddlers and preschoolers. *Language, Speech and Hearing Services in Schools*, *33*, 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*, 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*, 254–268.
- Gormley, W. (1999). Regulating child care quality. *Annals of the American Academy of Political and Social Science*, *563*, 116–129.
- # Gowen, J. W. (1987). Facilitating play skills: Efficacy of a staff development program. *Early Childhood Research Quarterly*, *2*, 55–66.
- # Griffin, M. E., Hare, P. D., Baggerly, P. A., Leftwich, M. M., Standifer, S., & Elkins, S. A. (1998). Improving the quality of child care in the rural south. *Journal of Sociology and Social Welfare*, *25*(1), 121–131.
- Harms, T., & Clifford, R. M. (1989). *Family day care rating scales*. New York, NY: Teachers College Press.
- Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. San Diego, CA: Academic Press.
- Honig, A. S., & Hirallal, A. (1998). Which counts more for excellence in childcare staff—years in service, educational level or ECE coursework? *Early Child Development and Care*, *145*, 31–46.
- * Horm, D. M., Caruso, D. A., & Golas, J. (2003). Head Start Teaching Center: Differential training effects for Head Start personnel. *Child & Youth Care Forum*, *32*(1), 23–46.
- * Horm-Wingerd, D. M., Caruso, D. A., Gomes-Atwood, S., & Golas, J. (1997). Head Start Teaching Center: Evaluation of a new approach to Head Start staff development. *Early Childhood Research Quarterly*, *12*, 407–424.
- Howes, C. (1997). Children's experiences in center-based child care as a function of teacher background and adult:child ratio. *Merrill-Palmer Quarterly*, *43*(3), 404–425.
- * Howes, C., Galinsky, E., & Kontos, S. (1998). Child care caregiver sensitivity and attachment. *Social Development*, *7*(1), 25–36.
- Howes, C., James, J., & Ritchie, S. (2003). Pathways to effective teaching. *Early Childhood Research Quarterly*, *18*, 104–120.
- Howes, C., & Stewart, P. (1987). Child's play with adults, toys, and peers: An examination of family and child-care influences. *Developmental Psychology*, *23*, 423–430.
- Howes, C., Whitebook, M., & Phillips, D. (1992). Teacher characteristics and effective teaching in child care: Findings from the National Child Care Staffing Study. *Child & Youth Care Forum*, *21*(6), 399–414.
- Hox, J. (2002). *Multilevel analysis: Techniques and applications*. Mahwah, NJ: Erlbaum.
- Hübner, A. J., Walker, J. A., & McFarland, M. (2003). Staff development for the youth development professional: A critical framework for understanding the work. *Youth & Society*, *35*(2), 204–225.
- # Jaeger, E. A., Shlay, A. B., & Weinraub, M. (2000). Child care improvement on a shoestring: Evaluating a low-cost approach to improving the availability of quality child care. *Evaluation Review*, *24*(5), 485–515.
- * Johnson, J. (1994). *Child care training and developmentally appropriate beliefs and practices of child care employees in Pennsylvania*. Harrisburg, PA: Center for Schools and Communities.
- * Kaplan, M. G., & Conn, J. S. (1984). The effects of caregiver training on classroom setting and caregiver performance in eight community day care centers. *Child Study Journal*, *14*(2), 79–93.
- Kellogg, L. (1999). The Kellogg child development center: High quality child care. *Annals of the American Academy of Political and Social Science*, *563*, 56–72.
- * 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.
- Kontos, S., & Wilcox-Herzog, A. (1997). Teachers' interactions with children: Why are they so important? *Young Children*, *52*(2), 4–12.
- Kontos, S., & Wilcox-Herzog, A. (2001). How do education and experience affect teachers of young children? *Young Children*, *56*(4), 85–91.
- # Landry, S. H. (2002). Supporting cognitive development in early childhood. In *Paper presented at the a summit on early childhood cognitive development: ready to read, ready to learn*.
- Larkin, E., & Newman, S. (2001). Benefits of intergenerational staffing in preschools. *Educational Gerontology*, *27*, 373–385.
- # Mathews, M. C., Thornburg, K. R., Espinosa, L., & Ispa, J. (2000). Project REACH: Training rural child care providers. *Young Children*, *55*(3), 82–87.

- McPhee, D. (1983). *Manual for the knowledge of infant development inventory*. Chapel Hill: University of North Carolina.
- Morris, S. B. (2000). Distribution of the standardized mean change effect size for meta-analysis on repeated measures. *British Journal of Mathematical and Statistical Psychology*, 53, 17–29.
- Morris, S. B., & DeShon, R. P. (2002). Combining effect sizes estimates in meta-analysis with repeated measures and independent-groups designs. *Psychological Methods*, 7(1), 105–125.
- Naber, M. B. (1995). *Increasing application of developmentally appropriate practice by childcare and Head Start staff following training* (ERIC Document Reproduction Service No. 389465).
- NICHD Early Child Care Research Network. (1996). Characteristics of infant child care: Factors contributing to positive caregiving. *Early Childhood Research Quarterly*, 11, 269–306.
- NICHD Early Child Care Research Network. (1999). Child outcomes when child care center classes meet recommended standards for quality. *American Journal of Public Health*, 89(7), 1072–1077.
- NICHD Early Child Care Research Network. (2000). Characteristics and quality of child care for toddlers and preschoolers. *Applied Developmental Science*, 4(3), 116–135.
- NICHD Early Child Care Research Network. (2002a). Child-care structure → process → outcome: Direct and indirect effects of child-care quality on young children's development. *Psychological Science*, 13(3), 199–206.
- NICHD Early Child Care Research Network. (2002b). Early child care and children's development prior to school entry: Results from the NICHD study of early child care. *American Educational Research Journal*, 39(1), 133–164.
- Norris, D. J. (2001). Quality of care offered by providers with differential patterns of workshop participation. *Child & Youth Care Forum*, 30(2), 111–121.
- Oberhümer, P., & Ulich, M. (1997). Staff in early childhood services: Patterns of training and provision in the European countries. *European Early Childhood Education Research Journal*, 5(1), 17–28.
- OECD. (2001). *Starting strong: Early childhood education and care*. Paris, France: OECD.
- Peters, D. L., & Kostelnik, M. (1981). Current research in day care personnel preparation. *Advances in Early Intervention and Day Care*, 2, 29–60.
- Phillips, D., Mekos, D., Scarr, S., McCartney, K., & Abott-Shim, M. (2000). Within and beyond the classroom door: Assessing quality in child care centers. *Early Childhood Research Quarterly*, 15(4), 475–496.
- Phillipsen, L. C., Burchinal, M. R., Howes, C., & Cryer, D. (1997). The prediction of process quality from structural features of child care. *Early Childhood Research Quarterly*, 12, 281–303.
- Powell, D. R., & Stremmel, A. J. (1989). The relation of early childhood training and experience to the professional development of child care workers. *Early Childhood Research Quarterly*, 4, 339–355.
- Raudenbush, S. W. (1994). Random effects models. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 301–322). New York, NY: Russell Sage Foundation.
- Rellon, L. (1998). *Minding your own business* (ERIC Document Reproduction Service No. 430710).
- # Restaino-Kelly, A., & Handler, J. M. (1996). *The DART mentor teacher model: Training early childhood supervisors to assist beginning teachers* (ERIC Document Reproduction Service No. 403071).
- * Rhodes, S., & Hennessy, E. (2000). The effects of specialized training on caregivers and children in early-years settings: An evaluation of the foundation course in playgroup practice. *Early Childhood Research Quarterly*, 15, 559–576.
- Rosenthal, M. C. (1994). The fugitive literature. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 86–94). New York, NY: Russell Sage Foundation.
- Ruopp, R., Travers, J., Glantz, F., & Coelen, C. (1979). *Children at the center: Final report of the National Day Care Study*. Cambridge, MA: Abt Associates.
- Saracho, O. N. (1993). Preparing teachers for early childhood programs in the United States. In B. Spodek (Ed.), *Handbook of research on the education of young children*. (pp. 412–426). New York, NY: Macmillan Publishing Company.
- Scarr, S., Eisenberg, M., & Deater-Deckard, K. (1994). Measurement of quality in child care centers. *Early Childhood Research Quarterly*, 9, 131–151.
- Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston, MA: Houghton Mifflin Company.
- Snider, M. H., & Fu, V. R. (1990). The effects of specialized education and job experience on early childhood teachers' knowledge of developmentally appropriate practice. *Early Childhood Research Quarterly*, 5, 69–78.
- Snow, C. W. (1982). In-service day care training program: A review and analysis. *Child Care Quarterly*, 11(2), 108–121.
- Sörensen, S., Piquart, M., Habil, D., & Duberstein, P. (2002). How effective are interventions with caregivers? An updated meta-analysis. *The Gerontologist*, 42(3), 356–372.
- Spencer, L. M., & Spencer, S. M. (1993). *Competence at work; Models for superior performance*. New York, NY: Wiley & Sons.
- Stott, F., & Bowman, B. (1996). Child development knowledge: A slippery base for practice. *Early Childhood Research Quarterly*, 11, 169–183.
- Taylor, A. R., Dunster, L., & Pollard, J. (1999). . . .And this helps me how?: Family child care providers discuss training. *Early Childhood Research Quarterly*, 14(3), 285–312.
- # Tennant, J., McNaughton, S., & Glynn, T. (1988). Increasing child's language in day care through caregiver conversations. *Early Child Development and Care*, 35, 39–51.
- Tietze, W., & Cryer, D. (1999). Current trends in European early child care and education. *Annals of the American Academy of Political and Social Science*, 563, 175–193.
- Venn, M. L., & Wolery, M. (1992). Increasing day care staff members' interactions during caregiving routines. *Journal of Early Intervention*, 16(4), 304–319.

- Wampler, K. S., Reifman, A., & Serovich, J. M. (2005). Meta-analysis in family therapy research. In D. H. Sprenkle & F. P. Piercy (Eds.), *Research methods in family therapy* (2nd ed., pp. 318–339). New York, NY: The Guilford Press.
- Wasik, B. A., & Bond, M. A. (2001). Beyond the pages of a book: Interactive book reading and language development in preschool classrooms. *Journal of Educational Psychology*, *93*(2), 243–250.
- Weinert, F. E. (2001). Concept of competence: A conceptual definition. In D. S. Rychen & L. H. Salganik (Eds.), *Defining and selecting key competencies* (pp. 45–64). Seattle, WA: Hogrefe & Huber Publishers.
- Whitebook, M. (1999). Child care workers: High demands, low wages. *Annals of the American Academy of Political and Social Science*, *563*, 146–161.
- Whitebook, M. (2003). *Early education quality: Higher teacher qualifications for better learning environments*. Retrieved January 1, 2005, from <http://iir.berkeley.edu/csece>.
- Whitebook, M., Howes, C., & Phillips, D. (1990). *Who cares? Child care teachers and the quality of care in America; Final report of the National Child Care Staffing Study*. Oakland, CA: Child Care Employee Project (ERIC Document Reproduction Service No. 323 031).
- Zaslow, M., & Martinez-Beck, I. (Eds.). (2006). *Critical issues in early childhood professional development*. Baltimore, MA: Brookes Publishing Company.