Using Observations to Evaluate Paid Child Care Settings.

Jacqueline Barnes

Published in

Petrogiannis, K. & Melhuish, E.C. (Eds.) (2001) *The pre-school period: care - education - development: findings from the international research*. Athens: Kastaniotis. pp. 395-440.

Introduction

As research on the impact of child care on children's development has progressed, the relevance of the quality of care has become paramount (Borge, Hartmann, & Strøm, 1996; Hagekull & Bohlin, 1995; Moss & Pence, 1994; Petrogiannis & Melhuish, 1996). Earlier work looked to factors such as the age of entry, the number of hours of care, or the type of child care for explanations of child outcomes. These studies led to a range of questions being asked and particularly - is it the amount of time children spend in care, who provides the care or something about the care that makes a difference? Several studies have concluded that the quality of care received is the most relevant factor, although not for all children. For instance, behavioural problems of children of middle class families who started in the first year of life were not predicted by day care quality (Deater-Deckard, Pinkerton & Scarr, 1996), but other studies found that quality may be more relevant for African-American than European-American children (Burchinal et al., 1995), that poor quality care is related to lower cognitive and language development (Burchinal et al., 1996) and that there may be an interaction with age of entry (Howes, 1990). Looking for positive rather than negative effects, there is some agreement that high quality day care can enhance development, particularly with respect to cognitive and language development (McCartney, 1984; Melhuish et al., 1990) and for children who are already at-risk for poor outcomes (Ramey, Bryant, & Suarez, 1989; Scarr & Eisenberg, 1993). The recent NICHD study (1999) in the USA has also found that good quality caregiving is associated with better cognitive and language development. Positive impact has been noted in relation to social development, better quality predicting more advanced social skills (Cost, Quality & Child Outcome Study, 1995) and less shyness in later peer relationships (Vandell, Henderson & Wilson, 1988).

The significance of good quality care has long been debated in relation to the equally important issue of cost, enabling families who need child care to find a service they can afford (Ruopp & Travers, 1982). There is currently discussion about the relative impact of day care quality on children's development in comparison with the influence of the home environment (NICHD, 1999; Scarr, 1999). If the cost of providing 'good quality' care is substantially greater than that of providing 'adequate care' then policy makers may question the additional funds. However, if it can be demonstrated that most care is not adequate then demands for additional funding are likely. Indeed, on the basis of investigations in several hundred randomly selected preschool classrooms in four states in the USA it was found that most child care is mediocre at best, poor enough to interfere with children's emotional and intellectual development (Cost, Quality & Outcomes Study Team, 1995), contrasting with the claims that group care is more uniformly of good quality in some European countries such as Germany (Beller et al., 1996), Norway (Borge & Melhuish, 1996), Sweden (Hagekull & Bohlin, 1995) and France (Balleyguier & Melhuish 1996). In order to place this issue in context it is important to think in detail about the constituents of quality and the varying ways that it has been assessed. If one cannot define a construct or measure it reliably then it is difficult to answer theoretical questions about its relevance.

Methodological Issues

The importance of advancing understanding about the concept of quality has been emphasised (Moss & Melhuish, 1991) and there are several important issues in relation to assessing quality. One is the content of the measure - what constitutes quality? A second is the methodology - how should one try to estimate whether these constituents are evident and for this review the particular concern is whether they can be observed? A third issue is the reason for measuring quality. Is it to answer a theoretical question about the impact of child care quality, perhaps in relation to the quality of the home, or the child's personality? Or is it so that the quality of a centre can be improved? These three issues are interlinked and interaction between them has driven methods.

The main area of contention that links the definition of quality with methodology is whether it is best to assess by direct observation or by evaluating structural aspects of the organisation such as staffing levels, educational qualifications or the salaries paid, or to combine both types of information into a comprehensive quality score. More particularly methods differ in the extent to which specific caregivers are described or global judgements are made about the overall care provided in any one centre? A child in a nursery of general good quality may be profoundly influenced by spending time with one or two insensitive caregivers, whose behaviour would be lost within an overall rating of quality.

The purpose inherent in different observational measures of quality can be grouped broadly into two contrasting approaches. One, principally from the USA, has been academic and research oriented, developing methods that can be used to test ideas about the relevance of quality for child outcomes. While the measures are multidimensional, they are often reduced in complex multiple regression models to a single quality score which is useful in dealing with the relationship between quality and child developmental outcomes, but possibly not so useful for practitioners who want to improve the quality of programmes (Cryer & Phillipsen, 1997). The second has been the need to measure quality for licensing, regulation and where necessary improvement of child care offered to the general public, either through the local authority services or as a purchased commodity.

The issue of administrators and practitioners examining quality may be even more important than that of academic researchers resolving theoretical questions in view of the findings of the USA Cost, Quality and Outcome study. There has recently been an emphasis in the UK (Munton, Mooney & Rowland, 1995; Pence & Moss, 1994) on developing a theoretical framework and consequently methods of assessing quality that can be incorporated into good practice that can inform day care providers and establish ways to improve services. This conceptualisation incorporates a 'children's rights' approach that puts into second place the theoretical research questions -- whatever is found in terms of differences in long-term outcomes young children have a right to the best quality day care that can be provided.

What is good quality?

Before looking in detail at how researchers and practitioners have assessed quality some clarification of the definition of quality is important. With good reason it has been described as a "slippery and multifaceted construct that requires careful measurement and interpretation" (Hwang, Broberg & Lamb, 1991, p.117). As in any other setting, children in child care need to be safe, healthy, responded to sensitively and appropriately, and encouraged to develop their potential. However, is it has been proposed (Pence & Moss, 1994) that quality is not an objective reality but a relative value that might vary depending on the informant, who might be a parent, a child care worker, or even a child? It has been found in the USA that parents generally rate their children's care as of high quality, giving higher ratings than independent observers (Cryer & Burchinal, 1997). The authors suggest that parents rated not according to reality but according to their hopes and desires. However, it is also possible that they have different expectations about what child care can provide from the professionals providing the care. The particular emphasis of parents was on health and safety while professionals may be more likely to stress the ways in which they can enhance children's development. It should be possible to have a universal framework so that different views can be compared with each other.

Munton et al. (1995) have identified three basic dimensions of quality: the structure, which includes material and human capital; the process, which includes the activities that constitute the care, and the outcomes or consequences of care. They suggest that in each of these three domains it should be possible to assess six dimensions - effectiveness, acceptability, efficiency, access, equity, and relevance. While this multidimensional model is useful, the positioning of outcomes within a model of quality present difficulties if these outcomes are then used to determine what impact quality has on children or their families. To include outcomes in a model of quality then makes it difficult to test the relationships between quality and children's development. The more traditional model of quality, inherent in most of the observational measures described in this chapter, incorporates only process and structure, and sometimes only process.

Bradley and Caldwell (1995a) have proposed a two-tiered model for classifying the process (acts and conditions) of caregiving. In the first tier they delineate the structural and functional aspects of the behaviour within five domains - sustenance, stimulation, support, structure, and surveillance. Structure in this model refers not to centre structure but to modulation of interactions, achieving a fit between a child's needs and inputs. They suggest that input in each of these five domains can come from one of four different sources (person, object, event, setting) through five different modalities. Their second tier describes the dynamic aspects of an interaction, specifically its intensity, reactivity and complexity. Their long-term goal is to identify types of caregiving environments that, while different in the details, have functional equivalence, so that quality of care can be planned for children with different home experiences and with different capabilities or disabilities. Theirs is an ambitious agenda, one they hope will lead to an objective system of classifying acts and conditions recognising that all children do not respond similarly to those acts

and conditions (Bradley & Caldwell, 1995b). However, it may not be readily transferable to observational assessments of the environment until more debate has taken place about issues such as defining optimal experiences or accounting for overlap between their proposed categories of interaction (McCartney & Black, 1995).

Is the effort involved in observing process worthwhile?

There is some foundation for using structural information, the regulatable aspects of child care, particularly when no other investigation is possible. Structure has more commonly been used as a measure of quality when resources are limited and costly observations are not possible since is it can be extracted from administrative records or elicited with an interview with the centre head. Studies have relied on information about the structure to infer process quality (Howes, 1988;Vandell, Henderson & Wilson, 1988). Aspects of child care such as staff qualifications, their years of experience, the highest wage paid, group size or the ratio of adults to children are known to be associated with better quality adult-child interactions, and better child outcomes (e.g Borge & Melhuish, 1995; Richman & McGuire, 1988; Scarr, Eisenberg & Deater-Deckard, 1994). Nevertheless, in a large study of 120 centres only one of six regulatable characteristics - the highest wage paid- was significantly correlated with observed quality, but did not capture enough of the variance in quality to be used as a surrogate measure of quality (Scarr, Eisenberg & Deater-Deckard, 1994). Another large study in the USA found that, as the adult: child ratio increased, so did process quality (Howes, Phillips, & Whitebook, 1992) but these were *observed* ratios, taken every 15 minutes over several hours of observation. Merely to know what the official ratio for the nursery is will not represent sufficient information since these are not routinely met in practice (McGuire & Richman, 1989; Scarr et al., 1994).

Observing process sits comfortably with the current tradition of developmental psychology, which has acknowledged that to understand family influences one needs to know about not just dyad but also triads and larger groups of children and adults. The major methodological concern when observing paid child care is how to reflect each individual carer within a centre or a carers home, each child's experience with a number of carers, but also arrive at a final estimate of quality that represents the centre or care home rather than one person's experience. The next section will describe the major observational measures that have been used to assess quality in child care, with particular reference to their strengths and weaknesses as research tools and as measures that staff in child care contexts could use to maintain or improve quality. They are divided into those that aim to describe the entire setting and those that focus on one adult within that setting, although this division is not clear-cut since some instruments attempt to accomplish both.

I. OBSERVATIONS TO ASSESS CARE SETTINGS

Early Childhood Environment Rating Scale (ECERS; Harms & Clifford, 1980)
Infant/Toddler Environment Rating Scale (ITERS; Harms, Cryer & Clifford, 1990)
Family Day Care Rating Scale (FDCRS, Harms & Clifford, 1989)

These scales are the most widely used observational measures, translated into a number of languages. They were developed to be comprehensive assessments of day care quality incorporating both process and structure, their difference being in the age of children for whom care is provided and the setting, centre or home. The ITERS is applicable for children up to 30 months while the ECERS concerns children from 30 to 60 months in groups such as nurseries and preschools. The FDCRS is applicable in family day care based in homes. Although the items are grouped into subscales these measures give one overall quality score. The ECERS has recently been revised (ECERS-R, Harms, Clifford & Cryer, 1998).

The ECERS has 37 items forming seven subscales - Personal care routines; Furnishings and display for children; Language and reasoning experiences; Fine and gross motor activities; Creative activities; Social development; and Adult needs. The ECERS-R again has seven slightly modified subscales, and 43 items. The new scales are - Space and furnishings; Personal care routines; Language-reasoning; Activities; Interaction; Program Structure; and Parents and staff. In both versions the items are rated on 7-point scales (1= inadequate, 3=minimal, 5=good, 7=excellent). Completion of the instrument usually involves 2-3 hours of observation but it is also necessary to talk to the child care providers about some aspects of routine (e.g. weekly music sessions, outings).

Face and construct validity for the ECERS were obtained through expert opinion and discriminant validity was based on the ability of items to distinguish between classrooms of varying quality as determined by trainers. Based on information from 25 centres, reliability was established through inter-rater agreement (.79 to .93) and by internal consistency of the subscales and the total scale. Subscale Cronbach alphas had some variability (.32 to .79) with a total scale alpha of ranging from 0.81 to 0.93. The test authors recommend that only the total score be used in research. However, work in the UK in 113 centres found more acceptable internal consistency for the subscales (range 0.65 to 0.91) and recommends that, with the exception of Personal care routines, subscale scores can be used (Munton, Rowland, Mooney & Lera, 1997). The total score of the ECERS can clearly differentiate between centres, with mean scores commonly ranging from barely adequate to excellent (e.g. 1.9 to 5.5; Howes & Hamilton, 1992b; 1.8 to 5.2; McCartney, 1984).

The ITERS has a similar format to the ECERS with 35 items and seven subscales, adjusted to describe the needs of younger children. Reliability, based on observations in 30 centres in North Carolina was established through inter-rater agreement (total score 0.84; subscales from 0.58 to 0.89). Again the internal consistency for the subscales varied substantially, but the total scale has a Cronbach Alpha of 0.83. Content validity was ascertained by comparing the ITERS

with other instruments and by the judgement of nationally recognised experts.

Additional validity information about both measures has been obtained by factor analysis, from which two subscales emerge (Howes, Phillips & Whitebook, 1992). The first, labelled 'Appropriate care-giving', included the items describing adult-child interactions, supervision and discipline, accounting for 52% of the variance in the ECERS, and 56% in the ITERS. The second subscale, labelled 'Developmentally appropriate activity', incorporated the procedural aspects of the curriculum such as materials, schedule and activities and accounted for the remaining variance. These subscales have been used in subsequent research and further validated (Stipek et al., 1992). Detailed observations of the instructional practices and social climate of preschool and kindergarten classes were conducted and on the basis of observations were rated as child centred, intermediate or didactic. The child centred classes had significantly higher mean scale scores than the other two types both for appropriate caregiving and appropriate activities.

Is it has been suggested that, at least for research purposes, the ITERS and ECERS have a great deal of redundancy (Scarr et al., 1994). In this study factor analysis revealed only one general factor in this study of 363 classes in 120 centres, including infants, toddlers and preschool children. Two factors could not be identified even when a forced analysis was conducted. Taking any 12 items at random a measure could be created that correlated with the total test (range 0.93 to 0.95), which has been replicated for the ECERS in a study of 30 German infant day care centres (Beller et al., 1996). Nevertheless the German study documented the relevance of three subscales. Multiple regression analyses showed that 'Language/reasoning' and 'Adult needs' predicted infant cognitive development, and 'Personal Care Routines' predicted gross motor development. They concluded that a care environment that fosters the needs of the staff and encourages communication with parents is important for children's intellectual development. This aspect may be more relevant than the superficially more pertinent aspects such as provision of books and stimulating toys, particularly if the centres being studied are in middle class communities and do not vary substantially in their layout or materials.

Most of the early measures of quality concentrated on group child care but as the research issues have broadened to look at quality, so measures have been adapted so that home environments used for child care - 'Family day care' in North America and Europe and 'Child minding' in the UK - can be assessed. The FDCRS is very similar in format to its companion measures, the ECERS and ITERS. The scale consists of 32 items covering the same six domains, with a very similar emphasis to the ITERS; fewer items than the ECERS on curriculum content and only one items on the schedule and program structure. There are 8 additional items for programmes enrolling children with disabilities. Inter-rater agreement was established (0.90 or above for subscales) and internal consistency of the subscales - space and furnishing, basic care, language and reading, learning activities, social development, and adult needs, ranged from 0.70 to 0.93 (Howes & Stewart, 1987). FDCRS ratings have been validated by comparison with home visitor's ratings and with caregiver education (Pepper & Stuart, 1985).

Quality of care assessed by the FDCRS has been shown to predict children's competence demonstrated in play (Howes & Stewart, 1987; Kontos, 1994). In addition there was a significant increase in FDCRS mean scores following a training programme (Kontos, Howes & Galinsky, 1996). A study of more than 100 childminders in the UK demonstrated its reliability in terms of inter-rater agreement but found that internal consistencies were acceptable only for Language and reading (0.89), Learning activities (0.84) and Social development (0.71) (Rowland, Munton, Mooney, 1996) with the remaining values below 0.70. Some items may not be as applicable to the UK context (e.g. if space to be alone is omitted the space and furnishings scale consistency is 0.73). However the total scale consistency was high (0.94) and the authors suggest that it might be most useful if aggregated into one score of global quality . Factor analysis confirmed this with one component accounting for 72% of the variance. While comments are often made about differences between countries in the quality of group care, examination of variations in the quality of home care may need even more thought in terms of selecting culturally appropriate measures.

Assessment Profile for Early Childhood Programs (Profile) (Abbott-Shim & Sibley, 1987)

The other major observational schedule developed in the USA is the Profile, designed to provide quantitative measures of classrooms and teaching practices for the purpose of programme regulation and improvement. It is organised in a series of levels of increasing specificity. The first level includes 5 programme components, the next level includes dimensions within each component and within each dimension is a set of standards - general statements representing values and expectations, supported by criteria which are observable procedures, behaviours, and documentation. For example under the dimension of 'Interacting' there is a component 'Provider is warm and nurturing with children' and one of the specific standards to be observed is 'Provider has a spontaneous sense of humour and engages children in laughter through verbal exchanges and/or playful activities'. The format is 'yes/no' and scores are summarised for each dimension as a percentage 'yes' to facilitate interpretation.

Scores that fall in the 90-100% range are considered exceptional, 70-89% is considered average and below 70% the programme would be considered weak. The dimensions and their content have been developed through consultation with early child care experts in terms of their face validity. There are a number of version to cover infants (0-2 years), preschool children (2-5 years) and older school children (6-10 years). The different versions have slightly different emphasis. The Infant (127 items) set covers safety and health, nutrition, learning environment, interacting between teacher and children, and individualising. The young toddler version has 152 items and the older toddler 151 items. The five areas covered are learning environment, scheduling, curriculum, interacting and individualising. A version has also been created for children cared for in homes.

A study conducted in centres from three states, including infants, toddlers and

preschoolers examined the construct validity of the dimensions (Scarr et al., 1994). Conducting factor analysis with items from the five sub-scales common to all versions - learning environment, scheduling, curriculum, interacting and individualizing one general factor was identified. Combining all age groups four factors emerged, reduced after examination of eigenvalues to a two factor solution. The authors label the first as 'Dynamic' including interactions, curriculum and individualizing, and the second as 'Prepare', including scheduling, planning and organising of play spaces and activities. Although the factors were identified by orthogonal rotation they were correlated, 0.59, 0.55, and 0.41 in the infant, toddler and preschool versions respectively. Both factors had high internal consistency (alphas of 0.97 and 0.95). The conclusion of Scarr et al. was that a single quality factor could be used for research purposes and that the length of the instrument could be radically reduced.

Research versions, slightly shorter than the original forms, are now available but factor analytic investigation indicates that the most useful construct may be overall quality (Scarr et al., 1994). While four factors initially emerged in two analyses using split-half versions of the new scale, explaining more variance than one general factor, two factors (the dynamic and prepare constructs identified in the longer version) explained the maximum amount of variance.

The Profile correlates highly with the ITERS and ECERS (Petrogiannis & Melhuish, 19 , Scarr et al., 1994) and with the CCFS (Petrogiannis & Melhuish, 1996). Thus it may be up to the researcher to decide which instrument has the better face validity for their particular purposes. The format of the Profile may be simple to train staff to use, with the yes/no ratings. While statistically it might be most effective to use one general factor, it will facilitate the work of observers to divide the aspects of the caregiving environment into conceptually distinct area such as safety, interactions and so forth. Researchers may indeed decide to select items themselves on the basis of specific content and its relevance to outcomes. McCartney et al. (1997) used only those items describing proximal aspects of the teacher-child interaction, which they found to be significantly associate with children's social approaches, although not to their behaviour problems. The NICHD study (1996; 1997a) also used an abbreviated form, taking items from two scales - safety and health and learning environment to create a measure of the physical environment, which had acceptable internal consistency. This observation of the environment correlated significantly with observations of specific children and adults in both home settings and centres (NICHD, 1996)

Child Care Facility Schedule (CCFS) (WHO, 1990)

There is not necessarily agreement either within or between cultures about which aspects of human development are the most desirable, or about the best ways to enhance any aspect of development (Moss & Melhuish, 1990). In view of the value laden nature of the concept of quality and the heavy emphasis in previous measures on the aspects of quality determined as important in the USA, the Division of Mental Health of the World Health Organisation (WHO) initiated the development of a new instrument for assessing quality in child care, one that could be used in a

broad range of developing and developed countries.

Pilot work was conducted in a number of countries including Greece, Nigeria, the Phillippines (Tsianis, Demenaga & Lambidi, 1988) and an 80-item schedule was developed, containing eight subscales: Physical environment, Health and safety, Nutrition and food services, Administration, Staff-family interactions, Staff-children interactions, Observable child behaviour, and Curriculum. Items are rated on a four point scale which was thought to be easier to complete than the 7-point scale used in the ECERS. Observers need approximately 2 hours in a centre to complete the schedule. Inter-rater agreement was established in Nigeria and Greece, with agreement from 0.91 to 0.94 (Tsiantis et al., 1991). Concurrent validity was established in 12 centres by comparing CCFS scores with ratings based on observation by a trained independent observer who spent one full day in each of 12 centres and then made ratings from one to four in each of the eight areas. Criterion validity was ascertained in 23 day care centres in Athens, comparing the CCFS with the ECERS (Dragonas, Tsiantis& Lambidi, 1995).However, factor analysis identified only two factors, one of which accounted for the majority of the variance and included items form all eight areas. These two factors include 43 of the original 80 items and the reduced scale has been recommended as a Short form CCFS, which correlates 0.76 with the ECERS and 0.77 with observer's ratings.

A comparison of private day care centres, state-run and those belonging to public welfare was conducted in Athens, using both the 80 item and the 43 item versions. (Dragonas, Tsiantis & Lambidi, 1995). The private centres had the highest quality, significantly greater than state run centres, with each version. Looking at the subscales, the Physical environment and Administration distinguished the private centres. They had good coordination and responsiveness between staff members and a lower staff:child ratio, while those run by public welfare had higher Health and safety.

The authors conclude that the CCFS provides a more balanced description of day care than the ECERS, despite their close correlation. In particular the relationship between the day care setting and the community are examined, child behaviour is noted and the administration is given greater emphasis. However, the inclusion of child outcomes in the measure will lead to difficulties if the total CCFS is used as a research instrument to predict child behaviour. This measure has not been used widely as yet but appears to be a useful addition to the range of observation schedules, written with cultural diversity uppermost rather than with the researcher's interests. Nevertheless, it has a strong relationship with existing measures such as the ITERS and the Profile suggesting that all these measures assess similar aspects of the day car environment (Petrogiannis & Melhuish, 1996) although the ITERS in fact provided the most differentiation between centres.

Group Day Care Observation Checklist (Mooney, Munton, Rowland & McGurk, 1997)

Developed in Britain, this instrument is uniquely designed for use by day care staff either to monitor or assess quality in their own group care centres leading to

improvement in day care provision (Mooney et al., 1997). They constructed the items on the basis of an inclusionary conceptual framework of quality that incorporates structure, process, and outcomes (Munton et al., 1995) but focussing in particular on process. A companion questionnaire addresses structure. The observational instrument lists 16 items concerned with adult-child interactions (e.g. Do adults talk positively to children?), coded on a 5-point scale after 5 minutes of observation. An additional 7 items are completed if any incident of distress or discipline is observed (e.g. Do adults encourage children to talk about their feelings and frustrations?). When ten 5-minute observations have been completed, including at least one mealtime and one observation of personal care routines a further 11 general items are completed (e.g. Do children get an opportunity for individual attention from adults?). A second part of the observation schedule asks questions about the indoor and outdoor environments, with some items specifically for centres with infants and toddlers.

It has been piloted in 3 local authorities in England to refine the content. Observers made ratings at the same time as staff and reliability was acceptable, although there was some variability. Looking at inter-rater agreement observers rated more stringently than staff members. Internal consistency for the 16 interactional items is good and the sum total of these core questions, completed ten times, correlated significantly with the total ECERS score completed by an independent observer.

More information is needed before any conclusions can be made about this instrument but its focus on providing a way for care providers to conduct observations is to be commended. However, it will face the difficulty of other global measures in that judgements have to be combined across several staff members, interacting with several children.

Child Care HOME Inventory (Caldwell & Bradley, 1991)

As the research questions about day care have been extended there are more studies of children in home day care. The FDCRS provides a strategy for observing child care in the home but there are many advantages to using an instrument that has been widely used to assess maternal child care. Thus a child care version of the HOME has been developed and was used in the NICHD study (1996). In most ways it is identical to the original HOME inventory (Caldwell & Bradley, 1984) with only minor changes to the wording. As with the HOME the Child Care HOME employs an interview and observation format to assess the caregiver's responsivity, acceptance, and involvement with the child, the organisation and learning materials in the home environment, and the variety of experienced offered to the child. The 45 items are rated 0 or 1 (present/absent) and the total scale has a Cronbach alpha of 0.81. It correlated significantly with observations of detailed caregiver behaviour assessed using the ORCE (see below, NICHD, 1996).

Quality of Day Care Rating (Hagekull & Bohlin, 1995)

Some European researchers have suggested that instruments such as the ECERS are more suitable for centres in the USA, but that in for instance Sweden the overall level of provision is such that there would not be sufficient variation between ratings, given the strict regulations (Hagekull & Bohlin, 1995). Instead, using the HOME as a basis, they developed an observational schedule looking at structure, stimulation, emotional tone, which was then reduced to a single five point rating. The low end (1) is a messy environment with few routines, little space, and negative behaviours between children while the high end (5) is given to an environment "highly adapted to the child's needs" with adequate space, play materials and discipline. Surprisingly, however, this rather gross measure was a significant predictor of children's current (29 month) externalising behaviour problems, positive emotions, and subsequent (4 year) internalising social withdrawal.

II. OBSERVATIONS TO DESCRIBE SPECIFIC CAREGIVERS.
Caregiver Interaction Scale (CIS, Arnett, 1989).

The Caregiver Interaction Scale (CIS) (Arnett, 1989) is based on Baumrind's theoretical model of socialisation which suggested that both warmth and the level of punitiveness and restriction will have important implications for children's development (Maccoby & Martin, 1983). The scale was designed to be completed as a global rating after at least 45 minutes of observation and the content was established during observations in a number of Head Start centres (Arnett, 1989). It has 26 items, each rated on a 4 point scale indicating how much the statement is characteristic of the caregiver (1=not at all to 4= very much). Four subscales - Positive interaction (10 items), Punitiveness (8 items), Permissiveness (4 items), and Detachment (4 items) - were derived from a principal components factor analysis. Nevertheless, subsequent investigators have suggested that a single factor score is a better indicator of the data, with a Cronbach alpha for the total scale of 0.93 (Peisner-Feinberg & Burchinal, 1997). Scores are usually expressed as the mean item score per subscale and inter-rater agreement is generally high (.89 to .98.) It was designed with staff training in mind and as able to distinguish caregivers with different levels of training. Those with more training were more positive and less punitive, detached and permissive than those with little or no training.

Subsequently in a study of day care centres for infants in Athens, Greece it was found that the Positive interaction score was significantly related to the Program Structure scale of the ITERS but not overall nursery quality, While punitiveness could be predicted by the total ITERS score and that adult:child ratio predicted both punitiveness and permissiveness (Petrogiannis & Melhuish, 1996).

Adult Involvement Scale (AIS, Howes & Stewart, 1987)

This is a rating scale that, like the CIS, is used to assess individual caregivers. It was developed from a scale devised in a study looking at factors influencing the development of play with toys and peers (Howes & Stewart, 1987) Its focus is on the intensity of the adult-child relationship and has six levels: (0)

ignores (1) routine caregiving making no verbal responses to the child; (2) minimal caregiving, with some verbal practical interaction, (3) answering social bids but not elaborating or extending, (4) extending and elaborating, and (5) intense caregiving with hugging and holding, playing and/or prolonged conversations. It is commonly completed after detailed time sampled observations. For example, Howes and Hamilton (1992a) observed for 2 hours, completing 20 second intervals for 5 minutes at a time. Three 5-minute observations of two target children in each group were conducted in the Cost, Quality and Outcomes Study (Peisner-Feinberg & Burchinal, 1997).

If the target child was deemed to be in proximity (3 feet from the caregiver) the intensity was rated. The percent of intervals receiving each code from 1 to 5 are used in data analysis. Clearly this scale confounds physical, affectionate contact with verbal stimulation and there might be difficulties in competing the coded if an adult hugged and comforted a child without much conversation. However, good inter-rater agreement has been reported (0.90 and above) and validity has been demonstrated by differences in intensity depending on attachment to the caregiver (Howes & Hamilton, 1992a). Staff caring for securely attached children, assessed using the Q-Set (Waters & Deane, 1985), were rated more intense and responsive than those caring for children who were either ambivalent or insecure. The AIS correlates significantly with other aspects of quality, measured but the ECERS and the CIS, and predictive validity is suggested by its significant association with children's language development and pre-academic skills (Peisner-Feinberg & Burchinal, 1997).

The AIS was used in an evaluation of a training program for family day care providers but failed to identify changes in the AIS scores although there were identifiable differences in the global FDCRS scale (Kontos, Howes, and Galinsky , 1996). The authors conclude that global aspects of care are more responsive to change than adult responsiveness, which would involve a more intensive and extended training programme.

Caregiver Sensitivity (Ainsworth, Bell & Stayton, 1974)

Goossens & Melhuish (1996) used as nine point sensitivity scale originally developed for studying maternal behaviour to show that sensitivity demonstrated by paid child care providers may be situation specific. The scale is rooted in attachment theory and based on four aspects of behaviour - awareness of signals, accurate interpretation of them, appropriate response and prompt response, and each level of the scale is defined in detail. Differences were found between sensitivity ratings of caregivers made after observation of free play in a laboratory, and a clean-up task and sensitivity observed in the caregiver's place of work. This reinforces the notion that observations should be made of both the context and the caregiver. This may be particularly important when evaluating the quality of the care of infants.

Interaction Style Measure (McGuire & Richman, 1989)

In a study of how local authority nurseries managed children with behavioural

problems, observers complete ratings of staff behaviour on ten dimensions after spending approximately 20 days making detailed observations of 79 children aged 2 to 4 years in six different centres. The staff rating (6-point scales) had previously been developed to indicate change in parenting behaviour in a clinical treatment setting (Oxby, 1984) and included ten different aspects of adult behaviour including tone of voice, the extent and nature of verbal and playful stimulation, and strategies used for discipline. When nurseries were ranked according to the mean staff ratings combining variability of tone, verbal stimulation, playful stimulation and use of praise there was a significant correlation with rankings based on specific child experiences observed during time-sampling, with targeted children as the focus (child in engaged play one-to one with an adult or in a group led by an adult). Positive verbal stimulation was also significantly associated with smaller group size (McGuire & Richman, 1989; Richman & McGuire, 1988).

III. OBSERVATIONS OF SPECIFIC CHILDREN

Detailed Observations. (Melhuish, Mooney, Martin & Lloyd, 1990)

Observations of individual children's experiences using frequency counts of predefined behaviours are often more adaptable as a means of assessing the nature and quality of child care, particularly when comparing home and group settings. In addition they do not have an inherent judgement about what is acceptable which means that observers' own expectations about what should happen may have less impact on the results. However, they usually require more complex statistical treatment in order to create summary scores. Melhuish et al. (1990) observed 159 18-month-olds cared for either at home, with a relative, with a childminder or in a nursery setting using time sampling. The behavioural catalogue was developed on the basis of observations of maternal behaviour (Clarke-Stewart, 1973; Dunn & Kendrick, 1980) using a similar continuous recording with a lined note-pad to show each 10 second interval. Some behaviours are momentary (e.g. kiss, give) and others extend over time (e.g. hold) and two (group activity, joint activity) show the extend of social involvement with adults. The target child was observed during free play for one hour. Rates per hour could be calculated for all the brief behaviours and contingency analysis was used to determine the responsiveness of the environment to the target child's communication, within the same or the next 10 second interval. Summary scores were calculated for Responsiveness, Affection and emotional behaviour, Communication to the child, Attention, and Contact.

It was found that several aspect of adult behaviour distinguished children in nurseries from the other types of care including less responsiveness to children, less affection, and fewer language utterances. Importantly, the study also identified differences in types of care based in the home. Children with childminders experienced less language than those at home with their mothers.

Observation du Lieu de Vie de l'Enfant (OLiVE) (Pierrehumbert, Ramstein, Krucher, El-Najar, Lamb & Halfon, 1996)

Research groups have been concerned about the ethnocentric, primarily USA

values inherent in the available measures of quality and the emphasis on the values of researchers and educators. Pierrehumbert et al. (1996) have created a new instrument suitable for observing infant care in either the home or in a centre. This measure differs from those described up to now that it focuses on one specific child's experience. The aim was to develop an instrument that was based more closely on theory, that was multi-dimensional and relevant to children from a range of social backgrounds. The instrument is relevant to children from infancy up to 5 year olds, cared for in a range of different settings, and is observational with an emphasis on adult-child interactions. In addition the authors wanted an instrument that described "characteristics" of child care, a less judgemental concept than "quality", arguing that different stakeholder groups may hold different value systems and expect different characteristics.

Starting with 500 items based on existing scales, semi-structured interviews with parents, brainstorming sessions with experts, and reviewing the literature, these were reduced through consultation with experts to 230 and then using factor analysis to 70 items in 8 factors. After developing a companion instrument to examine expectations (Cartes pour les Modèles Internes Éducatifs, CaMIE) the 70 items were reduced to 43 items covering seven factors common to both instruments: Attentiveness, Emulation (Encouragement and giving confidence), Firmness, Warmth, Autonomy, Stimulation, and Safety.

Observations are based on the experiences of a target child. There are three kinds of item in the OLiVE: 14 items describe brief events identified as present or absent (0/1) during fifteen 1-minute time samples. More extended interactive behaviours (17 items) are rated as present or absent in three 5-minute observations and scored 5 each time they occur to give a possible total from 0-15. Other stable aspects of the environment (12 items) are rated once at the end of approximately 90 minutes and scored 15 points each for their occurrence. Subscale scores can then be calculated, giving a mean item score for each scale ranging from 0-15 and, if required, a total item mean. Reliability has been shown by internal consistency, which is acceptable for all scales except autonomy. Inter-rater agreement is good, and test-retest observations were consistent for 4 of the 7 scales. Validity has been established by comparing the OLiVE with the ECERS and with the detailed observations used by Melhuish et al. (1990). The OLiVE correlated most strongly with the Adult Needs scale from the ECERS and Responsivity from the Melhuish behavioural catalogue.

The measure has been used with 106 children observed in their homes and in a child care setting. Apart from attentiveness the observations in homes were significantly correlated with those in child care suggesting that parents seek out care that will match their own behaviour. However, comparing the standardised scores child care staff were more attentive than mothers and the environment safer, while mothers displayed more encouragement towards competence and more warmth. Comparing with the companion CaMIE it was found that in neither home nor care setting did expectations correlate with observed behaviour. They suggest that people can be divided into those who do what they say is important and those who may do quite the opposite so that these two types cancel each other in the overall

correlations, which merits further investigation.

The observations can be used to construct a profile, using standardised scores for each scale, so that strengths and weaknesses can be identified. More work will be required before it is clear if this scale offers a uniquely different outcome from existing measures. However, face validity is important and European researchers or administrator concerned with regulating quality may be more comfortable with a measure that is developed in Europe. the companion questionnaire may also be useful, to resolve the issue of whether people behave in accordance with their stated beliefs.

IV. OBSERVATIONS OF SPECIFIC CHILDREN AND CAREGIVERS

Observational Record of the Caregiving Environment (ORCE) (NICHD Early Child Care Research Network 1996)

Comparisons of the effect of different types of care are sometimes complicated by the fact that different instruments have been designed to suit each type of care, such as the ECERS, ITERS and FDCRS or the different versions of the Profile, and between whether the child or the adult is the target of attention. To enable comparisons between different types of care and to incorporate both child and adult the NICHD study developed the Observational Record of the Caregiving Environment (ORCE). Its particular features are that it provides a measure of quality of care that focuses more precisely on a particular child's experiences, not on what happens in the classroom at large, but it also includes a rating of the caregiver. The method provides a record of specific acts while the qualitative ratings assess the nuances of the caregiver's behaviour in relation to the child.

The ORCE is based on 44 minute cycles, broken into 10 minute time sampling segments when the observer alternates between 30 seconds of observation of a target child and 30 seconds of completing the frequency checklist. After each 10 minutes notes are made so that at the end of the full 44 minutes the qualitative ratings are made. While the frequency count includes behaviour of any adult who interacts with the child the qualitative rating is made for the caregiver who has spent the most time with the child during the time sampling. In the NICHD study there were typically four such 44 minute sessions spread over 2 days. The observation schedule includes 17 adult behaviours, the setting of the activity, the child's activity, and the nature of the child's interaction with other children (positive or negative). The qualitative caregiver ratings are made on 8 characteristics rated from 1 (not at all characteristic) to 4 (very characteristic). The scales are: Sensitivity/responsiveness to distress; Sensitivity/responsiveness to non-distress; Intrusiveness; Detachment/disengagement; Stimulation of development; Positive regard for child; Negative regard for child; and Flatness of affect.

Reliability for observer training is required at or above 70% agreement for the behaviour scales and at 60% agreement or better for the qualitative ratings although the authors report that agreement was typically much higher than those minimal standards (NICHD, 1996). For the purposes of analysis combined scales have been

reported - Positive caregiving includes 9 specific behaviour codes: shared positive affect, positive physical contact, responsive to vocalisations, facilitates infant behaviour, stimulates cognitive development, asks questions, other talk, and reads, with a Cronbach alpha of 0.81. This scale was confirmed by factor analysis, with one factor accounting for 49% of the variance. A Positive caregiving composite has also been calculated using five of the eight qualitative ratings: Sensitivity to distress and to non-distress, Positive regard, Stimulation of cognitive development, Detachment, and Flatness of affect, with a Cronbach alpha of 0.89. This composite was confirmed by factor analysis by one factor accounting for 70% of the variance. The composite based on time sampling was significantly correlated with that derived from ratings (0.74). Validity has been further established by correlations with the child care HOME (Caldwell & Bradley, 1991), significant both for the ratings (0.63) and the behaviour composite (0.40). Multiple regression has indicated that both measures of positive caregiving are negatively predicted by group size, and child-adult ratio, and positively predicted by the caregivers non-authoritarian beliefs (NICHD, 1996).

Quality of care at 15, 24 and 36 months, indicated by the composite Positive caregiving rating, and by the frequency of language stimulation identified during time sampling, were both significant predictors of children's cognitive and language development at 36 months (Clarke-Stewart, 1999) and of their school readiness at 54 months (NICHD, 1999), although quality of care was not associated with attachment security (NICHD, 1997).

Conclusions

There are a number of observational strategies available for assessing the nature of paid child care. The importance of observations is emphasised by a recent study of the 'climate' in Swedish day care centres (Ekholm, Hedin & Andersson, 1995). Three different group climates were identified on the basis of observations: 'future focussed', more overtly cooperative and facilitative with children; 'present-focused', more authoritarian and controlling; and mixed focus. However, as with the French findings (Pierrehumbert et al., 1996) the staff's attitudes to their work (defined on the basis of questionnaire responses as Relaxed, Strained or Mixed) did not bear any relationship to observed behaviour. The authors propose that the climate as it relates to working conditions may not be strongly associated with the climate as it relates to child-rearing behaviours. This may be the case in centres with a basic level of provision that is acceptable, while it may not be so in countries with centres that vary more in quality and in staff training. Most measures developed in the USA, such as the ECERS, find that adult needs are important in assessing the overall quality of child care. The findings of Ekholm et al. (1995) highlight the necessity of observing what adults do rather than asking them.

The number of measures available may be disconcerting but it can be seen that one needs to think carefully about the purpose for observations before selecting methods. Most measures of quality in day care attempt to encompass all aspects of quality - the environment, the materials, ways they are used, the manner in which the staff behave and the behaviour of the children. However, to answer theoretical

questions about the influence of care quality on child outcomes it may be possible to create a global quality score from a small number of items taken from a larger observational schedule. To investigate over time why some children progress more than others the measure will probably need to focus more on the child's unique experience. To develop a staff training strategy it will, in contrast, be important to look at the behaviour of specific caregivers and track them over time. Measures of individuals such as the AIS or CIS may be particularly useful for managers to identify particular staff members in need of support and guidance or to evaluate the effectiveness of staff training since is it concentrates on the individual. The rating could also be used to make a complete picture of the range of behaviour demonstrated by adults in any centre, to supplement the global information from the ECERS/ITERS/FDCRS range of instruments.

Whatever the method used there will be questions about the values inherent in the development of the measure and the possibility of ethnocentricity. This is important if observations are to be conducted to make value judgements about care characteristics. Many European investigators have commented on the need to develop measures that reflect an orientation of child care that is cross cultural and not necessarily the North American ideal. The range of instruments being developed in Europe, such as the OLiVE and the CCFS, should enable cross-cultural research to move ahead more productively. Most of the observational methods described have excellent reliability. Construct validity is less clear for many measures and the issues are reminiscent of the ongoing debates about intelligence. For example can child care quality be represented by one global construct or should a multidimensional model be used? Similarly, what is the nature of good quality and does this remain the same for different cultures, or classes? The methods available should enable these and many other interesting questions to be clarified in the future.

Appendix

Below is an alphabetical listing of measures reviewed. This chapter has not attempted to cover every instrument developed to measure quality but brief details are given in the list below of other observational measures, not mentioned in the text. These are marked with an asterisk (*).

1. Adult Involvement Scale (AIS) (Howes & Stewart, 1987)

2. Assessment Profile for Early Childhood Programs (Profile) (Abbott-Shim & Sibley, 1987)

3. *Belsky and Walker Checklist.(Belsky & Walker, 1980; Lamb et al., 1988) A checklist of 13 positive and 7 negative events observed in 3-minute 'spot' samples. Three or four samples are collected and the numbers averaged.

4. *Caregiver-Child Interaction Scale (CCIS) (Cochran, 1977) Interactions between caregiver and children coded during five 10-minute observations. For each interaction the initiator, the developmental context (12 categories) and the technique used by the adult (14 techniques), and the success are recorded. Used with the QAS.

5. Caregiver Interaction Scale (CIS) (Arnett, 1989).

6. Caregiver Sensitivity (Ainsworth, Bell & Stayton, 1974)

7. Child Care Facility Schedule (CCFS) (WHO, 1990)

8. Child Care HOME Inventory (Caldwell & Bradley, 1991)

9. Climate of day care centre (Ekholm, Hedin & Andersson, 1995).

10. *Daycare Quality Assessment Inventory (DQAI). (Peterson & Peterson, 1986). Two days of observation required. A checklist covers equipment. Time sampling is used to assess the caregiver's involvement with children. There is also an interview component covering the curriculum and domestic arrangements.

11. Detailed Observations. (Melhuish, Mooney, Martin & Lloyd, 1990)

12. *Early Childhood Classroom Observation Scale (ECCOS) (Bredekamp 1986). 87 items are rated from 1 to 3 covering 10 dimensions of quality: staff-child interactions; curriculum, staff-parent interactions, staff training, management, staffing, physical environment, health and safety, nutrition, evaluation. Total score highly correlated with the ECERS (r=0.93, Hausfather et al., 1997)

13. Early Childhood Environment Rating Scale (ECERS; Harms & Clifford, 1980)

14. Family Day Care Rating Scale (FDCRS, Harms & Clifford, 1989)

15. Group Day Care Observation Checklist (Mooney, Munton, Rowland & McGurk, 1997)

16. Infant/Toddler Environment Rating Scale (ITERS; Harms, Cryer & Clifford, 1990)

17. Interaction Style Measure (McGuire & Richman, 1989)

18. Observation du Lieu de Vie de l'Enfant (OLiVE) (Pierrehumbert, Ramstein, Krucher, El-Najar, Lamb & Halfon, 1996)

19. Observational Record of the Caregiving Environment (ORCE) (NICHD Early Child Care Research Network 1996)

20. *Qualitative Analysis Scale (QAS) (Cochran, 1977) Children's behaviour coded sequentially during five 10-minute observations. Twenty two categories focus on social behaviour (e.g. to cooperate) and 22 are non social (to explore). Used with the CCIS.

21. *Quality of Day-Care Environment (QDCE). (Bradley, Caldwell, Fitzgerald, Morgan & Rock, 1986). 35 items (6 point response scale) assess 5 programme areas: building and grounds; organization of time and activities; materials and equipment; traits of caregiver; expectations of child.

22. Quality of Day Care Rating (Hagekull & Bohlin, 1995)

23. *Verbal Environment. (McCartney, 1984). Eight randomly chosen children are observed for six 10-minute segments. the number of functional utterances directed to children by caregivers and by peers are recorded in four categories: control, expressive, representational, social.

References

Abbott-Shim, M. & Sibley, A. (1987) Assessment Profile for Early Childhood Programs: Preschool, Infant and School Age. Atlanta, GE: Quality Assist.

Ainsworth, M.D.S., Bell, S.M. & Stayton, D.J. (1974) Infant-mother attachment and social development: Socialisation as a product of reciprocal responsiveness to signals. In M.P.M. Richards (Ed.) The integration of a child into a social world. Cambridge: Cambridge University Press. pp. 99-135.

Arnett, J. (1989) Caregivers in day-care centers: does training matter? Journal of Applied Developmental Psychology, 10, 541-552.

Balleyguier, G. & Melhuish, E.C. (1996) The relationship between infant day care and socio-emotional development with French children aged 3-4 years. European Journal of Psychology of Education, 11, 193-199.

Beller, E.K., Stanke, M., Butsz, P., Stahl, W. & Wessels, H. (1996) Two measures of the quality of group care for infants and toddlers. European Journal of Psychology of Education, 11, 151-167.

Belsky, J. & Walker, A. (1980) Infant-toddler center spot observation system. Unpublished manuscript. Pennsylvania State University, Department of Individual and Family Studies, University Park.

Borge, A.I.H. , Hartmann, E. & Strøm, S. (1996) The Norwegian perspective on issues of quality in day care. European Journal of Psychology of Education, 11, 129-137.

Borge, A.I.H & Melhuish, E.C. (1995) A longitudinal study of childhood behaviour problems. Maternal employment, and day care in a rural Norwegian community. International Journal of Behavioral Development, 18, 23-42.

Bradley, R.H. & Caldwell, B. (1995 a) Caregiving and the regulation of child growth and development: describing proximal aspects of the caregiving systems. Developmental Review, 15, 38-85.

Bradley, R.H. & Caldwell, B. (1995 b) The acts and conditions of the caregiving environment Developmental Review, 15, 92-96.

Bradley, R.H., Caldwell, B.M., Fitzgerald, J.A., Morgan, A.G. & Rock, S.L (1986). Experiences in day care and social competence among maltreated children. Child Abuse and Neglect, 10, 181-189.

Bredekamp, S. (1986) The reliability and validity of the Early Childhood Classroom Observation Scale for accrediting early childhood programs. Early Childhood research Quarterly, 1, 103-108.

Burchinal, M.R., Ramey, S.L., Reid, M.K. & Jaccard, J. (1995) Early child care experiences and their association with family and child characteristics during middle childhood. Early Childhood Research Quarterly, 10, 33-61.

Burchinal, M.R., Roberts, J.E., Nabors, L., & Bryant, D.M. (1996) Quality of center child care and infant cognitive and language development. Child Development, 67, 606-620.

Caldwell, B.M. & Bradley, R.H. (1984) Home Observation for Measurement of the Environment. Little Rock, Arkansas: University of Arkansas at Little Rock.

Clarke-Stewart, A. (1999) How child care relates to children's cognitive and language development. Paper presented top the AAAS Annual meeting of Science innovation Exposition, Anaheim, CA, January 23.

Cochran, M.M. (1977) A comparison of group day and family child-rearing patterns in Sweden. Child Development, 48, 702-707.

Cost, Quality & Outcomes Study Team (1995) Cost, Quality and child outcomes in child care centers: key findings and recommendations. Young Children, 50, 40-44.

Cryer, D. & Burchinal, M.R. (1997) Parents as child care consumers. Early Childhood Research Quarterly, 12, 35-58.

Cryer D. & Phillipsen (1997) A close-up look at child care program strengths and weaknesses. Young Children, 52, 51-61.

Deater-Deckard, K., Pinkerton, R. & Scarr, S. (1996) Child care quality and children's behavioral adjustment; a four-year longitudinal study. Journal of Child Psychology and Psychiatry, 37, 937-948.

Dragonas, T., Tsiantis, J. & Lambidi, A. (1995) Assessing quality day care: the Child care Facility Schedule. International Journal of Behavioral Development, 1995, 18, 557-568.

Ekholm, B., Hedin, A. & Andersson, B.E. (1995) Climates in Swedish day care centers: a methodological study. Journal of Research in Childhood Education, 9, 97-111.

Goossens, F. & Melhuish, E.C. (1996) On the ecological validity of measuring the sensitivity of professional caregivers: the laboratory versus the nursery. European Journal of Psychology of Education, 11, 169-176.

Hagekull, B. & Bohlin, G. (1995) Day care quality, family and child characteristics and socioemotional development. Early Childhood Research Quarterly, 10, 505-526.

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

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

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

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

Hausfather, A., Toharia, A., La Roche, C. & Engelsmann, F. (1997) Effects of age of entry, day-care quality, and family characteristics on preschool behaviour. Journal of Child Psychology and Psychiatry, 38, 441-448.

Howes, C. (1988) Relations between early child care and schooling. Developmental Psychology, 24, 53-57.

Howes, C. (1990) Can the age of entry into child care and the quality of child care predict adjustment in kindergarten? Developmental Psychology, 26, 292-303.

Howes, C. & Hamilton, C.E. (1992a) Children's relationships with caregivers: mothers and child care teachers. Child Development, 63, 859-866.

Howes, C. & Hamilton, C.E. (1992b) Children's relationships with child care teachers: stability and concordance with parental attachments. Child Development, 63, 867-878.

Howes, C., Phillips D.A. & Whitebook, M. (1992) Thresholds of quality: implications for the social development of children in center-based child care. Child Development, 63, 449-460.

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.

Hwang, C. P., Broberg, A. & Lamb, M.E. (1991) Swedish childcare research. In E.C. Melhuish & P. Moss (Eds.) Day Care for Young Children. International Perspectives. London: Tavistock/Routledge. pp. 102-120.

Kontos, S. (1994) The ecology of family day care. Early Childhood Research Quarterly, 6, 249-262.

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.

Lamb, M.E., Hwang, C.P., Bookstein, F.L., Broberg, A., Hult, G. & Frodi, M. (1988)

Determinants of social competence in Swedish preschoolers. Developmental Psychology, 24, 58-70.

Maccoby, E.E. & Martin, J. (1983) Socialization in the context of the family: parent-child interaction. In P.H. Mussen (Ed.), Handbook of Child Development, Vol. IV. New York: John Wiley & Sons. pp. 1-101.

McCartney, K. (1984) Effect of quality of day care environment on Children's language development. Developmental Psychology, 20, 244-260.

McCartney & Black (1995) In search of transactional processes in the caregiving system. Developmental Review, 15, 86-91.

McCartney, K., Scarr, S., Rocheleau, A., Phillips, D., Abbott-Shim, M., Eisenberg, M., Keefe, N., Rosenthal, S. & Ruh, J. (1997) Teacher-child interaction and child-care auspices as predictors of social outcomes in infants, toddlers, and preschoolers. Merrill-Palmer Quarterly, 43, 426-450.

McGuire, J. & Richman, N. (1989) Management of behaviour problems in day nurseries. Early Child Development and Care, 45, Special Issue.

Melhuish, E.C., Mooney, A., Martin, S, & Lloyd, E. (1990). Type of childcare at 18 months-I. Differences in interactional experience. Journal of Child Psychology and Psychiatry, 31, 849-859.

Mooney, A., Munton, A.G., Rowland, L. & McGurk, H. (1997) The development of materials to assess quality in group and family day care. Early Child Development & Care, 129, 27-42.

Moss, P. & Melhuish, E.C. (1991) Future directions for day care policy and research. In P. Moss & E.C. Melhuish (Eds.) Current issues in day car for young children. London: HMSO. pp.131-136.

Moss, P & Pence, A. (Eds.) Valuing quality in early childhood services: new approaches to defining quality. London: Paul Chapman Publishing Ltd.

Munton, A.G., Mooney, A. & Rowland, L. (1995) Deconstructing quality: A conceptual framework for the new paradigm in day care provision for the under eights. Early Child Development & Care, 114, 11-23.

Munton, A.G., Rowland, L., Mooney, A. & Lera, M. (1997) Using the Early childhood environment rating scales (ECERS) to evaluate quality of nursery provision in England: some data concerning reliability. Educational Research, 39, 99-104.

NICHD (1996) Characteristics of infant care: factors contributing to positive caregiving. Early Childhood Research Quarterly, 11, 269-306.

NICHD (1997a) Familial factors associated with characteristics of nonmaternal care

for infants. Journal of Marriage and the Family, 59, 389-408.

NICHD (1997b) The effects of infant child care on infant-mother attachment security: results of the NICHD Study of Early Child Care. Child Development, 68, 860-879.

NICHD (1999) Effect sizes from the NICHD Study of Early Child Care. Paper presented at the Biennial meeting of the Society for Research in Child Development. Albuquerque, New Mexico, April 17.

Oxby, L. (1984) An investigation of parenting behaviours in a child psychiatric setting. Unpublished MSc thesis, University of London Institute of Education.

Pence, A. & Moss, P. (1994) Towards an inclusionary approach in defining quality. In P. Moss & A. Pence (Eds.) Valuing quality in early childhood services: new approaches to defining quality. London: Paul Chapman Publishing Ltd.

Peisner-Feinberg E.S. & Burchinal M.R. (1997) Relationships between preschool children's child-care experiences and concurrent development: the Cost, Quality and Outcomes Study. Merrill-Palmer Quarterly, 43, 451-477.

Pepper, S. & Stuart, B. (1985) Informal family day care: a study of caregivers. Ontario Mental Health Foundation.

Peterson, C. & Peterson, R. (1986) Parent-child interaction and daycare: does quality of daycare matter? Journal of Applied Developmental Psychology, 7, 1-15.

Petrogiannis, K. & Melhuish, E.C. (1996) Aspects of quality in Greek day care centres. European journal of Psychology of Education, 11, 177-191.

Pierrehumbert, B., Ramstein, T., Krucher, R., El-Najar, S., Lamb, M.E.& Halfon,O. (1996) L'évaluation du lieu de vie du jeune enfant: Dévelopement d'un instrument. Bulletin de Psychologie, XLIX, no. 426, 565-584.

Ramey, C.T., Bryant, D.M., & Suarez, T. (1989) Preschool compensatory education and the modifiability of intelligence: a critical review. In D. Detterman (Ed.), Current topics in human intelligence Norwood, NJ: Ablex Publishing.

Richman, N, & McGuire, J. (1988) Institutional characteristics and staff behaviour in day nurseries. Children & Society, 2, 138-151.

Rowland, L., Munton A.G. & Mooney A. (1996). Can quality of family day care provision in England be assessed accurately using the Family Day Care Rating Scale? a report on reliability. Educational Psychology, 16, 329-334.

Ruopp, R.R. & Travers, J. (1982) Janus faced day care: perspectives on quality and cost. In E. Zigler & E. Gordon (Eds.) Day Care: Scientific and social policy issues. Boston: Auburn House. pp 72-101.

Scarr, S. (1999) What can we tell working parents? Unpublished paper. University of Virginia.

Scarr S. & Eisenberg, M. (1993) Child care research: issues, perspectives, and results. Annual review of Psychology, 44, 613-644.

Scarr, S., Eisenberg, M. & Deater-Deckard, K. (1994) Measurement of quality in child care centers. Early Childhood Research Quarterly, 9, 131-151.

Stipek, D., Daniels, D., Galluzzo, D. & Milburn, S. (1992) Characterizing early childhood education programs for poor and middle-class children. Early Childhood Research Quarterly, 7, 1-19.

Tsiantis, J., Demenaga, N, & Lambidi, A. (1988) Preliminary observations from the evaluation of the WHO Child Care Facility Schedule in Athenian day-care centres. Iatriki, 54, 57-62. (in Greek).

Tsiantis, J., Caldwell, B., Dragonas, T., Jegede, R.O., Lambidi, A., Banaag, C. & Orley, J. (1991) Development of a WHO Child Care Facility Schedule (CCFS): a pilot collaborative study. Bulletin of the World Health Organisation, 69, 51-57.

Vandell, D.L., Henderson, V.K. & Wilson, K.S.. (1988) A longitudinal study of children with day-care experience of varying quality. Child Development, 59, 1286-1292.

Waters, E. & Deane, K.E. (1985) Defining and assessing individual differences in attachment relationships. In I. Bretherton & E. Waters (Eds.), Growing points of attachment theory and research (pp. 41-65). Monographs of the Society for research in Child development, 50(1, Serial No. 209).

World Health Organisation (1990) WHO Child Care Facility Schedule with User's Manual. Geneva: WHO Division of Mental Health.