HANDBOOK OF
EARLY CHILDHOOD
INTERVENTION

SECOND EDITION

Edited by
JACK P. SHONKOFF
Brandeis University

SAMUEL J. MEISELS
The University of Michigan

2000

CHAPTER TWENTY-TWO

An Expanded View of Program Evaluation in Early Childhood Intervention

PENNY HAUSER-CRAM, MARJI ERICKSON WARFIELD, CAROLE C. UPSHUR, AND THOMAS S. WEISNER

Beginning with the first quantitative evaluations of the Head Start program (e.g., Westinghouse Learning Corporation, 1969), there have been lively discussions about the usefulness and limitations of quantitative evaluations of early childhood programs. Questions have emerged about the types of outcomes and magnitude of change that can be expected from a wide range of interventions and the value of such outcomes to children, parents, schools, and the society as a whole (Center for the Future of Children, 1995; Farran, 1990; Hamburg, 1994). Conflicting views and beliefs about the efficacy of early childhood interventions have made it necessary to broaden evaluations so that they provide meaningful information to the necessary audiences.

This challenge to expand evaluation studies has emerged from fundamental questions about 1) the nature of developmental change, 2) the inviolability of the traditional scientific paradigm, and 3) the dichotomy between quantitative and qualitative methodologies. Simultaneously, evaluations have benefited from a growing sophistication about how systems operate within political contexts and how programs themselves evolve over time. In this chapter, we explore each of these areas with a view toward expanding the scope of early childhood intervention program evaluations through the incorporation of multimethod approaches.

BELIEFS ABOUT DEVELOPMENTAL CHANGE

Program evaluators, similar to other researchers, are guided by their theoretical and metatheoretical perspectives and assumptions about the nature and process of developmental change (Lerner, Hauser-Cram, & Miller, 1998). Beliefs about key principles of development are likely to affect the perspective of the evaluation team and lead to decisions about the research design and methodology employed. In particular, evaluators' beliefs about four critical aspects of developmental functioning are likely to affect evaluation designs: 1) the extent to which mechanistic or constructivist theories adequately describe children's learning, 2) the degree to which the acquisition of cognitive processes is universal, 3) the extent to which plasticity occurs in developmental functioning, and 4) the degree to which development is embedded in a social and cultural context. We maintain that while seldom made explicit, these beliefs affect the way in which researchers conceptualize evaluations of early intervention (EI) and other early childhood programs.

Models of Learning

Different views of the process of children's learning have long existed in educational psychology. Mechanistic models tend to be elementalistic and

emphasize quantitative accretion of skills (Lerner & Tubman, 1989). Such models are well represented, especially in special education settings (e.g., Snell, 1987), by behavioral approaches (e.g., stimulus, response, and reinforcement) proposed by Skinner (1953) and applied to child development by Bijou and Baer (1966). In contrast, constructivist models tend to be broad in scope and focus on changes in the way children think about and understand phenomena (Lerner & Tubman, 1989). Such models guide early childhood settings that center on developmentally appropriate practice (Bredekamp & Copple, 1997). These models have different implications for the types of outcomes evaluators may consider appropriate. For example, evaluators working from the mechanistic tradition may choose to focus on specific skill acquisition, such as children's ability to count a series of objects. In contrast, those operating from the constructivist viewpoint might choose outcomes that illustrate how children "make meaning" in their school environment, such as the type of social play in which children engage.

Universality of Development

A longstanding debate in developmental psychology focuses on questions of the universality of the development of cognitive processes. Much of this debate has centered around the notion of invariant stages of development proposed by Piaget (1952) and whether the proposed stages are acquired in all cultures (Dasen & Heron, 1981; Jahoda, 1980). From this perspective, the concept of universality is viewed as a question of whether a phenomenon, such as the acquisition of conservation of mass, occurs within all cultures.

A different but equally important view of universality involves our understanding of the development of children with disabilities (Hodapp & Burack, 1990). In this regard, there is considerable debate, first generated in the late 1960s, about whether the development of children with mental retardation is structurally different or delayed but organized similarly to that of other children (Cicchetti & Beeghly, 1990; Weisz, Yeates, & Zigler, 1982; Zigler, 1969). Evaluators operating from a perspective that children with disabilities, such as mental retardation, are developmentally delayed tend to focus on the selection of different outcome measures from those

who consider the development of such children to be qualitatively different from that of other children. For example, the former may select measures assessing the organization and structure of development, such as Piagetian-based measures of cognition (e.g., Cicchetti & Pogge-Hesse, 1982), whereas the latter may select measures that are seldom used to evaluate typically developing children, such as measures of perseveration (Sandson & Albert, 1984).

Plasticity of Developmental Function

Another central and long-term debate in psychology focuses on questions about the plasticity of development. Historically, this discussion has engendered investigations on both biological and behavioral indicators of change in developmental trajectory (Lerner, 1984). Studies on the capacity of the central nervous system to recover function after insult, damage, or complications associated with premature birth have indicated the extent to which the organism can modify or adapt behavior. For example, investigations of children's recovery of function after head injury have been optimistic, depending on the age of the child and the extent of the injury (Spreen, Risser, & Edgell, 1995). Furthermore, studies of children born prematurely indicate that those from middle-income families display signs of "catching up" developmentally with chronological age peers by around 2 years of age (e.g., Greenberg & Crnic, 1988). Evaluators who believe that early development is quite malleable might choose to focus on such aspects of development as social development, in which great variability has been shown to occur (Damon, 1983). In contrast, those who subscribe to a view of early development as resistant to outside influence might choose to focus on psychomotor skills, in which maturational functions appear to outweigh experiential effects during the early years of life (McCall, 1987).

Sociocultural Context

In addition to these three issues, a fourth has emerged, however, that has critical implications for evaluating intervention programs: the question of how extensively development (and developmental change) is embedded within a social and cultural context. Throughout the 1980s and 1990s,

developmental psychologists increasingly recognized the importance of the multiple systems in which a child develops. Bronfenbrenner (1979), using an ecological perspective, maintained that development occurs within multiple contexts, from proximal interactions within the family to distal policies established by governmental agencies. Sameroff (1995) emphasized the interplay (i.e., transactions) among various forces in development in his description of the transactional model. Lerner (1991) advanced a bidirectional and transforming model, developmental contextualism, in which children both affect and are affected by the multiple and changing systems in which they develop.

One implication of the systems perspective of either Bronfenbrenner (1979) or Lerner (1991) is that any intervention program is just one of the many influences on children's development and may be best understood as expressing its effectiveness in interaction with other systems. Although some programs may have direct influences on children through individual services, they are more likely to have indirect effects on children through the microsystem of the family or the mesosystem of the family–program relationship (McCartney & Howley, 1991). From a systems perspective, viewing an intervention or early childhood program as the sole or most direct influence on development would be overly simplistic. Thus, researchers operating within a systems perspective would construct evaluation designs that incorporate the web of relations among various systems in which the child, family, and program operate. Weiss and Greene (1992) described several examples of program evaluations that reflect the complex relations among family support programs and the sociocultural milieu in which they are embedded. Understanding the effectiveness of such programs requires an understanding of the milieu, because change in one aspect of the system may involve changes in other aspects as well (Hetherington & Baltes, 1988). Guralnick (1997) has termed the refocusing of evaluation studies on this type of contextual understanding as *second generation* research.

In contrast, evaluators who view a program as having a discrete influence on children's development tend to approach evaluations from an experimental, rather than ecological, framework. Casto and White (1993) described a series of studies of EI undertaken to determine the effectiveness of specific program components. For example, Boyce, White, and Kerr (1993) studied the effect of an additional component of service, parent meetings, on children and families across two sites. Although this program component was found to have no direct effect on child or family outcomes, its relation to the multiple systems in which children and families develop was not a critical feature of the investigation. Thus, although no main effect was evident, the possibility of contextual effects was not explored.

The value of understanding culture as an integral part of the context in which children are nurtured has achieved increasing importance in current views of development (Rogoff & Morelli, 1989). This emphasis in part reflects the growing variability in the families served by many early childhood programs. Increasing numbers of Latino and Asian populations, as well as social and economic changes in communities (e.g., single parenthood and chronic unemployment; Hanson & Carta, 1995), have resulted in evolving and multidimensional child-rearing patterns that need to be understood by those who want to provide services to young children and their families.

Several researchers have proposed an ecocultural framework for understanding how development occurs within the activity settings of everyday life (Super & Harkness, 1986; Tharp & Gallimore, 1988; Weisner, 1984, 1996b; Weisner, Gallimore, & Jordan, 1988). They have stressed the importance of understanding parents' goals and aspirations for their children and how these are revealed through the activities in which children and parents engage. For example, Reese, Balzano, Gallimore, and Goldenberg (1991) described the determination of low-income parents of Mexican American children to keep their child on the *buen camino* (good path) and described some of the ways in which parents directed children's activities to achieve this goal. Some investigators have found that ecocultural factors (such as maternal and paternal workload, integration into both disabled and nondisabled social networks, or parents' information seeking about services for children with disabilities) predict children's developmental status above and beyond more traditional measures of the home environment (Nihira, Weisner, & Bernheimer, 1994).

At a macrolevel, the sociocultural context specifies the values and beliefs about what is important

and worthwhile. Institutions, such as the family, serve to organize behaviors around those values and beliefs (Gordon & Armour-Thomas, 1991). At a microlevel, the sociocultural context mediates the processes of teaching and learning through social interaction (Rogoff, 1990). Such interactions occur most frequently for young children within the family or neighborhood, although children in child care, early education, or early intervention programs also have these as additional settings in which teaching and learning occurs.

What are the implications for designing program evaluations if we view development from the perspective of the importance of the sociocultural context? First, evaluators need to consider the community in which the program is situated. This would mean understanding how a program is viewed within that culture (Nieto, 1992). Cultural views of the meaning of disabilities and the value attributed to such services as early intervention vary considerably (Lynch & Hanson, 1992). For example, families may regard services as a benefit, burden, or stigma. Next, the belief systems and values espoused by parents and other community members would be considered important. Sigel, Stinson, and Flaugher (1991) described how parents vary in their beliefs regarding how children learn and become socialized. Some parents believe that children learn best through direct instruction, whereas others believe that children learn best by figuring out solutions on their own. Such different beliefs have implications for how parents might work with service providers in structuring activities for children. Sociocultural discontinuities between children's home and school settings have been explored in the educational literature (e.g., Heath, 1982) and also require consideration by those evaluating any social program targeting children. Finally, evaluators need to consider the teaching–learning transactions that occur among children, parents, service providers, and other significant adults or siblings in the child's life.

Although evaluation of intervention programs has never been a simple task, evaluators who operate from a framework that incorporates the ecological and sociocultural systems that form the context of development take on an even more complex, yet in our view, necessary, task. They must attempt to understand the effect and meaning of a program within the cultural niches in which children and families

live (Weisner, Matheson, & Bernheimer, 1996). Their selection of measures must be true to the program's goals as well as to an understanding of parental aspirations and community culture. What is studied and how it is studied needs to integrate a range of perspectives. The assumption of "mainstream" – largely White, middle-class – goals and aspirations for all children and families does not hold in all contexts.

In summary, as theoretical models of human development attempt to incorporate the cultural embeddedness of human lives, the task of program evaluation becomes increasingly complex. Evaluators need to examine their own assumptions about development – as well as the assumptions of those implementing an intervention, delivering services, and participating in a program – if they are to design evaluations that will produce valid and useful information.

The move toward more contextual models of development within the child development community raises critical questions about the fundamental nature of designing evaluations. If we believe that development and change are embedded in context, the standard models of evaluation appear inadequate. Fundamental to those models are assumptions about the nature of scientific inquiry.

BELIEFS ABOUT THE NATURE OF SCIENTIFIC INQUIRY

The majority of evaluations of early childhood programs use traditional models of scientific inquiry, including experimental, quasi-experimental, and correlational designs. Although valuable information has accrued on the effectiveness of early childhood intervention programs in general, the methods of scientific inquiry used to generate that information are increasingly questioned by the public and policy makers. There is a growing skepticism about the truth of science and a concomitant belief that the positivist paradigm (i.e., traditional experimental design) is inadequate to reflect current reality (Humphries & Truman, 1994; Lincoln, 1994; Meenaghan & Kilty, 1994). The debate about the usefulness of the scientific–experimental method has been rooted in the empowerment of undervalued groups (Fetterman, 1994); discussions about contextual metatheories (Overton, 1998); new discoveries in the physical sciences that call into question

whether it is possible to conduct uncontaminated experiments (Meenaghan & Kilty, 1994; Shadish, 1995); and a recognition of value biases in all evaluation and research (Fischer, 1995; Lincoln & Guba, 1989).

The traditional primacy of the experimental paradigm, especially when entitlement to services has been established in most states or where ethical reasons dictate that services cannot be denied to eligible individuals, renders the conduct of evaluations of early childhood programs quite challenging. Random assignment to different treatment conditions for the purposes of experimental control is the basis of the scientific paradigm. Once held as sacrosanct, there is a growing movement within science to acknowledge the limitations of this feature of the traditional positivist paradigm.

A positivist framework for evaluation of early childhood intervention programs suggests that it is possible through scientific methods to determine definitively what works and does not work in intervention strategies, what effect sizes (standardized differences between intervention and nonintervention groups) are obtainable, and thus whether a program is worth continuing. This notion of objective truth, however, has been criticized by ongoing debates in the philosophy of science (Boyd, 1984; Dolby, 1979; Knorr-Cetina, 1984). What has come to be known as a postpositivist paradigm is now at the cutting edge of scientific discourse (Overton, 1998). The postpositivist paradigm acknowledges that science is not value free; that knowledge is socially constructed and constrained by history, culture, and time; and that there are multiple possible views of truth. In essence, the postpositivists argue that context and values are important in conducting and understanding scientific inquiry and are no less important in contributing to an understanding of the phenomenon under study (Fischer & Forester, 1987; Habermas, 1973).

Within the postpositivist framework, essential questions include for whom an evaluation is being conducted and for what purpose it is taking place. As Fischer (1995) pointed out, an intervention program can be evaluated at many different levels. At the first level, verification, the traditional scientific–technical paradigm, may be used to address the question of whether the program fulfills its stated objectives. For example, are the services

being delivered appropriately to the intended population and are proposed impacts being achieved? At the second level, validation, the question as to whether the objectives themselves are useful or valid is addressed. That is, do the specific objectives meet the identified need or problem? They may be fulfilled perfectly well but miss the mark in terms of what is most useful for the target population. A good example of this level of analysis comes from the debate about whether changing children's IQ scores or helping them be better prepared for the challenging tasks of the school environment is more important in promoting lasting school achievement among children with disabilities (Hauser-Cram & Shonkoff, 1995).

At the third level, one must examine the social vindication of the program. Does it have value for the participants, other stakeholders, and their collective community? Does it have value for the society as a whole? Does it promote commonly agreed upon values? In the case of early childhood education, strong arguments can be made in support of the broad and perhaps unmeasurable benefits gained by including children with disabilities in normative activities and by providing respite and support for parents, regardless of other measurable benefits. Is it enough that parents value the services and rate them positively (Upshur, 1991)?

Finally, programs can be viewed within a social choice framework, in terms of whether they support the broadly competing goals of equity, liberty, and community. To what extent does the program contribute to a vision of a civil society, and what precise vision is being supported? The debate here focuses on utilities and rights: whether, for example, disproportionate expenditures and services for children with disabilities compromise resources available for other children. Analysis at this level often engenders a theoretical debate but importantly links the mundane considerations of program documentation to considerations of the world view that certain interventions imply.

Fischer's (1995) framework suggests that there are many types of inquiry that are relevant to the evaluation of programs and policies. Several new models of evaluation practice have recently been proposed that address the concerns of the postpositivists, as well as acknowledge the fundamental values questions raised by Fischer's framework. Three examples

are a constructivist paradigm of evaluation, an empowerment evaluation, and a participatory evaluation. Guba and Lincoln (1989) argued for a constructivist paradigm of evaluation. Their approach suggests that the role of an evaluator is to identify carefully the goals, needs, and issues of staff, managers, funders, policy makers, and recipients and to produce information for all of these groups with the goal of obtaining consensus and agreement on what is happening in the program. The constructivist paradigm has direct applicability in early childhood programming because the field has moved away from professionally centered to parent-centered models of service delivery (Dunst, Trivette, & Deal, 1988; Friedman, 1996). For example, parents' goals for their children and parental assessments of success in achieving those goals may be as important a component of evaluating programs as using standardized developmental assessments. Parents may rate a program as highly successful if it helps a distractible child increase his or her attention span, even though a standardized assessment might not show measurable progress. In summary, the constructivist model acknowledges that there are various perspectives on how well a program may be performing. Each perspective is a valid and important viewpoint that should be included in the evaluation.

Similar and related evaluation strategies to the framework proposed by Guba and Lincoln (1989) are those of empowerment evaluation (Fetterman, 1994; Fetterman, Kaftarian, & Wandersman, 1995; Weiss & Greene, 1992) and participatory evaluation (Cousins & Earl, 1992; Upshur & Baretto-Cortez, 1995). Empowerment models seek to involve recipients in conducting their own evaluations. The essential focus is to give voice to those who are intended to be changed by the intervention so that they can identify their needs and the ways in which the intervention is or is not meeting those needs. Criteria for success are generated by the recipients, not outside researchers. Empowerment models have been slow to come to the field of disabilities but have been central to the family support movement (Weiss & Greene, 1992). Philosophically, empowerment evaluation recognizes the inherent ability of recipients to help themselves and to shape programs and resources to meet their own needs. Fundamentally, empowerment evaluation discounts the deficit

model for children and families who utilize early childhood intervention services.

Participatory evaluation is related to empowerment evaluation and has explicit roots in the work of Paolo Freire (Fals-Borda & Rahman, 1991) and W. F. Whyte (1991). Participatory evaluation is grounded in the experience of staff and recipients of services and seeks to provide practical, formative, and useful information to improve program outcomes. Like empowerment evaluation, participatory evaluation also results in the empowerment of participants but seeks to derive activities not just from the evaluation experience but through the development of a critical consciousness that explicitly deals with the issues of power and knowledge through the implementation of the evaluation findings. Participatory evaluation, similar to empowerment evaluation, values the views of the recipients of services more than those of funders or outside policy makers.

These new evaluation paradigms represent a variety of approaches that seek to address the inherent limitations of the traditional experimental designs and their derivatives in program evaluation. Another essential feature of these new approaches, however, is a rethinking of the presumed distinctions between quantitative and qualitative methodologies.

BELIEFS ABOUT THE DISTINCTION BETWEEN QUALITATIVE AND QUANTITATIVE METHODS OF INQUIRY

In current discourse, quantitative and qualitative methodologies are considered paired opposites. Qualitative research, in its everyday academic folk model, is typically opposed to its presumed opposite – quantitative research. Naturalistic research is paired with its presumed opposite – experimental research. Contextual, cultural, or comparative research is contrasted with its assumed opposite – monocultural work (which is, impossibly, somehow culture-free). In contrast, we propose a more useful way of conceptualizing these methods, one that assumes complementarity (Weisner, 1996a, 1997).

Qualitative, or holistic, research can more usefully be contrasted with *particularistic*, or specifically focused, research. Qualitative refers to capturing more of the whole of a phenomenon, in context, including the meanings and interpretations of actors.

Particularistic methods select out for analytical reasons, or for reasons of time, money, design, and so forth, specific aspects of a phenomenon, intentionally assessed out of context, in order to better understand that particular aspect of them. Quantitative is not the opposite of qualitative but has to do with the level of measurement available or appropriate for a study. Quantitative levels of measurement (ordinal, interval, and ratio) could be more accurately and usefully contrasted with *nominal* or categorical levels of measurement. For example, a community might divide the seasons into dry and rainy, or summer, fall, winter, and spring, or recognize a single diety, or call dogs and cats and some goldfish pets. These are all cultural categories with meaning and significance that are not inherently ranked or scaled.

One epistemological approach that requires complementary methods is outlined by Shweder (1996), who proposed that a useful and powerful assumption to make about the world is that it consists of "quanta" and "qualia." Quanta are objects, events, and processes in the world minus the subjective; they are things that have power independent of our experience and awareness of them. Examples of quanta include demographic circumstances (e.g., fertility rates, and migration cycles), resource availability (e.g., climate and ecology), social processes (e.g., attachment and self and other appraisal), cognitive or mental abilities (e.g., memory processes and neurological maturation), historical and path-dependent conditions (the prior existence in a community of clans, economic classes and castes, and trading networks, for instance, which, although social conditions with meaning and significance, are also in part fixed features from the past, shaping inevitably the way things are done in the present). In contrast, qualia are things that can only be understood with or through the subjective, that is, by what they mean, signify, or imply to persons in a particular place; how they are experienced, remembered, and enacted. Quanta typically are studied by using procedures such as pointing, counting, measuring, sampling, and calculating; qualia typically are studied by using empathy, interpretation, thematization/emplotment, narration, contextualization, and exemplification. Quanta and qualia are useful analytic distinctions, but they are not pure. For example, demographic and resource conditions (such as population density, gross domestic product,

and wealth and income distributions) are structural constraints but are always understood through our interpretations of them and actions in response to them. Qualia are influenced by social regularities and lawful general conditions, even if those regularities and conditions are not directly known to or experienced by actors. For example, memory processes and the varying kinds of memory constrain what we can recall and how we recall events, even though we are not directly aware of these processes. Public health conditions (e.g., lower infection rate due to better sanitation) influence how we protect children and what we worry about regarding their health and safety, but we may not be consciously aware of the public health environment. Indeed, we may take much of this environment for granted as if it were "natural." Hence, complementary methods are needed to represent and understand a world filled with both quanta and qualia.

Exclusive use of quantitative measurement can inadvertently close off understanding of phenomena vital to evaluation even before a study gets underway. Hence, we believe that complementary quantitative and qualitative methods should be the default standard (i.e., what is normally and routinely expected) for the design of evaluations of early childhood programs.

Paradoxically, then, the design of meaningful *quantitative* evaluations of early childhood programs may require methods other than those conventionally viewed as the traditional quantitative measurement strategies (i.e., questionnaires, tests and assessments, surveys, observer ratings, precoded behavioral observations, and the like). Not that these quantitative measures are not useful in evaluation work; to the contrary, they are essential. However, quantitative methods alone are unlikely to be sufficient without complementary methods that are effective in understanding subjective experience, context, representation, meaning, and interpretation.

Although perspectivistic, participatory, and empowerment models may highlight the importance of subjective experience, the general point applies to all evaluation models to varying degrees. Perspectivist, participatory, and empowerment models for evaluation have a common theme: the views of certain stakeholders matter most. Stakeholders include all individuals and groups who are affected, either

directly or indirectly, by the implementation and results of an evaluation (i.e., staff, managers, funders, policy makers, and recipients; Rossi & Freeman, 1993). The perspectivist, participatory, and empowerment models take most seriously the point of view of the participants in a program, compared to conventional evaluation designs and measures that have been motivated more often by the concerns of policy makers and funders. What do program participants experience? What cultural beliefs do staff and parents use regarding domains such as gender, ethnicity, or competence? What is the meaning and moral significance of program goals and personal involvement? How do different participants interpret the activities of a program? How reflexive and critical are staff and participants about aspects of the program? How ambivalent are they about some elements? These kinds of questions involve understanding at least three fundamental features of any social inquiry, including evaluations of early childhood interventions: 1) the subjective experiences of the participants in a program; 2) the local contexts in which these experiences occur; and 3) the representation of the program and experience in the form of cultural models and scripts (i.e., patterned shared schemas about why and how the world works) held in the mind of participants and others (D'Andrade & Strauss, 1992; Harkness & Super, 1996; Holland & Quinn, 1987).

Qualitative methods are holistic and include the qualia of events and experience. These can include any text, interview or interview transcript, vignette, or story or experience that is either told to a researcher or observed. Ethnographic methods are more comprehensive in topic as well as methods. Ethnographic methods include participant and nonparticipant observation, a variety of kinds of interviews (casual, informal, probing, structured, etc.), as well as the use of more formal methods such as network sampling, use of archival materials, and community analysis. An ethnographer attempts to capture the life of an individual, family, and community in context – to understand the social and cultural world surrounding the particular family or individuals or intervention program one is studying. Qualitative and ethnographic methods are especially useful in understanding these subjective experiences, contexts, and representations. Hence, it is no accident that qualitative and ethnographic methods are more

often associated with perspectivist and related evaluation models than are purely quantitative methods.

A similar point regarding these evaluation models holds for the use of naturalistic contexts for performing evaluations. Naturalistic studies, conducted in the everyday context of our lives, contrast with research that is in some way contrived by the researcher or others. Experimental work, which attempts to infer cause, is usefully contrasted with correlational studies, which attempt to discover coherence, relations, and patterns. Both qualitative and quantitative methods can include contrived, manipulated interventions that attempt to infer cause, and both include the analysis of patterns and relationships for which cause can perhaps be inferred but not directly.

Both quantitative and qualitative measures have a context. The most useful contrast is between studies in which there are context-examined procedures and those with little or no careful examination of the context in which the procedures were conducted. Questionnaires completed by a particular group have a context that is usually unexamined. It is rarely known in any depth what was going on in the mind of the informant or participant when he or she circled a number on a page or said yes to a predetermined question. Ethnographic studies have a visible context, usually much more carefully considered but much less contrived or controllable than other methods. All studies have an implicit comparative frame of reference of some sort – a meaning in a context relevant to some cultural place, whether for the purpose of cultural comparison or not. In this sense, all studies have an *ethnographic* component embedded in them, even if ethnography was not done or even considered by the researcher (Weisner, 1996a).

Complementarity has also been proposed by evaluators as one type of "mixed-methods" design. Caracelli and Greene (1997) described several types of mixed-method approaches to evaluation that attempt to integrate quantitative and qualitative methods to produce comprehensive evaluation plans. These styles include the following: triangulation, complementarity, development, initiation, and expansion (Greene, Caracelli, & Graham, 1989). Although only a few evaluations have been conducted using this framework (e.g., Mark, Feller, & Button, 1997), the concepts suggest useful new

evalu-
thods.
mod-
:ts for
, con-
ntrast
by the
ich at-
th cor-
coher-
e and
nanip-
e, and
lation-
ed but

s have
n stud-
edures
nation
e con-
ticular
nined.
ing on
: when
res to a
s have
ly con-
le than
ompar-
ng in a
her for
In this
onent
as not
eisner,

y eval-
design.
l types
n that
litative
luation
riangu-
iation,
1989).
n con-
ller, &
l new

strategies that should be explored in future evaluations of early childhood programs.

Triangulation involves identifying how results from different methodologies converge to produce similar findings. For example, in-depth, open-ended interviews with parents may indicate that they find some of the physical tasks of caring for a child with a motor impairment the most taxing of all child care tasks, whereas a more standardized instrument of caretaking burden may also indicate greater levels of impact related to this domain of parenting.

Complementarity of methods, as defined by Greene et al. (1989), describes designs in which one method is used to enhance or clarify the results of another. For example, suppose results of a quantitative analysis indicated that participation in a parent support group was associated with decreases in stress for some parents and increases for others. Observations of parent support group sessions combined with focus groups conducted with samples of families who experienced different stress outcomes may elucidate that finding.

Development designs are constructed so that different methods are used sequentially. For example, suppose an evaluator was interested in conducting participant observations of a range of home visits conducted by early intervention professionals who serve children with a variety of disabilities and their families. Quantitative data previously gathered on various aspects of the home visit sessions, such as the length of the session, the type of service provided (e.g., primarily counseling to the parent or predominantly therapy to the child), and the number of family members involved in the visit, could be analyzed so that a representative sample of visits could be selected for observation.

A different approach to the use of mixed methods is constructed when the two different methods are used to discover contradictions and paradoxical findings that generate hypotheses to be tested in future evaluations. Termed *initiation designs* (Greene et al., 1989), these are intended to use one method to provoke questions of data collected with a different methodology. For example, suppose parents report in interviews that greater learning occurs for their child with home-based services, but analyses of data gathered by means of standardized performance measures indicate greater increases with center-based services. Several hypotheses could be

generated from such a finding: parents might have greater opportunity to observe their child's learning during home visits, parents may themselves benefit from the interaction during home visits and believe that it is beneficial to the child as well, or both.

Finally, *expansion designs* (Greene et al., 1989) occur when either methodology is used to expand the breadth and depth of inquiry. As Caracelli and Greene (1993) pointed out, quantitative evaluations are usually the principal method for examining outcomes, whereas qualitative approaches examine processes. An evaluation could be expanded if both approaches were used for both types of examination. For example, quantitative analyses could elucidate the program components that parents and providers believe are most valuable for promoting peer interaction and interviews could probe why they consider these to be valuable. Furthermore, observational studies of peer interaction and content analyses of children's drawings might reveal the extent to which experience in child groups during early intervention relates to more successful peer interactions in preschool. Finally, analyses of "extreme cases" (e.g., children who are quite successful in making friends or who have great difficulty in peer relationships) could be selected for in-depth study, which will lend understanding to how children negotiate such relationships.

In summary, we take the scientific tradition seriously but include holistic, nominal, contextual, comparative, context-examined, subjectively rich data and conceptions in our notion of science. Such approaches provide knowledge essential to understanding children and the programs intended to assist them. Quantitative methods emphasize certain criteria for accuracy and believability, particularly reliability and validity. However, reliability and validity criteria can also be applied usefully to qualitative data as well (Bernard, 1988; Maxwell, 1992; Miles & Huberman, 1994). These are important for science but are not the only important standards for believability. Other criteria important to science are found in qualitative work but are often absent or weak in quantitative methods (Becker, 1996). These include breadth of coverage of a problem; depth of understanding across levels (including subjective experience), time, and multiple contexts; convergence of observation and interpretations across researchers and participants; veridicality or accuracy of data in

context (i.e., how phenomena actually appear in real context and real time); and precision of observations. We contend that early childhood evaluations that are designed from the outset to meet as many of these criteria as possible will produce greater understanding and better, truer science.

THE VALUE OF UNDERSTANDING THE POLITICAL CONTEXT OF EVALUATIONS

The different assumptions about human development and approaches to scientific inquiry made by the various individuals and groups interested in social programs are critical ingredients in designing evaluations. After all, evaluations are conceived, designed, and conducted in response to a demand made by program stakeholders to make some judgment about a social program or service delivery strategy. Although stakeholders often request "unbiased" evaluations of programs, they bring subjective judgment to the evaluation process.

Stakeholders include four main groups of individuals and organizations: 1) sponsors of the program; 2) program managers and staff; 3) participants in the program; and 4) evaluators and other members of the research, policy, and academic community (Rossi & Freeman, 1993). Each of these groups has something different "at stake" in an evaluation, be it their financial resources, their reputation as effective providers, their desire for a particular service to continue to be accessible and responsive to their needs, and their career aspirations, to name only a few (Guba & Lincoln, 1981). Given that these stakeholder groups have a different relationship to the program being evaluated, they often have conflicting perspectives regarding what the objectives of an evaluation should be and how an evaluation should be conducted (Thomas & Palfrey, 1996). Therefore, the ultimate form and design of an evaluation are influenced by the political context in which it is developed (Weiss, 1975).

Stakeholders and Their Information Needs

The political dynamics surrounding an evaluation will differ depending on the origin of the demand for the evaluation, the types of stakeholders who become involved, and the nature and extent of individual stakeholder participation. The demand for an evaluation may come from a variety of different sources. Historically, evaluation practice developed in response to the needs of high-level policy makers for proof of a program's efficacy (i.e., Did it achieve its goals?) and efficiency (i.e., Did it achieve its goals at a given level of expenditure, Weiss & Greene, 1992). Over the past twenty years, however, evaluation practice has expanded to consider questions regarding program implementation and operation in addition to efficacy and efficiency and to use a greater variety of methodological approaches. This expansion has encouraged other stakeholders to demand evaluations that meet their specific information needs. To understand the political context of evaluations, we must describe the way in which each of the four stakeholder groups is affected by the program being evaluated, their information needs, and the extent to which they are able to influence the objectives of an evaluation.

SPONSORS OF THE PROGRAM. This group of stakeholders includes three distinct sets of individuals or organizations. First, policy makers and decision makers are government officials who not only decide whether to implement a program initially but also make determinations regarding how long the program will be in operation and when it should be expanded or cut back. Second, program sponsors are organizations such as private foundations that initially fund the program to be evaluated. Third, evaluation sponsors are organizations such as private foundations or the federal government that initiate and fund the evaluation. In some cases, the individuals and organizations in these groups may overlap. For example, the original Head Start program and its evaluation were funded by the federal government, but the evaluation was conducted by the Westinghouse Learning Corporation and Ohio University (Westinghouse Learning Corporation, 1969).

This group of stakeholders is often interested primarily in knowing whether the services that it is paying for are being provided to the target population, whether the services are having the desired effect, and whether the program is economically efficient. Their focus on these bottom-line concerns stems from the financial resources they have invested and

the need to be accountable to the public, in the case of elected officials, or to board members of private organizations (Usher, 1995). These stakeholders can have a considerable amount of input in the design and structure of an evaluation by tying the receipt of future funds to certain performance standards or other evaluation requirements (Thomas & Palfrey, 1996).

PROGRAM MANAGERS AND STAFF. These stakeholders are responsible for both overall program operation and the actual delivery of services to the program's participants. Although both managers and staff may view evaluation as a way potentially to improve the services they provide, their specific information needs may differ because each is accountable primarily to a somewhat different audience (Chambers, Wedel, & Rodwell, 1992). Managers and administrators are accountable to program funders and must often focus on demonstrating positive outcomes that justify the program's expenditures and its allocation of staff and resources. In general, managers are interested in evaluations that enable them to compare the efficiency of alternative approaches to service delivery and monitor the operation of the program to ensure that particular models of service are being implemented and reaching those they are designed to serve (Usher, 1995).

Program staff are accountable not only to their supervisors but are also focused on how the intervention influences the participants both positively and negatively. These "street level bureaucrats" (Lipsky, 1980) are likely to be most interested in evaluations that concentrate on understanding whether services are acceptable to participants and relevant to their needs (Thomas & Palfrey, 1996).

Program managers and staff can influence the objectives of an evaluation in a variety of ways. As the individuals with the most knowledge about the day-to-day operation of the program as well as access to program records and other documentation, they possess information that is critical to the development of an evaluation plan. Therefore, these stakeholders provide needed input regarding those aspects of the program that can be studied and the types of questions that are most important to address. In addition, program staff hold beliefs about the principles of intervention, and such beliefs are

likely to affect the process of engagement with program participants. For example, in a study of utilization rates of early intervention services, Kochanek and Buka (1996) found that service providers had stronger beliefs about family-centered principles of service than did the mothers in their program. Finally, as gatekeepers of program information, staff cooperation is vital to any evaluation effort.

PROGRAM PARTICIPANTS. Although program participants may have the most at stake in an evaluation's outcome, they have often had little influence on the direction and nature of evaluations because they are typically not well organized as a group (Rossi & Freeman, 1993) and because traditional models of evaluation denied the importance of their voice. Most evaluations that have attempted to assess participants' concerns have focused simply on measuring client satisfaction (Thomas & Palfrey, 1996).

However, newer approaches to evaluation, as described earlier, have been designed to empower participants by involving them in the evaluation process (Greene, 1988). These evaluation approaches are particularly salient for early intervention services provided under the Individuals with Disabilities Education Act (IDEA) Amendments of 1990, 1991, and 1997 (P.L. 101-476, P.L. 102-119, and P.L. 105-17, respectively) and family support services in general because these programs emphasize participant empowerment and parent–professional collaboration as a central goal (Weissbourd & Kagan, 1989).

EVALUATORS AND THE RESEARCH AND POLICY COMMUNITY. This group of stakeholders includes the evaluator(s) responsible for the design and conduct of a particular evaluation as well as other individuals and organizations concerned with public policy, social science research, and program development. As discussed in prior sections, the assumptions of evaluators about the nature of developmental change, modes, and methods of inquiry, as well as their experience and skills, influence the scope and shape of any individual evaluation plan.

In summary, just as programs operate within sociocultural contexts, evaluations themselves reflect the political context in which they are developed and conducted. The main objective of an evaluation

will be influenced by the degree to which individuals within the various stakeholder groups advocate successfully for a study that meets their particular information needs. Evaluations may be undertaken to address questions of accountability, gather information for program improvement, develop and test model programs, gain a better understanding of how programs and services influence the behavior and development of children and families, or empower participants.

THE VALUE OF UNDERSTANDING THE EVOLUTION OF A PROGRAM

Programs vary not only in terms of the sociocultural and political contexts in which they operate but also in terms of their evolution. The fact that not every program categorized with a particular label (e.g., family support program and early intervention program) is the same in its organization, operation, level of resources, and population served suggests that evaluations must be tailored according to the characteristics of individual programs.

Weiss (1988) captured the vast differences in programs labeled as *family support programs* by distinguishing the few "flagship" programs from others in the larger "fleet." Flagship programs are those that have been set up as well-funded research and demonstration models whose purpose is to test the efficacy of specific service delivery strategies. In contrast, the fleet is made up of smaller community-based programs with uncertain funding that provide a variety of often innovative services in an attempt to meet the diverse needs of local families.

Although many early childhood programs may lie somewhere on the continuum between the flagship and the fleet, this metaphor emphasizes the importance of designing evaluations that take into consideration each program's unique characteristics such as how long it has been in operation, how clearly its goals and services are defined, and its data collection capabilities (Jacobs, 1988). This information is critical in determining whether a program is ready to be evaluated.

Evaluability Assessment

One systematic way to examine evaluation readiness is through evaluability assessment. Evaluability

assessment involves gathering information in three areas: 1) purpose, 2) program, and 3) technical feasibility (Chambers et al., 1992).

ASSESSMENT OF PURPOSE. The rationale for a program evaluation emanates from the information needs of different stakeholder groups. Conflicts among these groups, however, may arise. For example, Usher (1993) described a program in which the evaluability assessment found that the delivery of services was inconsistent. Therefore, although state and local policy makers had wanted to conduct an evaluation to assess program effectiveness, discussions with program managers resulted in a decision to postpone that type of outcome study in favor of designing a more precise model of service, developing a database to track implementation information, and assessing the targeting and delivery of services. Ideally, the purpose of an evaluation is determined by this type of negotiation between groups of stakeholders (Chen, 1989). The other two areas of evaluability assessment (i.e., program and technical feasibility assessment) can inform this negotiation process and help to resolve conflict regarding the objective of an evaluation.

PROGRAM ASSESSMENT. The purpose of program assessment is to identify how well program components are defined, how clearly program goals are specified, and whether there is a discrepancy between how the program was conceived and designed and how it actually operates in the field. This phase of an evaluability study provides the "situation-specific knowledge" and qualitative judgments that are important in determining the type of evaluation to conduct (Campbell, 1987, p. 349). Four iterative steps are used to gather data that allow for a complete program assessment. First, all written documentation (e.g., authorizing legislation, grant proposals, brochures, and so forth) about the program must be reviewed. These documents will identify broadly the mandate for the program and its mission and goals.

Second, a more detailed program description must be developed by interviewing policy makers as well as program managers and administrators. Key questions focus on understanding the informant's views of the short- and long-term objectives of the

program, what mechanisms are in place to make it possible to achieve these goals, why the program's inputs and activities will cause the goals to be achieved, what kinds of information are available, and what different types of information are unavailable but necessary to assess the program's performance.

Third, a flow model can be constructed to depict visually how each component of a program and the program as a whole are thought to work based on data gathered from documents and interviews. This model, which serves as a program's "theory of change," should spell out in great detail the assumptions that link each aspect of the program in order to chart how and why the intervention will produce the expected effects (Weiss, 1995). Fourth, the accuracy of the flow model should be tested by visiting the program, observing it in operation, interviewing staff responsible for its day-to-day operation, and interviewing program participants.

The program assessment phase should generate a clear definition of the program to be evaluated and an explication of the theory or theories behind the program. Although defining the program may be obvious in some cases, it is important to identify clearly the boundaries around what will and will not be evaluated. For example, EI programs often provide a variety of different services. It is important to establish whether the evaluation will focus on one component (e.g., home visits) or on all services that the program provides to children and families directly. Furthermore, EI programs under P.L. 102-119 are mandated to coordinate services with other local providers so that the full range of needs demonstrated by eligible children and families can be met. Whether the evaluation will focus on all of the services (both inside and outside of EI) that are received by a selected sample of children and families, or on only those services provided directly to the children and families by one program, has important implications for who the stakeholders are, what contextual features of the service system should be assessed, and how the evaluation results should be interpreted.

The flow model developed during the program assessment stage should underscore the theories on which the program is based. Evaluations grounded in a theoretical framework are more likely to provide important information for program improvement (Chen & Rossi, 1989). Improvement may occur in

two ways. First, by asking policy makers, administrators, service providers, and participants to make explicit what their assumptions are regarding how the program is expected to produce the desired changes, conflicting perspectives can be revealed. Discussions to resolve these conflicts can highlight whether they result in a lack of consistency in the delivery of services and what changes should be made (Weiss, 1995). Second, by designing the evaluation to examine the extent to which the different assumptions hold, the results can identify which specific aspects of the program should be modified (Weiss, 1995).

TECHNICAL FEASIBILITY ASSESSMENT. Four main technical feasibility issues must be examined. First, the availability of information that can be used in an evaluation must be assessed. It is critical to determine the type of data the program currently collects, the accuracy and completeness of those data, how long the data have been collected, on whom the data are recorded, and how relevant the data are to the evaluation questions being asked. Second, the feasibility of using different research designs and methodologies must be assessed in light of a variety of issues such as intrusion into program processes, burden on staff and participants, sampling procedures, and the integrity of control or comparison groups.

Third, the technical expertise and computer facilities required to analyze the data that will be gathered must be assessed. Finally, the resources necessary to conduct the program evaluation must be estimated. The cost of evaluation is influenced heavily by the amount of original data that must be gathered to supplement information already maintained by the program, the scope of the evaluation plan, and the technology required to conduct the study.

In summary, evaluability assessment is a useful technique for gaining a clear understanding of the goals of a program, how its day-to-day activities are organized to reach its goals, and the extent to which the program documents its activities in a systematic way. This information serves to identify unique aspects of a particular program so that an appropriate and useful evaluation can be designed and conducted. Evaluators using this approach recognize that programs are not static but change over time and that an evaluation of a program during its

maturity should differ from one conducted during its early phases.

WHAT HAVE WE LEARNED FROM EVALUATIONS?

While we contend that it is desirable for evaluations developed within a variety of political contexts to broaden their scope beyond traditional views based on positivism and the aura of numbers and to construct meaningful ways to recognize how change is embedded within a sociocultural context, we also recognize that no single evaluation, either currently or in the future, will be able to incorporate all of the ideal features we have delineated. In this section, we present a few examples of evaluations that, while largely quantitative, have used a range of methodologies and yielded valuable findings. These illustrations demonstrate how a broad array of strategies can be used to address similar questions in the following four topic areas: 1) the entrance of participants into a program, 2) the operation of a program, 3) program effectiveness, and 4) program cost and efficiency. We also identify newer techniques and approaches that can supplement and enhance the findings generated by future evaluations.

Entrance of Participants Into a Program

One critical question that concerns program staff is whether the target population is being reached. Individuals are considered to be underserved if they are eligible for services but do not enter the system or use services to the full extent of their eligibility (Arcia, Keyes, Gallagher, & Herrick, 1993). Because it can be difficult to estimate precisely the size of the target population and therefore compare who does and does not participate, studies have employed different strategies to get an estimate of those who are likely to be underserved.

Sontag and Schacht (1993) analyzed questionnaires on service utilization completed by 536 families enrolled in early intervention programs. They investigated differences in the type of services received based on ethnicity, income, and the age of the child and found barriers to early intervention service utilization for ethnic minorities, low-income families, and children under 18 months of age. These markers for underutilization are important because

Arcia et al. (1993) found that the percentage of children under 5 years of age from ethnically diverse backgrounds and the percentage of young children from low-income families are increasing. Furthermore, many families are experiencing increasing and persistent risk conditions (Dunst & Trivette, 1997). This suggests that a portion of the target population is both increasing in size and is at risk for underutilization of early intervention services.

One way to understand how to improve recruitment of families at risk of being underserved is to use a perspectivist model to identify those aspects of service provision that attract individuals to a program. For example, Pharis and Levin (1991) interviewed thirty mothers who were at high risk for difficulties in parenting about their participation in the Clinical Infant Development Program (CIDP). Using open-ended questions, researchers asked the mothers to describe how the program worked for them and what it meant to them. Using content analysis to categorize the responses, Pharis and Levin found that the mothers especially liked program components that gave them opportunities to talk to their primary clinician and to view infant assessment sessions. In general, participants preferred the relationship aspect of the services more than the concrete assistance that they received around transportation, housing, and money management.

Future evaluations that are designed to understand how programs can increase participation among all members of their target population might benefit from involving a greater variety of stakeholders (e.g., current participants and local community leaders) in developing, implementing, and assessing various outreach strategies. These strategies might use different methods to advertise the components of a program that participants view most favorably. This type of evaluation may also elucidate what successful participation means to those whom the program is trying to reach (McNaughton, 1994).

Program Operation

Evaluations of how programs for young children and their families operate can focus on a variety of important issues. The Early Intervention Collaborative Study (EICS) is a longitudinal investigation of the development of children with disabilities and the adaptation of their families (Shonkoff,

Hauser-Cram, Krauss, & Upshur, 1992). A sample of 190 children and families was followed during their first year in early intervention programs in Massachusetts and New Hampshire, and the type and amount of services they received each month were documented.

Analyses of the services received by the EICS sample revealed that the total amount of services received per month averaged a modest seven hours. Although children and families received, on average, three hours of home visits per month, they received less than two hours per month of each of the other service types (e.g., child group, parent support group, and center-based individual services). There was, however, a tremendous amount of variability in the amount and types of services received. For example, the average quantity received per month ranged from a few minutes to twenty-one hours. The child's type of disability and level of severity of impairment contributed to the observed variability in the intensity of services received. In addition, the analysis of variation in service types revealed substantial differences in both the intensity and the combinations of services received by a given child and family on a month-to-month basis over the course of the one-year study period (Erickson, 1991).

The modest level of services received and the variability in service provision offer important insights for other evaluations of early intervention that focus on estimating service impact. Although individualization of services is one of the hallmarks of early intervention programs, it strains traditional approaches to evaluation (Powell, 1988). Questions about the type of effects that can be expected from this level of intervention for children with different characteristics as well as queries about the types and amounts of services received outside of early intervention are important to consider in light of these findings (Guralnick, 1993).

Evaluations of program operation can also focus on observing how participants react to different features of an intervention in order to identify whether the model is working as designed. Brinker (1992) described an experimental program in which public aid–supported families were randomly assigned to either an intervention group that received tangible incentives when they attended the program (e.g., individualized help with housing, food, baby formula, baby clothes, and so forth) or a comparison group that received no tangible incentives. Preliminary results indicated that there was no difference in the rate of attendance between the incentive families and the comparison participants as families in both groups attended approximately 40% to 50% of the weekly sessions. Because incentives appeared to have no effect, staff attempted to make the program more receptive to participants' needs. For example, the program staffing was expanded to include community leaders and individuals with experience conducting self-help programs for families on public aid. Expanding the intervention team increased staff knowledge about the infrastructure of resources within the community and ways in which families could stretch their limited resources. These changes resulted in greater success in establishing relationships with families and in increasing the rate of participation in the program. This approach to evaluation contains some elements of the development style of a mixed-method design in that the lack of positive results from the experimental phase motivated the staff to better understand participants' needs through outreach into the community. The knowledge gained was then put to use in the next round of implementation and evaluation.

Another important aspect of program operation that evaluations can examine concerns the factors that influence how service providers interact with families. DeGangi, Wietlisbach, Poisson, Stein, and Royeen (1994) interviewed twenty-six early intervention professionals to assess their perceptions of the effects of cultural diversity and socioeconomic status (SES) on family–professional collaboration. The interview included both closed and open-ended questions as well as two case vignettes. DeGangi et al. found that service providers responded differently to families whose cultural background varied from their own and that families from lower, in comparison to higher, income groups differed in their response to providers. These insights indicate further how varied each family's experience is in early intervention.

These three studies underscore the complexities of measuring the actual treatment received by children and families. Beyond counting the number and types of services received, treatments between individuals vary in more subtle ways, depending on the training, attitudes, and goals of the provider(s); the evolving needs of the child and family; and the

interactions among the system and its providers, the community, and the family and child (Telzrow, 1993). If the evaluation is meant ultimately to elucidate the way services affect the multiple systems in which children and families develop, careful attention to the content of actual sessions with children and families will be needed.

Empowerment models of evaluation that focus on understanding the participants' point of view about how different interventions affect them and influence their behavior may help evaluations to identify specific aspects of the treatment that need to be measured in future studies. Such models are also critical in bringing together the assets needed to sustain an effective and valued program.

Program Effectiveness

Other chapters in this volume (see Farran, Brooks-Gunn, Berlin, & Fuligni) provide extensive reviews of studies assessing the impact of services on both child and family outcomes. This section, therefore, focuses briefly on the following three questions that are central to assessing program effectiveness: 1) In what domains will the program have an effect? 2) How will the outcomes be measured? and 3) How will change be measured?

IN WHAT DOMAINS WILL THE PROGRAM HAVE AN EFFECT? The ecological model of human development suggests that programs may influence child, family, and community-level outcomes (Bronfenbrenner, 1979; Bronfenbrenner & Morris, 1998). Until recently, evaluations of programs for young children and their families concentrated on child outcomes (Shonkoff & Hauser-Cram, 1987), although studies have begun to focus on family-level effects as well (e.g., Shonkoff et al., 1992; Trivette, Dunst, Boyd, & Hamby, 1995). Flow models that underscore the theories on which a particular program is operating should be developed and consulted in the process of identifying areas in which effects may be expected (LeLaurin & Wolery, 1992).

HOW WILL THE OUTCOMES BE MEASURED? Once the areas of potential program effect have been identified, decisions must be made regarding the appropriate instruments and techniques to

use to measure each outcome. The strengths and weaknesses of different assessment strategies that may be used with infants, parents, families, and communities are described in detail in other chapters in this volume (see Meisels & Atkins-Burnett; Kelly & Barnard; Krauss; and Earls & Buka). Evaluations may further benefit from current interest in early education on authentic assessment methods, which use children's typical work and activities as a way of recognizing change. Schwartz and Olswang (1996) proposed that early childhood special education programs consider developing portfolios of children's activities, including documents such as videotapes of children's play, observational data collected on children's peer interactions using time-sampling techniques, parent reports, and teachers' notes. Meisels's (1993) Work Sampling System offers a comprehensive approach to authentic assessment for children in early childhood and primary classrooms. Although not yet well used in program evaluations, such approaches offer promise to those who demand that outcomes of evaluation efforts have meaning for the program participants.

HOW WILL CHANGE BE MEASURED? Accurately quantifying the amount of change in an outcome measure over time and being able to attribute some portion of the total change to a particular intervention or treatment have always been a challenge for evaluators of early childhood programs. A variety of approaches to measuring change and then partitioning it into its various components have been proposed and applied. These methods include norm-referenced models, indices of change, difference scores, and residual change scores. The strengths and weaknesses of these approaches have been discussed elsewhere (Hauser-Cram, 1990; Hauser-Cram & Krauss, 1991).

A relatively new and useful approach to understanding change that is increasingly being applied to longitudinal studies is hierarchical linear modeling (HLM). Bryk and Raudenbush (1992) have contended that to really understand change, data must be collected over more than two time points (see also Baltes, Reese, & Nesselroade, 1977). HLM makes maximum use of data gathered over several time points and utilizes a two-step process. First, an estimate is made of the growth trajectory and overall

rate of growth for each individual in the sample being analyzed. Second, differences in these growth rates across individuals are assessed. This analysis involves examining whether differences in change are a function of selected characteristics of the child (such as gender and type of disability), the family (such as income), or services (such as intensity).

The advantage of HLM applied to the evaluation of early intervention services is that it allows ways of testing whether development occurs in a linear or curvilinear fashion. This is in contrast to other approaches that assume a linear model. Moreover, HLM allows researchers to test whether development approaches an asymptote (i.e., levels off) at a certain point, and if so, whether this occurs for certain subgroups and not others (Burchinal, Bailey, & Snyder, 1994).

Dunst and Trivette (1994) used HLM to examine the predictors of growth in mental age for children with chromosomal abnormalities (mostly children with Down syndrome) and physical impairments (mostly children with cerebral palsy), the majority of whom were receiving EI services. Mental age was measured using the Bayley Scales of Infant Development and the Stanford–Binet Intelligence Scale. Data were gathered on the subjects between birth and 48 months of age. Each child was assessed using the Bayley or Stanford–Binet between four and five times during this period.

Three sets of variables were included in the model as predictors of growth in mental age: 1) background variables, including mothers' education, marital status, occupation, race, and number of siblings; 2) stressor-related variables, including the number of family members or relatives with mental health problems and the number of child hospitalizations; and 3) intervention variables, including age of entry into EI and length of intervention. Controlling for both the background and stressor-related variables, age of entry was a significant predictor of growth in mental age for children with chromosomal abnormalities but not for children with physical impairments. Children with chromosomal abnormalities who entered EI at younger ages demonstrated a faster rate of progress than children who entered EI at older ages.

Future evaluations of early childhood programs may also benefit from utilizing another feature of HLM: assessing accurately the influence on development of specific characteristics of a child's environment. For example, Lee and Bryk (1989) employed HLM to examine whether differences in school outcomes for children in different classrooms or school systems may be accounted for by variations in specific features of those environments.

Program Cost and Efficiency

Most cost and benefit-cost analyses of early childhood programs have employed traditional experimental or quasi-experimental research designs in which children and families are assigned, using various methods, to intervention, control, or comparison groups and are provided a predetermined and unchanging package of services for a set period of time (Barnett & Escobar, 1990). Economic evaluations of services provided by programs operating with limited resources under federal or state legislative rules and regulations, however, face unique analytic challenges. The inability to standardize service delivery and maintain a control or comparison group requires that these evaluations address somewhat different but important questions. The purpose of this section is to describe the three objectives of a cost-effectiveness study that was conducted in conjunction with the Early Intervention Collaborative Study (EICS; Shonkoff et al., 1992). As noted earlier, EICS examined the services provided by publicly supported EI programs. Each child and family received services that were tailored to meet their particular needs and be responsive to their changing circumstances.

The first objective of the cost-effectiveness study was to identify the predictors of service cost (Erickson, 1992). Most cost studies focus simply on determining the average cost per client in a particular program or state. In a system in which services are provided in accordance with individualized needs, however, this approach can mask the potentially large differences in the cost required to serve children and families who vary on a wide range of characteristics.

Erickson (1992) developed models using actual service utilization data to identify the child and family characteristics that significantly influence service cost. The findings indicated that differences in the age at which a child enters an EI program and variation in the severity of a child's disability

significantly influence cost. The estimates produced for subgroups of children who differed on these two characteristics revealed a considerable amount of variability in the expenditures associated with providing one year of EI services.

The second objective was to compare the gains made by children across the various outcome measures to a set level of resources (Warfield, 1994). By holding the resource level constant, the estimated gains made by children with different characteristics were compared to identify the domains in which the greatest amount of change had occurred. The findings indicated that children classified as having mild disabilities experienced greater improvement in adaptive behavior, whereas children with more severe disabilities made greater gains in child–mother interaction. These results underscore the ways in which children with different characteristics move toward different goals relative to a common investment of resources. This information will help service providers and family members set priorities as to the relative importance of achieving certain outcomes.

The third objective was to compare the efficiency of home-visiting versus center-based group services for subgroups defined by severity of disability and age at entry (Warfield, 1995). The service identified as most cost-effective varied by subgroup and outcome measure. For example, home visits were more cost-effective in reducing parenting stress across all of the subgroups. In terms of improvements in mother–child interaction, however, group services were more cost-effective for children entering EI at less than 1 year of age, whereas home visits were more cost-effective for children entering after 1 year of age.

Therefore, the results associated with the three objectives of the EICS cost-effectiveness study indicate that critical questions about cost and efficiency can be addressed by disaggregating findings across subgroups that differ by selected child and family characteristics. Future studies conducted on more diverse samples utilizing a broader array of outcome measures will enhance the usefulness of this technique.

THE FUTURE OF EVALUATION RESEARCH

Our knowledge of evaluation results and our approach to evaluation design have grown immensely

since the days of the first Head Start studies. We have developed methods to model growth over time and to test the effects of critical predictors of development. We have learned to recognize the various audiences who attend to evaluation results and appreciate the political context in which evaluations occur. We have realized that programs themselves change and that evaluations need to fit with their stage of evolution. We have also come to question the limits of positivism, the foundation upon which quantitative evaluations have been built. We are on the threshold of change in evaluation approaches.

The challenge now is to find ways to incorporate these lessons successfully and consistently into evaluation plans. Meeting this challenge will require that new partnerships be forged among those who have different beliefs about developmental change, those who have different information needs, and those who champion different methods of scientific inquiry. Evaluations that involve a variety of stakeholders and take advantage of the vast array of methodologies will best be able to generate necessary and meaningful information about early childhood intervention programs in the future.

REFERENCES

Arcia, E., Keyes, L., Gallagher, J. J., & Herrick, H. (1993). National portrait of sociodemographic factors associated with underutilization of services: Relevance to early intervention. *Journal of Early Intervention, 17,* 283–97.

Baltes, P. B., Reese, H. W., & Nesselroade, J. R. (1977). *Life-span developmental psychology: Introduction to research methods.* Monterey, CA: Brooks/Cole.

Barnett, W. S., & Escobar, C. M. (1990). Economic costs and benefits of early intervention. In S. J. Meisels & J. P. Shonkoff (Eds.), *Handbook of early childhood intervention* (pp. 560–82). New York: Cambridge University Press.

Becker, H. S. (1996). The epistemology of qualitative research. In R. Jessor, A. Colby, & R. Shweder (Eds.), *Ethnography and human development: Context and meaning in social inquiry* (pp. 53–71). Chicago: University of Chicago Press.

Bernard, R. H. (1988). *Research methods in cultural anthropology.* Newbury Park, CA: Sage.

Bijou, S. W., & Baer, D. M. (1966). *Child development. Vol 2: The universal stage of infancy.* New York: Appleton-Century-Crofts.

Boyce, G. C., White, K. R., & Kerr, B. (1993). The effectiveness of adding a parent involvement component to an

.dies. We
over time
rs of de-
the vari-
sults and
l evalua-
ns them-
o fit with
e to ques-
on upon
en built.
valuation

incorpo-
ntly into
ll require
hose who
l change,
eds, and
of scien-
variety of
t array of
te neces-
rly child-

H. (1993).
associated
o early in-
3–97.
R. (1977).
to research

mic costs
sels & J. P.
tervention
y Press.
itative re-
.), Ethnog-
ning in so-
f Chicago

ral anthro-

velopment.
Appleton-

: effective-
ient to an

existing center-based program for children with disabilities and their families. Early Education and Development, 4, 327–45.

Boyd, R. N. (1984). The current status of scientific realism. In J. Leplin (Ed.), Scientific realism (pp. 41–82). Berkeley, CA: University of California Press.

Bredekamp, S., & Copple, C. (1997). Developmentally appropriate practice in early childhood programs. Washington, DC: National Association for the Education of Young Children.

Brinker, R. P. (1992). Family involvement in early intervention: Accepting the unchangeable, changing the changeable, and knowing the difference. Topics in Early Childhood Special Education, 12, 307–32.

Bronfenbrenner, U. (1979). The ecology of human development: Experiments by nature and design. Cambridge, MA: Harvard University Press.

Bronfenbrenner, U., & Morris, P. A. (1998). The ecology of developmental processes. In R. M. Lerner (Ed.), Theoretical models of human development: The handbook of child psychology (Vol. I, 5th ed., pp. 993–1028). New York: Wiley.

Bryk, A. S., & Raudenbush, S. W. (1992). Hierarchical linear models for social and behavioral research: Applications and data analysis methods. Newbury Park, CA: Sage.

Burchinal, M. R., Bailey, D. B., & Snyder, P. (1994). Using growth curve analysis to evaluate child change in longitudinal investigations. Journal of Early Intervention, 18, 403–23.

Campbell, D. T. (1987). Problems for the experimenting society in the interface between evaluation and service providers. In S. L. Kagan, D. R. Powell, B. Weissbourd, & E. F. Zigler (Eds.), America's family support programs: Perspectives and prospects (pp. 345–51). New Haven, CT: Yale University Press.

Caracelli, V. J., & Greene, J. C. (1993). Data analysis strategies for mixed-method evaluation designs. Educational Evaluation and Policy Analysis, 15, 195–207.

Caracelli, V. J., & Greene, J. C. (1997). Crafting mixed-method evaluation designs. In J. C. Greene & V. J. Caracelli (Eds.), Advances in mixed-method evaluation: The challenges and benefits of integrating diverse paradigms (pp. 19–32). San Francisco: Jossey-Bass.

Casto, G., & White, K. R. (1993). Longitudinal studies of alternative types of early intervention: Rationale and design. Early Education and Development, 4, 224–37.

Center for the Future of Children. (1995). The future of children: Long-term outcomes of early childhood programs. Los Altos, CA: David and Lucile Packard Foundation.

Chambers, D. E., Wedel, K. R., & Rodwell, M. K. (1992). Evaluating social programs. Boston: Allyn & Bacon.

Chen, H. T. (1989). The conceptual framework of the theory-driven perspective. Evaluation and Program Planning, 12, 391–6.

Chen, H. T., & Rossi, P. H. (1989). Issues in the theory-driven perspective. Evaluation and Program Planning, 12, 299–306.

Cicchetti, D., & Beeghly, M. (1990). An organizational approach to the study of Down syndrome: Contributions to an integrative theory of development. In D. Cicchetti & M. Beeghly (Eds.), Children with Down syndrome: A developmental perspective (pp. 29–62). Cambridge, England: Cambridge University Press.

Cicchetti, D., & Pogge-Hesse, P. (1982). Possible contributions of the study of organically retarded persons to developmental theory. In E. Zigler & D. Balla (Eds.), Mental retardation: The developmental-difference controversy (pp. 277–318). Hillsdale, NJ: Erlbaum.

Cousins, J. B., & Earl, L. M. (1992). The case for participatory evaluation. Educational Evaluation and Policy Analysis, 14, 397–418.

Damon, W. (1983). Social and personality development: Infancy through adolescence. Cambridge, England: Cambridge University Press.

D'Andrade, R., & Strauss, C. (Eds.). (1992). Human motives and cultural models. New York: Cambridge University Press.

Dasen, P. R., & Heron, A. (1981). Cross-cultural tests of Piaget's theory. In H. Triandis & A. Heron (Eds.), Handbook of cross cultural psychology: Vol. 4. Developmental psychology (pp. 295–341). Boston: Allyn & Bacon.

DeGangi, G. A., Wietlisbach, S., Poisson, S., Stein, E., & Royeen, C. (1994). The impact of culture and socioeconomic status on family–professional collaboration: Challenges and solutions. Topics in Early Childhood Special Education, 14, 503–20.

Dolby, R. G. A. (1979). Reflections on deviant science. In R. Wallis (Ed.), On the margins of science: The social construction of rejected knowledge. Sociological Review Monograph 27 (pp. 9–47). Keele, Staffordshire, England: University of Keele.

Dunst, C. J., & Trivette, C. M. (1997). Early intervention with young children and their families. In R. Ammerman and M. Hersen (Eds.), Handbook for the prevention and treatment of children and adolescents: Intervention in real world context (pp. 157–80). New York: Wiley.

Dunst, C. J., & Trivette, C. M. (1994). Methodological considerations and strategies for studying the long-term effects of early intervention. In S. L. Friedman and H. C. Haywood (Eds.), Developmental follow-up: Concepts, domains, and methods (pp. 277–313). San Diego, CA: Academic Press.

Dunst, C. J., Trivette, C. M., & Deal, A. (1988). Enabling and empowering families. Cambridge, MA: Brookline Books.

Erickson, M. (1991). Evaluating early intervention services: A cost-effectiveness analysis. Unpublished doctoral dissertation, Brandeis University, Waltham, MA.

Erickson, M. (1992). An analysis of early intervention expenditures in Massachusetts. American Journal on Mental Retardation, 96, 716–29.

Fals-Borda, O., & Rahman, M. A. (Eds.). (1991). *Action and knowledge: Breaking the monopoly with participatory action research*. New York: Apex Press.

Farran, D. C. (1990). Effects of intervention with disadvantaged and disabled children: A decade review. In S. J. Meisels & J. P. Shonkoff (Eds.), *Handbook of early childhood intervention* (pp. 501–39). New York: Cambridge University Press.

Fetterman, D. M. (1994). Empowerment evaluation. *Evaluation Practice, 15*, 1–15.

Fetterman, D. M., Kaftarian, S. J., & Wandersman, A. (Eds.). (1995). *Empowerment evaluation: Knowledge and tools for self-assessment and accountability*. Thousand Oaks, CA: Sage.

Fischer, F. (1995). *Evaluating public policy*. Chicago: Nelson-Hall.

Fischer, F., & Forester, F. (Eds.). (1987). *Confronting values in policy analysis: The politics of criteria*. Newbury Park, CA: Sage.

Friedman, D. (1996). *Parenting in public: A study of help-giving practices to support parenting and child well-being in Massachusetts' congregate family shelters*. Unpublished doctoral dissertation, Brandeis University.

Gordon, E. W., & Armour-Thomas, E. (1991). Culture and cognitive development. In L. Okagaki & R. J. Sternberg (Eds.), *Directors of development: Influences on the development of children's thinking* (pp. 83–99). Hillsdale, NJ: Erlbaum.

Greenberg, M. T., & Crnic, K. A. (1988). Longitudinal predictors of developmental status and social interaction in premature and full-term infants at age two. *Child Development, 59*, 554–70.

Greene, J. C. (1988). Stakeholder participation and utilization in program evaluation. *Evaluation Review, 12*, 91–116.

Greene, J. C., Caracelli, V. J., & Graham, W. F. (1989). Toward a conceptual framework for mixed-method evaluation designs. *Educational Evaluation and Policy Analysis, 11*, 255–74.

Guba, E. G., & Lincoln, Y. S. (1989). *Fourth generation evaluation*. Newbury Park, CA: Sage.

Guba, E. G., & Lincoln, Y. S. (1981). *Effectiveness evaluation: Improving the usefulness of evaluation results through responsive and naturalistic approaches*. San Francisco: Jossey-Bass.

Guralnick, M. J. (1993). Second generation research on the effectiveness of early intervention. *Early Education and Development, 4*, 366–78.

Guralnick M. J. (1997). Second generation research in the field of early intervention. In M. J. Guralnick (Ed.), *The effectiveness of early intervention* (pp. 3–20). Baltimore, MD: Paul H. Brookes.

Habermas, J. (1973). *Legitimation crisis*. Boston: Beacon Press.

Hamburg, D. A. (1994). *Today's children: Creating a future for a generation in crisis*. New York: Times Books.

Hanson, M. J., & Carta, J. J. (1995). Addressing the challenges of families with multiple risks. *Exceptional Children, 62*, 201–12.

Harkness, S., & Super, C. M. (Eds.). (1996). *Parents' cultural belief systems: Their origins, expressions, and consequences*. New York: Guilford Press.

Hauser-Cram, P. (1990). Designing meaningful evaluations of early intervention services. In S. J. Meisels & J. P. Shonkoff (Eds.), *Handbook of early childhood intervention* (pp. 583–602). New York: Cambridge University Press.

Hauser-Cram, P., & Krauss, M. W. (1991). Measuring change in children and families. *Journal of Early Intervention, 15*, 288–97.

Hauser-Cram, P., & Shonkoff, J. P. (1995). Mastery motivation: Implications for intervention. In R. H. MacTurk & G. A. Morgan (Eds.), *Mastery motivation: Origins, conceptualizations, and applications* (pp. 257–72). Norwood, NJ: Ablex.

Heath, S. B. (1982). Questioning at home and at school: A comparative study. In G. Spindler (Ed.), *Doing the ethnography of schooling: Educational anthropology in action* (pp. 102–31). New York: Holt, Rinehart and Winston.

Hetherington, E. M., & Baltes, P. B. (1988). Child psychology and life-span development. In E. M. Hetherington, R. M. Lerner, & M. Perlmutter (Eds.), *Child development in life-span perspective* (pp. 1–19). Hillsdale, NJ: Erlbaum.

Hodapp, R. M., & Burack, J. A. (1990). What mental retardation teaches us about typical development: The examples of sequences, rates, and cross-domain relations. *Development and Psychopathology, 2*, 213–36.

Holland, D., & Quinn, N. (Eds.). (1987). *Cultural models in language and thought*. New York: Cambridge University Press.

Humphries, B., & Truman, C. (1994). Rethinking social research: Research in an unequal world. In B. Humphries & C. Truman (Eds.), *Rethinking social research: Antidiscriminatory approaches in research methodology* (pp. 1–21). Aldershot, England: Avebury.

Jacobs, F. H. (1988). The five-tiered approach to evaluation: Context and implementation. In H. B. Weiss & F. H. Jacobs (Eds.), *Evaluating family programs* (pp. 37–68). Hawthorne, NY: Aldine.

Jahoda, G. (1980). Theoretical and systematic approaches in cross-cultural psychology. In H. C. Triandis & W. W. Lambert (Eds.), *Handbook of cross-cultural psychology* (Vol. 1, pp. 69–142). Boston: Allyn & Bacon.

Knorr-Cetina, K. (1984). The fabrication of facts: Toward a microsociology of scientific knowledge. In N. Steher & V. Meja (Eds.), *Society and knowledge: Contemporary perspectives in the sociology of knowledge* (pp. 223–4). New Brunswick, NJ: Transaction Books.

Kochanek, T. T., & Buka, S. L. (1996). *Utilization rates of early intervention services by infants/toddlers and their families*. Rhode Island College, RI: Early Childhood Research Institute.

ee, V. E., & Bryk, A. S. (1989). A multilevel model of the social distribution of high school achievement. *Sociology of Education, 62,* 172–92.

LeLaurin, K., & Wolery, M. (1992). Research standards in early intervention: Defining, describing, and measuring the independent variable. *Journal of Early Intervention, 16,* 275–87.

Lerner, R. M. (1984). *On the nature of human plasticity* (2nd ed.). New York: Random House.

Lerner, R. M. (1991). Changing organism–context relations as the basic process of development: A developmental contextual perspective. *Developmental Psychology, 27,* 27–32.

Lerner, R. M., Hauser-Cram, P., & Miller, E. (1998). Assumptions and features of longitudinal designs. Implications for early childhood education. In B. Spodek, O. N. Saracho, & A. D. Pelligrini (Eds.), *Yearbook in early education* (Vol. 8, pp. 113–138). New York: Teachers College Press.

Lerner, R., & Tubman, J. G. (1989). Conceptual issues in studying continuity and discontinuity in personality development across life. *Journal of Personality, 57,* 343–73.

Lincoln, Y. S. (1994). Tracks toward a postmodern politics of evaluation. *Evaluation Practice, 13,* 299–309.

Lincoln, Y. S., & Guba, E. G. (1989). *Fourth generation evaluation.* Newbury Park, CA: Sage.

Lipsky, M. (1980). *Street level bureaucracy.* New York: Russell Sage.

Lynch, E. W., & Hanson, M. J. (1992). *Developing cross-cultural competence: A guide for working with young children and their families.* Baltimore, MD: Paul H. Brookes.

Mark, M. M., Feller, I., & Button, S. B. (1997). Integrating qualitative methods in a predominantly quantitative evaluation: A case study and some reflections. In J. C. Greene & V. J. Caracelli (Eds.), *Advances in mixed-method evaluation: The challenges and benefits of integrating diverse paradigms* (pp. 47–59). San Francisco: Jossey-Bass.

Maxwell, J. A. (1992). Understanding and validity in qualitative research. *Harvard Educational Review, 62,* 279–300.

McCall, R. (1987). Developmental function, individual differences, and the plasticity of intelligence. In J. Gallagher & C. Ramey (Eds.), *The malleability of children* (pp. 25–35). Baltimore, MD: Paul H. Brookes.

McCartney, K., & Howley, E. (1991). Parents as instruments of intervention in home-based preschool programs. In L. Okagaki & R. J. Sternberg (Eds.), *Directors of development: Influences on the development of children's thinking* (pp. 181–202). Hillsdale, NJ: Erlbaum.

McNaughton, D. (1994). Measuring parent satisfaction with early childhood intervention programs: Current practice, problems, and future perspectives. *Topics in Early Childhood Special Education, 14,* 26–48.

Meenaghan, T. M., & Kilty, K. M. (1994). *Policy analysis and research technology: Political and ethical considerations.* Chicago: Lyceum Books.

Meisels, S. J. (1993). Remaking classroom assessment with the Work Sampling System. *Young Children, 48,* 34–40.

Miles, M., & Huberman, A. M. (1994). *Qualitative data analysis* (2nd ed.). Thousand Oaks, CA: Sage.

Nieto, S. (1992). *Affirming diversity: The sociopolitical context of multicultural education.* New York: Longman.

Nihira, K., Weisner, T. S., & Bernheimer, L. P. (1994). Ecocultural assessment in families of children with developmental delays: Construct and concurrent validities. *American Journal on Mental Retardation, 98,* 551–66.

Overton, W. F. (1998). Developmental psychology: Philosophy, concepts, and methodology. In R. M. Lerner (Ed.), *Theoretical models of human development: The handbook of child psychology* (Vol. I, 5th ed. pp. 107–88). New York: Wiley.

Pharis, M. E., & Levin, V. S. (1991). "A person to talk to who really cared": High-risk mothers' evaluations of services in an intensive intervention research program. *Child Welfare, 70,* 307–20.

Piaget, J. (1952). *The origins of intelligence in children.* New York: International Universities Press.

Powell, D. (1988). Challenges in the design and evaluation of parent-child intervention programs. In D. R. Powell (Ed.), *Parent education as early childhood intervention: Emerging directions in theory, research and practice* (pp. 229–37). Norwood, NJ: Ablex.

Reese, L., Balzano, S., Gallimore, R., & Goldenberg, C. (1991). *The concept of "educacion": Latino family values and American schooling.* Paper presented at the annual meeting of the American Anthropological Association, Chicago.

Rogoff, B. (1990). *Apprenticeship in thinking: Cognitive development in social context.* New York: Oxford University Press.

Rogoff, B., & Morelli, G. (1989). Perspectives on children's development from cultural psychology. *American Psychologist, 44,* 343–8.

Rossi, P. H., & Freeman, H. E. (1993). *Evaluation: A systematic approach* (5th ed.). Newbury Park, CA: Sage.

Sameroff, A. J. (1995). General systems theories and developmental psychopathology. In D. Cicchetti & D. J. Cohen (Eds.), *Developmental psychopathology, Vol. I: Theory and methods* (pp. 659–95). New York: Wiley.

Sandson, J., & Albert, M. L. (1984). Varieties of perseveration. *Neuropsychologia, 22,* 715–32.

Schwartz, I. S., & Olswang, L. B. (1996). Evaluating child behavior change in natural settings: Exploring alternative strategies for data collection. *Topics in Early Childhood Special Education, 16,* 82–101.

Shadish, W. R. (1995). The quantitative–qualitative debates "DeKuhnifying" the conceptual context. *Evaluation and Program Planning, 18,* 47–9.

Shonkoff, J. P., & Hauser-Cram, P. (1987). Early intervention for disabled infants and their families: A quantitative analysis. *Pediatrics, 80,* 650–8.

Shonkoff, J. P., Hauser-Cram, P., Krauss, M. W., & Upshur, C. C. (1992). Development of infants with disabilities and their families: Implications for theory and service

delivery. *Monographs of the Society for Research in Child Development, 57*(6, Serial No. 230).

Shweder, R. A. (1996). Quanta and qualia: What is the "object" of ethnographic method? In R. Jessor, A. Colby, & R. Shweder (Eds.), *Ethnography and human development. Context and meaning in social inquiry* (pp. 175–82). Chicago: University of Chicago Press.

Sigel, I., Stinson, E. T., & Flaugher, J. (1991). Socialization of representational competence in the family: The distancing paradigm. In L. Okagaki & R. J. Sternberg (Eds.), *Directors of development: Influences on the development of children's thinking* (pp. 121–44). Hillsdale, NJ: Erlbaum.

Skinner, B. F. (1953). *Science and human behavior.* New York: Appleton-Century-Crofts.

Snell, M. E. (Ed.). (1987). *Systematic instruction of persons with severe handicaps* (3rd ed.). Columbus, OH: Merrill.

Sontag, J. C., & Schacht, R. (1993). Family diversity and patterns of service utilization in early intervention. *Journal of Early Intervention, 17,* 431–44.

Spreen, O., Risser, A. H., & Edgell, D. (1995). *Developmental neuropsychology.* New York: Oxford University Press.

Super, C. M., & Harkness, S. (1986). The developmental niche: A conceptualization at the interface of child and culture. *International Journal of Behavior Development, 9,* 1–25.

Telzrow, C. F. (1993). Commentary on comparative evaluation of early intervention alternatives. *Early Education and Development, 4,* 359–65.

Tharp, R. G., & Gallimore, R. (1988). *Rousing minds to life: Teaching, learning, and schooling in social context.* New York: Cambridge University Press.

Thomas, P., & Palfrey, C. (1996). Evaluation: Stakeholder-focused criteria. *Social Policy and Administration, 30,* 125–42.

Trivette, C. M., Dunst, C. J., Boyd, K., & Hamby, D. W. (1995). Family-oriented program models: Helpgiving practices and parental control appraisals. *Exceptional Children, 62,* 237–48.

Upshur, C. C. (1991). Mothers' and fathers' ratings of the benefits of early intervention services. *Journal of Early Intervention, 15,* 345–57.

Upshur, C., & Barreto-Cortez, E. (1995). What is participatory evaluation (PE)? What are its roots? In *The evaluation exchange: Emerging strategies in evaluating child and family services.* Cambridge, MA: Harvard Family Research Project.

Usher, C. L. (1993). *Building capacity for self-evaluation in family and children's services reform efforts.* Paper presented at the annual meeting of the American Evaluation Association, Dallas, TX.

Usher, C. L. (1995). Improving evaluability through self-evaluation. *Evaluation Practice, 16,* 59–68.

Warfield, M. E. (1995). The cost-effectiveness of home visiting versus group services in early intervention. *Journal of Early Intervention, 19,* 130–48.

Warfield, M. E. (1994). A cost-effectiveness analysis of early intervention services in Massachusetts: Implications for policy. *Educational Evaluation and Policy Analysis, 16,* 87–99.

Weisner, T. S. (1984). A cross-cultural perspective: Ecocultural niches of middle childhood. In A. Collins (Ed.), *The elementary school years: Understanding development during middle childhood* (pp. 335–69). Washington, DC: National Academy Press.

Weisner, T. S. (1996a). Why ethnography should be the most important method in the study of human development. In R. Jessor, A. Colby, & R. Shweder (Eds.), *Ethnography and human development. Context and meaning in social inquiry* (pp. 305–24). Chicago: University of Chicago Press.

Weisner, T. S. (1996b). The 5 to 7 transition as an ecocultural project. In A. Sameroff & M. Haith (Eds.), *The five to seven shift: The age of reason and responsibility* (pp. 295–326). Chicago: University of Chicago Press.

Weisner, T. S. (1997). The ecocultural project of human development: Why ethnography and its findings matter. *Ethos, 25,* 177–90.

Weisner, T. S., Gallimore, R., & Jordan, C. (1988). Unpackaging cultural effects on classroom learning: Native Hawaiian peer assistance and child-generated activity. *Anthropology and Education Quarterly, 19,* 327–51.

Weisner, T. S., Matheson, C., & Bernheimer, L. (1996). American cultural models of early influence and parent recognition of development delays: Is earlier always better than later? In S. Harkness, C. M. Super, & R. New (Eds.), *Parents' cultural belief systems* (pp. 496–531). New York: Guilford Press.

Weiss, C. H. (1975). Evaluation research in the political context. In E. L. Struening & M. Guttentag (Eds.), *Handbook of evaluation research* (pp. 2–26). Beverly Hills, CA: Sage.

Weiss, C. H. (1995). Nothing as practical as good theory: Exploring theory-based evaluation for comprehensive community initiatives for children and families. In J. P. Connell (Ed.), *New approaches to evaluating community initiatives: Concepts, methods, and contexts* (pp. 65–92). Queenstown, MD: Aspen Institute.

Weiss, H. B. (1988). Family support programs: Working through ecological theories of human development. In H. B. Weiss & F. H. Jacobs (Eds.), *Evaluating family programs* (pp. 3–36). Hawthorne, NY: Aldine.

Weiss, H. B., & Greene, J. C. (1992). An empowerment partnership for family support and education programs and evaluations. *Family Science Review, 5,* 131–48.

Weissbourd, B., & Kagan, S. L. (1989). Family support programs: Catalysts for change. *American Journal of Orthopsychiatry, 59,* 32–48.

tion. *Jou*

is of earl
ations fo
s, 16, 87

e: Ecocul
ins (Ed.)
velopment
;ton, DC

ld be the
develop
), *Ethnog*
1ing in so
f Chicago

n ecocul-
. *The five*
(pp. 295–

f human
;s matter.

88). Un-
g: Native
activity.
1.
. (1996).
d parent
vays bet-
R. New
31). New

political
.), *Hand-*
lills, CA:

)od the-
prehen-
1ilies. In
commu-
pp. 65–

Norking
nent. In
nily pro-

nt part-
ms and

ort pro-
rthopsy-

J. R., Yeates, K. O., & Zigler, E. (1982). Piagetian
dence and the developmental-difference controversy.
E. Zigler & D. Balla (Eds.), *Mental retardation: The
developmental-difference controversy* (pp. 213–76). Hills-
le, NJ: Erlbaum.
inghouse Learning Corporation. (1969). *The impact of
Head Start: An evaluation of the effects of Head Start on*

children's cognitive and affective development. Athens, OH:
Ohio University.
Whyte, W. F. (Ed.). (1991). *Participatory action research*. New-
bury Park, CA: Sage.
Zigler, E. (1969). Developmental versus difference theories
of mental retardation and the problem of motivation.
American Journal of Mental Deficiency , *73*, 536–56.