This paper details the pathway of development of early intervention program evaluation research along with the often-enumerated problems with research results. Early intervention evaluation, which was initiated in 1969, involves those organized efforts designed to remediate and/or reconcile developmental disabilities in young children up to the age of 3 years. Previous research findings have resulted in considerable current research activity being conducted, much of which has attracted federal support. Also covered in the paper are the reasons that early intervention programming has not been proven, beyond all doubt, to be efficacious. A framework for program evaluators is provided to allow them to pinpoint the benefits of this kind of undertaking for the consumers of the services. The ultimate purpose of this examination is to stimulate the design of improved program evaluation models that use quantitative research methodology, supported by anecdotal information. (TJH)
A Program Evaluation Ontogeny Recapitulated: 
Looking Again at the Efficacy of Early Intervention

Terry R. Berkeley, Ed.D. 
University of Maine at Farmington
Abstract

The paper details the pathway of development (ontogeny) of early intervention program evaluation research (the beginning arguably being 1969 when the first such study was published in the United States), along with the often-enumerated problems with those results which, in turn, has yielded considerable research activity being conducted at the present time (recapitulation) with federal support. Early intervention, in this case, are those organized efforts designed to remediate and/or reconcile developmental disabilities in young children from birth to three years of age. Also covered in the paper are the reasons why early intervention programming has not been proven, beyond all doubt, to be efficacious; and, 2) a framework for program evaluators to use in order to pinpoint the benefits this kind of undertaking has for the consumers of those services. The purpose of this examination, ultimately, is to stimulate the design of improved program evaluation models which utilize quantitative research methodology, supported by anectodal data.
A Program Evaluation Ontogeny Recapitulated:
Looking Again at the Efficacy of Early Intervention

At the first national conference on early childhood special education (sponsored by the Council for Exceptional Children's Division for Early Childhood, and held in Denver, Colorado during October, 1985), a number of papers were presented highlighting new efficacy data espousing the benefits of early intervention programs for children with special needs between birth and three years of age and their families. These presentations were based upon different methodological reviews of large blocks of separately collected data from stratified samples of similar-in-age populations. It was understandable, then, that the groups engaged in this research would want to share their findings with an audience that worked with similar intended beneficiaries. However, several individuals, in particular the director of special education in a state noted for offering progressive services to young handicapped children, three researchers who have worked in this discipline for a number of years (the newest one for ten years), and two program providers who manage early intervention efforts in non-proximal states, asked the following questions, independent of each other, to the presentors at some of these sessions:

1) Is it still necessary to prove that early intervention is effective?

2) Is it necessary for the federal government to spend $5 million annually on efficacy studies instead of on services delivery?

It should not be assumed that those asking the questions were naive,
disinterested in research on efficacy, or desirous of just serving children without possessing a concomitant interest in child development or other relevant topics. Further, it should not be assumed that the papers delivered in Denver offered analyses of new variables or that new methodology was employed in the conduct of these particular national-level program evaluation efforts. Neither of these propositions are true, and such new advances do not appear to be forthcoming from most of the groups who are being sponsored by federal authorities. Rather, the questions were asked because data on early intervention does exist which proves these services are beneficial to the children and families who are enrolled in the programs, at least on a general level. The posers of the questions were trying to point this out and suggested that redundancy in this case is purposeless.

While program evaluation research in this area is perceived to be problematic by a number of individuals and, thus, is worthy of further study, the investigations reported upon in Denver seem to be missing the point. That is to say, evaluation data, generally, cannot prove beyond all doubt that the benefits derived from a program are causally-related to the treatments experienced by the intended beneficiaries in those programs. Also, meeting the objectives espoused in the policy that created such an effort in the first place is not likely to be specified given current program evaluation practices. Finally, the impact of the data, given the current state-of-the-art of program evaluation research (at the least the efficacy analyses), upon policymaking decisionmaker’s is hopeful, at best. The presenter’s in Denver, though, implied they were accomplishing these latter aims.
The purpose of this paper is to review the characteristics of early intervention in this country, albeit a history of only twenty-five years, as well as the ontologically recapitulated narrative of early intervention efficacy studies. Particular attention will be paid to those investigations of programs for children between birth and three years of age, although, at the outset, that will not appear to be the case since early intervention was originally viewed as an activity for children from three through five years of age. Additionally, those contributions to the literature from other disciplines that are not utilized in current early intervention evaluation research will be discussed as a means to end the present "overdone" scholarly work on early intervention efficacy which seems to be serving little purpose. Then, an alternative model that can be analyzed quantitatively (with accompanying narrative from anecdotal accounts), will be posited that is sound across the major theories of child development and from the standpoint of family dynamics research.

A Historical View

In 1961, J. McVicker Hunt published his landmark work Intelligence and Experience, which cast aside, for the American psychological and educational audiences, the notion that intelligence was fixed at birth and the development of the human organism was predetermined. Soon after, as a result of two federal policy statements, "educational" services for young children who happened to be disadvantaged, handicapped, or at-risk of becoming handicapped were begun in this country (Berkeley, 1984). These efforts were known as Project Headstart and the Handicapped Children's
Early Education Program (HCEEP, a.k.a. First Chance Network). Importantly, their evolution may be the most dramatic example of the oft-implied relationship between child development research and social policy (Berkeley, 1985), an interaction which has broad implications for program evaluators, child development researchers and policy analysts. In fact, White (1970) and Mosteller and Moynihan (1972) have specified that evidence existed which affirmed that child-centered benefits would accrue if Project Headstart were enacted:

There was social action dictated not by grass roots demands but by the social diagnostician; there was the urge to establish a solution outside the system that "failed"; there was scientific backing offered by the expert-turned-advocate (p. 49).

Similar statements were provided by individuals offering testimony to the Congress regarding the potential benefits for young handicapped children, and they resulted in the establishment of the HCEEP in 1969, just four years after Project Headstart was enacted (DeWeerd, 1984).

The call for program evaluation data or efficacy data began almost immediately after each of the programs was implemented. That is, their existed considerable interest in the intellectual and socialization gains of the children enrolled in Head Start (Westinghouse, 1969), as well as in the developmental progress young handicapped children made in the First Chance Programs (DeWeerd, 1984).

The demands for this data were exacerbated because significant federal expenditures were being made to support large-scale action programs, and in the case of Project Headstart, per Shapiro (1980), the "long-range goal....
was to contribute to a reduction in the economic and social attainment disparities among societal groups by increasing the level of economic and social attainment for the disadvantaged (through educational programs, present author contribution)" p. 59). Opposition to these kinds of expenditures of national tax revenues existed and was demonstrated in statements by Members of Congress who deemed that these programs would not be cost-effective since little return on the investment made in human services would be forthcoming (Mosteller and Moynihan, 1972), thereby, implying another reason for evaluative information. Thus, the proponents and opponents of these endeavors had sufficient interest to request the development and design of accountability measures. No parallel statement can be found in the testimony leading to the enactment of the HCEEP; however, Public Law 91-230, Part C. (the authorizing legislation) does require that grantees evaluate their efforts in terms of the intended beneficiaries served. Also, the requirement is for program providers to meet the needs of a variety of target audiences, a provision that may have come about due to the more general demand to analyze the results of social action programs emanating from the Great Society (Johnson, 1971).

Early Intervention

Early intervention services according to Meisels, Berkeley and Godfredsen (1980), are: "one or more interrelated activities designed to enhance the development of children between birth and three years of age and, directly or indirectly, to assist the families of those children" (p.1). These efforts are also known as infant intervention services. The
focus of this paper will be on programming for infants. However, the ontogeny of program evaluation for these services are grounded in the more general evaluation reports first published in 1974 in the wider area of early intervention. It was not until the 1980's, it seems, that interest in separating infancy from the entire age range of the early childhood period (birth to eight years and/or birth to five years) emerged, it can be speculated, as a result of the increased attention given to infants in child development research journals and the human services research literature.

Notes on Early Development

Kirk (1958), Caldwell (1964), and Stephens (1966) feel that the early experience of the handicapped infant is just as crucial as that for their non-handicapped peers. The difference(s) between the two groups of children is not necessarily the impairment with which they must live. Rather, the handicapped child progresses through the sequences of development in a different way than the non-handicapped child, at least in some way in at least one domain (e.g. motor, cognition, communications, and social-emotional). Piaget (1957) noted that this explanation of progressing differently through the development sequence is known as decalage. Along this same line of thinking, Schieflbusch (1978) explains, "The disabled child often seems to be normal in many respects, but will be markedly below the normal child in at least one critical parameter" (p. 430). Parkhurst (1980) contends the ability of the infant to reorganize conceptual information in its transactions with its ecology is "exceptional" for the developmentally delayed and at-risk child. And,
she adds "Exceptional outcomes of the individual's development are not seen simply as a function of an inborn capability to respond appropriately..." (p. 6). Solnit and Provence (1979) detail their explanation of developmental delay in terms of risk conditions, or characteristics in development that are not significantly delayed, rather they are somewhat below the norm. They suggest that these at-risk children must overcome problematic aspects of their environment in order to remediate their atypical developmental patterns. Sroufe (1979) notes that these children can move in and out of risk as their environment changes, and they are vulnerable for risk over extensive periods of time. It might be posited that just as other children experience reorganizations in their developmental schemes, these children have longer periods of discontinuity destabilizing, for longer periods, their development. Meisels and Anastasiow (1980), though, note: 

"...these impairments, whether real or suspected, must be identified and treated before...maturation...occurs. Moreover, when the treatment is provided by persons who have daily interaction with the infant or young child, the treatment is more likely to offset the impairment, regardless of whether the cause of the impairment is biological or environmental" (p. 29).

Foci of Early Intervention

The focus of treatment for these children can take one of two forms. The first form is directed toward improving the interactions between the
infant and its primary caretakers, usually the family. This method of providing services takes advantage of the child's first relationships which impact the child's developmental sequence. Sameroff and Chandler (1975) declare,

At each moment, month or year the characteristics of both the child and his environment change in important ways...The child is, in this view, in a perpetual state of active reorganization...In this view, the constants in development are not some set of traits but rather the processes by which these traits are maintained in the transactions between organism and environment (p. 234-235).

This model is important because it recognizes Piaget's (1957) contention that the human organism is an active participant in its own development, and it builds upon the mother-infant interaction and the impact the mother and baby have upon each other. Further, as the child becomes more competent within the context of the mother-infant dyad, he or she will venture forth into other relationships and the transactions continue.

Kagan, Kearsley and Zelazo (1978) noted three assumptions in their study of the infants earliest life which adds to the richness of the infant-primary caretakers focus in infant intervention. Their assumptions:

The first is that the experiences of the infant exert a profound influence on its contemporaneous development...The second, more serious, assumption is that the experiences during the first year extend far into the future and alter, in some mysterious way, the structures and processes that
will emerge in the 3-, 4-, and 5-year old...The mother-infant relation is at the center of the third presupposition (p. 2). These are not unproven theoretical assumptions.

Osofsky and Connors (1979) extensive review of the research points out that the interaction between the mother and child is complex. Bowlby's (1969) research led to a characterization of an attachment bond essential to the human organism. Ainsworth (1979) suggests, empirically, that the bond establishes a "responsiveness enabling an infant to form expectations, primitive at first, that moderate his or her responses to events, both internal and environmental" (p. 933). And, in extensive longitudinal work, Skeels and Dye (1939) and Skeels (1966) found that the dynamic of nurturing, usually offered by mothers to their infants, provides a situation which assists the child in acquiring knowledge and processing information. The caregiving, he concluded, must be constant.

The second focus for treatment in early intervention is a noninteractional view where the human organism's response to stimuli is more important than the interplay between the organism and the environment. Providers utilizing this biological focus in their approach to services delivery seem to be the exception to the rule at the present time. Fewell (1986) reasons that this state-of-affairs has to do with an increased understanding of a sophisticated knowledge base. She states,

It has become increasingly clear that the experiences provided in early intervention programs for young children
with handicaps are broad and varied. Programs no longer focus solely on the child's acquisition of identified competencies, but often address the family system, 'raining rents to acquire needed skills and facilitating positive interactions between the child and the family. Unfortunately, measurement systems have lagged behind these interventions, and even today many programs rely almost entirely on measures of child behavior as the sole outcome measure of effectiveness, regardless of the scope of intervention (p. xi).

**Early Intervention Efficacy**

Information about the effectiveness of early intervention efforts began being published in 1969. The first major evaluation of such programs was the Ohio-Westinghouse Study of Head Start. The results of that study showed children made short-term cognitive gains, and other indicators of development (e.g. socialization) were not as significant. Tjossem (1976) contends the results of the Ohio-Westinghouse Study demonstrated:

- The early initial sharp gains in cognitive development resulting from preschool intervention were not sustained;
- they were washed out as control children caught up with experimentals when programs were discontinued (p. 17).

Shapiro (1980), from his review of the Ohio-Westinghouse research, found there were problems in the program evaluation methodology used in the study. The problems, he suggests, follow along the lines of the experimental-quasi-experimental design debate, whereby several attempts
were made to imply significant causal inferences that decisionmakers could use in their deliberations.

In 1974, the second major review of the effectiveness of early intervention was published. In that study, Bronfenbrenner looked at a number of early intervention programs serving disadvantaged children from one to six years of age in either a home-based or center-based setting. He found: 1) children made significant gains in IQ; 2) structured curricula produced better IQ gains; 3) children tended to decline in IQ after being out of the program for one or two years; 4) the impact of the family was critical; 5) home-based programs were more beneficial; 6) parent-child intervention produced the greatest gains in IQ which were evident three or four years after the child left the program; and, 7) the most deprived child would not benefit from programming unless her or his environment changed.

In the next major report, Lazar et. al. (1977) reviewed the results of 14 programs for low income children who were enrolled in experimental infant and preschool programs in the mid- and late 1960's. Lazar and his colleagues found,

The combined results from all projects indicate that early intervention helps low income children to meet the minimal requirements of their schools. This can consist of reducing the probability of either being assigned to special education classes or being held back in grade...Thus, it appears that early education can result in cost savings by reducing the rate of assignment to special education and/or the rate of grade failure (p. 73).
Also, from the mid-1970's through late 1985, Weikart and his colleagues at the High/Scope Foundation in Yipsilanti, Michigan have evaluated their preschool programs which utilize a Piagetian-based curriculum. The findings of these researchers include: 1) effective programs are well-planned, highly organized and consistently implemented; 2) dramatic short-term improvements in development, attributed at least in part to the program, can be sustained if programming continues; and, 3) long term cost savings to the human services system through decreased dependency can occur with resulting cost-benefits.

Problems, of course, can and have been found with all of this work. Criticism has come from a variety of observers, and it has ranged from the consideration of program costs tied to developmental gains as a measure of benefit, to the limited information provided about developmental implications, to the use of IQ as an indicator of program outcome, to debate about the methodological shortcomings of the reports. However, since few of the children participating in the reviewed programs were handicapped or from birth to three years of age, generalization of the results is specious. As in any rapidly evolving field, then, enhancements can be made, and are to be expected. Yet, the early work did point to major benefits for intended beneficiaries as a result of the efforts stimulated by the initial Project Headstart work.

Infant Intervention Efficacy

In terms of handicapped children, Caldwell (1964) offers an excellent
early review of "efficacy" research relative to toileting, feeding, positioning, maternal conduct, the influence of social class, and early postnatal experiences between 1930 and 1960. Caldwell pointed out that these local efforts, some in the community and some in teaching hospitals, and some in other university laboratories or schools, were successful and young handicapped children did make positive gains in their development.

As the work of service providers and researchers continued in the 1970's due in large measure to the HCEEP, more and more reports of developmental gains in the children enrolled in those services efforts, were presented at conferences and some found their way in the child development and program evaluation literature. For example, between 1970 and 1977 studies of infants with sensory deficits appeared. Northcott (1971) reported that intervention programs with hearing impaired children prevented the development of inappropriate behaviors such as excessive bodily contact, grimaces, and other stereotypic behaviors that tend to limit the acceptance of deaf children by the non-hearing impaired population. Horton (1974) found that intervention with deaf children before the age of two resulted in their adaptation to normal classrooms, whereas deaf children who were not in intervention programs until the age of three years did not make these adaptations as easily.

Fraiberg (1977), in another example, noted that when congenital blind children were not provided with environmental stimulation to compensate for loss of sight, they may develop stereotyped hand behavior, rocking,
swaying, mutism, or echolalic speech. She also noted that blind-from-birth babies demonstrated severe problems of attachment and an insufficient amount of stimuli for the infants' tactile-kinesthetic-auditory development. After stimulation in an infant program, Fraiberg found they were closer to sighted infants ranges than to the ranges exhibited by blind-from-birth controls.

So far, all of the studies reported have been about sensory impaired infants. Keogh and Kopp (1976), Ross and Leavitt (1976), Silver (1979) and Field (1979) each reviewed the literature on infant intervention programs including sensory impaired and non-sensory impaired, but otherwise handicapped or at-risk infants, up to that point in time. They reported: 1) the most successful intervention programs were similar in delivery to the most sensitive parenting; 2) programs appearing to be the most successful had strong parent components and stressed formal and informal parental participation; and, 3) gains were made in most developmental domains.

In 1979, Bromwich and Parmalee repeated Caldwell's (1964) request to the field that it expand the variables covered in its program evaluation research. That is, they felt more needed to be done to determine the most appropriate instruments to use in assessing the quality of behaviors that parent exhibit in successful infant intervention experiences.

There was also debate in the field arising, in part, from Bronfenbrenner's (1974) work about where the best programming takes place, in a center or at-home. Masi (1979) reported "the strongest demonstrations of sustained improvements were studies involving longer term at-home
stimulation (p. 377). But, Kagan et al. (1978) found it difficult to determine if center-based programming or home-based programming for infants was the most effective. Too, Silver (1979) noted "both center-based and home-based infant intervention programs that deal with the child and the family have positive results" (p. 23).

The program evaluations conducted in the early 1980's have pointed to similar results and similar problems (Dunst and Rheingrover, 1981; and, Bricker, Bailey and Bruder, 1984); and, only the variables being analyzed have changed, and this has most usually reflected the family focus called for by Caldwell (1964) and illuminated by Fewell (1986) earlier in this paper. Dunst (1986), however, posits "Early intervention programs for biologically impaired infants have become widespread despite the fact that there are few well-conducted studies documenting the efficacy of these programs" (p. 79). This has resulted in the ontogeny recapitulated which was challenged by some participants at the Denver conference on early childhood special education.

Key Problems with Current Infant Intervention Efficacy

Each discipline in its history, it seems, goes through transitions to reach greater levels of maturity. Infant intervention, given the considerable attention it is receiving from Academe, government, and the human services, is in transition at the present time. Since January 1982, for example, three complete volumes of Topics in Early Childhood Special Education have been devoted to program evaluation and efficacy. Another issue was published attempting to link the work in program efficacy/program evaluation to social policy for young children. In addition, the federal
government is supporting research projects in this area at five major universities. Therefore, it should not be viewed in extraordinarily negative terms that program evaluation issues in early intervention will exist.

As previously stated, methodological concerns in the research used to analyze program evaluation data continue to revolve about the experimental-quasi-experimental debate. That is, threats to internal and external validity abound, in concert with incorrectly stated causal inferences that link a child's development to the interventions provided in the program in which the child is enrolled (Dunst 1986).

But, as Dunst contends, "It should be made explicit that... (author's) did not conclude that early intervention does not work. Indeed, this writer is of the opinion that there are many positive aspects of early intervention, but that we have failed to conceptualize and conduct experimental evaluations in a manner that has permitted us to document the efficacy of our intervention efforts" (1986), p. 80). Until these methodological problems are resolved, efficacy data may always be viewed with "suspicion" by other researchers and, most importantly, by decisionmakers. Given the history of methodological criticism, as well as the limits of quasi-experimental designs (Cook and Campbell, 1979) more precision will have to be drawn from improved explications of the programs being evaluated.

Meisels (1985) perceives the limitations in program evaluation in more fundamental terms. That is, he explicitly builds upon Bruner's classic contention by quoting the famed developmental psychologist: The importance of early childhood for the intellectual, social, and emotional growth of
human beings is probably...one of the most revolutionary discoveries of modern times (1980). In taking this tack, Meisels states:

In order to assess the efficacy of such efforts, investigators must explicitly clarify and take into account four major (but often overlooked) assumptions of the intervention program: its theory of human development, specific intervention strategies used, methods of measuring change, and criteria used for selecting participants. Without at least this much clarification, confusion will continue to reign in the arena of early intervention efficacy (p. 9).

The unintended, but crucial, consequence of this view is that program providers need to know more about what it is that they are doing, and they must be systematic in explaining what they do to a variety of audiences, including themselves. This line of inquiry and research is known as process evaluation (Weatherford, 1986).

Another limitation in program evaluation research in the field of early intervention rests in the too few reliable instruments which have been developed to measure developmental change in the human organism and in family systems (Mott, Fewell, Lewis, Meisels, Shonkoff, and Simeonsson, 1986). In a December 1985 conference sponsored by Utah State University, the above authors were asked to recommend one child-oriented instrument and one family-oriented instrument which could be utilized in the evaluation of an early intervention program. In only two instances was an instrument mentioned more than one time: The Battelle Developmental Inventory
Early Intervention

(Newborg et al. 1984) was noted twice for a child-oriented instrument; and, the Parenting Stress Index (Abidin, 1983) was mentioned with the same frequency. Each contributor, however, did specify similar themes when it came to the limitations of early intervention efficacy studies. These have been explicated previously in this paper.

Summary and Alternative Modeling

In this paper, a brief history of early intervention was traced from its roots in developmental psychology, its relationship to social policy, and in terms of efficacy studies that commenced with the Ohio-Westinghouse Study in 1969 of Project Headstart. The benefits of early intervention have also been explicated. This has provided an ontogeny (pathway) of early intervention efficacy which has been recapitulated since study after study seems to end in the same analysis: early intervention while efficacious, cannot be proven, beyond a reasonable doubt, to be beneficial. In other words, it is incorrect to make causal inferences linking a child's developmental gains to a set of program interventions, given the current research.

Another limitation widely cited in the literature is the absence of instrumentation designed to measure early intervention activities. However, it is more important that program providers first know that what it is that they do is what it is that they purport to do. Since early intervention is presently being offered in the context of child-family interactions, the need to measure quickly changing infant development and the dynamics of the families with which these children live is essential, but this needs to begin with a careful examination of the processes used by
the program; at that point, outcomes can be determined.

In spite of this criticism, Lewis and Starr's (1979) salient response model offers a conceptual base upon which a further ontogeny of early intervention program evaluation can be constructed, at least in terms of the development of the infant itself. Lewis and Starr are concerned with the stability of responses of the developing infant and note, "Salient attributes of responses include (the) quantity, quality, speed of acquisition, utilization, affective tone, generalizability, organizational properties, and intention" (p. 657).

The responses, which can cross all theories of child development, focus on how a child obtains and uses conceptual information across each developmental domain; in turn, it can be employed within two methodological contexts. First, an instrument needs to be devised to measure the responses at different points in time in a child's life so that a quantitative index of development can be obtained. This, of course, is a difficult and expensive undertaking. Second, a developmental assessment guide has been designed (Berkeley, 1985) that takes advantage of the Lewis and Starr model which provides a qualitative/ethnographic understanding of how infants develop. In this context, individuals utilizing the guide must be thoroughly familiar with at least one theoretical orientation to development in order to frame their observations according to the responses and, then, use that information to develop a service plan that is developmentally appropriate for the child. The major criticisms from the few individuals in early childhood special education classrooms, as well as staff in one infant program, who have used the Guide is that the training
is never really completed, and it is quite labor intensive. Therefore, the greater potential of this model lies in its use as a foundation for the development of an instrument which can measure a child's developmental status.

Additionally, formats need to be designed to measure what it is that early intervention program staff do in their interactions with families, instead of assuming that the gains children and families make are the result of that staff member's interventions. Instrumentation also needs to be developed that will measure the intended beneficiaries satisfaction, or lack of satisfaction, with a program in a meaningful manner. It is not enough to ask if a family is satisfied with the services they receive. Instead, early intervention program providers need to determine if they know what it is that parents want from a program, and, if not, ask them: families are the primary beneficiaries, the program staff are secondary beneficiaries.

In spite of these limitations, early intervention efficacy research has broad appeal for those interested in the theoretical foundations of program evaluation, as well as for individuals interested in advocating for the development of policy to assist those in need. Given the current trends in program evaluation, this is a refreshing change.
References


Meisels, S. J. (1985). The efficacy of early intervention: Why are we still asking the question. Topics in early childhood special
Early Intervention, 5 (2), 1-11.


