Designing meaningful evaluations of early intervention services

PENNY HAUSER-CRAM

The enterprise of evaluating human service programs has evolved considerably during the past two decades. Knowledge gained from previous evaluations of such programs as Head Start and from the growing body of research on young children with disabilities and their families enables evaluators to avoid the pitfalls of past evaluations by taking advantage of recent advances in analytical approaches and employing measures of child and family change that are based on a broad view of human development.

Evaluation differs from other forms of research in that it focuses primarily on questions of efficacy—that is, on how change has occurred as a result of a program or specific treatment. Although understanding the mechanisms and impact of change forms the basis of our knowledge of human development, in evaluation research the critical task is to disentangle the impacts of services from other effects. Nevertheless, the logic of inquiry is the same in evaluation research as in other research endeavors. Evaluation demands the conceptualization of research questions and hypotheses, development of an appropriate and feasible design, selection of valid and reliable measures, application of systematic and unbiased procedures for data collection, determination of precise analyses of data, and accurate interpretation of results.

Well-designed evaluations require investigators to progress through a series of steps, beginning with understanding the purpose of the evaluation, proceeding with conceptualization of the questions and study design, followed by selection of appropriate data analysis strategies, and culminating in interpretation of the results. Following these various steps, this chapter reviews the methodological decisions, dilemmas, and limitations that guide evaluation of early intervention programs and suggests new directions that evaluation research may take.

UNDERSTANDING THE PURPOSE OF EVALUATION

The functions of evaluation of service programs like early intervention are many and varied. Evaluation may serve to justify the existence or expansion of a program, improve its functioning, or demonstrate impact of a specific set of services on participants and the wider community. Each of the “stakeholders” (that is, individuals concerned with a program) may have different questions that evaluators could legitimately address. For example, state legislators may ask questions about a program’s cost-benefit ratio, program developers may be concerned with cost-effectiveness, service providers may wonder about the relative benefits of various program features, such as...
Designing meaningful evaluations of early intervention services

Conceptualizing the questions

Any human service program must respond to the question, Is this program successful? This question, which might appear simple and straightforward, at closer glance reveals its complexity. First, we need to be more precise about the definition of “this program.” What is early intervention, and is one early intervention program similar to another? Second, what is meant by “success” and how can it be measured? Third, “success for whom?” Is the program targeted only on children or do parents also benefit? Each of these questions will be considered in turn.

What is the program?

Descriptions of early intervention programs often do little more than differentiate between “home-based” and “center-based” services. Although such descriptions indicate the location of service delivery, they tell little about the theoretical bases underlying service provision or about the actual content of the services. One home-based program may differ substantially from another, as was revealed in a meta-analysis of studies on early intervention programs (Shonkoff & Hauser-Cram, 1987). Programs classified as home-based varied substantially in the number of hours of service, discipline of provider (e.g., educator, physical therapist, social worker), type of curriculum (developmentally based, language-based, behavioral), and model of service provision (transdisciplinary or multidisciplinary). Even more problematic was the finding that few studies reported sufficient information about program components to allow systematic study of these critical variables.

The question of defining services is not unique to early intervention programs and, to some extent, is relevant to all human service programs that are multifaceted in scope and multidimensional in purpose. One solution would be to code written information about a program’s goals or model of service provision. Yet, as was found in the evaluation of the Follow Through program, written information shares only some characteristics with the actual program (House, Glass, McLean, & Walker, 1978). Even similarly labeled models may have as much variation in service components as do models with different labels (Bock, Stebbins, & Proper, 1977). An additional problem occurs in defining services provided to participants in early intervention programs. Many programs may offer a similar array of features but individualize these services to families based on the family’s needs and desires, the child’s age and disability, and the program’s philosophy and staffing pattern. Individual variation may indeed be the hallmark of these programs, and suitable and efficient ways of documenting such variation need to be found.

Finally, as in all human service programs, because the services that are planned for a child and family are often quite different from those actually received, it is important that implemented services be considered. Patton (1978) delineates three aspects of implementation that require attention in an evaluation. The first, termed “effort evaluation,” concentrates on questions about whether there has been a sufficient quantity of service to allow the possibility of impact. Although the lack of effort seems clear in cases of individuals who terminate medical treatment before they are advised to do so, it is not as clear in cases where families enrolled in a program have an uneven rate of participation. The next phase, “process evaluation,” involves an analysis of how the program produces its results. For example, what actually occurs during a home visit that is likely to have an effect on a child or family? In the third phase, “treatment
Designing meaningful evaluations of early intervention services

Reliance on IQ as an outcome measure has overshadowed other critical areas of child development affected by early intervention programs. Such programs often seek to have an effect on a range of domains of development, such as functional skills, social competence, self-regulatory behaviors, motivation, and curiosity (Bricker & Kaminski, 1986). Although measures are not as well developed in these areas of development as in the area of cognition, they should not be neglected by evaluators (Zigler & Balla, 1982). Given the multifaceted goals of early intervention programs, an array of child outcomes is often required to understand program effects sufficiently (Meisels, 1985).

Evaluators may need to seek creative ways of measuring change in domains of development for which there are no standardized measures. A number of new instruments have been reported in the research literature that deserve careful consideration (Hauser-Cram & Shonkoff, 1988). Because such measures often lack sufficient documentation of their psychometric properties and standardization for atypical populations, admittedly there is an element of risk associated with their selection as outcome measures. Such risk is diminished if multiple measures are employed, with each sharing a portion of the theoretically relevant factors but each having different loadings of irrelevant factors or “noise” (Isaac & Michael, 1978). This approach, often referred to as triangulation of measurement (Miles & Huberman, 1984), offers more powerful evidence of effects than would reliance on a single measure alone.

For example, many early intervention programs aim to increase children's motivation and curiosity during self-initiated or self-directed activities. A measure, such as an IQ test, in which an examiner elicits a child's performance would be a poor choice as a means of measuring this outcome. Instead, a valid evaluation of this program goal might require observation of children during spontaneous play. The research literature on spontaneous play in normally developing children (Belak & Most, 1981) and in children with disabilities (Hill & McCune-Nicolich, 1981) offers some promising examples of measures of the level of play (ranging from mouthing to pretend play) in which a child may engage. Assessments of mastery motivation (Morgan & Harmon, 1984) in which task persistence is assessed during play with a problem-presenting toy, would further strengthen knowledge about children's self-directed activity. Finally, a questionnaire about task persistence and behavior during self-initiated activities (Morgan, Maslin, Harmon, Jennings, & Busch-Rosengren, in press) could be completed by parents and service providers to marshall more evidence of children's behavior in this domain.

In operationalizing “success,” evaluators need to consider both short-term, immediate gains and long-term advantages. Most evaluations of early intervention services for disabled infants focus on short-term skill development, and virtually none have considered sufficiently the long-term benefits of such programs (Shonkoff & Hauser-Cram, 1987). The value of short-term effects should not be minimized, and individual programs may well strive for short-term gains that evaluations should acknowledge. Equally important questions exist, however, about the stability of such gains and about latent effects that emerge after a child leaves an early intervention program.

Examples of investigations of persistent effects have emerged in the literature on early education programs for economically disadvantaged children. These programs usually include a range of policy-relevant variables, such as the avoidance of delinquency, successful employment, and decreased need for remedial services in school (Berretta-Clement, Schweinhart, Barnett, Epstein, & Weikart, 1984; Lazar & Darlington, 1982). Evaluations of early intervention services would benefit from careful thought about appropriate measures of their long-term effects. For example, it may be valuable to
investigate whether participation in early intervention affects later school adjustment, peer interaction, amount and type of additional services, placement in an integrated or segregated setting, parental advocacy, and parent-school relationships.

Success for whom?

Early intervention programs serve a wide range of children and families. Some children may have no documented delays but be "at risk" for problems, others may have delays of unknown etiology, and still others may have nervous system dysfunction related to a specific diagnosis. It would be naive to assume that the effects of early intervention programs would be similar across such a wide range of disabilities.

A central question facing evaluators is how best to define subgroups of children for purposes of elucidating the differential effectiveness of a program (Adelman, 1986). When subgroups have been defined in past evaluations of early intervention services, they have generally been based on diagnostic groups (e.g., Down syndrome, cerebral palsy, and so on). The value of this classification scheme is not yet known, and there may be better approaches. For example, Stein and Jessop (1982) recommended that the traditional categorial approach to the study of the impact of chronic illness on children and families be replaced by an approach that focuses on differences in functional status. It may be that for certain questions about the effectiveness of early intervention, differences may emerge by degree of the child's disability rather than by the type of disability. A hint of this possibility has occurred in the developmental literature, but has not yet been investigated in relation to program effects. In a study of a young group of Down syndrome, cerebral palsy, and developmental delay, Brooks-Gunn and Lewis (1984) found that maternal responsivity was more related to a child's mental age than to diagnostic status. Such data point to the value of thinking carefully about subgroup classifications and about possible interactions among interventions and characteristics of children and families.

In most early intervention programs, the participants are not only children but their parents as well. Despite a fair amount of literature on the impact of raising a handicapped child on families (Biachler & Meyers, 1982; Simeonsson & McHale, 1981), few evaluations of service programs have considered their effects on families aside from measures of satisfaction (Bailey & Simeonsson, 1986). Satisfaction measures may be useful for formative evaluations in helping to improve a program's functioning, but such measures do not lead to objective understanding of a program's effects. Measures of satisfaction can, however, help evaluators understand parents' views of the benefits of a program and how such views may vary based on characteristics of the family and child. For example, parents of children with mild disabilities may be satisfied if they believe the program helps to advance the child developmentally, whereas satisfaction for parents of children with severe disabilities may rely on a program's ability to provide support and respite from caretaking tasks (Sandow, Clarke, Cox, & Stewart, 1981).

A program's theoretical model should guide the selection of outcome measures of family impact. Several different models have been posited to explain how early intervention programs affect families. They share a reliance on the ecological view of the child and family (Bronfenbrenner, 1979) and on the transactional nature of development (Sameroff & Chandler, 1975). Marx and Kysela (1985) isolate three distinct approaches to the characterization of the role of parents in early intervention programs. In the first, the parent therapy model, parents are assisted through counseling or support groups to help resolve stress related to raising a child with disabilities. A more
Designing meaningful evaluations of early intervention services

features can be selected for systematic evaluation. Although beyond the scope of this chapter, studies of cost-effectiveness (see Barnett & Escobar, this volume) would complement such investigations.

Design issues are as critical in responding to the questions posed as in developing more general efficacy studies. As an example, suppose an evaluator were investigating the value of parent support groups in reducing stress in families. The true experimental design approach to this question would demand that subjects be randomly assigned to either a treatment group (i.e., to a parent support group) or a control group (i.e., the absence of a parent support group). Unless a large sample of participants were to be involved in the investigation, subjects should be paired (or matched) based on their stress scores on a pretest measure and possibly on demographic characteristics assumed to be related to outcome, such as gender. One member of each matched pair would be selected randomly to attend parent support groups. The evaluation would proceed to document aspects of the support group interaction, attendance levels, and other related process variables. A posttest measure of stress would then be given to members of each group. Finally, analyses would be developed to determine whether participation in support groups was associated with a reduction in level of stress.

When the value of a service is not well documented, as in the case of parent support groups (McGuire & Gottlieb, 1979; Wandersman, 1978), a true experiment, such as the one just described, may be feasible. However, if participants and/or service providers regard the service as valuable, then withholding it from some families touches on ethical issues. Moreover, a randomized study of one feature of an early intervention program is prone to certain weaknesses. First, it often establishes an artificial situation, which makes generalization of its results questionable. Perhaps parents are more or less likely to attend parent support groups when they are “assigned” to do so; thus, their actual behavior is quite different from what it would be if they could simply opt to select attendance at such a group. Second, whenever random assignment occurs within a program in which participants interact with each other, contamination can occur. For example, support group participants may carry on their support group discussion topics with members of the control group, creating, in effect, an underground treatment group.

Several creative options for designing evaluation studies that take into account the constraints of conducting a study within an ongoing program exist (Cook & Campbell, 1979). Although not yet well tested in the early intervention literature, quasiexperimental designs to situations in which treatments (or services) have been provided to participants in some nonrandom fashion (Cook et al., 1977). Four quasiexperimental designs would have particular relevance to investigations of the value of a particular feature of early intervention services, such as parent support groups: (1) the untreated control group with pretest and posttest scores, (2) the cohort design, (3) the nonequivalent dependent variables design, and (4) the planned variation design. An example of the application of each of the four designs to the question of the value of parent support groups in reducing stress is described briefly.

UNTREATED CONTROL GROUP DESIGN. This design would be feasible if more parents in an early intervention program wished to join parent support groups than there were staff members available to lead such groups. Since some parents would need to be placed on a waiting list, they would serve as a natural comparison group. Legitimate comparisons could be made only if members of each group were given a pretest and a posttest and analyses of changes could be made (Cook & Campbell, 1979).
Otherwise, differences between groups at the end of the support group intervention may be attributable to initial differences, rather than to the intervention itself. Other threats to validity exist with this design, such as regression to the mean; if either group has exceptionally high or low scores, their scores are likely to drift toward the mean even in the absence of intervention. A possible selection-maturation effect also exists if members of the two groups are likely to change at different rates irrespective of the intervention. Finally, threat to validity is lost for interventions that produce effects rather rapidly, for it is unlikely that programs will be willing to keep individuals on a waiting list for a specified service for a long period of time.

**COHORT DESIGN.** Cohort design takes advantage of knowledge of characteristics of families before a new service, such as parent support groups, is introduced. Changes in stress scores over time could be recorded for one cohort of participants in a program. Parent support groups could then be offered to the next cohort and changes in their stress scores compared to changes in those of the prior cohort. Such a design requires a great deal of patience and prior planning if sufficient data for a cohort are to be gathered before the introduction of a new service. Threats to validity, such as history (i.e., effects on the second cohort that did not exist for the first due to changes in legislation, public information, etc.), need to be considered in the application of this design.

**NONEQUIVALENT DEPENDENT VARIABLES DESIGN.** This design involves only one group, but comparisons on different outcome measures are made within the group. For example, suppose father support groups are hypothesized to reduce stress for fathers but not affect other aspects of their social support network. Pretest-posttest changes would be expected to occur for measures of stress but not for a measure of the size of the support network. This approach has the advantage of avoiding the problem of selecting a comparison group, yet it can still provide some evidence of change associated with participation in a support group. Dunst describes his use of this strategy in evaluating an intervention program in which changes were targeted on only certain aspects of development (Dunst & Rheingrover, 1981). A difficulty inherent in this design involves the determination of one set of discrete outcomes that are assumed to be affected and another set of outcomes assumed not to be affected by the intervention. Many interventions, such as parent support groups, may produce changes in one area of a child's or family's functioning that, in turn, affect changes in other areas.

**PLANNED VARIATION DESIGN.** Planned variation is appropriate for analyzing questions about the relationship between the intensity of a service and the outcome. Individuals are assigned to groups that receive different levels of service. For example, one parent support group might meet weekly and another only monthly. Changes in stress scores from pretest to posttest associated with these different levels of service would be analyzed. Ideally, the assignment to groups would be random (or would be a matched pair with random assignment), although some of the same concerns about random assignment discussed previously could occur, such as compensatory rivalry and resentment. If not random, selection-maturation effects and regression effects would need to be considered, as in the untreated control group design. Ideally, the evaluator would need to strive to ensure that the difference in level of treatment was large enough to be maintained. For example, a difference between weekly and biweekly group meetings may be inadequate because it is likely that parents in the weekly group may miss some meetings and, therefore, attend at about the same frequency as the biweekly group, obscuring the difference in level of intervention.

Each of the designs described has the potential of responding to questions about the value of a particular feature of early intervention services. Each also possesses potential weaknesses, as does even the randomized study. The evaluator's task is to anticipate such weaknesses and garner evidence that counters alternative explanations for findings.

**Statistical power**

Regardless of the design selected, evaluators need to consider issues of statistical power. Power is the ability to detect differences statistically when such differences exist in the population (Cohen, 1977). In statistical terms, it is the ability to reject the null hypothesis when, in fact, it should be rejected. In the two-group comparison case, the null hypothesis that there is no true difference in the population; (2) the within-group variation is large and the difference between group means is moderate or small; or (3) the sample size is small and the mean differences between groups in moderate. In the latter case, it is said that there was insufficient power to allow the analysis to detect the moderate differences that actually exist in the population.

Problems of insufficient power plague evaluation research on early intervention. Unless studies are done using particularly large programs, sample sizes are small, making it unlikely that moderate effects will be demonstrated. Samples of at least 20 subjects per group are needed to detect differences of one-half standard deviation 95% of the time (Shonkoff & Hauser-Cram, 1987). Yet, very few studies on early intervention have samples of that size. In a metaanalysis of early intervention studies, 82% of samples had 40 or fewer subjects.

Sample size questions to some extent intersect with questions about design. For example, designs that call for matching of subjects may mean that many potential sample participants are ineligible simply because no “match” can be found for them. This may necessitate summative evaluations to occur across programs in order to develop samples of sufficient size. An added benefit of this approach may be that the results of such studies can be generalized to a wider array of programs and not restricted to a particular site that may have an unusually dedicated staff or other idiosyncratic features that may account for differences.

**ANALYTICAL APPROACHES**

**Measuring change**

One of the greatest dilemmas for those evaluating early intervention programs is how to measure the effects of services in the absence of research designs with randomized or rigorously selected control groups. Documenting and analyzing change is indeed the central task of evaluation research, yet it is an exceedingly difficult undertaking when causal inferences are not readily apparent. Children and families naturally change over time, and there are many influences on such change, including environmental forces, genetic or biological influences, and support services. The question facing evaluators is how to disentangle the influences on change so that changes due to early intervention...
Designing meaningful evaluations of early intervention services

This approach has certain limitations, however. First, it is useful only for the measurement of change using standardized assessment measures in which developmental ages can be calculated. Second, it is based on the assumption that the ratio of developmental age to chronological age would be stable in the absence of intervention. This assumption is critical in the application of such indexes, yet lacks empirical verification.

The measurement of change in early intervention programs needs to be considered over a wide range of measures and not restricted to assessments of IQ and DQ. What methods are proposed for dealing statistically with such data?

DIFFERENCE SCORES. The most straightforward approach to the analysis of change is to use difference scores (also called gain scores). This involves subtracting the pretest score (before intervention) from the posttest score (after intervention). Although computationally appealing, there is debate among methodologists as to the wisdom of this approach. For some time it has been argued that a simple difference score lacks reliability (Cronbach & Furby, 1970; Linn & Slinde, 1977). Reliability indicates to what degree individuals can be distinguished from one another based on the measure. Moreover, a measure is said to be unreliable if it produces a different value for the same individual when the test is replicated. Some contend that difference scores are more unreliable than the scores themselves (Lord, 1956) and that in most cases, difference scores correlate negatively with pretest scores (Linn & Slinde, 1977). That is to say, those with low pretest scores would tend to have large difference scores even in the absence of effects of intervention.

Most recently, the assumption of the lack of reliability of difference scores has been questioned (Rogosa & Willett, 1983). Such scores may have value if external information about the test-retest reliability of the measure is employed. For example, Webster and Bereiter (1965) proposed a reliability-weighted measure of individual change. Such alternatives have not been used widely in the research literature, which instead has turned to the use of residual change scores as the standard approach to the measurement of change.

Cronbach and Furby (1970) defined the residual change score as “primarily a way of singling out individuals who have changed more (or less) than expected” (p. 70). The key to understanding such analyses is in how the “expected” (or predicted) scores are calculated. An equation (a regression equation) can be developed that describes the relation between the posttest scores and the pretest scores for the entire sample; a linear relation between pretest and posttest is assumed in developing the equation. For each individual a score can then be calculated that represents the difference between the actual posttest score that individual obtained and the score that would be predicted for the individual inserting the individual’s pretest score into the regression equation. This difference score is termed the “residual.” As with any regression equation, demographic and other critical variables can be put into the equation first to respond to questions of change above and beyond that accounted for by differences in characteristics of sample participants (such as income or education). The critical question for evaluators is whether there is a relationship between certain aspects of early intervention services (say, intensity of service) and the size of the residuals. In other words, did individuals who received home visits more frequently change more than predicted? The prediction in this case is purely empirical, not based on clinical judgment.

Even advocates of this approach to measuring change admit that residual regression has limitations. Two specific drawbacks are relevant to the application of this approach.
to the evaluation of early intervention programs. First, as Rogosa, Brandt, and Zimowski (1982) point out, residual regression tells us little about how an individual actually changes on a certain dimension. Instead, it tells us how a person would have changed on that dimension if all persons in the sample had had the same pretest scores. The lack of utility of this hypothetical situation applied to children (or families) in early intervention programs is obvious, as variation not only exists but, in fact, is the very question of interest: How do various program features affect children and families of varying characteristics? Second, regression relies heavily on group data and large samples and is relatively insensitive to the individualized nature of services provided in most early intervention programs.

GOAL ATTAINMENT SCALING. In response to this criticism, a method termed Goal Attainment Scaling (GAS) (Kiresuk & Lund, 1976) has been developed that takes advantage of the evaluation of progress of participants toward unique goals. GAS is a procedure for the development of a quantitative outcome measure based on levels of success in meeting individualized goals of program participants. The GAS score is in essence a measure of how close the individual’s progress on a particular goal comes to that predicted by service providers. It has the advantage of being sensitive to unique goals and of providing a quantitative outcome that lends itself to statistical analysis.

There are several different approaches to GAS (e.g., Bailey et al., 1986; Romney, 1976), but each involves a similar series of steps. The first four steps occur before an intervention commences. First, a set of goals is specified for each child and family. Next, each goal is weighted, based on its priority, although several goals may be given the same weight. Third, a continuum of expected outcomes is developed from “worse than expected” to “better than expected”), using criteria for the attainment of each level. Fourth, initial performance is assessed for each objective. Finally, after the participant has received intervention services for a specified period of time, performance is again assessed to determine level of outcome can be made. Outcomes can be standardized (i.e., converted to a common metric that has a specified mean and standard deviation) so comparisons can be made and statistical tests applied.

The apparent attraction of GAS — its ability to incorporate the reality of programs that individualize services — may at first glance mask its shortcomings. Programs that have attempted to use this method of evaluation have not always met with success. One consistent problem has been the question of the validity of goals, especially in areas of development where there is no agreed-upon measure of progress, such as socioemotional development (Meisels, 1987). An equally perplexing problem became apparent in the application of GAS to the evaluation of the Child and Family Resource Program (Nauta & Hewett, 1988), a federally funded demonstration project. Goals were not always a reflection of a family’s true needs, and reliability in goal determination was questionable. Some have found correlations between staff on goal attainment decisions to be as low as .12 (Woodward, Santa-Barbara, Levin, & Epstein, 1978). The advantage of SEM is that it offers a way of examining the multiple influences on change in children and families. It provides a way of moving from a simplistic notion of the impact of early intervention to one that more closely represents its multifaceted nature. A schematic representation of the multiple variables and their empirical relations is an appealing feature of this approach.
Currently, only a few examples exist of structural equation modeling applied to analyses of programs for young children, and none has focused exclusively on early intervention services. Tietze (1987) followed a group of 203 children through 4 years of elementary school and analyzed the effects of attending a preschool program on retention and assignment to a special class placement. He specified a model that included various context variables (e.g., socioeconomic characteristics, religion, etc.), several process variables (e.g., class size, hours of instruction per week, etc.), and criteria as outcomes (assignment to a retention class, retention in grade, and special class placement). He found that enrollment in a preschool program had the strongest influence on school success.

Although SEM has many advantages over more simplistic analytical approaches, enthusiasm for it needs to be tempered by a realistic understanding of its limitations. In order to specify a sufficient number of assumed relationships among variables in a model, a large number of cases is required. Tanaka (1987) reports that some contend that as many as 200 cases are required to provide adequate estimates using SEM, although based on simulations, Gebring and Anderson (1985) have questioned whether such a large sample is mandatory. There is general agreement that the more variables specified in the model, the larger the sample size required. Sophisticated questions about the multiple effects of early intervention programs on children and families with differing characteristics will clearly require studies with large samples if SEM is the analytic strategy of choice.

The most important limitation of SEM, however, is conceptual, not statistical. In a variation of the “correlation does not imply causation” theme, Biddle and Martin (1987) contend that confirmation of a model does not imply the model's validity. Even with models that appear to fit the data well, we do not necessarily know if we have failed to specify a critical variable that accounts for the relations among other variables. Therefore, even when researchers claim a model has been confirmed, they need to consider other possible explanations of the results.

Finally, some concern exists about whether SEM will replace the theoretical development of models (Connell, 1987). In the extreme, our knowledge of human development would be data-driven and atheoretical. Like the application of other analytical techniques, SEM can be misused. The critical steps in determining which variables to include, hypothesizing their relations in the model, and interpreting results are important challenges that can best be met by those working from theoretical frameworks. The possibility of using SEM for more “fishing” or data exploration should not restrict its application to appropriate data sets by researchers armed with a set of specific questions. Further refinements in analytic technique may permit its application to smaller data sets with well-specified hypotheses.

DERIVING MEANING FROM DIVERSITY

When the full impact of P.L. 99-457 is felt, there is bound to be a heightened focus on the need to understand the effects of programs for infants and toddlers with special needs. Past evaluations, for the most part, have been unidimensional in their view of participants, programs, and impacts. More meaningful evaluations of early intervention services will occur only if those undertaking the evaluation enterprise attempt to develop new ways of understanding and explaining diversity.

In evaluating early intervention services, diversity occurs at many levels. First, children with disabilities are not all the same. Rather than relying on a categorical view of disabilities, typologies need to be developed that take into account a range of important characteristics, such as severity of disability and temperament or behavior. Prior research can guide decisions about the selection of those critical variables, but others are likely to emerge as research results accumulate. Second, families rearing a disabled child are not all the same. Our view of families needs to be multidimensional, include “protective” and “regenerative” factors (see Werner, this volume; Whittacker & Garbarino, 1983), and incorporate theories of family development and adaptation (see Kraus & Jacobs, this volume). Third, all early intervention programs are not the same, and services vary for families within programs. Evaluators need to find ways of describing models of service provision and ways of classifying services that are individualized and vary over time. Finally, outcome measures need to be determined that are responsive to program goals and that reflect a broad view of child development and family adaptation.

It is unlikely that any one evaluation of early intervention services can enlist and fully incorporate such diversity. Instead, evaluators should aim for the development of a series of carefully designed studies with well-chosen outcomes on a tightly determined set of subgroups. We can then search for patterns across studies that take advantage of different ways of defining subgroups, services, and outcomes. Rather than more unidimensional evaluations, investigators should provide evidence of the different ways early intervention services affect children and families. Evaluations can play a critical role in such knowledge accumulation.

REFERENCES

PART VII

Policy issues and programmatic directions
Handbook of early childhood intervention

Edited by SAMUEL J. MEISELS and JACK P. SHONKOFF

CAMBRIDGE UNIVERSITY PRESS
Cambridge
New York New Rochelle
Melbourne Sydney
To our families

Contents

Foreword by Edward F. Zigler
Preface
Contributors

PART I. INTRODUCTION

1 Early childhood intervention: The evolution of a concept
   Jack P. Shonkoff, Samuel J. Meisels

PART II. CONCEPTS OF DEVELOPMENTAL VULNERABILITY

2 Biological bases of developmental dysfunction
   Jack P. Shonkoff, Paul C. Marshall

3 Adaptive and maladaptive parenting – Implications for intervention
   Leila Beckett

4 The human ecology of early risk
   James Garbarino

5 Protective factors and individual resilience
   Emmy E. Werner

PART III. THEORETICAL BASES OF EARLY INTERVENTION

6 Transactional regulation and early intervention
   Arnold J. Sameroff, Barbara H. Fiese

7 Comprehensive clinical approaches to infants and their families:
   Psychodynamic and developmental perspectives
   Stanley I. Greenspan

8 A behavioral-ecological approach to early intervention: Focus on cultural diversity
   Lisbeth J. Vincent, Christine L. Salisbury, Phillip Strain, Cecilia McCormick, Annette Tessier

9 Implications of the neurobiological model for early intervention
   Nicholas J. Anastasio