INTRODUCTION

Rethinking Child Care Research

DOUGLAS J. BESHAROV
JEFFREY S. MORROW
American Enterprise Institute for Public Policy Research, Washington, D.C.

This introduction summarizes the articles in this collection. It describes how the articles address one or more of the key elements of the child care research model: (a) selecting and measuring the independent variables *to determine the characteristics ("qualities") of the child care environment (and, in some studies, the characteristics of parents and family), (b) selecting and measuring the* dependent variables *to determine the child's physical and developmental status after a period of time in a particular child care arrangement (usually a school year) compared with that of children in other arrangements (or simply the same child before spending time in the arrangement), (c) establishing* causal links *between the independent and dependent variables that are either assumed in randomized experiments or estimated through statistical controls in nonexperimental studies, and (d) assessing impacts across* subgroups *to see whether the program benefits one particular group more (or less) than others. The collection closes with a proposal to develop a systematic federal research program to pursue improvements in child care and early childhood education programs.*

Keywords: *child care; quality; outcomes; causal links; subgroup; randomized experiments*

As the number of working mothers has risen sharply, so has the number of children being cared for by those other than their parents. A rough estimate (because of limited data) is that, each year, parents and government combined spend about $50 billion on the child care (University of Maryland [UMD] 2006) of about 12 million children (U.S. Census Bureau 1999). And as more children spend increasingly large amounts of time in nonparental care, research on child care has taken on even greater importance both in the ongoing policy debate and for day-to-day program planning.

Existing child care research, much of it a decade or more old, can be summarized in two related statements: (a) Children do best when then they are placed in "quality" child care, and (b) the quality of most child care is only "mediocre" and therefore threatens the well-being of many children.

EVALUATION REVIEW, Vol. 30 No. 5, October 2006 539-555
DOI: 10.1177/0193841X06291522

Unfortunately, as the articles in this collection describe, much of the research on which these conclusions are based is plagued by methodological weaknesses that undermine the confidence that can be placed on these conclusions. Studies are compromised by substantial methodological problems, including inadequate specification and measurement of independent and dependent variables, probable selection bias and omitted variables, and the failure to conduct fine-grained subgroup or target-group analysis.

One might ask why this research weakness should be a concern. After all, who could be against improving the "quality" of child care? But what if this body of research exaggerates the importance of "quality" for the *average* child's development? Or what if it understates the importance of quality care to those children who are otherwise at risk of developmental problems? Raising quality (as currently measured) is very expensive, so that attempts to increase quality across the board would price many families out of formal child care and cause them to use even lower quality care (unless the government stepped in with sharply higher subsidies). And what if the concept of quality is being misspecified, so that only some components of what is now considered quality are important to child development? Worse, what if the *wrong* characteristics are being identified as important to child development? That would surely be a public policy tragedy.

In the years since the most influential of these studies were published, various researchers (often with foundation or government help) have developed better tools for exploring the connections between the characteristics of child care and later child outcomes. The articles in this collection explain the problems in past research, trace more recent efforts to improve research of quality child care, and identify additional steps that should be considered. In doing so, each of the articles in this collection addresses one of the key elements of the research model:

- selecting and measuring the *independent variables* to determine the characteristics ("qualities") of the child care environment (and, in some studies, the characteristics of parents and family);
- selecting and measuring the *dependent variables* to determine the child's physical and developmental status concurrently or after a period of time in a particular child care arrangement compared with that of children in other arrangements (or simply the same child before spending time in the arrangement);
- establishing *causal links* between the independent and dependent variables that are either assumed in randomized experiments or estimated through statistical controls in nonexperimental studies; and
- assessing impacts across *subgroups* to see whether the program benefits one particular group more (or less) than others.

The collection closes with a proposal to develop a systematic federal research program to pursue improvements in child care and early childhood education programs.

INDEPENDENT VARIABLES

The first major element in the child care research model is the selection and measurement of the relevant independent variables. Usually, they are considered to be subsumed in the "quality" of the child care environment (and sometimes family and parental characteristics, as well). What elements of the child care environment are likely to contribute to child development or harm it?

In "The 'Quality' of Early Care and Education Settings: Definitional and Measurement Issues," Jean Layzer and Barbara Goodson of Abt Associates discuss the ways in which the quality of children's care settings is assessed. They focus on the most frequently used global measures of child care quality: the Early Childhood Environment Rating Scale (ECERS) and its companion measures, the Infant/Toddler Environment Rations Scale (ITERS), and the Family Day Care Rating Scale (FDCRS). Because these measures are so widely used, they have dominated the discussion of child care quality for many years. The authors argue that reliance on them has led to inadequate measurement of the developmental qualities of child care settings that most directly influence child outcomes. Consequently, the "conclusions drawn about the relationship between these measures and outcomes for children are frequently incorrect or overstated" (p. 556). The authors recommend that these measures be replaced or supplemented with measures that better capture the critically important aspects of the care environment and the conditions children actually experience while in care.

The authors give another reason for being concerned about the ECERS:

> In addition to [its] wide use in research studies, [it is] increasingly used by state and local child care agency staff to assess the quality of child care settings. ECERS scores have been used as evidence that child care is, on the whole, "mediocre" (not a term used in the scale itself) as the basis for differential payments to providers; and as justification for investing additional money in quality improvement, among other things. (p. 567)

None of these problems, however, preclude using the ECERS as originally intended—as a tool or program inventory for providing technical assistance to individual child care facilities seeking to improve their programs. In fact, it is still a good tool for that purpose. But those who use it need to ensure that improvements in point scores reflect changes in critical factors such as caregiver interactions with children rather than simple improvement in materials and resources.

The ECERS is a 37-item scale originally developed as a tool used to target areas for improvement in *individual child care centers*. As Layzer and Goodson describe,

> Its scope is broad, covering aspects of the environment important for the comfort of staff and parents as well as children. The scale is based primarily on expert opinion and reflects a generous and expansive vision of what is necessary to create a comfortable and nurturing center environment for children. (p. 567)

The authors' overarching criticism of the ECERS and its companion measures is that they do not adequately measure "the experiences that promote children's physical, social, emotional, and intellectual development" (p. 558). Here are their main points:

- The theory of learning that underlies the ECERS is outdated and too narrow, because it focuses on "independent play," whereas contemporary thought ascribes at least equal importance to "a much more active role for the teachers, especially in providing additional language stimulation and instruction in early literacy knowledge" (p. 569).
- The global score that the ECERS develops combines ratings of a facility's resources and its physical and administrative characteristics with the programmatic characteristics that more directly influence child development (such as teacher quality and classroom activities). As a result, conclude the authors, it is possible to give a highly favorable rating to programs that offer minimal support for language and literacy acquisition or have other developmentally insignificant shortcomings. The reverse could also happen, with developmentally strong programs receiving poor ratings.
- The ECERS measures activities at the classroom level (which may indicate significant amounts of child-caregiver interaction) but fails to capture the experience of individual children (who may receive minimal caregiver attention or none at all).
- The ECERS provides only a one-time, snapshot picture of what is happening to the child and hence misses important factors such as length of time in the facility and caretaker turnover.
- The measurement reliability of the ECERS is questionable when researchers use separate subscales as differentiated measures of quality. There is also a possible "halo" effect, where classrooms rated highly on one subscale receive high scores on others—as though the structure of the measure leads raters to generate a single, unified view of the classroom.

- The questionable psychometric qualities of the ECERS can lead to its inappropriate use. The scale was not designed for statistical comparisons over time or across programs.

The strongest argument in favor of continued use of the ECERS is that, because it has been so widely used in the past, continuing to use it facilitates comparisons across studies and across time. As David Blau wrote, "If everyone develops their own special child care quality measure (as the NICHD study did), it is more difficult to compare findings across studies" (personal e-mail communication, January 15, 2005). Layzer and Goodson conclude, however, that the ECERS should be replaced or supplemented with measures that "offer a better look at variables that matter." Already, some federal projects are "assessing the quality of care by using environmental rating scales as well as new measures that assess the quality of the interactions between providers and children and the intentional teaching and structuring activities that support early learning (e.g., Preschool Curriculum Evaluation Research (PCER) studies; Quality Interventions in Early Care and Education (QUINCE))" (U.S. Department of Health and Human Services [HSS], Administration for Children and Families, Child Care Bureau, personal e-mail communication, March 16, 2005).

DEPENDENT VARIABLES

The second major element in the research model is the selection and measurement of the relevant dependent variables, that is, child outcomes. What aspects of the child's condition (in the various domains of physical, emotional, and cognitive development) could have been affected by the child care experience?

In "Child Outcome Measures in the Study of Child Care Quality," Martha Zaslow (Child Trends), Tamara Halle (Child Trends), Laurie Martin (Child Trends), Natasha Cabrera (University of Maryland), Julia Calkins (Child Trends), Lindsay Pitzer (Child Trends), and Nancy Geyelin Margie (University of Maryland) find shortcomings that can seriously limit the ability of the research to assess the impact of child care quality on child outcomes. According to the authors, studies rarely cover all the important child outcomes, vary widely in the measures used and in what is measured, often measure outcomes only tangentially related to the program characteristics measured (and vice versa), use child-based outcome measures whereas quality measures tend to be classroom-based, and often pay insufficient attention to the measures' psychometric qualities (especially validity and reliability).

An overarching concern is that the research, perhaps because of the interests of funders, has "a tendency to focus on broad rather than on specific models of how quality is linked to child outcomes" (p. 587). Hence, they call for more careful and consistent selection of outcome measures and greater alignment of outcome and quality measures so that "specific types of inputs (e.g., health and safety practices in care) [can be] related to specific outcomes (e.g., children's health outcomes)" (p. 586).

Zaslow and her colleagues begin by identifying five domains of child development using the elements of "school readiness" established by the National Education Goals Panel:

> (a) children's physical well-being and motor development, (b) social and emotional development, (c) language development and early literacy, (d) approaches to learning (including task persistence, curiosity, creativity, and enthusiasm in engaging with tasks), and (e) cognition and general knowledge (including conceptual development, knowledge of math concepts, and knowledge of specific information such as full name and address). (p. 580)

They then review 65 studies completed between 1979 and 2005 and reach the following conclusions about how each domain is addressed:

- Studies rarely cover all the important aspects of child development, particularly "health outcomes and such motivational aspects of learning (i.e., 'approaches to learning') as task persistence or enthusiasm" (p. 579).
- Few studies use the outcome measures that surveys of kindergarten teachers indicate they think are important for school readiness, including the child's physical well-being, and approaches to learning.
- Studies vary widely in the outcomes they measure and in the instruments they use to measure them, especially in regard to children's socioemotional development, thereby compromising cross-study comparisons.
- Some of the measures used as child outcomes (such as secure attachment to the child care provider) could actually be measuring features of the environment, or the reciprocal engagement of the child and the environment, rather than child outcomes per se.
- Many studies do not collect data on child outcomes (such as motor skills) that are related to the measures of child care quality they use (which, for example, may measure activities oriented toward motor development).
- The psychometric properties of measures (especially validity and reliability) are often unaddressed.

To improve research, Zaslow and her colleagues recommend the more uniform use of established measures and, within each domain, a concomitant reduction in the number of constructs and measures used. For less established measures, they would like to see more information about the measures' psychometric qualities (especially validity and reliability). The authors also recommend greater use of measures developed for research on early childhood interventions. Their review of studies currently in the field suggests some that some have "the potential to strengthen our understanding" of the connection between children's development and child care quality.

The authors conclude by calling for a greater "alignment of the two parts of the equation, namely, moving from implicit logic models of how quality is linked to child outcomes to explicit models and from broad logic models to more differentiated and specific ones" (p. 599). This brings us to the next element in the research equation.

CAUSAL LINKS

The third major element of the model is causation. Even when measures of child care quality and of child outcomes are satisfactory (and sufficiently aligned), causation still must be established: Were the observed outcomes caused by the characteristics of the child care experience? This is a difficult question to answer because of the well-established association between family and parental characteristics and the type and quality of care they select; that is, parents with higher levels of income and education tend to have better child-rearing skills and attitudes as well as tending to select child care providers that score higher on various existing measures of quality.

Although hardly problem-free, randomized experiments are the most reliable way to establish causation. Unfortunately, as the articles in this collection reflect, relatively few randomized experiments have been conducted—largely because of the difficulties involved in randomly assigning children to different forms and qualities of child care and keeping them there. As a result, almost all major research on the impact of child care quality on child outcomes has been correlational in nature—with statistical methods used to control for selection bias and omitted variables.

In "Connecting Child Care Quality to Child Outcomes: Drawing Policy Lessons From Nonexperimental Data," Greg Duncan of Northwestern University and Christina Gibson-Davis of Duke University examine the use of nonexperimental studies to determine the impact of child care quality on child outcomes. They describe the problems that nonexperimental studies encounter concerning selection bias and omitted variables, attrition, and small effect sizes. They conclude that these problems need to be addressed more successfully if nonexperimental studies are to be used for policy making and make a series of suggestions about how to do so.

All nonexperimental studies raise "the nasty specter of the omitted-variables problem," write Duncan and Gibson-Davis. "The omitted-variable problem arises if difficult-to-measure characteristics of the child, mother, or family environment . . . are correlated with both choice of child care quality and children's cognitive development" (p. 613). Failing to control sufficiently for such variables opens the door to either upwardly or downwardly biased results. For example, parents who work hard to promote their child's development may, in doing so, select high-quality care. Conversely, although less likely, parents of children with the greatest developmental needs may seek out the highest quality care.

To illustrate this point, the authors use their experience with the National Institute of Child Health and Human Development's (NICHD's) Study of Early Child Care, the "most ambitious longitudinal study of child care quality." In ten different sites, the NICHD study has followed a cohort of about 1,200 children from birth to elementary school. It is "unique in its comprehensive measurement of the family and child care environments of children in its sample, and its results have been widely cited" (p. 612). However, as the authors describe, that has not made it immune from the problems encountered by most nonexperimental studies.

The authors find that controlling for all "family-based control variables," not just the ones NICHD researchers had previously used, "reduced the association between child care quality and 54-month cognitive development by roughly one half and that the bulk of the reduction occurred when maternal schooling was first entered into the regression" (p. 617). Because the added family factors are a random subset of potentially relevant factors, the authors conclude that "further controls for unmeasured selection factors may matter just as much" (p. 618).

Although there is "no truly convincing solution to the omitted-variables problem," the authors recommend a number of steps that can at least reduce the problem (p. 615):

- Measure as many family variables as possible and include them in the regression analysis (while being aware of the risk of multicollinearity).
- Use change models to hold family inputs constant (while being aware that family conditions can change over time and of the greater risk of measurement error).
- Use sibling models (sometimes called "family fixed effects") to difference out family inputs (while being aware that family, program, and environmental factors can change over time).

- Identify instrumental variables that are highly correlated with child outcomes and the quality of care but that affect these outcomes only through their impact on the quality of care.
- Use propensity scores to create "matched groups" of children for comparison who have similar characteristics other than their child care.
- Use regression discontinuity models to compare outcomes of children who did and did not receive services based on an exogenous characteristic not directly associated with those outcomes (e.g., children with birthdays before versus after the cutoff dates for prekindergarten).

Duncan and Gibson-Davis also discuss the sample attrition suffered by "virtually all longitudinal surveys" and recommend that "when study non-response is substantial, efforts should be made to investigate whether non-response might be biasing parameter estimates and, if so, to take measures to adjust for the bias" (p. 627). In the NICHD study, for example, the cumulative response rate fell to 52.5%, raising concerns about attrition bias.

Duncan and Gibson-Davis also discuss "the need to translate 'effect sizes' derived from these studies into the kinds of cost and benefit information needed by policy makers" (p. 611). That is, the need to answer the "key" question: Are the benefits of increased quality worth the costs? As they note, "small effect sizes that are inexpensive to generate may well be worth it, whereas big effects from expensive interventions may not be" (p. 625). (This becomes an even more important question as more effective corrections for selection effects result in smaller effect sizes.)

"In the absence of a natural metric," Duncan and Gibson-Davis recommend that "a useful method for expressing the impact of child care quality on child outcome is to scale both child care quality and child outcomes in [standard deviation (*SD*)] units" (p. 626). As an example, they apply this approach to the NICHD study. The result is enlightening: A 1 *SD* in quality results in only a 0.04 to 0.08 *SD* increase in school readiness. The policy question is whether the benefit of a 0.04 to 0.08 *SD* boost in cognitive scores is worth the cost of raising child care quality by 1 *SD*.

> If it cost $5,000 per year for several years to produce the gains observed in the NICHD study, then it may make sense to consider cheaper alternatives (e.g., supplementing family income directly or expanding highly targeted pre-K programs) for increasing children's test scores. (p. 626)

In closing, the authors warn against studies that claim causal findings when their design does not support them. Many studies couch their results in terms of "associations" and warn that it is "impossible to draw causal inferences from the analysis," but some are not that careful. Even when they are, policy makers have drawn causal links from the data. What to do? The authors give the following advice:

> No single approach will provide truly convincing estimates of causal impacts from nonexperimental data such as these, but policy makers cannot wait for the needed experimental studies. The best strategy . . . is to push the data as far as possible toward the goal of securing convincing conclusions about causation, search for robust findings across the set of studies, and then consider the costs and benefits associated with consensus estimates of the impact of child care quality. (p. 627)

Even this approach, of course, has its limits. A systematic or crosscutting error or bias, for example, may similarly affect the estimates of all or many studies.

SUBGROUP ANALYSIS

If, as Duncan and Gibson-Davis conclude, carefully controlling for as many family and parental characteristics as possible substantially reduces the estimated impact of child care (good or bad) on child outcomes, does that mean that the quality of child care is unimportant? What about those children who come from subpar environments? This brings us to the fourth element of the model: assessing impacts across subgroups.

In "Family Factors in Child Care Research," Anne Hungerford of the University of North Carolina at Wilmington and Martha Cox of the University of North Carolina at Chapel Hill review studies of the correlations among family characteristics, child care experiences, and child development (because of the paucity of randomized experiments for all but a few hothouse programs). They find that, generally, family and parental characteristics ("particularly socioeconomic status and parenting quality") are more predictive of child development than the "quality" of child care.

But, they also note, quality child care and early childhood education programs may be more important for identifiable subpopulations of children "who are at risk for poor outcomes because of unfavorable family environments." For such children, they recommend "intensive services" and the development of "effective interventions that are tailored to their needs."

Hungerford and Cox begin by examining the research on the influence of family and parental characteristics on child development. They warn that most of the research on the subject is correlational in nature. Even though efforts to control for selection bias are not as successful as they would like, certain patterns emerge with sufficient regularity for the authors to conclude that the most important elements in child development are family and parental factors.

According to the authors, the research also suggests that parenting during infancy and the preschool years is most predictive of the future parent-child relationship and that this relationship correlates to later child development. For example, stability and confidence in the context of the early parent-child relationship is correlated with the child's greater self-confidence and self-reliance. In addition, various family characteristics such as parental personality and social support networks are associated with child outcomes—as are income and maternal educational attainment.

Moreover, a growing body of research suggests that genetic factors, and especially the interaction between genes and the environment, "influence both parenting and child behavior." Consequently, particular kinds of parenting are likely to have different impacts on children, depending on the child's innate characteristics. And the feedback from child to parent is similarly likely to differ.

Having established the predominant influence of parent and family characteristics on child development, Hungerford and Cox compare these impacts against those of the child care environment. They conclude that "family experiences are typically a much stronger predictor of children's developmental outcomes than is child care" (p. 632). But—and here is the key to their thinking—"Evidence from both nonexperimental studies and experimental child care interventions suggests that child care experiences may have stronger effects for some children than for others" (p. 641). In other words, the quality of child care is likely to be more important "for young children who are at risk for poor outcomes *because of unfavorable family environments* [italics added]" (p. 632).

Hungerford and Cox recommend, therefore, that

> greater attention should be paid to the heterogeneity of low-income families . . . and that child care interventions need to be tailored to meet the needs of different groups of low-income families. For example, intensive services are likely to be warranted for families with multiple risk factors in addition to low income. (p. 648)

But identifying such children and families—and the services that they need—is a daunting task and probably impossible without randomized experiments. The next article takes on the question of randomized experiments.

GROUP RANDOMIZATION

Underlying both the Duncan/Gibson-Davis and Hungerford/Cox articles is a recognition that nonexperimental studies cannot satisfactorily resolve the problems of selection bias and omitted variables and that randomized experiments would do a much better job disentangling the causal relationships among the parental and family characteristics, the characteristics of child care, and child outcomes. But is it possible to mount more randomized experiments?

In "Randomize Groups, Not Individuals: A Strategy for Improving Early Childhood Programs," Robert St.Pierre of Abt Associates and Peter Rossi of the University of Massachusetts (emeritus) affirm that only randomized experiments can satisfactorily address the problems of selection bias and omitted variables. But recognizing the difficulty in mounting successful randomized experiments at the child or family level, they call for more projects that randomize at the group or site level. And because almost all disadvantaged children now receive some form of child care or early childhood education, the object would not be to discover whether a program works but whether it works better than other programs ("what works" versus "what works better").

The authors also decry research on child care because it "lacked coherence and paid little attention to what had worked (and not worked) in the past" (p. 656). Their "central message" is the need to "restructure" research efforts within a new federal office that conducts "an extended series of group-randomized experiments that compare promising variations of preschool programs" (p. 675).

St.Pierre and Rossi begin by describing how existing child care research and program evaluations are "deeply flawed." Studies with "weak program and evaluation designs" overclaim causation (and impacts) and fail to provide a platform for cumulative learning. According to the authors, "the research overly relied on weak, quasi-experimental approaches, such as posttest-only and pretest-posttest designs; correlational analyses of observational studies; and case studies" (p. 660).

Concerned about the same problem, federal Head Start and child care research agencies have recently increased their support for randomized experiments. Here is a sample:

- Preschool Curriculum Evaluation Research (PCER) evaluates an experimental preschool curriculum by randomly assigning children and classrooms to either the experimental curriculum or to the existing (control) curriculum (U.S. Department of Education n.d.).
- Quality Interventions for Early Care and Education (QUINCE) evaluates two programs that provide on-site consultation to early-childhood caregivers

by randomly assigning (a) children to teachers trained under the experimental program or not (the control) and (b) teachers who are trained under the experimental program to child care providers (centers and family-based) with the control being current teachers who continue with the existing program (HSS, personal e-mail communication, May 31, 2005).

- Project Upgrade–Miami-Dade County evaluates early language and literacy curricula, as well as the center's ability to implement such curricula, by randomly assigning child care centers to "one of the three different curricula or to a control group" (HSS, personal e-mail communication, May 31, 2005).
- Massachusetts Family Child Care Provider Study evaluates the Learningames curriculum in family child care settings by randomly assigning home-based care providers either to the experimental curriculum or to a continuation of their existing mentor visits (HSS, personal e-mail communication, May 31, 2005).
- Head Start and Early Head Start Evaluations evaluate the impact of each program by randomly assigning children on the waiting lists of individual programs to treatment and control groups (Advisory Committee on Head Start Research and Evaluation 1999; Love et al. 2002).

Like the other authors in this collection, St.Pierre and Rossi would like to see more such randomized experiments, but they are dubious of the ability to mount randomized experiments at the individual family or child level. They note the practical as well as ethical reasons why it is difficult to randomly assign children to different forms and qualities of child care (and keep them there) to see which does the most good and which might be harmful.

Moreover, random assignment off waiting lists, the most common approach to randomization, is an imperfect solution because, among other things, it (a) only serves families that want the program's services, (b) is plagued by high dropout rates, and (c) assumes that the members of the control group (and dropouts) will not receive similar services.

The traditional evaluation of a particular program requires that the group that received the treatment be compared with one that did not—and that also did not receive a similar treatment from any other program. However, as the authors point out, most communities now offer several forms of child care to all or most disadvantaged children, and consequently, many control group children receive similar services to those in the experimental group. (Recent experimental studies of Even Start, Early Head Start, and the Comprehensive Child Development Program [CCDP] all found children in the control group receiving other, often substantial, preschool education or child care.) The result is a control group that combines those receiving no treatment with those participating in competing programs, or, in the authors' words, a control group of individual children who "are *not* subject to 'no-treatment' but to a mixture of other treatments" (p. 664).

The result can be a reduction in the measured impact of the experimental treatment because it is determined relative to the control group's mixed-treatment experience. The mix-treatment experience of the control group is rarely reflected in cost analyses. Unfortunately, that provides "no systematic way of determining which approach is most effective or which program is the best use of taxpayer dollars" (p. 666).

Therefore, St.Pierre and Rossi recommend that the randomization be performed at the group or site level, not at the family or child level. A major advantage of group randomization is that it minimizes the problem of "crossover" and the resultant mixed-treatment-and-control groups. Although the problem is still present in group-randomized studies, it has less impact on them because they "include all children in a desired intervention as an integral part of the design" (p. 674). Group randomizations also reduce the problems caused by program dropouts, argue the authors, "because dropout is a normal program occurrence and, in fact, is an indicator of program acceptability. . . . Although studies based on group randomization may suffer from the dropout of entire sites, such an occurrence is rare" (p. 674).

Another advantage of group randomization, according to the authors, is that it reduces the usual antipathy of program operators towards randomized experiments that appear to deny services to eligible and needful children. Because

> every site is assigned to an intervention . . . all arguments about ethics and legal constraints are circumvented. Every site gets an intervention, no one is denied service, no parents are upset about their child being assigned to a no-treatment control group, and no program staff are upset about losing control over who participates. (p. 672)

The biggest obstacle to group randomization studies is cost. They are significantly more expensive than individual randomization studies because they require many more sites (and, as a result, more children) to obtain reliable results. As the authors note, "to achieve a given level of statistical power, more sites must be used and more individual children must be involved than in studies in which the individual child is the unit of analysis, multiplying the costs of data collection" (p. 673). Where five sites might be deemed acceptable for an individual randomization, fifty might be needed for a group randomization.

Despite higher costs, the advantages of group randomization often outweigh their disadvantages, and they are becoming more common. The authors note

> more than 200 'cluster randomized trials,' which use communities, schools, classrooms, or other organizations as the unit of allocation in randomized field trials of a wide variety of interventions (e.g., universal free breakfast,

> crime prevention, smoking cessation, substance abuse avoidance, violence reduction, nutrition education, fertility control, mathematics education, health education for the elderly, and reduced class size). (p. 671)

Shifting from individual to group randomization, however, "changes the research question from 'What works?' to 'What works better?' " (p. 656). The authors acknowledge that doing so means that it would not be possible to know whether "either program works better than no program," which they accept because they think the issue is no longer whether a child should be in child care or an early childhood education program but how to make such programs as good as possible.

A NATIONAL PLAN

St.Pierre and Rossi advocate a sustained and scientifically rigorous inquiry into the "comparative effectiveness" of various curricula, as well as a wide range of

> structural interventions for early childhood education centers, such as hours of operation, full-day versus part-day programs, 1- versus 2-year programs in which children enter either at age 3 or age 4, variation in the length of the school year (traditional 9-month versus full-year), variation in the size of preschool classrooms (paralleling work on class size done at the elementary level), variation in the training or formal education of early childhood education teachers, variation in the socioeconomic composition of classes (mixed versus homogeneous), or variation in the age mixture within classrooms (mixed versus homogeneous classes). (p. 657)

The authors are also strongly critical of the "haphazard fashion" in which the federal government has conducted the research in this area. Using Head Start as an example, they describe how "even when randomized experiments were used in studies of federal early childhood programs, they typically were conducted as stand-alone efforts, one large study at a time, with limited interconnections and, hence, little cumulative learning" (p. 660). This can be within programs as well across programs and is less about coordination and more about learning from the past. Head Start's CCDP in the 1990s, for example, largely duplicated the Child and Family Resource Program from a decade earlier, which had already been found ineffective.

To provide "credible information useful for designing effective programs," St.Pierre and Rossi conclude, "ongoing, systematic development and evaluation of alternative approaches for the improvement of large-scale early childhood programs" is required (p. 656). To mount such an effort, they propose an Early Childhood Education Systematic Improvement Effort (ESIE), a "long-term, carefully planned, systematic series of experiments firmly anchored in a powerful federal agency and designed to improve early education for preschoolers from low-income families" (p. 676).

Adopting the St.Pierre/Rossi plan would require substantial intellectual and political effort—because of the turf battles it would trigger, the scientific challenges involved in designing the program, the expense of so many multisite experiments, and the sustained monitoring and management needed. Nevertheless, without an effort on this scale and without intellectual clarity, it is difficult to see how research can identify the best approaches to child care and early childhood education.

CONCLUSION

Collectively, these articles provide a comprehensive overview of the weaknesses of the research on the "quality" of child care and early child education programs and of recent efforts to improve it. In effect, the authors call for a rethinking of all major aspects of this research. How realistic is it to hope that steps can be taken to improve child care research? With the passage of the No Child Left Behind law, the U.S. Department of Education embarked on the same kind of systematic research recommended by the authors. Why shouldn't early childhood services be next?

REFERENCES

Advisory Committee on Head Start Research and Evaluation. 1999. *Evaluating Head Start: A recommended framework for studying the impact of the Head Start program.* Washington, DC: U.S. Department of Health and Human Services. http://www.acf.hhs.gov/programs/hsb/research/hsreac/1999report/ (accessed June 1, 2005).

Love, J. M., E. E. Kisker, C. M. Ross, P. Z. Schochet, J. Brooks-Gunn, D. Paulsell, K. Boller, J. Constantine, C. Vogel, A. Fuligni, and C. Brady-Smith. 2002. *Making a difference in the lives of infants and toddlers and their families: The impacts of Early Head Start.* Executive summary. Report for the U.S. Department of Health and Human Services, Administration on Children, Youth and Families. Washington, DC: Mathematica Policy Research.

University of Maryland (UMD), Welfare Reform Academy. 2006. *Early education and child care (ee/cc) model.* College Park: UMD, Welfare Reform Academy.

U.S. Census Bureau, Population Division. 1999. *Who's minding the kids? Child care arrangements: Spring 1999 in Survey of Income and Program Participation.* Washington, DC: U.S. Department of Commerce.

U.S. Department of Education, Institute of Education Sciences, Preschool Curriculum Evaluation Research. N.d. *What is the Preschool Curriculum Evaluation Research (PCER) Program?* Washington, DC: U.S. Department of Education. http://pcer.rti.org/aboutPCER.htm (accessed June 1, 2005).

Douglas J. Besharov, a lawyer, is the Joseph J. and Violet Jacobs Scholar in Social Welfare Studies at the American Enterprise Institute for Public Policy Research in Washington, D.C. He is also a professor at the University of Maryland School of Public Policy, where he teaches courses on family policy, welfare reform, program evaluation, and the implementation of social policy. From 1975 to 1979, he was the first director of the U.S. National Center on Child Abuse and Neglect. His most recent book is Recognizing Child Abuse: A Guide for the Concerned, *a book designed to help professionals and laypersons identify and report suspected child abuse. He has written or edited fourteen other books and has contributed to the* Washington Post, Wall Street Journal, New York Times, *and* Los Angeles Times.

Jeffrey S. Morrow is a research assistant at the American Enterprise Institute for Public Policy in Washington, D.C. He completed his BA and MA in Politics at Brandeis University. His research areas include child care, welfare, and student aid. He has coauthored, with Douglas Besharov, reports on Head Start enrollment and on child care data in the Census Bureau's Survey of Income and Program Participation. In addition, he helped to conceive, design, and implement the Early Education/Child Care Model, a fully adjustable Excel-based model showing the eligibility and receipt of child care subsidies under current and user-created scenarios.