

# **WHAT TO MAKE OF FAMILY PRESERVATION SERVICES EVALUATIONS**

## **ABSTRACT**

This paper discusses the difficulties that have characterized Family Preservation Services (FPS) evaluations. These difficulties arise out of four challenges: 1) FPS's multiple purposes and the variety in the forms it takes in implementation; 2) the lack of explicit "theories of change" that encompass these many goals and beneficiaries; 3) the challenges of defining FPS and replicating particular models; and 4) the fact that many evaluations have suffered because of lack of access to these data, technical expertise, time, money, and political and administrative support. Reviews of the evaluations have noted their methodological limitations, especially issues of research design and program implementation.

The evaluations themselves fall into three "waves." The first consisted of small-scale efforts, focused on individual programs and then expanded to include several programs within particular states. These evaluations were nonexperimental, post-test only or pre- and post-test designs and boasted strongly positive findings. These evaluations, however, lacked comparison data and suffered from a reliance on placement rates as the measure of success. The second wave of evaluations responded to many of the criticisms leveled at earlier studies and is characterized by larger scale, controlled studies that also broadened the range of measured outcomes. In this wave, few statistically significant results surfaced, but other methodological challenges arose, including treatment fidelity and control group activity, targeting, heterogeneity of clients, sample size, and the differential attrition from treatment and control groups. The third wave of evaluations has brought further refinements in evaluation questions, research design, outcome measurement, and analysis, though large-scale external evaluation has diminished in anticipation of the multiyear, multisite, federally-funded National Evaluation of Family Preservation and Reunification Services, which focuses on the differential effects of FPS on subgroups, intending to identify the characteristics of "successful families" and the complement of services that are offered them.

Family preservation services to date do not appear to have a significant effect on placement rates. The results seem to be that, at best, FPS provides small and likely short-term effects on placement prevention and perhaps positive effects in certain discrete areas of family and child functioning. No substantial, enduring effects on children, families, communities, or agencies have been documented. Despite the very modest evaluative support for FPS, there continues to be broad-based, ongoing support for family preservation services. The paradox suggests that family preservation programs have yet to be evaluated adequately and the direct impact of research on policy has been limited.

Both FPS and its evaluations were repeatedly oversold, reflecting a broader context of scant resources. The paper concludes with some recommendations for conducting evaluations of this type of program and with observations on the relationship of evaluation or research to practice and policy.

Francine Jacobs

July 2001

## Chapin Hall Center for Children

The Chapin Hall Center for Children at the University of Chicago was established in 1985 as a research and development center dedicated to bringing sound information, rigorous analyses, innovative ideas, and an independent perspective to the ongoing public debate about the needs of children and the ways in which those needs can best be met.

The Center focuses its work on all children, while devoting special attention to children facing special risks or challenges, such as poverty, abuse and neglect, and mental and physical illness. The contexts in which children are supported—primarily their families and communities—are of particular interest.

Chapin Hall's work is shaped by a dual commitment to the worlds of research and policy. This requires that our work meet both the exacting standards of university research and the practical needs of policy and program development, and that we work to advance knowledge and to disseminate it.

Chapin Hall is committed to diversity not only of race, ethnicity, gender, sexual orientation, and physical ability but also of experience, discipline, and viewpoint.

Chapin Hall's commitment to all children, with special attention to those experiencing or at risk of serious problems, is reflected in the range of the Center's research projects and publications. The following represent the Center's major areas of endeavor:

- Children's services, covering the problems that threaten children and the services designed to address them, including child welfare, mental health, and the juvenile court
- Primary supports, concerning the resources in communities that enhance the development and well-being of all children, but may be especially important to children who are poor or otherwise disadvantaged
- Community building, focusing on the development, documentation, and evaluation of community-building initiatives designed to make communities more supportive of children and families
- Schools' connections with children's services, primary supports, and communities, examining the range of institutions that foster children's learning and development, including not only schools but also families, businesses, services, communities, and the relationships, actual and potential, among them
- International projects, covering Chapin Hall's recent and growing involvement with children's policy researchers and research centers in other countries
- Special activities and consultations, covering a range of projects, often undertaken in collaboration or consultation with other organizations

Discussion Paper CS-70

© 2001 by Chapin Hall Center for Children

A complete list of Chapin Hall publications is available at our Web site, [www.chapin.uchicago.edu](http://www.chapin.uchicago.edu).

Phone: 773/753-5900 Fax: 773/753-5940

ISSN: 1097-3125

## TABLE OF CONTENTS

Introduction .....	1
How FPS Challenges Evaluators .....	1
What We Know and When We Came to Know It. ....	4
The First Wave of Evaluations .....	5
The Second Wave of Evaluations .....	7
The Third Wave of Evaluations .....	10
The Bottom Line on Effectiveness .....	10
FPS Has Not Been Adequately Evaluated .....	11
Research Has Only Limited Direct Impact on Policy .....	13
The Future of FPS Evaluation .....	14



# What to Make of Family Preservation Services Evaluations

---

Francine Jacobs

This essay is part of a set of three that examines the history and legacy of family preservation services. The essays—*What Is Family Preservation and Why Does It Matter?* by Jacquelyn McCroskey, *The Shifting Policy Impact of Intensive Family Preservation Services* by Frank Farrow, and *What to Make of Family Preservation Services Evaluations* by Francine Jacobs—grew from the realization that, whereas family preservation services have matured and become a standard part of child welfare, if not human services, they continue to evolve and to elude clear description. It is our hope that these essays add to family preservation’s development and to a shared understanding of its importance to the field.

## INTRODUCTION

It shouldn’t have been such a hard job—evaluating family preservation services. Many of the necessary elements for successful evaluation were in place: on the program side, Homebuilders and its immediate descendents appeared as promising, innovative, and standardized interventions. They boasted clearly stated goals, a specific clientele, willing and able program personnel (local and state)—many with considerable experience implementing the model—and enthusiastic supporters able to secure ample funds for program implementation and expansion. Willing, able, and experienced researchers stood ready on the evaluation side, with substantial funding for their efforts and numerous audiences eager for their products. A match made in heaven, one might imagine. Yet instead, FPS has bedeviled evaluators, leaving the field without a definitive answer

to the fundamental question—“does it work?”—posed across hundreds of evaluations over the past two decades.

Why has this task proven so difficult, virtually impossible? The answer is multi-determined, having to do with the program itself, the political climate in which the program was developed and has operated, the nature of the evaluations that have been undertaken, and the state of practice in the evaluation field. Although the body of research reviewed here is disappointing to many in its inability to offer conclusive findings, there are valuable lessons to take from these brave, but sometimes naive, attempts to assess family preservation services; I hope to highlight several of these in this essay.

This paper begins with a brief review of the challenges to evaluating FPS, then describes three eras, or “waves,” of FPS evaluation, summarizing and interpreting the findings to date. It then offers some recommendations for conducting evaluations of this type of program, and concludes with a few final observations on the relationship of evaluation or research to practice and policy.

## HOW FPS CHALLENGES EVALUATORS

Family preservation services present a host of significant challenges to researchers. Some of these challenges are inherent in the program and pertain to its goals, rationale for services, and implementation. Others pertain to the context in which the evaluations are undertaken and the resources available for the job. Several of these challenges are noted below.

Fashioned after the prototype program—Homebuilders of Tacoma, Washington—FPS

## Acknowledgments

*Behind this essay and the two others that share in its examination of family preservation are numerous people who helped shape this work. First, the editors thank Susan Notkin of the Edna McConnell Clark Foundation for supporting and guiding the project. Her patience and judgment were indispensable as the project continually transformed in our effort to define family preservation and capture what it means for human services. Thanks, too, to the many scholars and practitioners who offered their insights in the initial steps of the project: Douglas Besharov, Barbara Blum, Charlotte Booth, Charles Bruner, Steven Budde, Frank Farrow, Peter Forsythe, Mark Fraser, David Haapala, Robert Halpern, Francine Jacobs, Susan Kelly, Jill Kinney, Duncan Lindsey, Tony Maluccio, Peter Pecora, Peter Rossi, Lisbeth Schorr, Matthew Stagner, Gary Stangler, Cecelia Sudia, Elizabeth Tracy, Lynn Usher, Kenn Visser, Michael Wald, Michael Weber, Carol Weiss, Kathleen Wells, James Whittaker, Carol Williams, and Ying Ying Yuan. In preparation for an ensuing symposium, five authors wrote papers that were critical in shaping the days’ discussions: Frank Farrow, Francine Jacobs, Jacquelyn McCroskey, Martha Shirk, and James Whittaker. The symposium participants included Charlotte Booth, Francine Jacobs, Susan Kelly, Jacquelyn McCroskey, Kristi Nelson, Susan Notkin, Harold Richman, Lisbeth Schorr, John Schuerman, Martha Shirk, Gary Stangler, and James Whittaker. Their perceptiveness and engagement provided for lively debates and for an enhanced, if not always shared, understanding of family preservation.*

*The project also owes a great debt to Pattie Bengston who provided administrative and logistical support for the project. The project also thanks Susan Campbell, Jeff Hackett, and Laurie Juliana for their careful editorial support.*

*Finally, this work owes its existence to the Edna McConnell Clark Foundation, which supported this effort as a way of examining family preservation’s history and legacy, including but also beyond its own role.*

was initially advertised as a singular, programmatic entity with a clearly defined mission. This set of intensive, short-term, home-based services was to be provided to qualifying families in which serious child abuse or neglect had been substantiated. It was intended to prevent the otherwise inevitable protective placement of children in these families into care arrangements outside their own homes (Kinney et al., 1991). Guaranteeing child safety while maintaining the family unit was seen as the primary responsibility of family preservation workers. In order to satisfy these goals, specific guidelines were developed about how FPS was to be delivered, to which families, with the expectation that they would, and could, be followed.

In truth, FPS has never been, and certainly is not now, so simply wrought and faithfully implemented. To begin, there are multiple purposes for the program. Of course, ensuring child safety is paramount, but other goals have surfaced insistently over the years. These include, for example, enhancing child development, increasing parental competence, improving family functioning, reducing child welfare expenditures, and enhancing cross-agency collaboration. This expanded list of goals suggests that a range of groups or entities may profit from FPS. Thus children, mothers, and fathers (both in their parenting roles and as adults in need of a range of supports), families,

*... why should a particular program, configured in a particular way and targeted to a specific population, achieve the desired results? In the case of FPS, we might ask for the assumptions about child and adult development and about how change occurs in individuals and families that would support core aspects of the program model.*

communities, public welfare agencies, and even statewide child-serving agency systems, are all potential beneficiaries of the program. No doubt these multiple goals with multiple beneficiaries more accurately reflect what FPS programs across the country are actually trying to accomplish; they also have con-

siderable political value, since there looks to be something for everyone in FPS. However, this expansion of mission has also exponentially complicated the design and execution of evaluation.

A second challenge is the lack of explicit “theories of change” that encompass these many goals and beneficiaries. In the past several years, Weiss (1995, 1998), Rossi & Freeman (1993), and other evaluation theorists have focussed on the need for program developers and staff to explicate the intervention theory on which the program is built. (See also Bickman, 1987; Chen, 1990; and Chen & Rossi, 1987.) That is to say, why should a particular program, configured in a particular way and targeted to a specific population, achieve the desired results? In the case of FPS, we might ask for the assumptions about child and adult development and about how change occurs in individuals and families that would support core aspects of the program model (e.g., intensive, short-term services; enrollment of families at imminent risk in which a minimal level of safety can be assured; services delivered at home). This information is crucial to evaluators engaged in outcome research in the human services. To the extent that programs can articulate this theory—A happens to cause B, which usually yields C—evaluators can track the course of change, identifying intermediate indicators or markers of effect. Should the desired, ultimate change not occur even though intermediate effects are noted (for example, families are functioning better, but placement rates remain high), then the theory can be revised and other programmatic approaches tried.

The proliferation of goals for FPS has occurred without the accompanying explication of intervention theory. For example, the pathways through which receipt of FPS for a certain percentage of neighborhood residents should result in better outcomes for the community have not been proposed, yet some believe that FPS increases community safety. FPS is thought to enhance child development, but the mechanism by which this occurs, say for infants in marginal, neglectful homes, is nowhere stated. And indeed, one might argue that even the initial goals and service delivery structure of FPS were not sufficiently informed by developmental or systemic theory. In any case, the

greater number of goals and beneficiaries, the worse the problem becomes for evaluators, who are committed to identifying effects—even intermediate effects—if they are occurring.

A third challenge resides in the numerous definitions of “family preservation services,” and in the belief that, when there is agreement about what it is, that precise entity can and will be replicated. The field learned late that critical components of FPS, for example enforcing the condition of “imminent risk” as a criterion for enrollment, were not being achieved across cases, communities, and states. This program implementation issue is at the core of many of the contested evaluations in the field.

The fact that FPS is not the straightforward, controlled intervention that many program proponents insisted it was, or could be, is particularly bad news for evaluators. Multiple goals and beneficiaries usually yield a dizzying array of evaluation questions, most of them legitimate. If the community is to benefit from FPS, for example, it may be important to ask whether and to what extent community residents have had a voice in how services are provided, or whether families enrolled in the program think FPS is useful. These questions suggest the importance of process evaluation. Private funders, on the other hand, might only be willing to pay for outcome studies. Program administrators might be most interested in questions that pertain to program coverage and utilization; in this case monitoring- and implementation-oriented evaluations are in order. And with multiple questions come multiple audiences: for FPS these include providers, policy makers at all levels of government, citizens, program participants, researchers and funders. With so many masters, evaluators inevitably satisfy one group and disappoint others.

Finally, the success of any evaluation depends on the availability of a set of critical research resources, including *data, technical expertise, time, administrative and political support, and funds* (Jacobs & Kapuscik, 2000). Certain FPS evaluations have enjoyed access to these resources, but most have been constrained by their absence. For example, FPS evaluations have suffered from the lack of adequate and appropriate data; for many of the early years of research, state child welfare

agency MIS systems were poorly organized or simply nonexistent. This situation has improved, but there are still many states with inadequate MIS systems. Appropriate instruments to measure progress in outcome areas outside the usual—for example, to examine family functioning or community well-being—are scarce. And numerous FPS eval-

*The fact that FPS is not the straightforward, controlled intervention that many program proponents insisted it was, or could be, is particularly bad news for evaluators. Multiple goals and beneficiaries usually yield a dizzying array of evaluation questions, most of them legitimate.*

uations were undertaken prematurely, before the programs had an opportunity to refine their services and put their “best foot forward.” In these instances the lack of sufficient time, to allow the programs to mature and then to conduct the research, have conspired against the evaluation enterprise. These limitations have added to the difficulty of conducting FPS evaluations.

It is reasonable to ask why these challenges were not anticipated by evaluators and program administrators; after all, similar stories have emerged from generations of evaluation research in early childhood education (e.g., Barnett, 1995; Consortium for Longitudinal Studies, 1983; McKey et al., 1985; Ramey & Ramey, 1992; Westinghouse Learning Corporation and Ohio University, 1969), family support programming (e.g., Barnes, Goodson, & Layzer, 1996; St.Pierre et al., 1997; Quint, Bos, & Polit, 1997), home visitation interventions (e.g., Olds & Kitzman, 1993), literacy programs (e.g., St.Pierre et al., 1995; Tao, Gamse, & Tarr, 1998), and so forth. In fact, as will be evident in the review of studies that follows, evaluators did attempt, prospectively, to address a number of these issues in the design and conduct of their evaluations. But they missed opportunities at other critical junctures. I return to this question at the end of the essay.

## WHAT WE KNOW AND WHEN WE CAME TO KNOW IT

There is a large body of family preservation evaluations dating back to the late 1970s and early 1980s. Although research on FPS has been ongoing from that time to the present, its heyday appears to have been the decade from the mid-1980s to the mid-1990s. Over the past five years or so, the number of evaluation reports, at least in the published literature, has tailed off quite dramatically, suggesting some loss of interest among funders and researchers alike in evaluating FPS.

There are several likely explanations for this phenomenon: First, in 1995, the federal government initiated a series of national, multiyear evaluations of family-oriented pro-

*It appears that “evaluation fatigue” has set in in the FPS field. The incontrovertible difficulties inherent in FPS evaluation, the limitations of conventional research in finding solutions to these challenges, and the enduring questions about the utility of evaluation to participants, program personnel, and policy makers, have surfaced with force.*

grams. One of those evaluations—the National Evaluation of Family Preservation and Reunification Services—is an experimentally designed impact evaluation of family preservation services. Given that this evaluation was intended to be a definitive study of the effectiveness of FPS, others in the field may simply be waiting for its results before launching what could be redundant or unnecessary research.

Secondly, although FPS continues to be a popular service strategy, a string of bad press stories and an earlier round of equivocal evaluation findings appear to have dimmed FPS’s bright light. The National Family Preservation Network in fact argues that a successful “smear campaign” was launched against family preservation; it also claims that FPS has been scapegoated by “politicians and self-proclaimed ‘child advocates’” (National Family Preservation Network, 1999). These tactics, the Network believes, has eroded much of the support FPS had previously enjoyed. Whether or not

the Network’s assessment of blame is correct, it does seem that a restless public has moved on to other possible remedies for decreasing out-of-home protective placements; the current policy darlings are community-based child protection initiatives (Farrow & the Executive Session on Child Protection, 1997; Shirk, 1998). The clamor to evaluate is now being appropriately heard in that quarter.

Finally, it appears that “evaluation fatigue” has set in in the FPS field. The incontrovertible difficulties inherent in FPS evaluation, the limitations of conventional research in finding solutions to these challenges, and the enduring questions about the utility of evaluation to participants, program personnel, and policy makers, have surfaced with force (Jacobs & Kapuscik, 2000). External evaluators appear to be seeking their challenges elsewhere.

Over the past decade, several excellent reviews of the extent research have made their way to print. In the early 1990s, Peter Rossi’s (1991, 1992a, 1992b) trenchant analyses seem to have inaugurated the “critical” phase in the history of FPS evaluation. Rossi called into question the quality of evaluations conducted to that point, suggesting that most claims of effectiveness were unsupported, given the design and execution of the studies. Schuerman, Rzepnicki, and Littell (1994) took a similarly discouraging view of the state of knowledge about FPS effectiveness and the manner of practice among evaluators. Subsequent reviews had increasingly more material to consider, and generally took a more positive view of the situation. These reviewers argued that there were, indeed, individual studies worth serious consideration, that important lessons had been learned about how to undertake even more trustworthy and illuminating evaluations, and that future evaluations could, and would, do a better job. (See, for example, Blythe, Salley, & Jayaratne, 1994; Fraser, Nelson, & Rivard, 1997; Pecora et al., 1995.) Other reviews seemed a bit less sanguine about whether the evaluation field would, in the foreseeable future, be up to the task altogether (Jacobs, Williams, & Kapuscik, 1997; Chalk & King, 1998).

Each of these reviews discusses the methodological limitations of the majority of the studies available for examination.

These include issues of research design, for example, the absence of comparison or control groups, the small sample sizes of many of the studies, the lack of knowledge of comparison group/control group activity, the narrow range of outcomes measured and the choice of measures, and the choice of data analysis strategies, among others. Shortcomings in program implementation are also identified; these include the inability of program personnel to target FPS to families appropriately (to families at “imminent risk” of placement), poor FPS service implementation by the program, and so forth.

Based on findings from the more rigorous studies, the reviews draw essentially the same conclusions about the effectiveness of FPS. At best, FPS is seen to have small and likely short-term effects on placement prevention. FPS also may produce positive effects in certain discrete areas of family and child functioning. No substantial, enduring effects on children, families, communities, or agencies have been documented.

The individual studies that provide the fodder for these reviews cluster, by the complexity of their designs and by the period in which they were undertaken, into three groups, or “waves,” of activity. In essence, these waves suggest a developmental trajectory for the field of FPS evaluation. Below I review this history briefly.<sup>1</sup>

### **The First Wave of Evaluations**

The first wave of evaluations included those undertaken from the late 1970s through the early 1990s. These studies began as small scale efforts, focused on individual programs; later, several programs within particular states were combined for investigation. A substantial percentage of these evaluations were of the Homebuilders model in Washington and Utah. Placement rates and other system-based indicators (for example, length of stay in out-of-home care, if placed) were selected as the central outcome measures; nonexperimental, post-test only or pre- and post-test designs generally were employed. For those studies relying on placement rates, the post-test only designs were believed to be stringent enough because, given the “imminent risk” eligibility criterion for entry into family preservation, the assumption was that without the program the placement rate would be close to 100 per-

cent. Thus, for example, any forestalling or elimination of the need to place a child could be legitimately claimed as a success.

This group of evaluations boasted strongly positive findings, suggesting that FPS was successful in avoiding placement for at least two-thirds of all families involved. These rates ranged from 75 to 93 percent of families immediately after treatment (Callister, Mitchell, & Tolley, 1986; Nelson, 1991; Pecora, Fraser, & Haapala, 1992; Spaid & Fraser, 1991; Thieman & Dail, 1993); 86 to 100 percent of families three months post-intervention (Bartsch & Kawamura, 1993; Kinney et al., 1990); and 69 to 96 percent of families at one year after the cessation of services (Bartsch & Kawamura, 1993; Bath & Haapala, 1993; Bath, Richey, & Haapala, 1992; Cunningham et al., 1993; Haapala & Kinney, 1988; Kinney et al., 1990; Nelson, 1990; Thieman & Dail, 1992a, 1992b). FPS was also found to be less costly when compared to providing foster care and other types of out-of-home placement (Bartsch & Kawamura, 1993; Cunningham et al., 1993; Kinney et al., 1990; Kinney et al., 1977; Nelson, 1990; New York City Department of Juvenile Justice, 1992, 1993).

Child welfare reformers were heartened by these results, and it appears that they added fuel to the growing movement in the United States toward home-based child protection strategies. Toward the end of this wave, however, methodological criticisms surfaced. The lack of any, or of any reliable, comparison data and the reliance on placement rates as principal outcome measures were among the most frequently cited (Frankel, 1988; Rossi, 1992b). To the first, the usually noted limitations of nonexperimental design were applied. Without comparison data, claims of causal effects were suspect. Rossi (1992b) was the first to suggest that all families “at imminent risk” might not actually have experienced placement, and to ask how, in that case, the net effect of the program could be determined. Other factors could have accounted for the impacts observed.

Rossi (1991, 1992a, 1992b), among others, also raised concern about the choice of placement rates as the measure of success. He argued that these rates are not independent measures of program effectiveness

because of the many influences “outside the program,” including political and judicial forces, that affect policy and placement decisions. For example, they may more accurately reflect top-down agency directives (“we are going to reduce placements”) than “true” improvements in family life that obviate the need for child removal (Berlin, 1992; Kirk, 1993). In addition, the focus on reduced placement rates neglects the fact that placement can sometimes be in the best inter-

*Reducing placement rates was to have direct effects on child welfare costs; indeed, cost savings was a significant “selling point” in the early days of promoting FPS.*

est of the child (Berlin, 1992), and that increased contact between workers and clients during the course of FPS may actually make family problems more evident and placement more likely (Bath & Haapala, 1994). Later critics of FPS evaluation made similar arguments (Schuerman, Rzepnicki, & Littell, 1994; Littell, 1997; Fraser, Nelson, & Rivard, 1997).

Even were these placement rates to be considered appropriate indicators, the studies did not measure them consistently or comprehensively. Blythe, Salley, and Jayaratne (1994) noted that some studies considered placement with a family member an “out-of-home” arrangement while others did not; many concerned themselves with foster care placement only, failing to consider placements in other systems of care (i.e., mental health) or those placement decisions made outside public systems altogether (i.e., child runaways).

Reducing placement rates was to have direct effects on child welfare costs; indeed, cost savings was a significant “selling point” in the early days of promoting FPS. Several studies in the first wave did report positive results on this index. (See, for example, Kinney et al., 1990.) However, the cost studies were relatively primitive in design. For example, they assumed that all comparable cases would have experienced placement had they not been enrolled in FPS (Nelson, 1990). They only considered typical FPS program and out-of-home placement costs, not the other services and resources that may

have been utilized by the family (particularly after termination from the program). And because they generally only tracked families for a short period after program termination, they could not account for the costs incurred by families in which placement was delayed but not eliminated (Frankel, 1988).

Even with the gift of hindsight, it is difficult to fault evaluators on their selection of outcome indicators during this first wave. Most of the programs in these early years claimed reduced out-of-home placements and child protective services costs as their central, often only, goals. And perhaps because program proponents were so convinced that the case for FPS would be made overwhelmingly, concern over the “finer points” of design (e.g., the availability of comparison data) was muted. Markers of mediating effects—improvements in child and family functioning, for example—likely were not followed for the same reasons. This, then, was a missed opportunity to generate program theory, which would have come in handy in later outcome evaluations.

In a larger sense, the most crucial missed opportunity in this first wave of studies was the decision not to initiate serious qualitative investigations of FPS. Had critical assumptions about the program been studied—the extent to which it was enrolling families at “imminent risk” of placement; the extent to which families outside the intervention were, indeed, being placed; the extent to which the program was operating as planned (for example, in the hours and type of services being offered)—more accurately designed evaluation in the next phase might have resulted. Qualitative studies might also have provided valuable demographic data on program participants, anticipating later questions about the differential effects of FPS. What was the overall profile of families in FPS, and for which kinds of families, with which kinds of children, located in which kinds of communities, did it seem to work? By “skipping over” many of the knotty process and implementation questions so essential to understanding the program, this wave of studies disadvantaged, in a significant way, those that followed. The field played “catch up” from that point on.

In summary, then, this first wave of evaluations produced conditions that promote Type I errors (false positives). The

studies demonstrated significant program effects, but the designs were relatively weak and the choice of outcome measures narrow. While it is likely that FPS did affect placement rates, based on these data, it is not possible to determine accurately the extent of that effect.

### **The Second Wave of Evaluations**

The second wave of evaluations responded to many of the criticisms leveled at earlier studies. During the 1990s, larger scale, controlled studies were undertaken, usually by state agencies or groups of programs.

A range of approaches to controls was taken in these studies. Some used enhanced nonexperimental designs, combining pre- and post-testing of subjects with small comparison groups (Fraser, Pecora, & Haapala, 1991). More studies used quasi-experimental methods studying several types of control groups (see Rossi, 1991, 1992a, 1992b), including unserved families (Fraser, Pecora, & Haapala, 1991; Yuan et al., 1990), matched or constructed comparison groups (Schwartz, AuClaire, & Harris, 1991; University Associates, 1993), and retrospective “randomly selected” groups of program graduates and others (Collier & Hill, 1993). Experimental designs with randomized control groups were attempted by several evaluators (Feldman, 1991; Henggeler, Melton, & Smith, 1992; McCroskey & Meezan, 1997; Schuerman et al., 1993; Schuerman, Rzepnicki, & Littell, 1994); the largest being the statewide experiment with Illinois’ Families First conducted by Schuerman et al. (1993, 1994).

Although placement rates and other agency-based indicators continued as the central measures of effects, these studies also broadened the range of outcomes examined to include measures of child, parent, and family functioning. For example, various standardized measures were utilized to assess child outcomes such as well-being (Berry, 1993; Feldman, 1991; McCroskey & Meezan, 1997; Yuan et al., 1990), behavior problems (Henggeler, Melton, & Smith, 1992; Hornick, Phillips, & Kerr, 1989; Wells & Whittington, 1993), and social competence and peer relations (Henggeler, Melton, & Smith, 1992). Functional measures such as juvenile rearrest, reconviction, and reincarceration rates (Collier & Hill, 1993;

Henggeler, Melton, & Smith, 1992) were also used. In addition, evaluations began including parent and family measures, tapping domains such as parent-child conflict (Wells & Whittington, 1993), overall parental functioning (Feldman, 1991; McCroskey & Meezan, 1997; Schuerman et al., 1993; Schuerman, Rzepnicki, & Littell, 1994), stress and support (Feldman, 1991; Fraser, Pecora, & Haapala, 1991; McCroskey & Meezan, 1997; Schuerman, Rzepnicki, & Littell, 1994; Wells & Whittington, 1993), and family relations (Berry, 1993; Feldman,

*In a larger sense, the most crucial missed opportunity in this first wave of studies was the decision not to initiate serious qualitative investigations of FPS. Had critical assumptions about the program been studied—the extent to which it was enrolling families at “imminent risk” of placement; the extent to which families outside the intervention were, indeed, being placed; the extent to which the program was operating as planned (for example, in the hours and type of services being offered)—more accurately designed evaluation in the next phase might have resulted.*

1991; Fraser, Pecora, & Haapala, 1991; Henggeler, Melton, & Smith, 1992; Scannapieco, 1993; Thieman & Dail, 1992a, 1992b; Wells & Whittington, 1993).

The findings appear to have surprised and disappointed program proponents. Few statistically significant results, across all the domains studied, surfaced. Furthermore, even when effects were identified in a number of studies, they were countered by the results of others (Chalk & King, 1998; Pecora et al., 1995; Schuerman, Rzepnicki, & Littell, 1994). For example, several studies showed significant differences in placement rates between treatment and control groups (Henggeler, Melton, & Smith, 1992; Schwartz, AuClaire, & Harris, 1991), while others did not (Schuerman, Rzepnicki, & Littell, 1994; Schuerman et al., 1993; University Associates, 1993; Yuan et al.,

1990). Others, such as Feldman (1991), found initial differences in placement rates but the strength of the effect dissipated over time. The only somewhat consistent statistically significant positive results were modest improvements in certain discrete aspects of child and family functioning.

In their fine review, Fraser, Nelson, and Rivard (1997) suggest that the heterogeneity of clients in FPS may obscure the placement effects that are achieved for select subgroups.

*Although placement rates and other agency-based indicators continued as the central measures of effects, these studies also broadened the range of outcomes examined to include measures of child, parent, and family functioning.*

For example, their meta-analysis finds that “FPS appears to be moderately effective in preventing the placement of children who are in early adolescence and who are referred for truant, oppositional, or delinquent behavior” (p. 147). This conclusion is based on a small subset of studies of family preservation programs primarily designed for older children (Fraser, Nelson, & Rivard, 1997). It does, however, offer support, to the opinion that had evaluators in this wave entered their investigations searching for differential effects—determined both by the characteristics of the participants and the nature of the actual intervention received—a greater array of them might have appeared.

Although this wave of evaluations addressed some of the more glaring methodological shortcomings of the earlier evaluations, others came to the fore. These included, for example, the lack of attention to treatment fidelity and control group activity (Littell, 1997; Schuerman et al., 1993; Schuerman, Rzepnicki, & Littell, 1994; Yuan et al., 1990), the inability to target services appropriately to families “at imminent risk of placement” (Blythe, Salley, & Jayaratne, 1994; Chalk & King, 1998; Schuerman, Rzepnicki, & Littell, 1994), the failure to address heterogeneity within the target population (Bath & Haapala, 1993; Fraser, Nelson, & Rivard, 1997; Littell, 1995; Rossi, 1991; Wells & Biegel, 1991), sample sizes

too small to detect significant effects (Bath & Haapala, 1993; Littell, 1995; Pecora et al., 1995), and the differential attrition from treatment and control groups (Blythe, Salley, & Jayaratne, 1994; Feldman, 1991; Pecora, Fraser, & Haapala, 1992).

There are at least two important aspects to the issue of treatment fidelity. First, expectations of effects from family preservation initially were based on the Homebuilders model. To the extent that evaluators sought to measure effects from different, usually less intensive, program models, they were, in essence, changing the terms of the bargain. Yet several, including the often-noted experiment evaluating Illinois’ Families First, were based on program models that departed substantially from that of Homebuilders (Littell, 1997; Schuerman et al., 1993; Schuerman, Rzepnicki, & Littell, 1994; Henggeler, Melton, & Smith, 1992). Second, even in those studies of programs that purported to be implementing Homebuilders or another specific model of family preservation, there was insufficient attention to documenting the extent to which the model was implemented faithfully; wide variations across sites in multisite studies were common.

Documentation of the services comparison groups families received also was lacking. While all children with substantiated maltreatment are served in some way by the public child protective service system, several studies in this wave failed to specify which child protective, and other, services were being used by clients in the control groups. These services were simply described, for example, as “traditional services” (Blythe, Salley, & Jayaratne, 1994), “existing community services” (Feldman, 1991), “usual services” (Henggeler, Melton, & Smith, 1992), or as out-of-home placements such as “foster care” (Collier & Hill, 1993; Schwartz, AuClaire, & Harris, 1991; University Associates, 1993). It may be, in fact, that the treatment and control groups were receiving similar services, particularly if the model of family preservation being promoted by the child welfare agency was a less intensive one, and other similar support programs were available within the community (Yuan et al., 1990).

Lively debate has followed this set of studies. For example, Julia Littell, a member of Illinois’ Family First research team,

takes issue with many of the methodological criticisms levied against the experimental studies of FPS programs (Littell, 1995). In particular, she argues that (1) treatment integrity is a “non-issue” since there currently exists no evidence of the superiority of one treatment model over another, and (2) undocumented control group activity is of little concern since higher caseload sizes in control groups limited opportunities for workers to provide similar amounts of service to control group members. In her view, client heterogeneity has a minimal effect on aggregate evaluation results, since few studies, including that of Families First, have found differential effects for subgroups of clients. She also views placement as an appropriate outcome measure, so long as FPS is seen as a means to prevent placements.

On the other hand, Howard Bath and David Haapala, who participated in the development and evaluation of Homebuilders, believe that targeting problems reduced the number of children and families that could be studied, compromising the strength of statistical tests and the findings from subgroup analyses (Bath & Haapala, 1994, 1995). In addition, they argue that treatment inconsistency and the practice of aggregating evaluation data across sites and programs may have decreased the researchers’ ability to test the effectiveness of one specific type of intervention, or may have “muddied” the difference between interventions provided to treatment and control groups. This, then, would have decreased the chance of finding true differences between these groups (Bath & Haapala, 1995).

Despite their differences of opinion, both sets of researchers highlight the need for a more in-depth examination of families, program processes, and the service context in which FPS programs operate, and for a solution to the problem of how best to target the right families for service.

Perhaps the most damaging shortcoming of the second wave studies as a group, was their assumption that families “at imminent risk of placement” were the ones being served in FPS. As it turns out, this criterion has proven difficult, in fact virtually impossible, to apply consistently across cases and workers. On the one hand, workers appear reluctant to use it for many families they perceive as needing services often only available

through family preservation. On the other hand, families thought to be at imminent risk of placement by one worker often are not ultimately placed (Chalk & King, 1998; Feldman, 1991; Schuerman et al., 1993; Schuerman, Rzepnicki, & Littell, 1994; University Associates, 1993; Yuan et al., 1990). The placement rates of control groups bear out this fact.

First wave research found that treatment group families had low placement rates. Studies conducted during the second wave found that control group families also tended to remain intact, with placement rates ranging only from 7 to 24 percent (Feldman, 1991; Schuerman et al., 1993; Schuerman, Rzepnicki, & Littell, 1994; University Associates, 1993; Yuan et al., 1990). Even when caseworkers paid special attention to the level of imminent risk of placement of families included in evaluation studies, it remained extremely difficult to assure that all entrants were at a high level of risk, and that treatment and control groups were equivalent in terms of their risk. For example, while caseworkers in the study of Michigan’s Families First claimed that 96 percent of families in treatment and control groups were at risk, only 24 percent of matched comparison children were placed during a one-year period following the intervention (University Associates, 1993). This program implementation problem has held serious consequences for evaluations seeking to establish the effectiveness of FPS.

Even with hindsight, the continued lack of serious focus on collecting critical implementation data through this second wave of studies is mystifying.<sup>2</sup> Program administrators must have suspected, at least, that there were problems with the “at risk” classification system. And they also might well have known that some sites were more closely implementing the specific model than others. Evaluators, for their part, must have known that programs are rarely implemented faithfully, and in a program’s diversions from its model are useful pieces of information for practitioners and evaluators. Furthermore, good evaluation requires one to “unpack” the black box of services—to understand exactly what has been provided to whom, in what intensity over what duration. Otherwise, the assertion that a certain program did or did not work is a specious

one. And given that so much of the success of these evaluations rested on the correct families being enrolled, attention to program sensitivity (including all possible members

*The findings from the second wave of evaluations suggest that family preservation does not have broad, significant effects on children, families, or child protective service system behaviors. However, given the inability of the programs to implement the targeting component faithfully, and the additional methodological problems that emerged, these results may represent Type II errors (“false negatives”).*

of the target population) and specificity (excluding those outside the target population) would have necessarily been of paramount importance. The time to discover that many local FPS programs were not enrolling families at imminent risk of placement was not at the end of a series of costly evaluations, but in their early phases.

The findings from the second wave of evaluations suggest that family preservation does not have broad, significant effects on children, families, or child protective service system behaviors. However, given the inability of the programs to implement the targeting component faithfully, and the additional methodological problems that emerged, these results may represent Type II errors (“false negatives”). That is, there may well have been many significant effects from the programs, but the research has been unable to capture them. This appears a relatively common situation for complex social programs (Posavac & Carey, 1989).

### **The Third Wave of Evaluations**

Recent evaluation efforts have brought further refinements in evaluation questions, research design, outcome measurement, and analysis (Bell et al., 1996; McCroskey & Meezan, 1997; Pecora et al., 1995; Schuerman et al., 1995; Schuerman, Rzepnicki, & Littell, 1994, Littell, 1997).<sup>3</sup> But large-scale external evaluation of FPS has tailed down some, perhaps in anticipation of the multi-

year, multisite, federally-funded National Evaluation of Family Preservation and Reunification Services.<sup>4</sup> The national evaluation is designed to measure effects in the areas of placement prevention, length of stay and subsequent placements, and family functioning. It focuses on the differential effects of FPS on subgroups, intending to identify the characteristics of “successful families” and the complement of services that are offered them (Schuerman, et al., 1995). Data on a wide range of program and child welfare system variables are being collected, so that a fuller picture of influences on FPS success can be produced. The evaluation also includes a series of cost studies to determine which combination of services appear most cost-effective.

Five sites were selected to participate in the national evaluation. These include four FPS program sites—Louisville/ Lexington, KY; Memphis, TN; seven counties in New Jersey; and Philadelphia, PA—and one family reunification program site in New York state. In an effort to compare different service models, the evaluators chose three FPS sites that are implementing the Homebuilders model and one that is using another approach. Control group members receive ordinary child welfare services, so there is no “no treatment” group, but the services used by control group members will be tracked in the same way as services for the treatment group (Cook, personal communication). To ensure that targeting problems do not interfere with the evaluation design, sites were required to develop and implement strategies to ensure that all cases meet the criteria for “imminent risk.” Over the course of the evaluation, Westat is monitoring sites for compliance with these criteria and with random assignment procedures (Schuerman, et al., 1995; Bell et al., 1996).

### **THE BOTTOM LINE ON EFFECTIVENESS**

Family preservation services do not appear to have a significant effect on placement rates, although future subgroup analyses by child, family, and program characteristics may yield more promising findings on this score. Several studies found modest positive effects in discrete areas of child and family functioning, and it appears likely that some improvements in these areas do redound to FPS. Powerful

effects generally have emerged from less “powerful” designs, but the more powerful designs have their own methodological limitations, and are cumbersome and expensive to implement. FPS does not appear, at this point, to be “the” solution to the problems besetting the child welfare system or the families within it.

These evaluation findings provide only very modest support for FPS. Nonetheless, there continues to be broad-based, ongoing support for family preservation services as a necessary, in fact critical, offering in the child protection armamentarium. For example, at the same time as the National Research Council/Institute of Medicine’s 1998 report concludes that it is “difficult to determine their [IFPS] impact on children’s outcomes as well as on placement rates and levels of family functioning” (Chalk & King, 1998, p. 8), it also asserts that, “intensive family preservation services may provide important benefits to the child, family, and community in the form of emergency assistance, improved family functioning, better housing and environmental conditions, and increased collaboration among discrete service systems” (p. 8). Therefore, intensive family preservation services “represent an important part of the continuum of family support services...” (p. 302).

I agree with these seemingly paradoxical conclusions. I assume that FPS works for many families, and fails others. These results likely depend on complicated interactions among characteristics of *the enrolled families*—their particular strengths, needs, hopes, and life circumstances; *their contexts or “holding environments”*—their neighborhoods, informal support systems, the network of accessible community-based services; and *the FPS program itself*—its duration, intensity, quality, and the nature of the relationship established between FPS caseworker and family. These interactions have not been studied sufficiently to provide much guidance to policy makers and program administrators; perhaps the current national evaluation of FPS will shed more light on them.

How are we to understand this paradox: equivocal findings, on the one hand, and continued support for programming, on the other? Two plausible answers come to mind: First, family preservation programs have yet

to be evaluated adequately; they should be given the benefit of the doubt. And second, research has only limited direct impact on policy anyway. I discuss each of these ideas briefly below.

### **FPS Has Not Been Adequately Evaluated**

Evaluation theorist Carole Weiss (1998) suggests two core explanations for “no effects” findings in evaluation research. The first is “theory failure”—that the theory on which the program is built is simply incorrect, and the second is “implementation failure”—that the program was not implemented faithfully enough to demonstrate whatever effective-

*Family preservation services do not appear to have a significant effect on placement rates, although future subgroup analyses by child, family, and program characteristics may yield more promising findings on this score.*

ness, if any, does exist. In earlier work (Jacobs, Williams, & Kapuscik, 1997), we added a third possible explanation, “evaluation failure.” In this instance, the research was not conducted in such a way as to surface whatever effects might have been present. The first explanation of “no effects” would suggest that the findings are correct, while the other two suggest that the findings may not be. Let’s consider these three explanations for the case of family preservation services.

*Is there a theory failure in FPS? Is this service model based on erroneous conceptions of how individuals, families, and systems change, and what is needed to sustain change? Are there fundamental errors in the attendant assumptions about what constitutes a minimally acceptable home environment for children?* The answer to these questions is “maybe, we don’t know.”

FPS is rooted in crisis intervention theory, which is essentially disequilibrium theory in system-theory terms. The notion is that a system (for example, a family), is most amenable to real, sustainable change when it is in dire and disorienting circumstances. Having a child at “imminent risk of placement” is such a situation for a family. Brief, intensive adult-oriented casework, combined

with instrumental supports, help to construct and stabilize the “new order.” The reconstituted family then may be quickly and respectfully given back its autonomy. There is a great deal of face validity to this theory. Indeed, it informs popular approaches to psychotherapeutic practice (for example, brief strategic therapy), and to organizational consultation and management. But applied to families in family preservation services, it lacks an essential ingredient: there is little acknowledgement of child and family development in it. Not all families at imminent risk of placement are alike. Those with infants might need a different constellation and duration of services than those with acting-out adolescents. Younger parents may need more support, over a much longer period of time, than older parents. Getting “in and out” quickly may be precisely the wrong strategy for certain families. And using the standard model of service in communities where there are few supports available to pick up where FPS leaves off might be particularly discouraging to families that are starting to manage well but need continued help. Coleman and Collins’ (1997) qualitative study of over 100 parent participants in family preservation services makes a similar point. When asked in interviews what they did not like, or did not find helpful, in FPS (a six-week program), the response was as follows:

Not surprisingly, the most common complaint about the program concerned the length of the service. While voiced slightly more often by parents of behavior problem children, both groups [behavior problem children and abused children] disliked the brevity of the treatment, with all agreeing that the program was too short... “In a while, things were okay, more positive, but after a month or so, the same problems recurred. One reason I don’t think the problem got better is that the program is not long enough. It doesn’t affect long-term resolution or healing” (pp. 268-269).

Therapists applying crisis intervention theory would argue that “helpers” who stay too long become inducted into whatever system they are trying to help change, ultimately disabling the members and undermining the change process. This notion certainly sounds sensible, but given how notoriously under-researched psychotherapy is, one could hardly claim it as fact. It might be true for certain

families, under certain conditions, and not for others. And perhaps it is precisely those parents wishing to modify the model, who change their parenting conduct according to a different calculus. Until family preservation can explicate its theories of change for different types of children and families in a way that can be studied—or at least discussed and contested—one has to question its application to all families at imminent risk.

*Have there been implementation failures in FPS?* The answer here is an unequivocal “yes.” Numerous commentators have also drawn this conclusion, so I will not dwell on it here. Suffice it to say that there are few aspects of the model that have been documented as consistently implemented across the programs that have been studied. Applying the criterion of imminent risk, maintaining the recommended intensity of service, garnering follow-up support services for families after FPS, are among the most often mentioned in this regard. This is not to say that many organizations or programs are not striving to implement their models faithfully and are doing so successfully along a variety of program dimensions. But that evidence is not available in published outcome studies.

FPS proponents and researchers might also have to entertain the notion that FPS actually *cannot* be replicated as planned across localities, counties, and states. After reviewing the evaluations of six nationally-implemented demonstration home visiting programs, Gomby, Culross, and Behrman (1999) come to this rather disappointing, but likely realistic, conclusion for those enormously popular interventions. After reporting that once families were enrolled, “they received on average half the number of scheduled visits, no matter what the intended frequency of visits was” (p. 16), and that enrolled families did not attend group meetings as they were meant to do, the authors conclude that,

[T]he consistency with which this occurs across the models suggests that this [low levels of involvement] is a real phenomenon in the implementation of home visiting programs, not just the result of poor implementation of particular program models or specific program sites. Families are either not willing or not able to take as much of the service as is intended by program designers (p. 16).

Of course, much more must be known about the implementation of FPS before one would arrive at this conclusion. But if the expected results rest on implementation of a model of service that cannot realistically be standardized, then one of two changes is in order: Either the model should be altered and a new approach tried, or the expectations should be altered, acknowledging the strengths and limitations of the service as it is currently being implemented. But these recommendations pertain to a time in the future. There is no question that at present, widespread implementation failures have occurred.

*Have there been evaluation failures in FPS?* We can offer an unqualified “yes” here as well. There have also been evaluation successes—that is, there are evaluations with reliable and interesting results, evaluations that raised important questions, and evaluations with thoughtful and creative designs that moved the field forward. Some of what are considered failures are understandably flawed early attempts to hit a moving target—to capture the essence of an exciting, innovative intervention that was launched and replicated quickly. But there have been some persistent blind spots in evaluation efforts, and some reluctance to admit the limitations inherent in conducting certain kinds of FPS evaluations at particular moments in the program’s development. The current national evaluation promises to correct many of these problems, but one would have to decline to decide the fate of FPS based on the body of available evaluations.

### **Research Has Only Limited Direct Impact on Policy**

Although researchers might want it to be different, direct, instrumental use of evaluation for policy decision-making is not commonplace. The policy process and the research process operate on different schedules, with different goals, different masters, and different criteria for success. Weiss’s (1983) paradigm for understanding how policy makers use information is instructive here. She identifies three forces that influence policy—ideology, interests, and information—and suggests that these forces need to be in some consonance in order for policy to develop. The converse also seems to be true. It is

unlikely that a policy direction would be derailed by virtue of pressure from a single source. In this context, then, research—even high quality research—would not hold sway on a policy issue if it flew in the face of popular beliefs or powerful political interests. Rogers (1995) similarly considers “observability” (the extent to which results of innovations are observed and communicated to others) as only one of five attributes that determine the adoption of innovations.

Evaluators have long been aware of the danger of irrelevance. Indeed, in an attempt to increase the pertinence of evaluation to policy makers and program administrators, Patton (1997) and others developed a more “user-friendly” orientation-utilization-focused evaluation. This approach puts a premium on use; evaluators are urged to identify specific evaluation consumers, work with them on developing an evaluation agenda, and then make certain that the evaluation produces results that fit their needs for information. This seems like good advice, though high profile evaluations with many potential users might find it difficult to satisfy them all. Nonetheless, the fact that evaluations, by themselves, usually suffer from lack of use, not from overuse, is the point to make here.

The stunning policy breakthrough effected by FPS proponents in the early 1990s was based in politics, not science; to the minimal extent that research helped to make the case, it was used selectively and opportunistically. Our continued, virtually unblinking support for out-of-home placement as a prime child protection strategy is further evidence of the relatively minor role that evaluation plays in supporting even large-scale human service interventions. Many commentators of different stripes (for example, Chalk & King, 1998; National Family Preservation Network, 1999) have noted that the “effectiveness” of foster care has never even been scrutinized. One might argue that it is so consonant with broad social values and is represented by such powerful interests, that evaluation research is seen as either wholly unnecessary, or simply impossible to undertake. It is, then, easy to understand why broad support for family preservation can operate alongside evaluation findings that might suggest caution and reappraisal.

## THE FUTURE OF FPS EVALUATION

Notwithstanding the many real signs of FPS evaluation fatigue, rumors of its death are premature. In fact, while external evaluation efforts have fallen off in recent years, demands for outcomes-based accountability in child welfare have, if anything, increased since the mid-1990s. For example, the 1997 Adoption and Safe Families Act requires states to maintain data systems that can measure performance and track outcomes. This is not so much the paradox it appears to be. It has been established as valuable practice for public agencies to collect data, over time,

*Now that FPS is institutionalized as a core component of child protection services across the country, it is time to tone down the rhetoric about its impact.*

that can be useful both for monitoring services and for capturing what may possibly be program effects at a systems level (such as changes in child protection recidivism rates). This is due in part to the creative and persistent efforts of child welfare evaluators such as C. Lynn Usher, Peter Pecora, and their colleagues. (See, for example, Pecora et al., 1996; Usher, Gibbs, & Wildfire, 1995; Usher, Wildfire, & Gibbs, 1999.) State child welfare agencies have accepted the need for thoughtfully developed and well-maintained management information systems, and the value of using data to reflect on practice and policy. Reports of these evaluation activities likely will not be published in scholarly journals, but it appears that many “site based” evaluations will ensue. This is a positive legacy of the era of FPS evaluation.

My final bit of advice on FPS evaluation is the following: **Set realistic expectations for both programs and their evaluations.**

As Monday morning quarterbacks, we know now that FPS was greatly oversold as a program strategy. Given the range and diversity of families involved, the child welfare system as it was constructed at the time, the absence of community-based services to pick up where family preservation services left off, FPS could not possibly have achieved the benefits that were touted to the

public. And, as we have learned with other social programs, the higher the expectations, the more intense the feelings of disappointment when they are not realized.

Yet proponents of FPS did achieve some remarkable results. At the time that FPS burst on the scene, federal funding for child protection services for seriously troubled families was almost wholly for out-of-home placement, “back-end,” services that were not preventive in any sense. The early program-initiated evaluations of family preservation suggested that it was possible to intervene earlier, in a different way, with families, and that those interventions were better for the family members and potentially less expensive to the system. The opportunity to support a successful, new genre of child protective service, more respectful of family and community, attracted proponents from all quarters. In the swell of optimism about family preservation, a new service option, lodged within the generally conservative, inflexible child welfare system, was institutionalized.

Now that FPS is institutionalized as a core component of child protection services across the country, it is time to tone down the rhetoric about its impact. Modest, realistic statements of effects are more likely to garner additional support than the wildly positive statements made by program advocates fifteen years ago.

Evaluators should also take the vow of modesty. As a field, evaluation has come to acknowledge, however slowly, the limitations of both conventional and newer approaches to establishing the effects of social programs. FPS evaluators should reflect this refreshing turn to modesty in all aspects of their work. There are certain effects we can attempt to measure, but for others, no sound measures exist. There are certain questions we can reasonably hope to answer, but others require technical expertise or support that may be beyond us personally, or beyond the present capacity of the field. There are conclusions based on our data that are reasonable to make, and others we are tempted to make that overstate what we know. In a field as contested as this one, less is more.

It is perhaps true that modesty suits certain developmental periods better than others. Modest promises on the part of program

advocates might not convince legislators to initiate funding for seemingly worthy efforts; likewise, modest promises on the part of evaluators might not garner the support, financial and political, for their important long-term work. Bravado rather than caution may be necessary in the early stages of development. Thankfully, both family preservation and the field of evaluation are past their childhoods; realistic appraisals of the value in each of these efforts should now dominate.

There is, of course, a larger context for this discussion that must be noted. Some of the impetus for overpromising results from family preservation, or home visiting, or family support, or even community-based

child protection initiatives is that, as a society, we care so little for poor children that we refuse to invest in them without the “hype.” We want to yield extraordinary benefits at a bargain price, and we want scientists to validate that we got good value. We don’t want to confront poverty directly—that would require fundamental, structural changes—so we approach it weakly and obliquely—through social services of one type or the other. But all these programs are sorely constrained in their bids for effectiveness until we make a decision truly to invest in these families and children. And what is needed for that to occur is not better science; it’s better values.

## REFERENCES

- Barnes, H. V., Goodson, B. D., & Layzer, J. I. (1996). *Review of the research on family support interventions*. Cambridge, MA: Abt Associates.
- Barnett, W. S. (1995). Long-term effects of early childhood programs on cognitive and school outcomes. *The Future of Children*, 5 (3), 25-50.
- Bartsch, D., & Kawamura, G. (1993). *Family preservation services in Colorado*. Denver: Colorado Division of Mental Health/Department of Institutions
- Bath, H. I., & Haapala, D. A. (1993). Intensive family preservation services with abused and neglected children: An examination of group differences. *Child Abuse & Neglect*, 17, 213-225.
- Bath, H. I., & Haapala, D. A. (1994). Family preservation services: What does the outcome research really tell us? *Social Service Review*, 68, 386-404.
- Bath, H. I., & Haapala, D. A. (1995). Evaluation outcomes of family preservation services and the way ahead: A reply to Littell. *Social Service Review*, 69 (2), 351-359.
- Bath, H. I., Richey, C. A., & Haapala, D. A. (1992). Child age and outcome correlates in intensive family preservation services. *Children and Youth Services Review*, 14, 389-406.
- Bell, J., Kaye, E., Ensign, K., Schuerman, J., Rossi, P., & Cook, R. (1996). *Evaluation of family preservation and reunification services: Cost study design*. Rockville, MD: Westat, Inc.
- Berlin, G. (1992). Choosing and measuring interventions. *Children and Youth Services Review*, 14, 99-118.
- Berry, M. (1993). The relative effectiveness of family preservation services with neglectful families. In E.S. Morton & R.K. Grigsby (Eds.), *Advancing family preservation practice* (pp. 70-98). Newbury Park, CA: Sage Publications.
- Bickman, L. (Ed.). (1987). Using program theory in evaluation. *New Directions for Program Evaluation*, No. 33. San Francisco: Jossey-Bass.
- Blythe, B., Salley, M. P., & Jayaratne, S. (1994). A review of intensive family preservation services research. *Social Work Research*, 18 (4), 213-224.
- Callister, J. P., Mitchell, L., & Tolley, G. (1986). Profiling family preservation efforts in Utah. *Children Today*, 15 (6), 8-11, 23-25.
- Chalk, R., & King, P. A. (Eds.). (1998). *Violence in families: Assessing prevention and treatment programs*. Washington, DC: National Academy Press..
- Chen, H. T. (1990). *Theory-driven evaluations*. Newbury Park, CA: Sage Publications.
- Chen, H. T., & Rossi, P. H. (1987). "The theory-driven approach to validity." *Evaluation and Program Planning*, 10, 95-103.
- Coleman, H., & Collins, D. (1997). The voices of parents: A qualitative study of a family preservation program. *Child & Youth Care Forum*, 26 (4), 261-278.

- Collier, W. V., & Hill, R. H. (1993). *Family ties: Intensive family preservation services program: An evaluation report*. New York: Department of Juvenile Justice.
- Consortium for Longitudinal Studies. (Ed.). (1983). *As the twig is bent—lasting effects of preschool programs*. Hillsdale, NJ: Erlbaum.
- Cunningham, M. L., Homer, K. S., Bass, A. S., & Brown, M. G. (1993). *Family preservation in Tennessee: The home ties intervention*. Knoxville, TN: University of Tennessee College of Social Work/Office of Research and Public Service.
- Drisko, J. W. (1998). Utilization-focused evaluation of two intensive family preservation programs. *Families in Society: The Journal of Contemporary Human Services*, 79, 62-74.
- Farrow, F. with the Executive Session on Child Protection. (1997). *Child protection: Building community partnerships: Getting from here to there*. Cambridge, MA: John F. Kennedy School of Government, Harvard University.
- Feldman, L. H. (1991). *Assessing the effectiveness of family preservation services in New Jersey within an ecological context*. Trenton, NJ: New Jersey Division of Youth and Family Services, Bureau of Research, Evaluation, and Quality Assurance.
- Frankel, H. (1988). Family-centered, home-based services in child protection: A review of the research. *Social Service Review*, 62 (1), 137-157.
- Fraser, M. W., Nelson, K. E., & Rivard, J. C. (1997). Effectiveness of family preservation services. *Social Work Research*, 21 (3), 138-153.
- Fraser, M. W., Pecora, P. J., & Haapala, D. A. (1991). *Families in crisis: The impact of intensive family preservation services*. Hawthorne, NY: Aldine de Gruyter.
- Fraser, M. W., Walton, E., Lewis, R. E., Pecora, P. J., & Walton, W. K. (1996). An experiment in family reunification: Correlates of outcomes at one-year follow-up. *Children and Youth Services Review*, 18 (4/5), 335-361.
- Gomby, D. S., Culross, P. L., & Behrman, R. E. (1999). Home visiting: Recent program evaluations—analysis and recommendations. *The Future of Children*, 9 (1), 4-26.
- Haapala, D. A., & Kinney, J. M. (1988). Avoiding out-of-home placement of high-risk status offenders through the use of intensive home-based family preservation services. *Criminal Justice and Behavior*, 15 (3), 334-348.
- Henggeler, S. W., Melton, G. B., & Smith, L. A. (1992). Family preservation using multi-systemic therapy: An effective alternative to incarcerating serious juvenile offenders. *Journal of Consulting and Clinical Psychology*, 60 (6), 953-961.
- Hornick, J. P., Phillips, D. M., & Kerr, N. (1989). Gender differences in behavioral problems of foster children: Implications for special foster care. *Community Alternatives*, 1 (1), 35-52.
- Jacobs, F.H. & Kapuscik, J.L. (2000). *Making it count: Evaluating family preservation services*. Medford, MA: Eliot-Pearson Department of Child Development, Tufts University.

- Jacobs, F. H., Williams, P. H., & Kapuscik, J. L. (1997). Evaluating family preservation services: Asking the right questions. In W. Hellinckx, M. J. Colton, & M. Williams (Eds.), *International perspectives on family support* (pp. 206-223). Brookfield, VT: Ashgate Publishing.
- Kinney, J., Haapala, D. A., Booth, C., & Leavitt, S. (1990). The Homebuilders model. In J. K. Whittaker, J. Kinney, E. M. Tracy, & C. Booth (Eds.), *Reaching high-risk families: Intensive family preservation in human services* (pp. 37-64). New York: Aldine de Gruyter.
- Kinney, J., Haapala, D. A., Booth, C., & Leavitt, S. (1991). The HOMEBUILDERS Model. In E. M. Tracy, D. A. Haapala, J. Kinney, & P. J. Pecora (Eds.), *IFPS: An instructional sourcebook* (pp. 15-49). Cleveland, OH: Mandel School of Applied Social Sciences, Case Western Reserve University.
- Kinney, J. M., Madsen, B., Fleming, T., & Haapala, D. A. (1977). Homebuilders: Keeping families together. *Journal of Consulting and Clinical Psychology*, 45 (4), 667-673.
- Kirk, R. S. (1993). Methodological issues in measuring outcomes. *Protecting Children*, 10 (2), 7-9.
- Littell, J. H. (1995). Evidence or assertions? The outcomes of Family Preservation Services. *Social Service Review*, 69 (2), 338-351.
- Littell, J. H. (1997). Effects of the duration, intensity, and breadth of family preservation services: A new analysis of data from the Illinois Family First Experiment. *Children and Youth Services Review*, 19 (1/2), 17-39.
- McCroskey, J., & Meezan, W. (1997). *Family preservation and family functioning*. Washington, DC: CWLA Press.
- McKey, R., Condelli, L., Ganson, H., et al. (1985). *The impact of Head Start on children, families, and communities: Final report of the Head Start Evaluation, Synthesis, and Utilization Project*. Washington, DC: U.S. Department of Health and Human Services.
- National Family Preservation Network. (1999). *Best evidence: An analysis of the findings and conclusions concerning family preservation found in the report of the National Research Council Committee on the Assessment of Family Violence Interventions*. [On-line]. Available: [www.nfpn.org/bestevid.htm](http://www.nfpn.org/bestevid.htm).
- Nelson, D. (1990). Recognizing and realizing the potential of "family preservation." In J. Whittaker, J. Kinney, E. Tracy, & C. Booth (Eds.), *Reaching high-risk families: Intensive family preservation in human services* (pp. 13-30). Hawthorne, NY: Aldine de Gruyter.
- Nelson, K. E. (1991). Populations and outcomes in five family preservation programs. In K. Wells, & D.E. Biegel (Eds.), *Family preservation services: Research and evaluation* (pp. 72-91). Newbury Park, CA: Sage Publications.
- New York City Department of Juvenile Justice. (1992). *Family ties: The first eighteen months*. New York: Department of Juvenile Justice.

- New York City Department of Juvenile Justice. (1993). *Family ties intensive family preservation program: A financial analysis*. New York: Department of Juvenile Justice.
- Olds, D. L., & Kitzman, H. (1993). Review of research on home visiting for pregnant women and parents of young children. *The Future of Children*, 3 (3), 53-92.
- Patton, M. Q. (1997). *Utilization-focused evaluation: The new century text* (3rd ed.). Newbury Park, CA: Sage Publications.
- Pecora, P. J., Fraser, M. W., & Haapala, D. A. (1992). Intensive home-based family preservation services: An update from the FIT project. *Child Welfare*, 71 (2), 177-188.
- Pecora, P. J., Fraser, M. W., Nelson, K. E., McCroskey, J., & Meezan, W. (1995). *Evaluating family-based services*. Hawthorne, NY: Aldine de Gruyter.
- Pecora, P. J., Seelig, W. R., Zirps, F. A., & Davis, S. M. (1996). *Quality improvement and evaluation in child and family services: Managing into the next century*. Washington, DC: CWLA Press.
- Posavac, E. J., & Carey, R. G. (1989). *Program evaluation: Methods and case studies*. Third Edition. Englewood Cliffs, NJ: Prentice-Hall.
- Quint, J. C., Bos, J. M., & Polit, D. F. (1997). *New Chance: Final report on a comprehensive program for disadvantaged young mothers and their children*. New York: Manpower Demonstration Research Corporation.
- Ramey, C. T., & Ramey, S. L. (1992). Early educational intervention with disadvantaged children-To what effect? *Applied and Preventive Psychology*, 1, 130-140.
- Rogers, E. M. (1995). *Diffusion of innovations*. New York: The Free Press.
- Rossi, P. H., & Freeman, H. E. (1993). *Evaluation: A systematic approach* (5th ed.). Newbury Park, CA: Sage Publications.
- Rossi, P. H. (1991). *Evaluating family preservation programs: A report to the Edna McConnell Clark Foundation*. New York: Edna McConnell Clark Foundation.
- Rossi, P. H. (1992a). Assessing family preservation programs. *Children and Youth Services Review*, 14, 77-97.
- Rossi, P. H. (1992b). Strategies for evaluation. *Children and Youth Services Review*, 14, 167-191.
- Scannapieco, M. (1993). The importance of family functioning to prevention of placement: A study of family preservation services. *Child and Adolescent Social Work Journal*, 10 (6), 509-520.
- Schuerman, J. R., Rossi, P. H., Cook, R., & Kaye, E. (1995). *Evaluation of Family Preservation and Reunification Services: Research Design*. Rockville, MD: Westat, Inc.
- Schuerman, J. R., Rzepnicki, T. L., & Littell, J. H. (1994). *Putting families first: An experiment in family preservation*. Hawthorne, NY: Aldine de Gruyter.

- Schuerman, J. R., Rzepnicki, T. L., Littell, J. H., & Chak, A. (1993). *Evaluation of the Illinois Family First placement prevention program: Final report*. Chicago: Chapin Hall Center for Children, University of Chicago.
- Schwartz, I. M., AuClaire, P., & Harris, L. J. (1991). Family preservation services as an alternative to out-of-home placement of adolescents: The Hennepin County experience. In K. Wells, & D. E. Biegel (Eds.), *Family preservation services: Research and evaluation* (pp. 33-46). Newbury Park, CA: Sage Publications.
- Shirk, M. (1998). *We are in this together: Community child protection in America*. New York, NY: Edna McConnell Clark Foundation.
- Spaid, W. M., & Fraser, M. (1991). The correlates of success/failure in brief and intensive family treatment: Implications for family preservation services. *Children and Youth Services Review, 13*, 77-99.
- St. Pierre, R. G., Goodson, B. D., Layzer, J. I., & Bernstein, L. S. (1997). *National evaluation of the Comprehensive Child Development Program: Final Report*. Cambridge, MA: AN Associates.
- St. Pierre, R. G., Swartz, J. P., Gamse, B., Murray, S., Deck, D., & Nickel, P. (1995). *National evolution of Even Start Family Literacy Program: Final report*. Cambridge, MA: Abt Associates.
- Staff, I., & Fein, E. (1994). Inside the black box: An exploration of service delivery in a family reunification program. *Child Welfare, 73* (3), 195-211.
- Tao, F., Gamse, B., & Tarr, H. (1998). *Second national evaluation of the Even Start Family Literacy Program: Final report*. Arlington, VA: Fu Associates, Ltd.
- Thieman, A. A., & Dail, P. W. (1992a). Family preservation services: Problems of measurement and assessment of risk. *Family Relations, 41*, 186-191.
- Thieman, A. A., & Dail, P. W. (1992b). *Iowa's family preservation program, FY 1991 evaluation*. Ames, IA: Child Welfare Research and Training Project, Iowa State University.
- Thieman, A. A., & Dail, P. W. (1993). *Iowa's family preservation program, FY 1992 evaluation*. Ames, IA: Child Welfare Research and Training Project, Iowa State University.
- Thieman, A. A., & Dail, P. W. (1997). Predictors of out-of-home placement in a family preservation program: Are welfare recipients particularly vulnerable? *Policy Studies Journal, 25* (1), 124-139.
- University Associates. (1993). *Evaluation of Michigan's Families First program: Summary report*. Lansing, MI: Author.
- Usher, C. L., Gibbs, D. A., & Wildfire, J. B. (1995). A framework for planning, implementing, and evaluating child welfare reforms. *Child Welfare, 64* (4), 859-876.
- Usher, C. L., Wildfire, J., & Gibbs, D. A. (1999). Measuring performance in child welfare: Secondary effects of success. *Child Welfare, 78* (1), 31-51.

- Walton, E., Fraser, M. W., Lewis, R. E., Pecora, P. J., & Walton, W. K. (1993). In-home family-focused reunification: An experimental study. *Child Welfare, 72* (5), 473-488.
- Weiss, C. H. (1995). Nothing as practical as good theory: Exploring theory-based evaluation for comprehensive community initiatives for children and families. In J. P. Connell, A. C. Kubisch, L. B. Schorr, & C. H. Weiss (Eds.), *New approaches to evaluating community initiatives* (pp. 65-92). Washington, DC: Aspen Institute.
- Weiss, C. H. (1983). Ideology, interests and information. In D. Callahan, & B. Jennings (Eds.), *Ethics, the social sciences and public policy analysis* (pp. 213-245). New York: Plenum Press.
- Weiss, C. H. (1998). *Evaluation: Methods for studying programs and policies* (2nd ed.). Upper Saddle River, NJ: Prentice Hall.
- Wells, K., & Tracy, E. (1996). Reorienting intensive family preservation services in relation to public child welfare practice. *Child Welfare, 75*, 667-692.
- Wells, K., & Biegel, D. E. (Eds.). (1991). *Family preservation services: Research and evaluation*. Newbury Park, CA: Sage Publications.
- Wells, K., & Whittington, D. (1993). Child and family functioning after intensive family preservation services. *Social Service Review, 67* (1), 55-83.
- Westinghouse Learning Corporation and Ohio University. (1969). *The impact of Head Start: An evaluation of the effects of Head Start on children's cognitive and affective development*. Report to the Office of Economic Opportunity. Athens, OH: Westinghouse Learning Corporation and Ohio University.
- Yuan, Y. T., McDonald, W., Wheeler, C., Struckman-Johnson, D., & Rivest, M. (1990). *Evaluation of AB 1562 in-home care demonstration projects: Final report*. Sacramento, CA: Walter R. McDonald & Associates.

## ENDNOTES

<sup>1</sup> Significant portions of this discussion of the three waves of FPS evaluation are presented in two other documents: Jacobs, Williams, & Kapuscik (1997), and Jacobs & Kapuscik (2000).

<sup>2</sup> To suggest that additional process evaluation is needed does not diminish the fact that there have been a number of interesting process-oriented studies of aspects of FPS over the years. For example, Bath & Haapala, (1993), Coleman & Collins (1997), Drisko (1998), Staff & Fein (1994), Thiemann & Dail (1997). Evaluations have addressed a variety of process questions related to program implementation, characteristics of the client population, group differences, and other issues. Unfortunately, what appears to be missing are in-depth process studies that assess the faithful implementation of program models (with regard to targeting, service intensity/duration, etc.) and that have had an observable, subsequent impact on the conduct of further evaluation or the structure of programs. We hope that the national evaluation underway will be able to put their process studies to good use in examining outcomes and contextualizing evaluation findings.

<sup>3</sup> Evaluations of family reunification programs using FPS technology, while not the subject of this review, have contributed to this progress in research as well.

<sup>4</sup> The study is being conducted by Westat, Inc., in association with the Chapin Hall Center for Children at the University of Chicago and James Bell Associates, Inc.

## AUTHOR

**Francine Jacobs** is an Associate Professor with a joint appointment in the Eliot-Pearson Department of Child Development and the Department of Urban and Environmental Policy at Tufts University. Her research and teaching interests are primarily in the area of child and family policy—child welfare and protection, childcare and early childhood education, family support, and community-based initiatives—and in program evaluation.

Dr. Jacobs graduated from Brandeis University, and received her Master's and doctoral degrees from Harvard University. Prior to joining the Tufts faculty in 1986, Dr. Jacobs was the Director of Research at the Harvard Family Research Project, and the program director for two early childhood programs. She also maintained an active program evaluation practice, consulting to numerous organizations on the planning and conduct of evaluation activities. During her time

at Tufts, she has been the principal investigator for almost a dozen grants, with projects ranging from coordinating an early childhood community-planning process in Boston to the development of an evaluation process for state family preservation programs. Her current project is the large-scale, multiyear evaluation of a universal home visiting program for teen mothers—Healthy Families Massachusetts. Her co-principal investigators for this project are Professors Ann Easterbrooks and Jayanthi Mistry of the Child Development Department.

Dr. Jacobs has lectured and written extensively about program evaluation and about issues in child and family policy. Her two co-edited volumes—*Evaluating Family Programs* (with Heather Weiss), and *More Than Kissing Babies: Current Child and Family Policy in the United States* (with Margery Davies)—focus in these areas.

## RELATED PUBLICATIONS

*A Synthesis of Research on Family Preservation and Family Reunification Programs*  
Julia Littell, John Schuerman 1995

*A Review of Family Preservation and Family Reunification Programs*  
John Schuerman 1995

*Evaluation of the New York City HomeRebuilders Demonstration*  
John Schuerman, Ronna Cook 1998

“What Works Best for Whom? A Closer Look at Intensive Family Preservation Services”  
Julia Littell, John Schuerman, forthcoming

“Decisions on Placement and Family Preservation: Agreement and Targeting”  
John Schuerman, Peter Rossi, Stephen Budde 1999

*Innovations in Child Welfare: Preventing Out-of-Home Placement of Abused and Neglected Children*  
Julia Littell, John Schuerman 1999

*Best Interests and Family Preservation in America*  
John Schuerman 1997

“Effects of the Duration, Intensity, and Breadth of Family Preservation Services: A New Analysis of Data from the Illinois Family First Experiment”  
Julia Littell 1997

*An Experimental Study of Family Prevention Services: Early Findings from a Parent Survey*  
Tina Rzepnicki, John Schuerman, Julia Littell, Amy Chak, Marva Lopez 1994

*Putting Families First: An Experiment in Family Preservation*  
John Schuerman, Tina Rzepnicki, Julia Littell 1994

*What Works Best for Whom in Family Preservation? Relationships Between Service Characteristics and Outcomes for Selected Subgroups of Families*  
Julia Littell, John Schuerman, Amy Chak 1994

*Shifting Objectives in Family Preservation Programs*  
Julia Littell, John Schuerman, Tina Rzepnicki, Jeanne Howard, Stephen Budde 1993

“Implementation Issues”  
John Schuerman, Tina Rzepnicki, Julia Littell, Stephen Budde 1992

“From Chicago to Little Egypt: Lessons from an Evaluation of a Family Preservation Program”  
John Schuerman, Tina Rzepnicki, Julia Littell 1991

*Issues in Evaluating Intensive Family Preservation Services*  
John Schuerman, Tina Rzepnicki, Julia Littell 1991

*Facing Uncertainty: Reuniting High-Risk Families*  
Tina Rzepnicki, John Schuerman, Penny Ruff Johnson 1997