ELSEVIER

Evaluation and Program Planning 30 (2007) 136–148

EVALUATION and PROGRAM PLANNING

www.elsevier.com/locate/evalprogplan

The relationship between implementation fidelity and educational outcomes in a school-based family support program: Development of a model for evaluating multidimensional full-service programs

John Kalafat[a,*], Robert J. Illback[b], Daniel Sanders Jr.[b]

[a]*Rutgers Graduate School of Applied and Professional Psychology, 152 Frelinghuysen Road, Piscataway, NJ 08854-8085, USA*
[b]*REACH of Louisville, Inc., 501 Park Ave., Louisville, KY 40208, USA*

Abstract

There is increasing recognition of the need for evaluations that identify program processes or mediators and assess degree of program implementation rather than focusing solely on outcome evaluation. This paper describes the application of complementary qualitative and quantitative evaluation procedures to assess the degree of implementation of multi-component family support programs for improving educational outcomes for at risk youth, and to assess the relationship between program implementation and outcomes. The qualitative evaluation involved prolonged engagement to identify common program domains or mediators. Using a method called Innovation Configuration Analysis, levels of implementation of program domains were explicated as well as an overall Implementation Fidelity Index. Strong positive relationships were found between overall program implementation and program-level outcomes achieved by student participants.

Keywords: Implementation; Outcomes; Family support

1. Introduction

Recent reviews of prevention programs for children and youth have reported advances in evidence-based practice, particularly for coordinated community and school-based initiatives that promote competencies and connections among students, schools, and families (Catalano, Berglund, Ryan, Lonczak, & Hawkins, 2002; Durlak & Wells, 1997; Weissberg, Kumpfer, & Seligman, 2003). These same reviews call for more process research on the contributions of specific program components, contexts, and implementation processes that affect program outcomes. Weissberg et al. concluded:

> Our understanding of mediating and moderating variables that influence program effects—especially in larger systems like schools and communities—is limited. Research studies should pay greater attention to process measures of program quality and fidelity, rather than focusing primarily on the more typical strategy of outcome evaluations (p. 430).

In fact, the current evaluation field reflects a growing emphasis on process and implementation evaluation and recognizes the lack of utility in solely conducting black box outcome evaluation. As Dobson & Cook (1980) note, "Simply put, if treatments are not clearly specified and if the services implied by those treatments are not delivered in a way which is consistent with program objectives, it is likely that evaluation results will be less than useful, or, perhaps, meaningless" (p. 270). This has been called Type III error, which consists of concluding that a program is not effective, when the measurement may have reflected inadequate implementation of the program. Several studies have documented a relationship between levels of program implementation and program outcomes (Dane & Schneider, 1998; Dumas, Lynch, Laughlin, Smith, & Prinz, 2001;

*Corresponding author. Tel.: +1 732 445 2000x121; fax: +1 732 445 4888.
E-mail addresses: kalafat@rci.rutgers.edu (J. Kalafat), illbackr@reachoflouisville.com (R.J. Illback), sandersdh@reachoflouisville.com (D. Sanders Jr.).

0149-7189/$ - see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.evalprogplan.2007.01.004

Harachi, Abbott, Catalano, Haggerty, & Fleming, 1999; Moncher & Prinz, 1991).

In addition to the potential for Type III errors, the dissemination of effective programs, which involves their adaptation to new settings, cannot be accomplished without a clear understanding of how the program accomplishes its effects. Reinvention of a program is acceptable, but only if its causal mechanisms are understood and preserved (Bauman, Stein, & Ireys,1991; Elias, 1997). In addition, the original mission or intent of the program must be understood and preserved (Blakely et al., 1987; Schorr, 1997).

Complex community and school-based programs are particularly susceptible to Type III errors because they are composed of multiple components and thus multiple potential causal elements whose contributions to outcome may not be known (Durlak & Wells, 1997; Goodman, 2000). Knapp (1995) referred to this as the elusive independent variable.

While current prevention science is not yet sufficient to provide evidence-based guidance regarding program implementation and adaptation (Weissberg et al., 2003), the evaluation literature reflects a growing consensus regarding methods for evaluating multicomponent prevention programs.

Assessing the contribution of program implementation to outcomes begins with a clear explication of the conceptual model underlying the program and how this informs interventions (Dumas et al., 2001; McGrew, Bond, Dietzen, & Salyers, 1994). This has been referred to as the program's theory of action (Patton, 2001) or program logic. Elias (1997) noted that what is transferred when a program is implemented in a new site includes not only procedures but also an understanding of the principles that undergird the practice. Without an understanding of the conceptual basis for a program, it is not possible to identify its core components or mediators and to distinguish these from more changeable or adaptable features.

Once hypothesized program mediators have been identified, operational definitions of these components must be developed (Harachi et al., 1999; McGrew et al., 1994). These must be observable or verifiable through such procedures as interviews and surveys of program personnel and recipients (Blakely et al., 1987). For complex multicomponent interventions, program components must be disaggregated (Altman, 1986; Durlak & Wells, 1997; Goodman, 2000). Scheirer (1994) called for micro-evaluations involving dosage measures of hypothesized mediators.

Dane and Schneider (1998) and Moncher and Prinz (1991) noted that dichotomizing interventions can result in substantial loss of information and that regressing continuous implementation variables on salient outcome measures is likely to result in more powerful analysis. In this regard, Hall and Loucks' (1978) methodology for quantifying levels of implementation (as opposed to placing them into dichotomous categories) has merit. Consistent with this approach, several investigators have described a participatory action research approach involving collaboration between evaluators and program personnel, extended grounded assessment of programs in context (Glaser & Strauss, 1967), and a combination of qualitative and quantitative methods that draw on multiple data sources (Bond & Halpern, 1988; Kalafat & Illback, 1998; Lipps & Grant, 1990; Roitman, Gotschalk, Mayer, & Blakely, 1983).

This article describes the application of such procedures in the evaluation of a widely disseminated school-community family support program aimed at improving educational outcomes for high-risk children and youth. This evaluation consisted of a 4-year collaboration between evaluators and program personnel that started with a qualitative exploration and explication of program components and processes. This formed the basis for a quantitative evaluation of the degree of implementation of program components, leading to an analysis of the relationship between levels of implementation and program outcomes. First, the family support program will be described along with the evaluation challenges that it presented. The qualitative process evaluations have been previously reported (Kalafat & Illback, 1998; Kalafat, 2003) and are briefly summarized here. The remainder of the paper describes the implementation fidelity and outcome evaluations.

1.1. The Family Resource Center program

The Family Resource Center (FRC) program is a multisite statewide initiative consisting of school-based centers that are an essential component of Kentucky's Educational Reform Act (KERA). Their overriding mission is to ensure that children come to school healthy, safe, and prepared to learn. The inclusion of family support programs in educational reform initiatives across the country was prompted by the confluence of several findings. The grim reality for a growing number of families is reflected in such statistics as, a quarter of children under six in the USA live in poverty (Children's Defense Fund, 1996), and the growing prevalence of what Dryfoos (1994) called the "new morbidities" of youth such as interpersonal violence, substance abuse, sexually transmitted diseases, and unwanted pregnancies. These conditions have overwhelmed service systems for families and children, which have been characterized as fragmented, inaccessible, disrespectful, and ineffective (Illback, 1994; Knitzer, 1997). It is now recognized that such socially disintegrative conditions are significant barriers to children's learning (Adelman & Taylor, 1997; Dryfoos, 1994). And, there is growing evidence for the positive impact of parental involvement on school performance in the education process (Christenson, Rounds, & Gorney, 1992).

Consistent with other family support programs, FRCs operate on a number of levels. They provide wraparound services to address the physical, psychological, and social

needs of children and their families; they seek to enable or empower families to participate more fully in the educational process and to access a variety of community services, and they strive to forge cooperative links among families, schools, and communities as expressed by the FRC motto, "Bridges Over Barriers" (Kagan & Weissbourd, 1994; Kalafat & Illback, 1998).

While FRCs were mandated to provide a set of core services in Kentucky, the enabling legislation called for the development of service profiles adapted to local community needs and contexts. Thus, in addition to core services, centers provide a wide variety of services at individual, group, and community levels. Center services are characterized as "family-centered" rather than "agency centered" (Trivette, Dunst, & Hamby, 1995) in that, rather than offering eligibility-based or categorical services, centers tailor services to identified client needs in wrap-around fashion. Legislation also provided for rapid proliferation to over 500 sites throughout the state during the first 4 years of operation. An initial group of about 40 coordinators participated in a 1-week immersion orientation that emphasized an empowerment mission (Dunst, Trivette, & Deal, 1988), and bonded them into a mutual support network. Coordinators who came into the project subsequent to this group received little or no orientation or initial training, aside from an annual FRC conference and a nascent grassroots coordinator support network. The FRC diffusion process had, in effect, resulted in a set of "field experiments" in which local committees conducted a community needs assessment, and selected center coordinators who were given broad mission and service guidelines.

Data from the management information system developed for the centers revealed that the populations being served reflect the program's goals, as family demographics showed high levels of undereducated and economically disadvantaged persons. Children and youth exhibited diverse, multiple, and interrelated difficulties, with health, behavior, emotional, and learning problems most frequently noted at intake. Family and setting risk factors were also numerous and complex, including financial needs, clothing, childcare, food, family conflict, and divorce (Illback, 1998). These multiple needs and stressors can be seen as barriers to learning, and are hypothesized to be amenable to interventions consistent with the aims of the program.

1.2. Process evaluation

Rather than adhering to specified program parameters, FRCs provide a range of flexible, needs-based services to clients who choose their level and type of involvement. As such, it is not only impossible to standardize the treatment in such programs, it is difficult to specify actual program parameters (Weiss & Jacobs, 1988). Self-selected participants also compromise sample integrity and represent differing units of analysis. Since full-coverage programs are mandated to serve all families, these programs leave little room for comparison conditions, given ethical, political, and logistical considerations (Powell, 1994). Moreover, factors that influence participation are difficult to secure in comparison groups, which attenuates the logic of such designs (Knapp, 1995).

In the absence of controlled implementation of specified interventions, the task was to gain an understanding of what these complex, dynamic FRC programs were *doing* to meet a mandate that was both flexible and ambitious. Thus, as with other youth-oriented prevention programs, there have been calls for process and implementation evaluations of family support programs. For example, Bond and Halpern (1988) reported that their "most radical strategic decision" (p. 353) was to break the summative evaluation mold and to carry out formative evaluation to identify important activities and domains; and implementation evaluation to describe what varying levels of implementation would look like. Knapp (1995) stated that if comprehensive, collaborative services can be so many different things, it makes sense to initially put a great deal of emphasis on qualitative "thick descriptions" (Geertz, 1973).

The evaluation of FRCs was carried out in four waves, each consisting of reviews of program documents and annual work plans, and interviews of program staff, school principals, community agency staff, and families. In the first wave, open-ended orientation interviews were conducted by the first author with an intensity sample of 10 information-rich centers from urban and rural regions of the state. In the second wave, focused observation and interviews moved from general description to more targeted probing, as analysis of the data from the first phase allowed general themes to emerge that required further investigation. A purposeful random sample (Patton, 2001) of 12 centers was selected and 3 centers from the first wave were re-visited to examine developmental trends. Two interviewers conducted site visits together at two centers, supplementing open-ended questions with an interview guide for pursuing themes from the first wave. They compared the content of their field notes to ensure agreement and split the remaining site visits between them. Each wave was followed by member checks of results with a standing committee of center directors and staff from the sponsoring state agency and presentations at annual coordinators' conferences. Feedback from coordinators provided anecdotal evidence for the validity of findings. Comments such as "You did not visit my center, but you described it exactly" were common.

In sum, qualitative evaluations consisting of observation, document review, and interviews of multiple sources identified common cross-site program characteristics (Kalafat & Illback, 1998). Themes that emerged and held across the first two qualitative waves included: (1) the critical importance of coordinator characteristics and competencies to the operation and success of the center (a finding similar to those of Schorr (1997) and Elias (1997)

who characterized programs as 'operator dependent'); (2) the need for principal support; and (3) the salience of the degree of center relationship with the school, community, and families. Barriers to effective programming were also identified, including an overemphasis on the provision of direct services to student and family "clients" rather than on the development or brokering of collaborative, empowering, preventive programming and resources, and "mission drift" consisting of the provision or brokering of a wide variety of programs and resources that taxed the limited resources of centers and whose relationship with or impact on educational outcomes for students was questionable.

These qualitative evaluations began to flesh out the hypothesized mediators of the stated program goals of FRCs. That is, center coordinators appeared to cross boundaries and establish ongoing relationships or "clout" with families, schools, and communities in order to facilitate more cooperative relationships among them and to enable families to effectively access school and community resources. In addition to these cross-site commonalities, some variations in approach associated with the centers' contexts were identified. For example, urban centers operated in a relatively "resource rich" environment compared to rural centers; and the urban centers tended to broker services, whereas rural centers were more likely to develop and/or directly provide services to families. Additional detail as to the enabling and empowering manner in which coordinators interacted with families was provided by a multiple case study procedure (Kalafat, 2003).

Patton (2001) suggested that it is possible to convert detailed qualitative descriptions into quantitative scales. In the third and fourth waves, this was done employing a method called Innovative Configuration Analysis (Hall & Hord, 1987) to assess the degree or levels of implementation of program components hypothesized to address barriers to learning. The next section describes this methodology. The sections that follow then describe two core elements of the overall evaluation of the FRYSC program: (1) assessment of implementation fidelity; and (2) exploration of the relationship between center-level implementation and educational outcomes of program participants.

2. Implementation evaluation

2.1. Method

2.1.1. Measures

Hall and his associates at the Texas R&D Center developed a methodology based on the assumption that social programs consist of a finite number of components (Hall & Loucks, 1978; Heck, Steigelbaur, Hall, & Loucks, 1981). The methodology, called Innovation Configuration Analysis (Hall & Hord, 1987), consists of drawing upon detailed process findings that identify program activities in specific domains and quantifying the presence of activities or elements associated with each domain. Applications of this methodology have included the demonstration of the success of a program diffusion approach that called for high fidelity implementation of a reading instruction program (Reid, 1980), and demonstrations of a positive relationship between degree of implementation and program outcomes (Blakely et al., 1987; Pratt, Winters, & George, 1980).

For this study, a committee of FRC evaluators, center coordinators, and state administrative personnel held several meetings to operationalize various levels of implementation of program domains that were based on the process findings. The seven domains that emerged were Needs Assessment, Relationship With School, Relationship With Community, Relationship With Families, Advisory Council (involvement of), Mission Focus (degree to which activities addressed barriers to learning), and Evaluation (carried out by coordinators). Levels of implementation for each domain were explicated. Successive versions of the analysis were submitted to a number of center coordinators for review and modification. This iterative process resulted in the Innovative Configuration Analysis measure employed in this study (see Appendix A for Relationship With School Domain).

Overall domain scores, representing the degree of implementation, can be derived by dividing the number of points earned in a domain by the total number of points (yielding a percentage score). Hall and Hord (1987) called this implementation measure an Innovation Components Configuration (ICC) map. These ICC maps profile the degree of implementation of identified program domains for individual centers or across groups of centers. In sum, the Innovation Configuration Analysis captured common cross-site activities of the FRC initiative in specific program domains, and described different levels of implementation in each domain.

A second implementation measure was developed, consisting of a survey of teachers at the respective schools (some centers served more than one school) served by each center. The survey assessed the degree of knowledge, support of, and involvement with the center in their school. In effect, the survey measures "market penetration" for an important constituency in the centers' efforts to address barriers to learning.

2.1.2. Participants

The implementation evaluation was also carried out in two waves. In each, 10 centers were randomly selected for site visits, resulting in a total sample of 20 centers. Table 1 provides descriptive information about the 20 centers, including setting characteristics and number of schools served. As in the first two wave samples, a review of MIS data on family demographics and family and children problems most frequently noted at intake revealed that the populations being served reflect the program's goal.

Table 1
Center characteristics

Center	Setting[a]	Number of schools served	Total enrollment	% free/reduced lunches	County ranking for educationally at risk[b]
1	U	1	946	35.0	34
2	R	3	1722	40.4	59
3	R	3	1245	40.9	61
4	R	1	377	73.7	118
5	S	1	531	43.1	45
6	R	1	600	84.6	115
7	R	2	534	54.6	98
8	R	2	1117	59.1	72
9	U	1	427	46.7	56
10	U	1	1306	62.3	56
11	R	1	579	58.2	59
12	R	2	815	58.2	80
13	R	4	1622	26.2	6
14	S	2	368	43.1	45
15	R	3	1055	44.4	70
16	S	3	1141	44.2	41
17	R	1	198	77.2	119
18	R	2	680	76.4	114
19	U	1	649	56.4	56
20	R	1	475	84.7	98

For the purposes of this research, classification of a particular center along the urban, suburban, rural dimension was made in relation to both population density and distance from the city center (per US Census Bureau definitions). Generally, centers were considered urban in areas where the population density was greater than 2000 per square mile and relatively close to the city center; rural where population density was less than 2000, and suburban when density was greater than 2000, but the features of the area did not qualify as urban relative to distance from the city center.

[a]U = urban, S = suburban, R = rural.

[b]Ranking out of 120 Kentucky counties where 1 represents the lowest proportion and 120 the highest proportion of at-risk students, based on data supplied by the Kentucky Department of Education.

2.1.3. Procedure

The procedure employed to collect data for the current evaluation study was similar to that employed within the first two waves. Evaluators conducted site visits that included reviews of annual work plans required for centers and other materials developed by centers, and semi-structured interviews with the same categories of individuals (coordinator, staff, parents, and principal). Intending to maintain a collaborative, less formal atmosphere, interviews were guided by the ICC instrument, but were not completely structured around its items. Field notes and some tentative ratings were made during the site visit, and the ICC ratings were completed after the site visit. Again, two evaluators conducted joint site visits of two centers in each of the two samples of centers. For the first sample of centers, the evaluators agreed on 85% of their ratings of two centers, and for the second sample of centers, they achieved 84% agreement on their ratings of two centers. They then conducted individual site visits with the remaining centers.

2.2. Results

2.2.1. ICC

Centers achieved varying levels of implementation as depicted in Table 2. The evaluation domain did not prove to be informative because most centers did not conduct evaluations aside from participant feedback. This domain was not included in subsequent analyses. Centers were rated highest on formation and utilization of Advisory Councils (87.5) (a state-mandated component) and Needs Assessment (79.0); followed by Relationship with Families (76.7) and Mission (70), and then Relationship With School (59.1) and Relationship with Community (56.1).

2.2.2. Teacher surveys

Return rates of teacher surveys for the first and second samples of centers were 75.7% (334/411) and 77.6% (308/397), respectively. Mean responses across the 20 centers to teacher surveys are presented in Table 3. Results indicated that most of the teachers' involvement with the centers consisted of responding to needs assessments (81.3%), receiving services (66.3%), and referring students and families to the centers (73.9%). The referral result is supported by 1-year data from the MIS, which revealed that, after self-referrals (32.6%), teachers were the most common source of referrals to the centers (25.4%). The majority of teachers rated themselves as familiar or very familiar with centers (67.7%), having visited the centers occasionally or often (67.3%). They considered the centers a very important or indispensable resource (76.7%).

Responses to an open-ended survey item, "What do you see as the main purpose or mission of the FR/YSC?" were placed into one of five categories (see Table 3). A 92%

Table 2
Center implementation by domain

Center	Advisory council	Mission	Needs assess	Relate W/family	Relate W/community	Relate W/school
1	90.9	62.5	70	86.7	64.3	63.9
2	72.7	75.0	63.3	66.7	57.1	42.2
3	95.5	62.5	90.0	90.0	78.6	77.8
4	81.8	75.0	76.7	80.0	78.6	61.6
5	100.0	75.0	80.0	80.0	78.6	61.6
6	68.2	50.0	80.0	93.3	53.6	61.6
7	90.9	75.0	80.0	80.0	64.3	61.6
8	95.5	87.5	86.7	93.3	78.6	86.1
9	81.8	62.5	80.0	66.7	60.7	61.6
10	81.8	75.0	80.0	63.3	60.7	50.0
11	90.9	62.5	83.3	93.3	67.9	86.1
12	90.9	50.0	76.7	86.7	75.0	61.6
13	81.8	87.5	73.3	80.0	60.7	80.6
14	72.7	75.0	80.0	80.0	57.1	61.6
15	95.5	50.0	80.0	86.7	60.7	44.4
16	90.9	62.5	70.0	86.7	64.3	63.9
17	81.8	75.0	80.0	66.7	67.9	61.6
18	95.5	100.0	90.0	93.3	78.6	97.2
19	100.0	87.5	86.7	86.7	64.3	77.8
20	90.9	50.0	73.3	73.3	50.0	69.4
Mean	87.5	70.0	79.0	81.67	66.08	66.61

Table 3
Mean responses to teacher surveys across twenty centers[a]

1. Teacher perception of FRYSC mission ($N = 500$)	
No response	23.3%
Don't know/not sure	4.3%
General help students	30.4%
Describe specific help for students	35.9%
Connect help to school performance	14.6%
2. Involvement with FRYSC	
Responded to needs assessment	81.3%
Received service/assistance	66.3%
Referred to FRYSC	73.9%
Attended program/activity	48.5%
Assisted/participated in program	36.1%
Provided program/activity	10.5%
Served on advisory council	12.2%
3. Familiarity with FRYSC ($N = 634$)	
Not very	6.0%
Somewhat	26.1%
Familiar	43.4%
Very familiar	24.3%
4. Visited FRYSC ($N = 636$)	
Never	13.8%
Once or twice	18.1%
Occasionally	37.5%
Often	29.8.%
5. Role of FRYSC	
Not sure	7.8%
Nice to have	14.6%
Very important resource	22.7%
Services school should not do without	54.0%

[a]Number of responses is 642, unless noted.

interrater agreement rate by the two site visitors was achieved on their first attempt to rate survey responses from two centers. While the majority of teachers indicated that centers generally helped students or identified specific assistance that centers provided to students, only 14.6% connected the centers' mission to improved school performance. These results must be interpreted with caution as they may represent what teachers were willing to write more than what they actually know.

2.2.3. *Validity checks*

Validity of the measures was assessed, first, by triangulation of results of the ICC and the teacher survey. Given the centers' mission to enhance students' readiness to profit from their educational experience, the Relationship With School domain is of particular importance. For the teacher survey, overall estimates of teachers' involvement with and views of the centers were obtained for each center by deriving the average rating for each survey item for that center. Evidence for the validity of the Relationship With School domain is provided by high positive correlation between implementation ratings (based on the site visits) and teacher surveys. Ten of the teacher survey variables correlated .70 or higher with the Relationship With School domain scores for their respective FRCs. The highest correlations were between the centers Relationship With School Domain and responses of teachers to How familiar are you with FRYSC services/activities ($R = .978$), Received services from FRYSC (.909), Describe the role of FRYSC in your school (.904), How often visited FRYSC (.896), Referred child or family to FRYSC (.887). These

correlations depict a convergence between (primarily) coordinator estimates and teacher reports of the linkages between centers and schools.

Second, teacher survey scores were averaged across schools that served single and multiple schools, respectively. These scores differed between centers that served single and multiple schools in expected ways. That is, the site visit interviews with coordinators and principals had revealed that centers serving more than one school had difficulty establishing solid relationships with all of their school sites. Subsequently, independent samples *t*-tests were used to compare multiple and single school centers on mean ratings for teacher survey variables. The mean scores for single school centers were significantly higher on eight of the 10 variables, indicating better connection with teachers at single school centers (Table 4).

3. Implementation fidelity as a predictor of outcome evaluation

This component of the evaluation sought to test a central hypothesis of the FRC program theory: that targeted, multidimensional interventions with educationally at-risk students can reduce barriers to learning, and thereby improve student educational outcomes. One means to achieve at least some preliminary understanding of program functioning and effectiveness is to determine how key implementation variables correlate with (predict) educational gains for students served by the program. While this approach does not directly test whether program-specific processes enable educational gains (a difficult task with such a complex program and setting), it does facilitate knowledge about whether program processes and outcomes conform to the theory of change presupposed by proponents of integrated services (e.g., Adelman & Taylor, 1997; Dryfoos, 1994).

3.1. Method

3.1.1. Participants

The center sample consisted of the 20 sites that comprised the purposeful random sample of centers that were visited in the two waves of the implementation evaluation. Again, Table 1 provides descriptive information about the 20 centers. The student sample was comprised of 5161 students within the 20 centers that were served during the 2-year period corresponding to the implementation evaluation. The average duration for targeted interventions was 237 days for students served within FRCs.

In order to conduct exploratory analyses, decisions also had to be made as to which students within the 20 centers should be included *within* each analysis. Because the reason for referral to centers was idiosyncratic as to student educational needs, not all participating students received ratings from the teachers on every educational outcome variable (i.e., some were irrelevant to the focus of the specific intervention). Teachers recorded information relevant to the problem for which the student was referred to the center, and did not always complete items that did not seem to apply to that particular student. Therefore, cases with missing data were not excluded from the analysis. It was also decided that the most compelling measure of program effectiveness is the degree to which program implementation was associated with change within specific educational problem areas (recognizing that integrated service programs tend to address a range of issues and problems concurrently). Therefore, for each outcome variable, cases were excluded from specific analyses when an initial satisfactory rating was given by the teacher, assuming that amelioration of this particular problem area was not at issue in the intervention. Therefore, the number of students included in analyses involving different educational outcome variables (gain scores) ranged from 200 to 1770 (see Table 6 for the *n* within each outcome analysis).

Table 4
Comparison of teacher responses to single- and multiple-school centers

Item[a]	Single school centers		Multiple school centers			
	M	SD	*M*	SD	*t*	*P*
Mission ($N = 500$)	2.81	.832	2.52	.818	3.75	.001
Responded to survey questions	.824	.381	.816	.389	.270	NS
Received service/assistance	.740	.439	.584	.490	4.03	.001
Referred a child of family	.829	.378	.636	.482	5.28	.001
Attended a program or activity	.572	.495	.373	.485	4.96	.001
Assisted/participated in program	.455	.499	.232	.423	6.01	.001
Provided program or activity	.127	.334	.101	.261	2.28	.031
Served on Advisory Council	.103	.304	.098	.274	.903	NS
How familiar with center ($N = 634$)	3.05	.784	2.65	.878	5.79	.001
How often visited center ($N = 636$)	3.18	.810	2.45	1.10	8.83	.001

[a]Number of responses is 642, unless noted.

It will also be noted on Table 6 that the number of centers varies. This occurs because, for a given education variable, one or more centers did not have a sufficient number of students who were initially rated as deficient to justify statistical treatment. While the above strategy of flexible sample selection represents a methodological compromise dictated by real-world circumstances, it may also serve as a useful approach to exploratory data analysis within broad-based, integrated service program evaluation where multiple and complex problems are of concern.

3.1.2. Measures

The specific independent variables for the analysis consisted of center-level implementation data for the six implementation domain scores (Needs Assessment, Relationship with School, Relationship with Community, Relationship with Families, Advisory Council, and Mission Focus) and 11 teacher survey items from the implementation evaluation data set.

Given the number and co-linearity of these implementation variables, a principal components factor analysis with varimax rotation was performed on the implementation and survey data for data reduction purposes. As shown in Table 5, factor analysis yielded five factors accounting for 88% of the variance in the 17 constituent variables. Based on visual inspection, the five factors were labeled as follows:

Factor 1: Teacher/principal knowledge and familiarity with center.
Factor 2: Teacher active involvement with center.
Factor 3: Center involvement with families and community.
Factor 4: Center program mission focus (degree of focus on educational readiness).
Factor 5: Teacher awareness of center mission.

Table 5
Factor analysis of implementation variables (varimax rotation)

*F*1	*F*2	*F*3	*F*4	*F*5	
.94					Teacher referred child/family
.90					Teacher received service or assistance
.86					Teacher rating of importance
.85					Teacher familiar with services and activities
.83					Teacher visited center
.74					Implementation—relationship with School
.57					Teacher responded to needs assessment
	.88				Teacher provided activity for center
	.84				Teacher served on Advisory Council
	.82				Teacher assisted in center activity
	.81				Teacher attended center activity
		.88			Implementation—Advisory Council
		.88			Implementation—relationship with families
		.74			Implementation—needs assessment
		.63			Implementation—relat. with community
			.85		Implementation—mission
				.85	Teacher recognizes center mission

Each of these factors can be seen as a proxy for a more complex set of process variables, but they are useful for exploratory purposes. In order to estimate the relationship between program implementation and educational outcomes, these five factors were used in subsequent multiple regressions to predict student improvement at the center level on each of the educational outcomes (gain scores).

Also, using the unstandardized regression coefficients generated in the above regression analyses, the five factors were combined into a single latent variable (second-order factor) termed *Overall Implementation Fidelity Index*. The Overall Implementation variable embodies the contribution of each of the five factors. The contribution of each factor to the prediction equation differed as a function of the educational outcome being predicted in the regression. Thus, for the purpose of the exploratory analyses that follow, *each* of the 17 education outcome variables has an *Overall Implementation Fidelity Index* variable associated with it.

The dependent variables consist of gain scores on the items from a locally developed teacher-rating scale called the Educational Status Checklist (see Table 6). Classroom teachers were asked to judge student status on the checklist at the beginning and at completion of center intervention with the student. Each rating is dichotomous (yes–no), with the exception of At Risk Educationally (High, Moderate, Low), Academic Achievement (performance at, above, or below grade level), and teachers' report of Academic Proficiency achieved on the statewide KERA assessment (Novice, Apprentice, Proficient, Distinguished). Indicators

Table 6
correlations between center success rates and overall implementation variables

Educational outcome	R^a	*N* Students	*N* Centers
Classroom variables			
Remains on task	.692	927	18
Obeys rules	.670	493	18
Follows directions	.666	654	17
Completes homework	.604	799	18
Completes classwork	.541	640	17
Tardiness	−.519	212	14
Attends regularly	.295	238	17
Peer relations *variables*			
Relates appropriately	.810	489	17
Cooperates	.705	428	17
Participates	.702	281	17
Has friends	.564	222	18
Risk variables			
Drop out risk	−.662	903	18
At risk for retention	−.623	472	18
At risk educationally	−.555	447	17
Global variables			
Achieve above/below grade level	−.570	1020	19
Retained previously	.475	200	17
Academic proficiency (KERA)	.401	1770	17

[a]All significant at $<.001$.

can be grouped with four general educational outcome areas: (1) classroom performance; (2) peer relations; (3) perceived risk; and (4) global school success.

A review by Hoge and Coladarci (1989) provided support for the use of teacher-based judgments of student performance. Assessing 16 studies of the relationship between teacher judgments and students' actual performance, they concluded that there is evidence of moderate to strong correspondence (with a median correlation of .66). Most of the studies reviewed used teacher ratings. The classroom-based teacher judgment variables selected for pre-post assessment in the present study (shown in Table 6) were chosen on the basis of teaching and learning research that suggests their predictive power (see review by Hoge and Andrews (1987) for discussion of the relationship of academic performance and classroom behaviors to academic achievement). They include on-task behavior (academic engaged time), rule compliance, work completion, school attendance, peer relations, retention, and general risk perception (Elliott & Shapiro, 1990; Kamphaus, Yarbrough, & Johanson, 1990).

Changes of teacher perceptions for one or more problem areas are considered gain scores for the purpose of this study. Improvement was considered to have occurred when a higher level of teacher-rated performance on a given educational variable was attained at post-intervention. For dichotomous variables, this consists of moving from the deficient to the non-deficient category (e.g. "does not remain on task" to "remains on task"). For the non-dichotomous variables, improvement consists of movement of 1 or 2 levels for the achievement variable (from "below", to "at" or "above" grade level), and at risk educationally (from high to moderate or low); or 1, 2, or 3 levels on the KERA proficiency variable (improve from Novice to Apprentice, Proficient, or Distinguished, or from Apprentice to Proficient or Distinguished, or decline from Apprentice to Novice).

As noted earlier, for the purpose of this exploratory analysis, gain scores were aggregated within the centers and expressed as a *success rate* for each center on each educational variable. This success rate reflects the proportion (percentage) of students who were below criterion at the pre-test and reached criterion level on a given education outcome (i.e. educational improvement) at the time of post-test. For the dichotomous variables, this measure is effectively the percent of students who improved from a deficient to a non-deficient category for each of the categorical educational outcomes. For the non-dichotomous variables, this measure is effectively the average number of levels of student improvement for a center.

3.1.3. Analysis

Multiple regression analysis of the independent (predictor) and dependent (criterion) variables described above was conducted. To reiterate, predictor and criterion variables included in regression analyses were school-level, but each center success rate variable was computed for a different group of students (those who had unsatisfactory ratings for each educational outcome variable at intake).

3.2. Results

The Overall Implementation Fidelity Index variables were used as predictor variables in multiple correlational analyses of the relationship between implementation and educational outcomes. Table 6 portrays the resulting multiple correlations (R), with accompanying columns showing the number of students and centers included in that particular analysis (see decision rules for analysis discussed earlier). For the most part, the number of centers included is between 17 and 19, with only 14 having sufficient data for analysis of the variable "tardiness". Across the centers, the number of students included varies as a function of the extent to which that particular problem area is noted at intake. All of the multiple Rs are arrayed in the predicted direction, indicating that, in general, the more fully a center program is implemented (relative to our scale), the higher its educational success rate. For those educational outcomes that are stated in negative terms (e.g. at risk, tardiness), the correlation is negative, as would be expected. Also, the achievement variable (at/below grade level) is coded inversely, so the correlation is in the predicted direction. That is, the greater the implementation, the better the gains with respect to achievement.

Table 7 depicts center success rates for each education variable, showing that there was considerable variability across centers in success rates (aggregated gain scores). In this context, a large standard deviation is considered to indicate that the distinction between greater implementation and lesser implementation represents meaningful differences in the educational improvements among students served by the centers.

4. Discussion

Strong positive relationships between overall program implementation level and program-level outcomes achieved by program participants, as measured by teacher ratings of relevant social and academic areas as well as performance on a statewide proficiency exam, add to the emerging body of research exploring relationships between program implementation and outcomes. The findings also provide preliminary evidence in support of the program logic of the FRCs: that school based family support programs that address barriers to learning may alter the trajectories of educationally at risk youth.

We characterize the findings as preliminary because this study was not designed as a controlled experiment but instead as an exploration of the relationship between identified process and outcome variables. Our intent was not to validate the program's efficacy or make definitive statements about causal factors. Rather, this study demonstrates the utility of deriving measures of hypothesized program mediators from detailed qualitative descriptors,

Table 7
Descriptive overview of center success rates

Educational outcome	Minimum implementation center	Maximum	Mean	Standard deviation
Classroom variables				
Remains on task	21.88	89.29	36.89	15.50
Obeys rules	23.53	100.00	43.81	18.54
Follows directions	32.00	95.65	45.51	15.25
Completes homework	0.00	95.45	37.67	16.12
Completes classwork	27.27	100.00	46.63	15.06
Tardiness	8.33	80.00	55.19	19.97
Attends regularly	30.00	100.00	59.24	15.76
Peer relations *variables*				
Relates appropriately	21.54	100.00	49.08	17.36
Cooperates	25.00	100.00	51.64	19.88
Participates	33.33	100.00	61.21	17.09
Has friends	31.82	100.00	60.36	16.90
Risk variables				
Drop out risk	7.69	75.00	24.74	15.15
At risk for retention	0.00	100.00	47.25	19.60
At risk educationally	12.50	141.67	52.13	30.31
Global variables				
Achieve above/below grade level	0.00	112.50	29.71	18.13
Retained previously	16.67	100.00	75.50	22.59
Academic proficiency (KERA)	−33.77	46.15	8.36	11.84

*All significant at $<.001$.

based on prolonged engagement and an iterative collaborative process evaluation. Through this procedure, the evaluation served to open the "black box" of a comprehensive school-community-based program to a certain degree. That is, common cross-site program activities in certain domains and measures of the degree of implementation of these domains individually and in the aggregate were identified. The findings provide evidence that these domains can be used reliably, as well as evidence of convergent and predictive validity in that multiple measures of a given domain (e.g. Relationship to School) relate to each other and to outcomes in expected ways.

Additional support for their validity as measures of relevant program mediators is provided by similar findings from other evaluations. For example, Gager and Elias (1997) examined variables hypothesized to influence degree of implementation of school-based prevention programs and the relationship of these variables to program goal attainment. They found that goal attainment was best predicted by program visibility in the school, principal support, characteristics and performance of program implementers, and the linkage of the program to school goals or mission. Our evaluation identified the centrality of the program coordinators, and both the teacher survey and Connection to School domain measured similar variables to those identified by Gager and Elias. Also, the correlation between our overall implementation fidelity index and academic achievement as measured by performance on the statewide KERA Proficiency exam (.40) was similar to the correlation between program fidelity and effectiveness across seven educational and criminal justice programs (.44) reported by Blakely et al. (1987).

Undoubtedly, the domains identified in this evaluation are proxies for a variety of specific actions on behalf of students and families that vary according to the emerging needs of students and families and their individual ecologies. In this regard, the specific make up and contribution to outcomes of each domain requires further research.

However, while the elements that make up the domains require further explication, delineating the specific activities within each domain is not the most salient consideration for understanding and replicating broad scale programs. That is, we are not proposing that if one provides such services as a given amount of parent training along with brokering specific community services and quid pro quo parent involvement, one will achieve these educational outcomes. Regarding efforts to scale up effective prevention programs, Schorr (1997) noted that successful programs were "tight about their mission and simultaneously loose about how their mission was carried out" (p. 8). Thus, the present study provides preliminary support for the proposition that the implementation of individual, group, and systemic interventions in a mission-focused and empowering manner may be associated with positive social and educational outcomes for at risk children. The specific activities through which these interventions are implemented will vary across different

contexts. The overall level of implementation fidelity, and perhaps, levels of implementation of individual domains, are candidates for more systematic analyses of their relationship to program outcomes.

In their overview of the special issue of the *American Psychologist* on prevention that works for children and youth, Weissberg et al. (2003) concluded that effective prevention programs for youth seek to structure settings that systematically and regularly provide services, supports, and opportunities for families and children as an integral part of standard practice. The evaluations of Kentucky's FRCs describe one such initiative and its proposed mediators. To the extent that successive implementation and outcome evaluations support this proposition, an empirical basis for the program logic of this initiative and for its diffusion will be more firmly established.

5. Lessons learned

Program evaluators must consistently be mindful of their primary stakeholders, who have a substantial influence on the course and emphasis of the evaluation. The major stakeholder for this evaluation was the state agency that sponsored the Family Resource initiative. This had both positive and negative influences on the evaluation. On the positive side, the state committed to a rapidly expanding program that emphasized local control and adaptation to local conditions within a broad conceptual framework. The program logic was clear and permitted the application of several recommended qualitative evaluation strategies, including prolonged engagement, member checks, and a collaborative inductive approach, which in turn formed a solid basis for exploring the relationship of the program implementation to program outcomes. This utilization-focused approach provided data that was useful to program personnel. However, had we been more mindful of the general field of school-based supports as a stakeholder, we might have dedicated resources to identify comparison groups in the schools consisting of at risk students who had not been served by the FRCs. This would have yielded stronger evidence as to program mediators.

On the negative side, the state management staff for the program turned over in each of the first 3 years of the evaluation, which attenuated their involvement with the evaluation. As we have learned in other evaluations, this could have led to problems if the program sites had been resistant to the evaluation. While we were fortunate that they were not, the evaluation findings were rarely used to inform and highlight this innovative program. When a more knowledgeable and involved management team was in place, they utilized the evaluation results and staff, as well as program personnel, to implement training and guidelines for new site directors.

Appendix A

Innovation Components Configuration Instrument Sample Sections

Center_____.
Coordinator_____.

I. **Introduction**

1. Who developed the original proposal?_____.
2. How long has the program been operating?____years.
3. How long have you been coordinator?_____years. Original?___Y___N.

III. **Relationship With School.**

1. Connectedness
   1. Negative (turf/hostile).
   2. Uninformed (inappropriate or no requests/referrals)
   3. Customers (school personnel make appropriate requests/referrals; starting to be impressed with center capabilities).
   4. Advocate (enthusiastic about center; see it as needed resource/capability of school).
   5. Team (school works collaboratively with center; buys into the family involvement education mission).

2. Principal support
   1. Negative (either micromanaging or turf/hostile).
   2. Laissez-faire
   3. Supportive/customer
   4. Advocate
   5. Involved/collaborative

3. Principal's attitude toward family involvement
   1. Not open to it
   2. Neutral
   3. Appreciates outreach to families
   4. Promotes family involvement
   5. Has track record and strategies for family involvement.

4. Teacher acceptance of parent involvement.
   1. None
   2. Some (very little)
   3. Many (somewhat)
   4. Most/all (very much)

5. Teacher interaction with center
   1. Refer to center: (3) all, (2) most, (1) some, (0) none. (multiply "1" by (#).
   2. Drop by center: (3) all, (2) most, (1) some, (0) none. (multiply "2" by (#).
   3. Participate in center programs/activities: (3), (2), (1), (0). (multiply "3" by (#).

4. Assist in center programs/activities: (3), (2), (1), (0). (multiply "4" by (#).

6. Perceived effects on school performance
   1. Center is not seen as improving student attendance, classroom performance, or student achievement.
   2. School personnel are unsure about the effect of the center on student attendance, classroom performance, or student achievement.
   3. Center seen as important, but school personnel unsure of impact.
   4. Center is clearly seen as improving student attendance, classroom performance, and achievement.

Note: This is one domain; the full instrument is available from the first author.

References

Adelman, H. S., & Taylor, L. (1997). System reform to address barriers to learning: Beyond school linked services and full service schools. *American Journal of Orthopsychiatry*, *67*, 408–421.

Altman, D. G. (1986). A framework for evaluation of community based heart disease prevention programs. *Social Science and Medicine*, *22*, 479–487.

Bauman, L. J., Stein, R. E. K., & Ireys, H. (1991). Reinventing fidelity: The transfer of social technology among settings. *American Journal of Community Psychology*, *19*, 619–639.

Blakely, C. H., Mayer, I. P., Gottschalk, R. G., Schmitt, N., Davidson, W. S., Roitman, D. B., et al. (1987). The fidelity-adaptation debate: Implications for the implementation of public sector social programs. *American Journal of Community Psychology*, *15*, 253–268.

Bond, I. T., & Halpern, R. (1988). The cross-project evaluation of the child survival/fair start initiative: A case study of action research. In H. B. Weiss, & F. H. Jacobs (Eds.), *Evaluating family programs* (pp. 347–370). New York: Aldine de Gruyter.

Catalano, R. F., Berglund, L., Ryan, J. A. M., Lonczak, H. S., Hawkins, J. D. (2002). Positive youth development in the United States: Research findings on evaluations of positive youth development programs. *Prevention and Treatment*, *5*, Article 15. Retrieved August 10, 2004, from 〈http://journals.apa.org/prevention/volume5/pre0050015a.html〉.

Children's Defense Fund. (1996). *The state of America's children: yearbook 1996*. Washington, DC: Author.

Christenson, S. L., Rounds, T., & Gorney, D. (1992). Family factors and student achievement: An avenue to increase students' success. *School Psychology Quarterly*, *7*, 178–206.

Dane, A. V., & Schneider, B. H. (1998). Program integrity in primary and early secondary prevention: Are implementation effects out of control? *Clinical Psychology Review*, *18*, 23–45.

Dobson, D., & Cook, T. J. (1980). Avoiding type III error in program evaluation: Results from a field experiment. *Evaluation and Program Planning*, *3*, 269–276.

Dryfoos, I. G. (1994). *Full-service schools*. San Francisco: Jossey-Bass.

Dumas, I. E., Lynch, A. M., Laughlin, J. E., Smith, E. P., & Prinz, R. J. (2001). Promoting intervention fidelity: Conceptual issues, methods, and preliminary results from the early alliance prevention trial. *American Journal of Preventive Medicine*, *20*, 38–47.

Dunst, C. J., Trivette, C. M., & Deal, A. G. (1988). *Enabling and empowering families: Principles and guidelines for practice*. Cambridge, MA: Brookline Books.

Durlak, J. A., & Wells, A. M. (1997). Primary prevention mental health programs for children and adolescents: A meta-analytic review. *American Journal of Community Psychology*, *25*, 115–152.

Elias, M. J. (1997). Reinterpreting dissemination of prevention programs as widespread implementation with effectiveness and fidelity. In R. P. Weissberg, T. P. Gullotta, R. L. Hampton, B. A. Ryan, & G. R. Adams (Eds.), *Establishing preventive services* (pp. 253–289). Thousand Oaks, CA: Sage.

Elliott, S. N., & Shapiro, E. S. (1990). Intervention techniques and programs for academic performance problems. In T. B. Gutkin, & C. R. Reynolds (Eds.), *Handbook of school psychology* (pp. 635–660). New York: Wiley.

Gager, P. J., & Elias, M. J. (1997). Implementing prevention programs in high-risk environments: Application of the resiliency paradigm. *American Journal of Orthopsychiatry*, *67*, 363–373.

Geertz, C. (1973). Thick description: Toward an interpretive theory of culture. In C. Geertz (Ed.), *The interpretation of cultures* (pp. 3–30). New York: Basic Books.

Glaser, B. G., & Strauss, A. L. (1967). *Discovery of grounded theory: Strategies for qualitative research*. Chicago: Aldine.

Goodman, R. M. (2000). Bridging the gap in effective program implementation: From concept to application. *Journal of Community Psychology*, *28*, 309–321.

Hall, G. E., & Hord, S. M. (1987). *Change in schools: Facilitating the process*. Albany, NY: State University of New York Press.

Hall, G. E., & Loucks, S. F. (1978). *Innovation configurations: Analyzing the adaptation of innovations*. Paper presented at the annual meeting of the American Educational Research Association, Toronto, Ont., Canada.

Harachi, T. W., Abbott, R. D., Catalano, R. F., Haggerty, K. P., & Fleming, C. B. (1999). Opening the black box: Using process evaluation measures to assess implementation and theory building. *American Journal of community Psychology*, *27*, 711–731.

Heck, S., Steigelbaur, S., Hall, G. E., & Loucks, S. F. (1981). *Measuring innovation configurations: Procedures and applications*. Austin: Research and Development Center for Teacher Education, University of Texas.

Hoge, R. D., & Andrews, D. A. (1987). Enhancing academic performance: Issues in target selection. *School Psychology Review*, *16*, 228–238.

Hoge, R. D., & Coladarci, T. (1989). Teacher-based judgments of academic achievement: A review of the literature. *Review of Educational Research*, *59*, 297–313.

Illback, R. J. (1994). Poverty and the crisis in children's services: The need for services integration. *Journal of Clinical Child Psychology*, *23*, 413–424.

Illback, R. J. (1998). *Bridges over barriers: Kentucky's family resource and youth services centers*. Louisville, KY: R.E.A.C.H. of Louisville, Inc.

Kalafat, J. (2003). Enabling and empowering practices of Kentucky's school-based family resource centers: A multiple case study. *Evaluation and Program Planning*, *1*, 65–78.

Kalafat, J., & Illback, R. J. (1998). A qualitative evaluation of school-based family resource and youth service centers. *American Journal of Community Psychology*, *26*, 573–604.

Kagan, S. L., & Weissbourd, B. (Eds.). (1994). *Putting families first*. San Francisco, CA: Jossey-Bass.

Kamphaus, R. W., Yarbrough, N. D., & Johanson, R. P. (1990). Contributions of instructional psychology to school psychology. In T. B. Gutkin, & C. R. Reynolds (Eds.), *Handbook of School Psychology* (pp. 143–174). New York: Wiley.

Knapp, M. S. (1995). How shall we study comprehensive, collaborative services for children and families? *Educational Researcher*, *24*, 5–16.

Knitzer, J. (1997). Service integration for children and families: Lessons and questions. In R. J. Illback, C. T. Cobb, & H. M. Joseph (Eds.), *Integrated services for children and families* (pp. 3–22). Washington, DC: American Psychological Association.

Lipps, G., & Grant, P. R. (1990). A participatory method of assessing program implementation. *Evaluation Review*, *14*, 427–434.

McGrew, J. H., Bond, G. R., Dietzen, L., & Salyers, M. (1994). Measuring the fidelity of implementation of a mental health program model. *Journal of Consulting and Clinical Psychology*, *62*, 670–678.

Moncher, F. J., & Prinz, R. J. (1991). Treatment fidelity in outcome studies. *Clinical Psychology Review*, *11*, 247–266.

Patton, M. Q. (2001). *Qualitative evaluation and research methods*. Thousand Oaks, CA: Sage.

Powell, D. R. (1994). Evaluating family support programs: Are we making progress? In S. L. Kagan, & B. Weissbourd (Eds.), *Putting families first* (pp. 441–470). San Francisco, CA: Jossey-Bass.

Pratt, H., Winters, S., & George, A. (1980). *Effects of a concerns-based implementation plan on the achievement of elementary science students.* Paper presented at the annual meeting of the American educational Research Association, Boston.

Reid, E. R. (1980). *Exemplary Center for Reading Instruction (ECRI). Educational programs that work*. San Francisco: Far West Laboratory for Educational Research and Development.

Roitman, D., Gotschalk, R., Mayer, J., & Blakely, C. (1983). Implementation of social program innovations in public sector organizations: A test of the modified RD and D model. *IEEE Transactions on Engineering Management*, *30*, 68–75.

Scheirer, M. A. (1994). Designing and using process evaluation. In J. S. Wholey, H. P. Hatry, & K. E. Newcomer (Eds.), *Handbook of program evaluation* (pp. 40–68). San Francisco: Jossey-Bass.

Schorr, L. B. (1997). *Common purpose: Strengthening families and neighborhoods to rebuild America*. New York: Doubleday.

Trivette, C. M., Dunst, C. J., & Hamby, D. H. (1995). Family-oriented program models of help giving practices, and parental control appraisals. *Exceptional Children*, *62*, 237–248.

Weiss, H. B., & Jacobs, F. (1988). *Evaluating family programs*. Hawthorne, NY: Aldine.

Weissberg, R. P., Kumpfer, K. L., & Seligman, M. E. P. (2003). Prevention that works for children and youth: An introduction. *American Psychologist*, *58*, 425–432.