

Community-Based Interventions and Services

Christina J. Groark, Ph.D. and Robert B. McCall, Ph.D.
University of Pittsburgh Office of Child Development

The purpose of this chapter is to summarize the process by which one can conceive, design, implement, monitor, evaluate, and sustain empirically-based community-operated interventions and services. The focus is on the *process*, not the content, of such interventions and services. As such, the discussion accepts (and will not cover) the existing literature on the problem or target behaviors as well as on approaches to preventing or treating those behaviors as the starting point of the process.

There is relatively little scholarly analysis and even less empirical validation of this process per se. Consequently, this chapter relies to a greater extent than is typical on the experience of the authors blended with those of others and what scholarly work may exist.

Scope and Definition

The scope of this chapter can be further specified by considering the working definitions of certain key concepts.

Community-based. Community-based interventions and services are those that are implemented and operated by service professionals in the community. Often, but not always, such services are part of a demonstration and evaluation project investigating whether the intervention or service can be created, organized, implemented, and produce benefits in participants. If successful, the same type of professionals will provide the service in the community on a routine basis. In contrast, innovative or experimental services conducted in a specialized environment (e.g., laboratory school) by specialized practitioners (e.g., academic educators or social workers) do not fall in the chapter's purview. Thus, community-based services represent the most ecologically valid version of the service and the last step in the basic-to-applied research continuum.

Interventions and services. The services to be considered consist of organized supports and activities designed to promote the health, education, and welfare of children, youth, and families. They focus on target behaviors that tend to be public rather than strictly intrapsychic. As such, they consist of problem behaviors that come to public attention and that require public resources to prevent or treat, such as abuse and neglect; drug and alcohol problems; antisocial behavior and delinquency; poverty, unemployment, and public assistance; poor parenting; and mental health prevention, detection, diagnosis, and referral. While participants may receive services individually or in groups, the services of concern do not include traditional private or group psychotherapy (see APA Presidential Task Force on Evidence-Based Practice, 2006; Weisz, Jensen-Doss, & Hawley, 2006). Instead, they are closer to services within the domains of community psychology and psychiatry, social work, public health, and education.

Prevention and treatment/remediation. The services include both prevention and treatment/remediation, although the distinction between these categories often blurs. For example, individuals come to public attention and are referred to services because they have already displayed certain problematic behaviors or are substantially at risk of doing so; and the services focus on treating/remediating the manifest behaviors, preventing additional or more extreme undesirable behaviors, and promoting more positive behavioral alternatives.

Evidence based. Services that are evidence-based have some research supporting their potential or actual effectiveness. While some people reserve the phrase “evidence-based program” for a well articulated service program that has been empirically shown to be effective and could be replicated, we will also use the term to refer to new programs based on principles or consisting of components that have theory and evidence that support their likely effectiveness. Unfortunately, no clear standard prescribes how much or what kind of evidence is required to substantiate the claim of “evidence-based” (see below) so that use of the phrase has become commonplace and less meaningful in an era in which policy makers and practitioners value and require evidence-based practices (Groark & McCall, 2005; McCall, 2007).

Universal, quasi-universal, and targeted services. Strictly speaking, *universal* services are available to everyone. For example, public education and some form of social security or public retirement benefits are universal in many countries. Universal programs tend to be most appropriate when essentially everyone in society can benefit (e.g., public education, retirement benefits) or when the target group is broadly dispersed in the population (e.g., smokers, automobile drivers, obese children) or not readily identifiable and/or without clear, predictive risk factors (e.g., abusive parents). They often have the advantages of reaching large numbers of people, some who otherwise would not be identifiable, and satisfying broad, crucial social needs (e.g., education). But only a few social needs merit the cost of very intense services (e.g., education); some are inefficient because many people are reached or served who do not need the service; and some programs benefit the educationally, financially, and behaviorally advantaged members of society more than those of relative disadvantage, because such individuals are more likely to take advantage of the program and already have a head start on achieving its goals (Ceci & Papierno, 2005). Truly universal programs are not within this chapter’s focus.

Instead, we will emphasize “*quasi-universal*” programs, which are ones that are available to all individuals but within a geographic area (e.g., neighborhood, town) or institution (e.g., an entire school, orphanage), and such programs may or may not have additional eligibility criteria, such as income level (e.g., Head Start early childhood care and education programs in the USA). These programs tend to have broad goals (e.g., improvement of family functioning, parental psychological and economic self-sufficiency; improvement of child development, school readiness, and life success), but some have more specific aims (e.g., school readiness).

Conversely, *targeted* programs (discussed in Chapter 62 by Vitaro and Tremblay) are often restricted to participants having certain specific characteristics that may exist in relatively small percentages of the population (e.g., drug addiction, blindness, physical disabilities) and may require intense (e.g., detoxification) or extensive services (e.g., for children with severe disabilities).

Quasi-universal programs tend to be most appropriate for conditions and services in between those for truly universal and highly targeted programs, for example, when the target group is concentrated in low-income neighborhoods and when the needed service is moderately intense or extensive and clearly needs to be targeted to some extent to be cost-efficient and to benefit those most in need (e.g., Sure Start in Great Britain, Head Start and offered to all in low-income neighborhoods in the USA). They have the advantage of having a large percentage of participants who need the service, but they may not serve all in society who need it (e.g., low-income families who live in unserved neighborhoods).

Examples of Different Quasi-Universal Programs

We describe below several examples of quasi-universal programs that illustrate some of their commonalities and differences, many of which will be discussed more fully later.

Nurse-Family Partnership (formerly Nurse Home Visitation Program). The Nurse-Family Partnership (NFP) program (Olds et al., 1999) is aimed at first-time, low-income, higher-risk mothers to improve pregnancy outcomes by reducing health-related adverse behaviors (e.g., smoking, alcohol consumption, drug use); improve child health and development by teaching parents to provide more responsible and competent health and behavioral care; and promote the economic self-sufficiency of families by encouraging planful pregnancies (i.e., fewer and more widely spaced) and supporting education, secure employment, and linkages with other health and human services within the community. The program is being implemented as a service in nearly half the USA states and some other countries. It is quasi-universal for high-risk, first-birth women located in the geographic region in which it is offered.

The program relies on nurses to visit the homes of participants, preferably during pregnancy and throughout the child's first two years. The program is heavily centrally prescribed, with specific and concrete goals and activities, although some tailoring to the needs of individual families must occur. The focus on small, concrete, specific goals from visit to visit and throughout the program likely contributes to the ease of implementing the program and positive outcomes.

The program was evaluated in three sequential studies in three USA cities (Olds et al., 1999). Pregnant women who had not previously given birth and who had at least one additional risk factor (e.g., low-income, unmarried, low education, unemployment) were randomly assigned to a treatment group (nurse home visitors, transportation to periodic health visits, and home visits through the child's second birthday) versus a variety of control conditions that included fewer or shorter services. Samples ranged from 400 to 1139 with a predominance of either white, African-American, or Hispanic women in different cities. Results showed benefits to mothers and children, but these were not pervasive and tended to occur under specific experimental and non-randomized circumstances—namely, if nurses rather than paraprofessionals did the home visiting, if home visiting lasted through the child's second birthday, and especially if the mother was at higher risk (e.g., unmarried, smoker, very low income/education). For these groups, NFP mothers were more likely to be employed and have better life skills, fewer problem behaviors, better reproductive histories, improved parenting, and less abuse; their children improved in health, social/emotional, language, and cognitive development as children and with reduced

criminal/behavioral problems, sexual activity, and substance use as teenagers. A similar but more intensive and extensive home visiting program in New Zealand produced similar (but not identical) results (Fergusson, Grant, Horwood, & Ridder, 2005, 2006).

Early Head Start (EHS). Early Head Start (EHS), is an attempt by the US government to provide health, child care, family support, and parent education to primarily low-income families. EHS is quasi-universal in that its services are available to all low-income families with at least one child approximately birth – 3 years of age in the geographical area in which the program is located. A small percentage of families who are not low income and others who have children with disabilities are also enrolled.

The program is a combination of a centralized set of program requirements and standards for implementation with some local flexibility to create a program that matches the needs of local participants. For example, child development services are required, but they can be provided in centers, in the homes of participants through home visiting, or in a mixture of the two.

Seventeen programs were selected from applicants to provide EHS services and to participate in a research and evaluation consortium. The total evaluation followed 3,001 families who were randomly assigned to EHS or whatever other program parents chose. EHS was unusual in having local evaluation specialists at each site supervise data collection consisting of core assessments common across the 17 sites plus additional assessments to answer specific research and evaluation questions pertinent to that site. The Consortium met periodically to advise the cross-site evaluation team and participate in subsets of sites that had similar programs and similar research interests (e.g., the role of fathers).

When EHS children were 3 years old, they performed better on measures of cognition, language, and social-emotional functioning than control peers, and EHS parents were more supportive of children's emotional, cognitive, and language development and more likely to be in education or job training. However, the benefits of EHS varied with non-randomized characteristics of participants and implementation. For example, benefits were greater for African-American families, families who enrolled during pregnancy versus after delivery, and families with a moderate rather than a substantial number of demographic risk factors. Further, programs with a mixed center and home visiting approach and those that more fully implemented the comprehensive performance standards had a wider range of larger benefits. Also, children who transitioned from EHS to formal early care and education programs between ages 3 and 5 had better early reading-related skills at age 5 (Administration for Children and Families, 2002; Love, Kisker, Ross, Raikes, Constantine, Boller, et al., 2005; http://www.acf.hhs.gov/programs/OPRE/EHS/EHS_resrch/index/html).

Comprehensive Child Development Program (CCDP). The Comprehensive Child Development Program (CCDP) was a multi-sited demonstration project in 34 sites around the USA to provide comprehensive health, education, mental health, and welfare services to low-income families with children birth to 1 year of age at the beginning until the child reached approximately age 5. Its purpose was to increase the financial and psychological self-sufficiency of those families and improve the children's development. It was quasi-universal, allowing all community members to participate, but sites were targeted to low-income areas.

The program was a combination of central prescription and local innovation. Services were required to be comprehensive, because high-risk families tend to have a variety of different needs, including obtaining basic necessities (shelter, food, medical care), preparation for employment (education, training), mental health services (e.g., drug and alcohol), child care and parenting education, etc. (for further information about parent based programmes in general see Chapter 66) While specific services were to be tailored to the needs of each family and thus could vary between families and sites, certain procedures, including the frequency of contact, devising individual family plans, etc. were specified by the federal government.

A central evaluator conducted an evaluation on 21 sites, which had randomly assigned either individuals or site locations within their area. An “intent to treat” strategy was employed. However, the government had awarded separate contracts for the central evaluator and the management information system developer, and these two databases were never merged across sites so that the nature and extent of services as well as specific family goals could not be related to outcomes. Results indicated that twice as many mothers in CCDP treatment sites became employed and improved their financial status as comparison mothers, but there were few other benefits for parents or children (St. Pierre, Layzar, Goodson, & Bernstein, 1997a, 1997b). However, sites varied substantially in the nature and extent of the services they provided, families in the comparison group obtained on their own almost as many services as those in the CCDP groups, families differed in their individual goals and the services they used but the intent to treat strategy assessed each outcome on all participants, regardless of their individual goals, nature and extent of services, or length of participation in the program (Gilliam, Ripple, Zigler, & Leiter, 2000; McCall, Ryan, & Plemons, 2003).

School Development Program (SDP). James Comer’s School Development Program (SDP) seeks to improve schools, primarily in low-income neighborhoods, and student academic performance and skills by mobilizing adults to support student learning and development (e. g., Comer, 1988; Comer, Haynes, Joyner, and Ben-Avie, 1996; Joyner, Comer, & Ben-Avie, 2004). It is quasi-universal because an entire school—administrators, teachers, students, parents, and local organizations—and increasingly an entire school system may be involved. SDP is operating in hundreds of schools in dozens of USA states and other countries.

The program is planned and implemented within each school and/or school system, rather than by some central body. The program consists of a structure and process, rather than a set of specified actions, services, or directions. Three structures comprise the basic system on which the Comer process is built. The *School Planning and Management Team* is composed of administrators, teachers, support staff, and parents. It is responsible for developing a comprehensive school plan; setting academic, social, and community relations goals; coordinating all school activities, including staff development programs; and monitoring the progress to identify needed changes. The *Student and Staff Support Team* consists of the principal and professionals in child development and mental health, such as school counselors, social workers, psychologists, and nurses. It promotes the social conditions and relationships necessary to connect all of the school’s student services; it facilitates the sharing of information and advice; and it addresses individual student needs, obtains resources outside the school, and develops prevention programs. The *Parent Team* is composed of parents. It selects

representatives to serve on the School Planning and Management Team and develops activities for the parents to support the school's social and academic programs. Central to the process are three school operations including development of a comprehensive school plan, provision of staff development opportunities, and assessment and modification.

Evaluations conducted by SDP and independent professionals have shown the program can improve school climate, student behavior, and student achievement (e.g., Cook, Hunt, & Murphy, 1998; Haynes, 1995), but results have not been uniformly positive (e.g., Cook et al., 2000; Neufield & LaBue, 1994). Success seems to depend on how extensively and vigorously the process is pursued and implemented. This result is concordant with research conducted years earlier (Maughan, Pickles, Rutter, & Ouston, 1990; Ouston, Maughan, & Rutter, 1991) revealing that non-experimental changes that produced better attendance and school performance in students rested heavily on the extent there was a clear vision of and process to guide change and a focus by all of its administrators, staff, students, and parents on improving the entire school—precisely what SDP intends to guide.

Sure Start Local Programmes (SSLPs). Sure Start Local Programmes (SSLPs) is an attempt by the United Kingdom to reduce child poverty and social exclusion (the gap between rich and poor) with a quasi-universal program within certain geographic areas that are relatively disadvantaged (Rutter, 2006). It is aimed at infants or young children. Support for the program is widely available, and numerous sites have been continued, modified, or created across the UK.

Programs and services were largely locally created and unspecified. The central government required five core service themes: 1) Outreach and home visiting; 2) support for families and parents; 3) high-quality play, learning, and child care; 4) primary and community health care including advice about child and family health; and 5) support for children and parents with special needs. Services within these domains were to be “evidence-based” but their nature and extent were up to local agencies. It was expected that existing services would be improved and coordinated, and that facilities for early childhood care and education would be expanded, remodeled, and otherwise improved. The government's rationale was that a minimally targeted program open to all in the area would minimize any stigma associated with using the program, and localities should have the opportunity to tailor services within the five areas specifically to fit the needs of their local participants and existing services.

An extensive national evaluation (National Evaluation of Sure Start Team, 2005a, 2005b) was conducted, but with at least two constraints imposed by government—first, neither families nor sites could be randomly assigned to Sure Start or a comparison condition because Sure Start was expected to be successful and random assignment would unnecessarily deny services to certain families, and second, that the services would not be “manualized,” which meant that no site was required to specify, and no monitoring information would be gathered on, the nature and extent of services provided. These constraints meant that there was no uniform Sure Start program, sites were extremely heterogeneous in the nature and extent of services provided, it was nearly impossible to relate any characteristics of the nature and extent of services to outcomes, and the opportunities for a comparison group were limited.

Nevertheless, the national evaluation was able to study 150 SSLP sites in which the program had been initiated for at least three years and 50 comparison sites that had not yet begun to implement SSLP. Extensive data were collected on participants, administrative and broad characteristics of the SSLP services, and parental report, observations, and developmental assessments on a variety of family and child (between 9 months and 36 months old) outcomes. The evaluation used an “intent to treat” strategy (see below) in which all families eligible within an SSLP area with 9- or 36-mos.-old children were included in the sample regardless of the nature and extent of services received.

Given the large number of possible outcomes for parents and children, SSLP was associated with consistent benefits for very few. For example, SSLP sites operated by health agencies and other local authority agencies did better than those led by voluntary agencies. In contrast to the NFP, there were some adverse effects the more disadvantaged the family (e.g., mothers who were teenagers, single parents, unemployed parents), including slightly lower scores of children on verbal ability, social competence, and more behavioral problems; similar to EHS, there were slight benefits for families who were not so disadvantaged, including less behavioral problems, more social competence, and less negative parenting (Belsky et al., 2006; National Evaluation of Sure Start Team, 2005b).

Conclusion. While not intended to be representative of quasi-universal programs, these projects varied in how specific and concrete the program was prescribed centrally versus the amount of local flexibility allowed and in how the programs were implemented and evaluated. The remainder of this chapter focuses on these aspects of program development, implementation, and evaluation.

Plan of the Chapter

The next section of the chapter considers various issues in bringing evidence to the creation of evidence-based programs in the community. Then, principles of collaboration are offered, because most community-based programs are or should be developed, implemented, and evaluated by a collaborative team whose members are often unaccustomed to working with each other and whose values and roles may conflict. We next consider program design, which may occur at the national, state/province, or local level and be very specific and detailed or consist of broad principles or service domains. Then we consider how such programs may be implemented, followed by a discussion of certain issues in the monitoring and evaluation of such programs. Finally, we discuss sustainability, which includes maintaining program effectiveness year after year as well sustaining funding.

Evidence and Evidence-Based Interventions and Services

Policy makers and funders in many countries are demanding that the services they initiate or support be “evidence-based,” that is, be rooted in empirically supported principles and preferably have already been demonstrated to produce benefits for participants (see Chapter 60 for an overview of treatment evaluation) . The simplest and presumably most expeditious and effective strategy for creating and implementing service programs is to replicate “proven programs.” Indeed for some policy makers, this is the only definition of an “evidence-based program.”

Requirements for Services Replication

While the rationale for service replication is straightforward, it makes certain assumptions that must be met for this strategy to be effective, and such premises are often not fulfilled (Groark & McCall, 2005; McCall, 2007; McCall, Groark, & Nelkin, 2004).

One or more programs must have been demonstrated to have been effective. Obviously, there must be a service program that has been implemented, evaluated, and found to be effective to be a candidate for replication. Such demonstrated programs exist for teenage problem behaviors including school failure, risky sexual activity, substance abuse, delinquency and violence (Weissberg & Kumpfer, 2003) and for lowering rates of abuse and improving parenting skills for high-risk mothers (Olds & Kitzman, 1993), but this is not the case for many if not most areas needing services. For example, nearly 50 years of research on early childhood education indicates such programs are beneficial, but much less research exists on specific curricula that might be replicated.

The service must be packaged in detail. Relatively few service programs, even demonstration programs intended to be replicated if found successful, have a “manual” describing all of the elements of the program and its implementation that practice professionals elsewhere can use to replicate the program.

Local service providers must be willing to replicate the original service. A major tenant of good therapeutic practice is to match the treatment to the specific characteristics of the client, so it is not surprising that agencies often want to adapt a model program to fit their own specific type of clientele, local social and policy circumstances, and each individual client. Indeed, a major principle of family support programs, for example, is that each individual family specifies their strengths and their needs and a program of referrals, services, and supports is tailored to their individual profile. But, even for services that are comprehensively and specifically prescribed, the service program may differ substantially from one incarnation to another (Gilliam, Ripple, Zigler, & Leiter, 2000; McCall & B. L. Green, 2004), and often no research exists suggesting whether such modifications improve, harm, or have no influence on program effectiveness.

Will a replicated program also replicate the benefits of the original demonstration? Even if the program is faithfully replicated to the extent possible, it is not clear that it will have the same benefits for participants in its replicated incarnation as it did in its original demonstration (L.W. Green, 2001). Demonstration projects are typically implemented by professionals who are highly energized, passionate, and committed to the new program and who may have substantial expertise in creating relationships with clients and motivating them to participate and persist in the service regimen. The new practitioners who volunteer or are assigned to replicate the program, however, may not possess the same level of motivation, commitment, and interpersonal expertise. Also, the intended participants may be different than in the original demonstration, and some services are not necessarily equally effective with all types of participants. For example, physical punishment is often discouraged in parenting improvement programs; but physical punishment is very common among African-American families, and

research indicates that it is not as deleterious to African-American as it is to middle- and upper-class Caucasian children (Baumrind, Larzelere, & Cowan, 2002).

For these reasons, replicating proven programs may not produce replicated benefits; often the original demonstration is more effective than its replicates, which is why policy makers and funders should evaluate replications and not assume that the program already has been shown to be effective and its replication does not need evaluation.

The Nature of the Evidence

From a practical standpoint, a service program is never “proven;” rather, the evidence for its effectiveness will always be more or less “persuasive.” Further, not all research studies and demonstrations are equally persuasive; some designs and results are more compelling than others. Finally, from the perspective of a policy maker or practitioner faced with a mounting behavioral or social problem, one picks the “best available service approach” almost regardless of the persuasiveness of evidence.

Assessing the evidence. It would be ideal to have a simple scheme to assess the research literature on a service program, and such a scheme has been proposed (Chambless & Hollon, 1998; Mrazek & Haggerty, 1994) that literally provides a “grade” from 1 to 7 (1 is highest) for the persuasiveness of the evidence. But this particular approach seems too narrow, considers only a few types of research designs (e.g., randomized trials and replications), emphasizes internal over external validity, and ignores a great deal of information (e.g., effect size, cost/benefit ratio) that would be valuable in making policy and service program decisions (Groark & McCall, 2005; McCall & B. L. Green, 2004; McCall et al., 2004).

Table 1 presents an alternative scheme (McCall, 2007) based on a broader representation of the standards of program evaluation research methodology. It has the advantages of being substantially more comprehensive, but it has the apparent disadvantages of being much more complex and not providing a simple grade for the research literature as a whole (although such a numerical rating scheme could be developed). In practice, the disadvantages may not be important, because research literatures are typically uneven and contain gaps, and they are best evaluated by academic research specialists who are capable of weighing different kinds of evidence and making judgments regarding the persuasiveness of the corpus of research.

Some groups, such as the Campbell Collaboration (C2) (Cooper, 1998) and the US Centers for Disease Control (2006) provide reviews of evidence on services for a variety of physical and behavioral health programs, and governments and foundations often operate consensus groups that bring together academic authorities to review literature and make specific recommendations regarding treatment approaches and services. The scheme in Table 1 makes more explicit the types of evidence and to some extent the standards that might be used by such groups to evaluate the literature.

Insert Table 1 about here

When the evidence is not sufficient. Except for the most thoroughly researched service programs (e.g., Nurse-Family Partnership), the research literature is not likely to cover all of the elements necessary to decide what kind of service will be most effective and how it should be implemented and operated. There may be gaps, uncertainties, and technical limitations; no specific program may have a persuasive research track record; and doubts may emerge about how a program will need to be modified to fit the local clientele, service staff, and circumstances.

An alternative to prescribing a specific service program is to provide *guidelines* for service creation and operation that are supported by research and that permit creativity and flexibility in matching the service to local circumstances. Sure Start was extreme in this regard, requiring only domains of services and the admonition that they be evidence-based. Fortunately, the general characteristics of successful service programs are remarkably similar across services aimed at preventing or treating a variety of different problems of children, youth, and families. For example, Table 2 (Groark & McCall, 2005) presents an integration across several lists of characteristics of successful programs in early childhood care and education (McCall, Larsen, & Ingram, 2003), family support (Layzar, Goodson, Bernstein, & Price, 2001; Schorr, 2003), and the prevention of adolescent problem behavior (Kumpfer & Alvarado, 2003; Nation, Crusto, Wandersman, Kumpfer, Seybolt, Morrissey-Kane, & Davino, 2003). This list represents a beginning set of general guidelines that could be supplemented with additional characteristics that are unique to one versus another specific type of service and problem area.

Insert Table 2 about here

While such characteristics are useful as guidelines to local service professionals and funders, the characteristics themselves are rarely researched directly to assess their role in contributing to service effectiveness (but see Olds et al., 1999). Instead, they tend to be common characteristics of programs that have been demonstrated to successfully produce benefits in participants.

Nevertheless, coupled with other characteristics supported by research and best practices of service professionals, such guidelines can communicate what research and professional practice have learned and permit modifications to fit local needs.

Consensus groups. To produce such guidelines, consensus groups need to be composed not only of relevant scholars to assess the available research literature but also service professionals, policy makers, and representatives of the intended clientele, and they to be able to fill in the research literature's gaps and come out with an action plan based upon the best available evidence and practice. The Pathways Mapping Initiative (PMI) developed by Lizbeth B. Schorr and colleagues (Schorr, 2003) represents a structured process by which such a group can create a "map" or set of recommended guidelines for the processes and characteristics that are effective in reaching the outcome under consideration. PMI recognizes that the traditional knowledge regarding a service program or a type of service typically is too limited, comes from a small number of interventions that have been adequately evaluated, and usually fails to identify all of the elements that actually make the service effective. The PMI process broadens the knowledge base by permitting reasonable judgments and plausible interpretations of a preponderance of evidence based on professional experience as well as program evaluations coupled with strong

theory. The process also attempts to curb idiosyncratic opinion by insisting upon broad consensus among its diverse members. The product consists of a combination of actions needed to produce the desired outcome, the key ingredients that are likely to make those actions effective, and the community contexts that will influence program effectiveness (for examples, see www.PathwaysToOutcomes.org).

Proven program or principles? In practice, the best solution may be somewhere between replicating a well specified proven program (e.g., Nurse-Family Partnership) and a set of broad service domains (e.g., Sure Start), a structure and process (e.g., School Development Program), or principles of successful programs (e.g., Table 2). True replication of a highly specified program is likely a myth, Friesen & Koroloff, 1990; Green, Rodgers, & Johnson, 1999; McCall & Green, 2004—it does not really take place; but vague or unspecified characteristics, such as in Sure Start, permit too much local variation and unsubstantiated program characteristics. A compromise of as much specification as research substantiates coupled with reasonable local flexibility perhaps is most desirable (and is likely to occur anyway).

Limits on the Transfer of Research and Best Practices to Communities

While a set of guidelines reflecting the research literature and professional best practices may represent a necessary first step in creating and implementing effective service programs at a local level, it is not likely to be sufficient. First, simply communicating this information has not resulted in localities adopting the most effective program strategies (Ringwalt, Ennett, Vincus, Thorne, Rohrbach, & Simons-Rudolph, 2002; Wandersman & Florin, 2003). For example, until recently, the DARE program to prevent drug abuse had been adopted in 80% of the school districts in the United States despite a research literature that showed it is relatively ineffective (e.g., Ennett, Tobler, Ringwalt, & Fewling, 1994; General Accounting Office, 2003).

The research-to-practice gap. This gap may occur for several reasons: 1) The research knowledge is generated and communicated unidirectionally from researchers to practitioners and thus may be less relevant to the policy and practice context than is needed; 2) The research and practice knowledge may be inadequately and ineffectively communicated to local practitioners and policy makers; and 3) factors other than research and best practice contribute to service selection and effectiveness, such as the availability of a packaged program, ease of implementation, unique characteristics of local clientele, and local personnel and financial resources.

For example, Chinman, Hannah, Wandersman, Ebener, Hunter, and Imm (2005) identified four broad factors that influence whether communities adopt, create, and implement evidence-based services. First, implementing high-quality service programs is a complex process that requires considerably more knowledge and skill than is needed to simply follow a program manual. Second, systems factors pertaining to coordination among different agencies and community readiness to adopt and maintain new strategies need to be considered. Third, communities must have sufficient financial, technical, and personnel resources; and fourth, local clienteles and other circumstances may pose unique difficulties.

Technology transfer. One approach to building such community capacity is technology transfer (Backer, David, & Soucy, 1995), which emphasizes training and technical assistance. But these strategies alone have been only partly successful. Chinman et al. (2005) reviewed a variety of training programs in substance abuse prevention, for example, and concluded that while such programs were helpful, their content was not always appropriate to the specific local context and there were often local barriers to incorporating the information into practice. Alternatively, having an intermediate set of professionals provide direct, hands-on technical assistance similarly had limited effectiveness (Chinman et al., 2005). Such technical assistance, even when provided without cost, was not always welcomed at the local level, some minimal level of community capacity was required to make full use of such assistance, and community organizations seemed better able to utilize some forms of assistance (e.g., planning, implementation, organizational maintenance) than others (e.g., evaluation procedures, data analysis). While these results are based on experience with training and technical assistance in one area, namely substance abuse prevention, and their generality to other areas is unknown, many funders and service professionals believe that training and technical assistance as typically provided may be necessary but are not often sufficient to improve or create service programs.

Toward Effective Community Processes

The ability of local communities to adopt, create, and implement effective services often requires a collaborative community process (L. W. Green, 2001), not simply technology transfer or a ready-made program to be replicated. Chinman et al. (2005) suggest that this process must have genuine community involvement and commitment; the community must possess skills in a variety of domains; resources must be identified, acquired, and managed; and there must be a collective sense of community efficacy or power to manage the skills and resources and direct them toward successful outcomes.

More specifically, such a community process requires 1) collaboration among a variety of stakeholders, 2) a strategy for designing services that fit local needs and circumstances, 3) effective implementation and operational strategies, 4) appropriate monitoring and evaluation, and 5) a plan to sustain the effectiveness and financing of services. We consider each of these elements below.

Collaborations

While a specific service may be adopted or created by a single agency, many contemporary services are multisited and/or comprehensive, requiring collaboration across agencies and involving policy makers, funders, academics, evaluators, the media, and members of the intended clientele in their planning, implementation, and operation. Such diverse collaborations have the potential benefit of converging the complementary expertise of such individuals to produce a better and more comprehensive service and to create involvement (i.e., “buy in”) of major participants and groups that are necessary for the successful funding, roll-out, operation, and sustainability of the service. Conversely, collaborations may take more time, require skilled leadership, and sometimes involve individuals who obstruct progress for a variety of personal and professional reasons.

Membership. The members of a collaboration are typically *stakeholders* in the service to be created, that is, individuals who have an interest in the project; something to gain or lose as a function of its success or failure; and can contribute in some manner to its creation, implementation, operation, and sustainability. Ideally, every member of the collaboration should be necessary to its success, and no smaller subset of the collaborators would be sufficient to do the project as well or at all. Even collaborators whose contribution will occur late in the process should nevertheless be present from the beginning so they are knowledgeable about and involved in the project.

More specifically, members of a community planning and implementation collaboration might include:

- Relevant service professionals representing key agencies to be involved in the future service;
- Policy makers and funders who have administrative responsibility for the type of proposed service, who will or could help to fund it, and who will be necessary to overcome local political barriers the service may face;
- Academics who may have the research and professional knowledge behind the guidelines that can be used to steer the creation of the service;
- Evaluators who will work with service providers to monitor and evaluate the implementation, service delivery process, and outcomes for participants;
- Potential participants in the service who can provide the perspective of the client in designing a user-friendly service; and
- Others who may be necessary to succeed (e.g., the media if public awareness will be needed, celebrities if a public spokesperson will be helpful, business leaders if the service is to be integrated with or benefit private enterprise, and community leaders if the service will influence the wider community or will need community support to operate).

Members should be selected for their ability to fulfill their specific roles in the project and for their personal ability to function in a collaborative group. Collaborations often leave individual members with less control over the process and product than they are accustomed to exercising and require them to make compromises with cherished values and principles (e.g., evaluators not having random assignment). They should be committed to attend every meeting, represent their organization, bring its resources to the project, listen and understand divergent points of view, communicate their perspective honestly and clearly, and accept a group decision even if it conflicts with their own self-interest.

Models of university-community relations. In the past, academics and program evaluators, generally from a local college or university, often have not been part of this process; but in the era of evidence-based practices and increased emphasis on evaluation, they have become more crucial to the planning and implementation process. While each member of the collaboration is likely to have his or her own professional values and attitudes, those of academics and program evaluators may be less similar to, even at odds with, the values and attitudes of other collaborators (Groark & McCall, 1993, 1996, 2005; Mattessich & Monsey, 1992; McCall, B. L. Green, Strauss, & Groark, 1997; Shonkoff, 2000).

Historically, academics have worked with community professionals in different ways (Denner, Cooper, Lopez, & Dunbar, 1999). For example, often academics simply serve as *expert consultants*. In this case, academics provide their knowledge in a rather unidirectional fashion. Sometimes the *community sets the agenda*. In this case, the community tends to pick the problem, the methods, and the services, selectively taking what information the academic provides that is useful to its purposes. The result is highly relevant to the community and thereby ecologically valid, but it tends to marginalize the role of evidence, monitoring, and evaluation making it less scientifically valid and difficult to replicate.

A better approach is the *true collaboration among stakeholders*, in which university and community members develop the project together, blending evidence and program evaluation with practitioner expertise and the resources and limits of the community. The interventions are developed in the community context with the community using its own resources, but research and evaluation priorities are selected collaboratively. In addition, consideration is given to broader policies and processes so that findings are relevant for general theory development. This model also requires the inclusion of various disciplines and diverse groups, thereby being more relevant to the ordinary lives of a wide variety of clientele.

One example of this collaborative approach is practiced by the University of Pittsburgh Office of Child Development (OCD) (e.g., Groark & McCall, 1993, 1996, 2005; McCall et al., 1997, 1999, 2004). OCD employs a) academic specialists in various aspects of child development who can bring research knowledge to the process; b) service professionals who are experienced in creating, implementing, and managing innovative service demonstrations in collaboration with community agencies; c) program evaluation specialists who train service providers in monitoring and evaluation and who conduct evaluations on community-based programs in a collaborative and participatory manner with the service agency; and d) specialists experienced in policy and governmental service administration who work with policy makers to create better policies and administrative services for children and families. Some or all of these types of OCD staff may work on a specific project in collaboration with other academics, community service agencies and professionals, and administrators and policy makers and each component of this team influences the others. The essential ingredient, however, is an attitude that emphasizes partnership—sharing of responsibility, power, control, and credit among all stakeholders who collaborate to achieve a common goal.

Characteristics of successful collaborations. Successful collaborations tend to have certain characteristics in common (Butterfoss, Goodman, & Wandersman, 1996; Grobe, Curnan, Melchior, & The Center for Human Resources, 1993; Kegerise, 1999; Mattessich & Monsey, 1992; Wandersman & Goodman, 1993). First, they have a *common purpose*. Each participant must share the value for the common purpose, which focuses each participant on a single set of criteria, and the mutual dependency helps to keep the coalition together and produce mutual respect. The common purpose must be articulated early and clearly so that it becomes the touchstone that helps members focus on the reason for the collaboration during debates. Therefore, it is helpful to write down the common purpose in the manner of a vision or mission statement that can be reviewed periodically to keep members operating in the same direction and also to constitute criteria for deciding whether each major decision contributes to accomplishing the stated purpose.

Further, the collaboration should establish *common, clear, achievable, and specific goals; a plan with concrete and realistic steps; and a timetable for achieving them*. This plan becomes the roadmap for all future activities, and a monitoring method that helps institute controls on priorities for each partner by specifying who is responsible for which tasks and how they will be measured.

Strong, balanced, sensitive leadership is required. Although all participants should share in the rights, responsibilities, and credit for the collaboration's activities and products, strong leadership is required to make the process work. The leader may be a local participant or an independent facilitator, but he or she needs to have or earn the respect and trust of all group members. The ideal leader needs to be sensitive and fair, hearing all sides of a debate, and at the same time be neither dictatorial nor benign. The leader must structure the process and move it along at an appropriate pace without restricting participation by each member and encouraging participation by all. Further, the leader needs to deal in a balanced and fair manner with disagreements and conflicts within the group while respecting the diverse contribution of each participant. Few people have all these characteristics, yet such leadership is crucial to an effective collaboration (Groark & McCall, 2005; McCall et al., 1997).

Program Development and Design

Effective program development requires establishing and operating a successful collaboration, preparing and planning, and then conducting a logic model program development process. Many of the principles described in this section appear obvious or common sense, but in our experience they are not commonly followed or enacted.

Operating successful collaborations. Successful collaborations tend to operate in a specific way (Butterfoss, Goodman, & Wandersman, 1996; Groark & McCall, 2005; Grobe, Curnan, Melchior, & the Center for Human Resources, 1993; Kegerise, 1999; Mattessich & Monsey, 1992; Wandersman & Goodman, 1993). They have regular meetings with required attendance, and they begin by identifying the values and perspectives of each participant and what they bring to the collaboration. Skills, resources, roles, and responsibilities of each participant should be identified, as well as major resources and time commitments that will be needed.

If the collaboration is large and complex, written policies or even bylaws and procedures should be considered. The group should decide how decisions will be made, whether votes will be taken, and how ties will be resolved. Ultimately, the collaboration should develop a "business plan" that specifies potential resources and expenses as well as goals, procedures, responsibilities, and legislative, legal, and administrative needs. Finally, the collaboration itself may need to be monitored and periodically reviewed, occasionally procedures need to be modified, and sometimes even the leadership needs to change.

Planning and preparation. An initial step in preparing to develop a new service program is to *understand the current situation*; namely, the nature of the intended participant group and its cultural and social values, political opportunities and constraints, and the history of the problem and previous attempts to remediate it (see Chapter 73 on service panning in general). For example, in the case of an intervention project to improve caregiving in the orphanages of St.

Petersburg, Russian Federation (i.e., Groark, Muhamedrahimov, Palmov, Nikiforova, & McCall, 2005; Muhamedrahimov, Palmov, Nikiforova, Groark, & McCall, 2004), this meant understanding 1) how caregiving was currently provided, the rules and regulations governing the orphanage; 2) why various practices exist (e.g., caregivers do not get social-emotionally close to children to avoid the pain of separation later); 3) that caregivers work 24-hour shifts to minimize transportation expenses and to have three days off to be with their own children or work other jobs); and 4) a long tradition of adult-directed teaching.

Next, learn what has been tried before, both locally and nationally. This step includes a review of the service literature, which may focus on the guidelines offered by a consensus group or brief reviews provided by local scholars plus knowledge of what has been tried specifically in the target locale. In the case of the orphanage project, there were few precedents, but there was a literature on early social-emotional development (e.g., attachment) and on the components of early childhood care and education programs that correlate with children's development.

A needs assessment, conducted community-wide or specific to a particular location, agency, or program, is very useful. Community needs assessments may be conducted by an independent organization with the assistance of a diverse committee of practice professionals, funders, policy makers, scholars, and service participants, or it could be the collaborative community group conducting this program development process. The needs assessment consists of reviewing the scholarly literature on the issue (e.g., after-school programming, adolescent crime prevention, comprehensive family services), determining the frequency and severity of the problems to be prevented or remediated in the target area, specifying the short- and long-term consequences to individuals and costs to society if nothing is done or current services were continued, the nature and extent of existing services, the geographic distribution of problems versus services, available personnel and their training and preparation, and resources available or needed.

Creating the planning document. The program planning process should produce a written planning document. This is sometimes a grant proposal in which a government or private funder has issued a request for proposals that specifies the target population, goals, and services it wants to fund. Alternatively, the diverse community-collaborative group may plan the services with or without a set of guidelines or request for proposals.

Several structured formats are available to facilitate program development. For example, the Strengths, Weaknesses, Opportunities, and Threats (SWOT) analysis (Kearns, 1992) provides an organized approach for general strategic planning, perhaps for an entire service agency. More specific to program development is the logic model. While research on the effectiveness of the logic model approach is limited, people who use it recommend it and it is being widely advocated. For example, the Kellogg Foundation (1998, 2000) provides a detailed guide for conducting a generic logic model that could be applied to any service need. Benson (1997) presents an asset-based needs assessment pertaining to adolescents, and Catalano and Hawkins (1996) describe risk and protective factors and suggestions for program development for preventing adolescent antisocial behavior. More recently, the logic model concept has been packaged into the Getting to Outcomes (GTO) process (Wandersman, Imm, Chinman, & Kaftarian, 1999, 2000) that is available in several formats pertinent to particular problems (Chinman, Imm, & Wandersman, 2004; Fisher, Imm, Chinman, & Wandersman, 2006;

<http://www.RAND.org/publications/TR/TR101>). The Getting to Outcomes approach will soon be available in an interactive web-based technology system iGTO.

The logic model approach. The logic model approach consists of a series of questions that provide a structure and logic to the program planning process. The logic model process is common sense, but not commonly conducted. It forces the collaborative to think clearly, specifically, and realistically about needs, goals, participants, services, rationale, measures of process and outcome, and resources, and to make sure these components are aligned. Proponents of the logic model approach argue that it produces a clear, integrated, logical program that is more likely to be effective and an evaluation that is more likely to be relevant and to contribute to program improvement and understanding how the program works (e.g., Armstrong & Barsion, 2006; Axford, Berry, & Little, 2006; WK Kellogg, 1998, 2000). While the specific questions vary from one approach to another, generally logic models address the following questions (Groark & McCall, 2005).

1. *What is the ultimate goal of the services?* Specifically, what is the problem or need and what should be the situation if the new intervention or services were completely successful? This is the “blue sky question”—in the ideal case, what should the intervention or services accomplish? In the case of the orphanage project, the ultimate goal was that children should develop at typical levels in all domains. 2. *What measures would indicate that this long-term goal has been achieved?* Often the goal stated in number 1 above is phrased in very general terms (e.g., children develop typically) so in this step that goal must be specified in measurable terms. For example, the measurable long-term goal for the orphanage children was that they should have heights, weights, and developmental test scores for personal-social, communication, and cognition that average the same as non-orphanage parent-reared children.

2. Care must be taken to insure that the indices of success indeed reflect the long-term goal and are reasonable and potentially attainable. Some policy makers and funders prefer aggressive goals, such as “cutting the teenage pregnancy rate by 50%.” The goal is worthwhile, but not likely attainable; should a program that cuts the rate by 25% be deemed a failure relative to such a lofty goal? Instead, determine both the minimum level of improvement that would be considered “success” as well as the maximum level likely to be attained? Thus, a goal of reducing the rate by 20-40% might be appropriate (no one will be upset if the program achieves a 45% reduction). The goal also should be well-specified—a reduction of 20-40% relative to what (e.g., current levels, a comparison group), in what length of time (e.g., 2 years, 5 years), for what ages of adolescents, etc.? Measuring instruments should also be specified when different assessment tools, reflecting different emphases, are available.

3. *Who are the targeted participants?* Who do you want to help, how will they be identified, what are their characteristics, and how many are there? In the orphanage project the participants were all caregivers and children in three different orphanages, but in the case of Sure Start the services would be located in low-income neighborhoods and serve anyone living there. Should outreach efforts be made to recruit families, who should be employed to do this, and how should this be done? For instance, should workers be hired from the neighborhood to recruit families because target families can relate to them, or would that cross the line of privacy?

4. *What services can be provided that will produce the long-term goal?* First, determine a promising theory of change (Chen & Rossi, 1983, 1987), then identify services that follow from the theory that have been tried by others and found to be successful. This is the step in which theory, research evidence, and best practice guidelines enter the process. For example, what are the elements and characteristics of successful services, and what modifications need to be made to fit the local clientele and social and political circumstances? This step should have at least two major characteristics: First, there needs to be a clear rationale for why the proposed services should produce the intended outcome (i.e., theory of change), and second, there should be some level of evidence (e.g., Table 1) to justify the potential effectiveness of the services, service components, and characteristics. What frequency and intensity of services are sufficient to produce the intended benefits? Do the services match the risk level of participants? In the orphanage project, for example, attachment theory provided a theoretical rationale for reducing the number of caregivers, making them more consistent in the lives of children, and having them behave in a warm, caring, sensitive, and responsive manner. Also, there was an empirical rationale for smaller group size, better caregiver-child ratios, age and disability integration, and child-directed caregiver-child interactions. Together, these changes should promote all aspects of children's development.

5. *How will you know that the intended services are actually delivered and match the plan in character and intensity?* This step requires a plan to monitor the delivery of services, not just how many participants receive which services but whether the nature and extent of the services are consistent with the plan (i.e., "program fidelity"). Perhaps a management information system (MIS) needs to be developed. In the case of the Comprehensive Child Development Program, did the case manager work with the family to identify their strengths, did the family identify their needs and put them into a priority sequence, did the case manager develop a relationship of trust and support with the family, did the family participate in services relevant to the family's goals?

6. *What are the short-term goals and measures?* What should be the first beneficial outcomes of such services, and is there evidence that achieving the short-term outcomes predicts achieving the long-term outcomes? In the orphanage project, did HOME Inventory scores of caregivers' behavior on the wards increase, because there is evidence that higher scores are related to better developmental outcomes for children? Short-term goals should be ladderred from goals that are obvious and easily attained to goals that are more difficult to achieve, so that the project will have some success.

7. *What is an appropriate timeline both to deliver services and expect outcomes?* It is important to be realistic in planning the timeline for both implementation and when short- and long-term outcomes are likely to be achieved. Sometimes a neighborhood is not ready for a needed service, and this should be anticipated. For example, a funder may suggest establishing a family support center in a particular low-income community. To do so, there must be community "will" for such a center, a recognized and respected leader, an agency willing to administer the program, and other resources. If not, program developers may need a year or more to hold town meetings to assess the community's true interest and capacity and provide recreational services to gain trust.

8. *What is a reasonable budget to deliver these services and measure their implementation and effectiveness?* This step begins with assigning costs to all of the elements in the previous steps, determining if the total budget is feasible, and if not, revisiting each of the above steps and making modifications to meet a specified budget. Ultimately, the group needs to decide whether a program that is likely to be effective can be operated for the amount of money available. This is also one of several reasons why it is helpful to have policy makers and funders at the table during this planning process.

Implementation

Implementation is as crucial a process to successful services as program development, but it is often forgotten and rarely studied. Further, implementing new services in new communities may require months or even years to build trust among potential participants, and several cohorts of participants and much trial-and-error may transpire before even experienced teams of providers refine procedures sufficiently to produce benefits in participants.

Community collaborative. The same community collaborative that designs the program may also be used to oversee its implementation. Such a committee should have the characteristics and operate according to the principles for collaborations described above, and implementation should strive to have the characteristics of successful programs given in Table 2.

Characteristics of good implementation. Crucial to good implementation is to have a strong leader who relates to the diverse organizing committee plus a well-educated, trained, and experienced staff who are supervised in a supportive and encouraging manner on a regular basis. Studies show that training alone is relatively ineffective unless it is accompanied by such supervision (e.g., Morrow, Townsend, & Pickering, 1991). Management needs to involve the staff in understanding the purpose of the services, their roles, and the intended outcomes, perhaps by involving some of them in the logic model process described above or reviewing that process so that they can contribute modifications in the design of the services. If staff will collect data, they need to understand why it is necessary and how they can use the data to guide and improve their services.

Staff also need structural supports, including reasonable work loads and the availability of resources, such as technology, consultants, and training necessary to do their jobs. Studies also show, that teacher education and training tend not to be related to positive outcomes for children in early care and education settings, for example, unless group size and teacher:child ratios are small enough so that teachers can exercise the best practices of their training (Love, Schochet, & Meckstroth, 1996).

A management information system may be needed for large and multi-sited projects that systematizes in a uniform manner participant information, services delivered, and short- and long-term outcome measures. Directors and supervisors need to monitor such information to determine that the services are being delivered according to plan, including whether the participants are those who were originally targeted; whether services are delivered in a timely and consistent way and at an intensity or frequency sufficient to produce benefits; and whether short-term goals are being achieved.

Mid-course corrections may be necessary. For example, one program aimed at providing services and supports to drug and alcohol abusing adolescent mothers found after two years of operation that the average age of the mothers was 26 years, not the adolescents the program was designed to target. Staff reported that drug and alcohol abusing adolescents did not perceive that they had a problem, because drug and alcohol use and abuse was rampant in their peer group; it seemed to take nearly ten years for such mothers to recognize that they had a problem and to seek help. The project could change its target group to older mothers who were motivated to use the services and/or it could make greater efforts to recruit adolescents. It was decided to do both, using the older mothers to convince the adolescents they needed help and enroll them in the service.

Evaluation

Most large, comprehensive, multi-sited services have a required monitoring and evaluation program. The funder may contract with a central and independent evaluation team, which may prescribe and conduct monitoring and evaluation for all sites, such as in Sure Start and CCDP. Smaller programs should also have a monitoring and evaluation plan, and select appropriate indices of participants, services, and outcomes during the logic model process.

Program evaluation is a specialized skill, and relatively few service professionals have the appropriate training. Furthermore, simply providing local service professionals with technical assistance in evaluation does not work very well (Chinman et al., 2005). Consequently, large-scale and multi-sited service projects often have a central team of independent evaluators who should be part of the diverse community planning and implementation group and contribute to the program guidelines as well as collaboratively design the monitoring and evaluation plan, specify measurements, conduct the assessments, analyze the data, and report on the process and outcome effectiveness of the project. Alternatively, some communities have local evaluation teams that work with local agencies in a participatory manner to design and conduct local evaluations, even the evaluation of the local site of a multisited national intervention, such as in Early Head Start (e.g., McCall, et al., 1997; McCall, B. L. Green, Groark, Strauss, & Farber, 1999).

Modern service programs, especially services that are tailored to client needs and characteristics at both the site and individual participant levels, pose several issues for traditional methods of evaluation. We consider a few such issues below.

Independent versus participant/collaborative evaluation. Historically, innovative services were often created by a university scholar who both implemented the intervention and evaluated its effectiveness. Conversely, funders feel it is a conflict of interest for the people who create and implement a service to also evaluate it, so they prefer an independent evaluation team.

In the past, independent evaluators have sometimes been distant from the program and the services staff, have “done the evaluation to the project,” and sometimes issued reports that policy makers and practitioners regarded as having missed the point of the service program. A modern and compromise strategy is participatory or empowerment evaluation (Fetterman, 1993; Fetterman, Kaftarian, & Wandersman, 1996; McCall et al., 1997, 1999), in which evaluators and

service professionals work together from the beginning of program development to monitor and evaluate the process and outcome of the service program. Participatory evaluation is more likely to produce useful results, because the measures and design are determined collaboratively with service professionals.

Central versus decentralized evaluations. Large multi-sited service programs have often had a single evaluation team that designs and collects the data for all sites (e.g., Sure Start, CCDP). Such an approach has the benefit of a uniform evaluation of all sites, and it works best when the service program is prescribed in great detail and expected to be implemented in a uniform manner at each site. However, many modern service programs permit flexibility and modification to fit local circumstances, and even programs intended to be uniform across sites nevertheless are often implemented in different ways at different sites (Gilliam, Ripple, Zigler, & Leiter, 2000; McCall, Ryan, & Plemons 2003). A compromise strategy was used by the Early Head Start Research Consortium, which had a common core of service themes and a central evaluator, but each site had local evaluators who conducted the core assessments plus measures more pertinent to their local program emphases (Administration for Children and Families, 2002; Love et al., 2005).

The rush to evaluate outcomes. Funders and policy makers understandably are anxious to determine the effectiveness of programs they fund, and often there is an emphasis on assessing outcomes (i.e., benefits to participants) in the first cohort of participants. But the first “outcome” of any program should be that the services were implemented as planned for the targeted participants with the frequency and intensity that is likely to produce benefits. Until the services are delivered with such “fidelity,” it is unreasonable to expect outcome benefits to participants; and it may well take two or three cohorts of participants before the program operates smoothly and according to plan. This fact prompted Donald Campbell, a patriarch of program evaluation, to admonish: “Evaluate no program before it’s proud” (Campbell, 1987).

Over-emphasis on gold standard methodology. Scholars, policy makers, funders, and practice professionals have come to value the randomized clinical trial as the gold standard of applied research and program evaluation, so much so that often entire literatures using other methodological approaches are completely dismissed (e.g., as in review articles that only include randomized trials). While in some fields (e.g., education), there has been an avoidance of randomized trials, and some scholars (Cook, 2002) have called for greater use of this gold standard, in other fields randomized trials are over-emphasized (McCall & B. L. Green, 2004).

For example, while acknowledging the potential benefits of random assignment, McCall and B. L. Green (2004) argue that randomized trials work best in a double-blind design in which participants do not know which treatment they are being given and have relatively little influence over the success of the treatment (e.g., drug vs. placebo studies). However, in behavioral interventions, participants are almost never blind to which treatment group they have been assigned, and participants’ values and perceptions of the treatment can have substantial influence on how effective the treatment is for them. As a result, some participants display “reactive disappointment” to being in the control group and go get the treatment services on their own, shrinking the difference between the “treatment” and the “control/comparison” groups. For example, in one evaluation of the School Development Program (Millsap et al., 2000, reported

by Datta, 2003) in which schools were assigned to treatment versus control conditions, several of the control schools implemented some or all of the Comer principles on their own. The result was that there was minimum difference between treatment and control schools in the amount of SDP “treatment” and in outcomes. However, the extent to which schools’ implemented the SDP principles within the control group was related to outcome. A gold-standard purist would discount this finding because the beneficial outcomes were confounded with school motivation. Technically true, but a policy maker or superintendent might be sufficiently persuaded to adopt the SDP strategy if it can be effectively implemented

Further, after a demonstration project is completed successfully and services are offered routinely, they are never randomly assigned. So “participant bias” will always be in effect; from a practical standpoint, the “biased” demonstration sample is the group from which generalizations to future clients should be made (of course, one wants to know if such people would improve without a service program).

Also, such “bias” might be critical for program success. For example, if divorcing couples are randomly assigned to divorce mediation versus court-ordered settlements, it is possible that some couples cannot cooperate sufficiently for mediation to work whereas other couples would be dissatisfied with a divorce arrangement that was prescribed for them by a judge. The result of a strictly randomized trial might be no difference in parental satisfaction and adjustment of children between these two approaches. However, if couples were allowed to choose which approach they wanted, both strategies might be shown to be effective for those who choose them, more effective than for those who are randomly assigned to each approach. Notice that this is a case in which the randomized design produced a low effect size, contrary to those who sometimes claim that randomized designs provide a maximum effect size.

Consequently, McCall and B. L. Green (2004) advocated for a greater balance of and respect for different methodological approaches which provide different kinds of information about an intervention or service. Results from a comprehensive set of approaches are more likely to provide a more complete and better understanding of the effectiveness of that service. Different research designs and statistical strategies are described elsewhere (e.g., McCall & B. L. Green, 2004; Rossi, Lipsey, & Freeman, 2004).

Intent to treat analyses. Another gold standard approach to program evaluation is to conduct “intent to treat analyses” in which the outcomes are assessed on all participants who were *assigned* to the treatment conditions regardless of whether they actually *experienced* them, such as in Sure Start. This approach preserves the random assignment feature; it “avoids” the problem of selective dropouts, which is common among the high-risk populations that many services target, by including the dropouts as if they were treated participants. This strategy works well when the treatment program is uniformly delivered to all participants and the number of actual dropouts is small.

But including dropouts basically inflates the error within the treatment but not the control condition; and as the number of dropouts and partially treated participants increases, detecting significant differences between treatment and comparison groups becomes less likely. Further, it is logically absurd to pretend that dropouts received the treatment when they did not. In this case,

it seems reasonable to conduct both the intent to treat analysis as well as an analysis that is restricted to fully-treated subjects (perhaps limiting the control subjects to those who were similar to the fully-treated participants through propensity score analysis; Rosenbaum & Rubin, 1983).

The intent to treat strategy also seems limited if not inappropriate when the treatment itself is different from participant to participant. In the Comprehensive Child Development Program, families chose their goals, and only approximately 15% of the treatment families chose more education and 10% of the control families decided on their own to get more education. Yet the evaluation asked the percentage of *all* participants in both treatment and control groups who attained more education. Not surprisingly, few did (i.e., certainly less than 15%-10%), and there was a very small and insignificant difference in the rates between the two groups (McCall et al., 2003; St. Pierre, Layzer, Goodson, & Bernstein, 1994, 1997a, 1997b). Clearly, evaluation analyses must consider different goals and services delivered within the treatment and control groups with designs and analyses that consider that modern services are often individualized (see McCall & B. L. Green, 2004).

Sustainability and Replication

Sustainability refers to at least two outcomes: 1) maintaining the quality of services and benefit to participants over time, and 2) keeping the service funded after the initial demonstration grant expires. *Replication* refers to attempts to reproduce the program in new sites with similar populations with the intent of replicating the original benefits and outcomes.

Sustaining program benefits. It is not uncommon for there to be unusual passion, commitment, and dedication among program directors and staff during the implementation of a new service program, and such commitment often wanes over subsequent years. Also, staff turnover and changes in the participant group may also lessen the effectiveness of the services over time.

A management and information system (MIS) and a good plan for monitoring participants and services can provide feedback to directors regarding changes in the characteristics of participants and the manner and intensity with which services are being delivered. The director and supervisory staff in turn need to provide continuous supervision and monitoring, even of highly experienced staff, to maintain standards and program effectiveness. Supervision must be formative and reflective (e.g., Johnson & Tittnich, 2004) so that continuing improvements are made in services with staff input and enthusiasm for the program is maintained.

It also helps to design the original demonstration in such a way that the intervention can be maintained. Building intellectual and skill capacity in the community is essential in sustaining the quality of services. For example, the St. Petersburg (Russia) orphanage project (Groark et al., 2005; Muhamedrahimov et al., 2004) used a train-the-trainer approach in which a written curriculum was produced and a specialist trained the professionals in each orphanage. Those professionals were available to train new caregivers who replaced those who left and thus keep the level of performance high in the orphanage over time.

Sustaining funding. Many interventions and services are initiated with a demonstration project, which is often funded at relatively generous levels. But once the demonstration is completed, it can be much more difficult to obtain funding to continue the program, even when the evaluation has clearly demonstrated program benefits to participants. Also, when governments assume funding after demonstration programs, it is often at substantially lower levels but with the same expectations for success—it is the government’s penchant to expect champagne benefits on a beer budget.

Programs that must be trimmed to fit smaller budgets need to analyze and break down the program into its components, including personnel, participants, the types and extent of services provided, as well as the costs involved. The research literature on this type of service and characteristics of successful programs should be used as the criteria to determine which aspects of the service are likely to be crucial and which aspects appear less important and are likely candidates for trimming. Sometimes the number of participants may need to be reduced to maintain program quality within the available budget. This is always a very difficult choice for service professionals, but it seems obvious that it is better to be effective with fewer participants than to be ineffective with many.

One strategy in getting programs sustained after their initial demonstration is to invite potential long-term funders (e.g., the director of children and youth services for the geographic area in which the service program resides) to be on the community collaborative committee that designs, implements, and oversees the management of the demonstration program. By involving such funders at the beginning, they are more likely to be invested and to feel a sense of ownership of the program after the initial demonstration and therefore more likely to provide funds to sustain it.

The results of a successful evaluation should be disseminated widely and specifically to potential funders. Further, results need to be packaged differently for different audiences, from detailed technical reports for scholarly journals to one-page summaries for policy makers.

It is also helpful to have cultivated one or more “champions,” who may be policy makers, funders, influential citizens, celebrities, or media professionals. They also should be part of the community collaborative that designs and oversees the project from the beginning, so they are fully informed and feel partial ownership of the service. Eventually, their job is to “sell” the program to policy makers, funders, and even the general public.

Sometimes successful participants can be the most powerful advocates for sustaining a service. Consider a drug and alcohol abusing mother of two young children who is assisted by a family support program to become substance free, develop marketable skills, become employed, and whose children wind up in honors classes. If she is willing to tell her story publicly, she can do more to sell a program than all the carefully collected evaluation results. It is difficult for legislators to look such a success story in the eye and then vote against funds to offer the same opportunities to others.

Replication. Replicating a program in a new locale with new staff and clientele requires the same planning process and ingredients for successful implementation as for the original—i.e.,

leadership, committed staff, perceived need for change, program development with the logic model process, etc. For example, does the new director have sufficient commitment, energy, and confidence to “pull this off?” Is the staff capable and receptive to change? Are the local social, cultural, and economic conditions conducive to change? If not, these elements need to be resolved before or during the program planning process.

Conclusion

There is no “silver bullet” or step-by-step manual that will guarantee that community-based interventions services will be successful. However, there are principles and strategies suggested by research that help to create highly effective programs. They start with a thorough process of planning that customizes the service and its delivery to the clientele and context. It includes developing a solid collaboration of diverse but relevant stakeholders who become knowledgeable, loyal partners. Together these stakeholders develop a common purpose, goals, clearly agreed-upon rules, and are led by a fair, balanced, and respected leader. They employ a structured planning process that designs a program founded on theoretical and practice evidence and based on characteristics that work, while always being sensitive to future sustainability and replication issues. The program is continuously monitored for service improvement and evaluated for interim and long-term outcomes. Moreover, all steps are documented and frequently reviewed by senior management and the diverse oversight group for “lessons learned.” Finally, outcomes and lessons learned are disseminated to policy makers and practitioners for sustainability in the current sites and replications to improve services and lives elsewhere.

References

- Administration for Children and Families (2002). *Making a difference in the lives of infants and toddlers and their families: The impacts of Early Head Start*. Washington DC: U.S. Department of Health and Human Services.
- APA Presidential Task Force on Evidence-Based Practice (2006). Evidence-based practice in psychology. *American Psychologist*, *61*, 271-285.
- Armstrong, E. G., & Barsion, S. J. (2006). Using an outcomes-logic-model approach to evaluate a faculty development program for medical educators. *Journal of the Association of American Medical Colleges*, *51*(5), 483-488.
- Axford, N., Berry, V., & Little, M. (2006). Enhancing service evaluability: Lessons learned from a programme for disaffected young people. *Children and Society*, *14*, 287-298.
- Backer, T., David, S., & Soucy, G. (Eds.) (1995). *Reviewing the behavioral science knowledge base on technology transfer*. Washington, DC: National Institute of Drug Abuse.
- Baumrind, D., Larzelere, R. E., & Cowen, P. A. (2002). Ordinary physical punishment: Is it harmful? Comment on Gershoff, 2002. *Psychological Bulletin*, *128*, 588-589.
- Belksy, J., Melhuish, E. Barnes, J. Leyland, A. H., Romaniuk, H., et al. (2006). Effects of Sure Start local programmes on children and families; Early findings from a quasi-experimental, cross sectional study. *British Medical Journal*, doi: 10.1136/bmj:38853.451748, 2F
- Benson, P. (1997). *All kids are our kids: What communities must do to raise caring and responsible children and adolescents*. San Francisco, CA: Jossey-Bass Inc.
- Butterfoss, F. D., Goodman, R. M., & Wandersman, A. (1996). Community coalitions for prevention and health promotion: Factors predicting satisfaction, participation and planning. *Health Education Quarterly*, *23*, 65-79.
- Campbell, D. T. (1987). Problems for the experimenting society in the interface between evaluations and service providers. In S. L. Kagan, D. R. Powell, B. Weissbourd, and E. F. Zigler (Eds.), *America's family support programs* (pp. 345-351). New Haven, CT: Yale University Press.
- Catalano, R., & Hawkins, J. D. (1996). The social development model: A theory of antisocial behavior. In J. D. Hawkins, (Ed.), *Delinquency and crime: Current theories*. New York, NY: Cambridge University Press.
- Ceci, S. J., & Papierno, P. B. (2005). The rhetoric and reality of gap closing: When the "have nots" gain but the "haves" gain even more. *American Psychologist*, *60*(2), 149-160.
- Centers for Disease Control (2006). Guide to community preventive services. www.thecommunityguide.org.
- Chambless, D. L., & Hollon, S. D. (1998). Defining empirically supported therapies. *Journal of Consulting and Clinical Psychology*, *66*, 7-18.
- Chen, H., & Rossi, P. (1983). Evaluating with sense: The theory-driven approach. *Evaluation Review*, *7*(3), 283-302.
- Chen, H., & Rossi, P. (1987). The theory-driven approach to validity. *Evaluation and Program Planning*, *10*, 95-103.
- Chinman, M., Hannah, G., Wandersman, A., Ebener, P., Hunter, S. B., & Imm, P. (2005). Developing a community science research agenda for building community capacity for effective preventive interventions. *American Journal of Community Psychology*, *35*(3/4), 143-157.

- Chinman, M., Imm, P., & Wandersman, A. (2004). *Getting to Outcomes 2004: Promoting accountability through methods and tools for planning, implementation, and evaluation*. Santa Monica, CA: RAND Corporation Technical Report, Available at: <http://www.rand.org/publications/TR/TR10>.
- Comer, J. P. (1988). Educating poor minority children. *Scientific American*, 259(5), 42-48.
- Comer, J. P., Haynes, N. M., Joyner, E. T., & Ben Avie, M. (1996). *Rallying the whole village: The Comer process for reforming education*. New York: Teachers College Press.
- Cook, T. D. (2002). Randomized experiments in educational policy research: A critical examination of the reasons the educational evaluation community has offered for not doing them. *Educational Evaluation and Policy Analysis*, 24, (3), 175-199.
- Cook, T. D., & Campbell, D. T. (1979) *Quasi-experimentation: Design and analysis issues for field settings*. Chicago, IL: Rand McNally.
- Cook, T. D., Hunt, H.D., & Murphy, R. F. (1998). *Comer's School Development Program in Chicago: A theory-based evaluation*. Evanston, IL: Northwestern University Institute for Policy Research.
- Cook, T. D., Habib, F. N., Phillips, M., Settersten, R. A., Shagle, S. C., & Degirmencioglu, S. M. (2000). Comer's School Development Program in Prince George's County, Maryland: A theory-based evaluation. *American Educational Research Journal*, 36(3), 543-597.
- Cooper, H. (1998). *Synthesizing research: A guide for literature reviews*. Thousand Oaks, CA: Sage.
- Datta, L. E. (2003). Avoiding unwarranted death by evaluation. *The Evaluation Exchange*, 9(2), 5.
- Denner, J., Cooper, C. R., Lopez, E. M., & Dunbar, N. (1999). Beyond "giving science away": How university-community partnerships inform youth programs, research, and policy. *Society for Research in Child Development Social Policy Report*, 13(1).
- Early Head Start: www.acf.hhs.gov/programs/OPRE/EHS_resrch/
- Ennett, S. T., Tobler, N. S., Ringwalt, C. L., & Fewling, R. L. (1994). How effective is drug abuse resistance education? A meta-analysis of Project DARE outcome evaluations. *American Journal of Public Health*, 84, 1394-1401.
- Fergusson, D. M., Grant, H., Horwood, L. J., & Ridder, E. M. (2005). Randomized trial of the Early Start Program of Home Visitation. *Pediatrics*, 116, 803-809.
- Fergusson, D. M., Grant, H., Horwood, L. J., & Ridder, E. M. (2006). Randomized trial of the Early Start Program of Home Visitation: Parent and family outcomes. *Pediatrics*, 117, 781-786.
- Fetterman, D. M. (1993). Empowerment evaluation, *Evaluation Practice*, 15(1), 1-15.
- Fetterman, D. M., Kaftarian, S. J., & Wandersman, A. (1996). *Empowerment evaluation: Knowledge and tools for self-assessment and accountability*. Thousand Oaks, CA: Sage.
- Fisher, D., Imm, P., Chinman, M., & Wandersman, A. (2006). *Getting to Outcomes with Developmental Assets: Ten steps to measuring success in youth programs and communities*. Minneapolis, MN: Search Institute, www.search-institute.org.
- Friesen, B. J., & Koroloff, N. M. (1990). Family-centered services: Implications for mental health administration and research. *Journal of Mental Health Administration*, 17, 13-25.
- General Accounting Office (2003, January 15). *Youth illicit drug use prevention: DARE long-term evaluations and federal efforts to identify effective programs* (GAO-03-172R).

- Gilliam, W. S., Ripple, C. H., Zigler, E. F., & Leiter, V. (2000). The Comprehensive Child Development Program: What went wrong? *Early Childhood Research Quarterly*, 15(1), 41-59.
- Green, L. W. (2001). From research to "best practices" in other settings and populations. *American Journal of Health Behavior*, 25(3), 165-178.
- Green, B., Rodgers, A., & Johnson, S. (1999). Understanding patterns of service delivery and participation in community-based family support programs. *Children's Services: Social Policy, Research, and Practice*, 2(1), 1-22.
- Groark, C. J., & McCall, R. B. (1993, Spring). Building mutually beneficial collaborations between researchers and community service professionals. *Newsletter of the Society for Research in Child Development*, 6-14.
- Groark, C. J., & McCall, R. B. (1996). Building successful university-community human service agency collaborations. In C. D. Fisher, J. P. Murray, and I. E. Sigel (Eds.), *Applied developmental science: Graduate training for diverse disciplines and educational settings* (pp. 237-251). Norwood, NJ: Ablex.
- Groark, C. J., & McCall, R. B. (2005). Integrating developmental scholarship into practice and policy. In M. H. Bornstein and M.E. Lamb (Eds.), *Developmental psychology: An advanced textbook (5th Edition)* (pp. 570-601). Mahwah, NJ: Lawrence Erlbaum Associates.
- Groark, C. J., Muhamedrahimov, R. J., Palmov, O. I., Nikiforova, N. V., & McCall, R. B. (2005). Improvements in early care in Russian orphanages and their relationship to observed behaviors. *Infant Mental Health Journal*, 26(2), 96-109.
- Grobe, T., Curnan, S., Melchior, A., & The Center for Human Resources (1993, October). *Synthesis of existing knowledge and practice in the field of education partnerships*. Washington, DC: U. S. Department of Education Office of Educational Research and Improvement.
- Haynes, N. M. (Ed., 1995). School Development Program research monograph, ERIC document # ED 371 091.
- Johnson, B., & Tittnich, E. (2004). Are you ready for an infant mental health consultant? *Zero to Three* (24)6, 39-46
- Joyner, E. T., Comer, J. P., & Ben-Avie (Eds.; 2004). *Transforming school leadership and management to support student learning and development*. Thousand Oaks, CA: Corwin Press.
- Kearns, K. P. (1992). From comparative advantage to damage control: Clarifying strategic issues using SWOT analysis. *Nonprofit Management & Leadership*, 3(1), 3-22.
- Kegerise, K. E. (1999, March). Keys to successful collaboration. *Developments, Special Report*. Pittsburgh, PA: University of Pittsburgh Office of Child Development.
- W. K. Kellogg Foundation (1998). *W. K. Kellogg Foundation evaluation handbook*. Battle Creek, MI: W. K. Kellogg Foundation.
- W. K. Kellogg Foundation (2000). *W. K. Kellogg Foundation logic model development guide*. Battle Creek, MI: W. K. Kellogg Foundation.
- Kumpfer, K. L., & Alvarado, R. (2003). Family-strengthening approaches for the prevention of youth problem behaviors. *American Psychologist*, 58, 457-465.
- Layzar, J. I., Goodson, R. D., Bernstein, L., & Price, C. (2001). *National evaluation of family support programs. Vol. A: The meta analysis*. Cambridge, MA: Abt Associates Inc.

- Love, J. M., et al. (2005). The effectiveness of Early Head Start for 3-year-old children and their parents: Lessons for policy and programs. *Developmental Psychology*, 41(6), 885-901.
- Love, J. M., Schochet, P. Z., & Meckstroth, A. L. (1996). Are they in real danger? What research does and doesn't tell us about child care quality and children's well-being. *Child Care Research and Policy Papers*. Princeton, NJ: Mathematica Policy Research, Inc.
- Mattessich, P. W., & Monsey, B. R. (1992). *Collaboration: What makes it work? A review of research literature on factors influencing successful collaborations*. St. Paul, MN: Amherst H. Wilder Foundation.
- Maughan, B., Pickles, A., Rutter, M., & Ouston, J. (1990). Can schools change? I Outcomes at six London secondary schools. *School Effectiveness and School Improvement*, 1(3), 188-210.
- McCall, R. B. (2007). Toward more realistic evidence-based programming and policies. *American Psychologist*.
- McCall, R. B., & Green, B. L. (2004). Beyond the methodological gold standards of behavioral research: Considerations for practice and policy. *Society for Research in Child Development Social Policy Reports*, 18(2), 1-19.
- McCall, R. B., Green, B. L., Groark, C. J., Strauss, M. S., & Farber, A. E. (1999). An interdisciplinary, university-community, applied developmental science partnership. *Journal of Applied Developmental Psychology*, 20, 207-226.
- McCall, R. B., Green, B. L., Strauss, M. S., & Groark, C. J. (1997). Issues in community-based research and program evaluation. In I. E. Sigel and K. A. Renninger (Eds.), *Handbook of child psychology, Vol. 4, 5th Edition*, (pp. 955-997). New York: Wiley.
- McCall, R. B., Groark, C. J., & Nelkin, R. (2004). Integrating developmental scholarship and society: From dissemination and accountability to evidence-based programming and policies. *Merrill-Palmer Quarterly*, 50, 326-340.
- McCall, R. B., Larsen, L., & Ingram, A. (2003). The science and policy of early childhood education and family services. In A. J. Reynolds, M. C. Wang, and H. J. Walberg (Eds.), *Early childhood programs for a new century: Issues in children's and families lives* (pp. 255-298). The University of Illinois at Chicago Series on Children and Youth. Washington, D.C.: CWLA Press.
- McCall, R. B., Ryan, C. S., & Plemons, B. W. (2003). Some lessons learned on evaluating community-based, two-generation service programs: The case of the Comprehensive Child Development Program (CCDP). *Journal of Applied Developmental Psychology*, 24, 125-141.
- Millsap, M. A., Chase, A., Obeidallah, D., Perez-Smith, A., Brigham, N., & Johnston, K. (2000). *Evaluation of Detroit's Comer schools and families initiative*. Cambridge, MA: Abt Associates.
- Morrow, A. L., Townsend, I. T., & Pickering, L. K. (1991). Risk of enteric infection associated with early child day care. *Pediatric Annals*, 20(8), 427-433.
- Mrazek, P. J., & Haggerty, R. J. (Eds.) (1994). *Reducing risks for mental disorders: Frontiers for prevention intervention research*. Washington, DC: National Academy Press.
- Muhamedrahimov, R. J., Palmov, O. I., Nikiforova, N. V., Groark, C. J., & McCall, R. B. (2004). Institution-based early intervention program. *Infant Mental Health Journal*, 25(5), 488-501.

- Nation, M., Crusto, C., Wandersman, A., Kumpfer, K. L., Seybolt, D., Morriseey-Kane, E., & Davino, K. (2003). What works in prevention: Principles of effective prevention programs. *American Psychologist*, *58*, 449-456.
- National Evaluation of Sure Start Team (2005a). *Report for team: Variation in Sure Start local programme effectiveness: Early preliminary findings*. London, UK: HMSO.
- National Evaluation of Sure Start Team (2005b). *Report 13: Early impacts of Sure Start local programmes on children and families*. London, UK: HMSO.
- Neufield, B., & LaBue, M. (1994). *The implementation of the School Development Program in Hartford: Final evaluation report*. Cambridge, MA: Education Matters, 1994.
- Olds, D. L., & Kitzman, H. (1993). Review of research on home visiting for pregnant women and parents of young children. In R. E. Behrman (Ed.), *Home visiting. The Future of Children*, *3*, 53-92. Los Angeles, CA: The David and Lucile Packard Foundation.
- Olds, D. L., Henderson, C. R. Jr., Kitzman, H. J., Eckenrode, J. J., Cote, R. E., & Tatelbaum, R. C. (1999). In R. E. Behrman (Ed.). *Home visiting: Recent program evaluations. The Future of Children*, *9*(1), 44-65.
- Ouston, J., Maughan, B., & Rutter, M. (1991). Can schools change? II Practice in six London secondary schools. *School Effectiveness and School Improvement*, *2*(1), 3-13.
- Ringwalt, C. L., Ennett, S., Vincus, A., Thorne, J., Rohrbach, L. A., & Simons-Rudolph, A. (2002). The prevalence of effective substance use prevention curricula in U.S. middle schools. *Prevention Science*, *3*, 257-265.
- Rosenbam, P. R., & Rubin, D. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, *70*, 41-55.
- Rossi, P. H., Lipsey, M. W., & Freeman, H. E. (2004). *Evaluation: A systematic approach (7th edition)*. Thousand Oaks, CA: Sage.
- Rutter, M. (2006). Is Sure Start an Effective Preventive Intervention? *Child and Adolescent Mental Health*, *11*, (3), 2006, 135–141.
- Schorr, L. B. (2003, February). Determining “what works” in social programs and social policies: Toward a more inclusive knowledge base.
www.brookings.edu/views/papers/sawhill/20030226.htm.
- Scott, K. G., Mason, C. A., & Chapman, D. A. (1999). The use of epidemiological methodology as a means of influencing public policy. *Child Development*, *70*, 1263-1272.
- Shonkoff, J. P. (2000). Science, policy, and practice: Three cultures in search of a shared mission. *Child Development*, *71*, 181-187.
- St. Pierre, R. G., Layzer, J. I., Goodson, B. D., & Bernstein, L. S. (1994) *National evaluation of the Comprehensive Child Development Program: Interim report*. Cambridge, MA: Abt Associates.
- St. Pierre, R. G., Layzer, J. I., Goodson, B. D. & Bernstein, L. S. (1997a, September). *The effectiveness of comprehensive case management interventions: Findings from the national evaluation of the Comprehensive Child Development Program*. Cambridge, MA: Abt Associates.
- St. Pierre, R. G., Layzer, J. I., Goodson, B. D. & Bernstein, L. S. (1997b, June). *National impact evaluation of the Comprehensive Child Development Program: Final Report*. Cambridge, MA: Abt Associates.
www.surestart.gov.uk

- Wandersman, A., & Florin, P. (2003). Community interventions and effective prevention. *American Psychologist, 58*(6/7), 441-448.
- Wandersman, A., & Goodman, R. (1993). *Understanding coalitions and how they operate: An "open systems" organizational perspective*. W. K. Kellogg Foundation Community-Based Public Health Initiative.
- Wandersman, A., Imm, P., Chinman, M., & Kaftarian, S. (1999). *Getting to Outcomes: Methods and tools for planning, evaluation and accountability*. Rockville, MD: Center for Substance Abuse Prevention.
- Wandersman, A., Imm, P., Chinman, M., & Kaftarian, S. (2000). Getting to Outcomes: A results-based approach to accountability. *Evaluation and Program Planning, 23*, 389-395.
- Weissberg, R. P., & Kumpfer, K. L. (Eds.) (2003). Prevention that works for children and youth. Special Issue, *American Psychologist, 58*, 425-490.
- Weisz, J. R., Jensen-Doss, A., & Hawley, K. M. (2006). Evidence-based youth psychotherapies versus usual clinical care: A meta-analysis of direct comparisons. *American Psychologist, 61*(7), 671-689.

Table 1. Elements of a Scheme for Assessing the Research Evidence
in Support of a Service Program

1. Program effectiveness (internal validity): To what extent does the research demonstrate that the program per se produces its intended benefits in participants? (Generally $a > b > c$ in persuasiveness).
 - a. *Randomized experimental and quasi-randomized evidence*. Studies using random assignment of individuals or larger units to program vs. comparison conditions plus experimenter-selected units that are arbitrarily assigned to treatment vs. comparison conditions;
 - b. *Non-randomized quasi-experimental designs* (e.g., Cook & Campbell, 1979; Rossi, Lipsey, & Freeman, 2004). Studies lacking random or experimenter-controlled assignment (i.e., participants self-select to conditions) with some comparison provision (e.g., comparison groups, propensity score comparison individuals, instrumental variable estimation) that supports the inference of causality. Also, interrupted time series with three or more time points before and after the program.
 - c. *Non-random single-group designs with a pre- and post program assessment*. These studies are more persuasive if they provide some additional evidence consistent with causality (e.g., dose-response effect; relations between program elements hypothesized to be causal mediators, such as fidelity of implementation or success of program implementation in individuals, and outcome benefits).
2. Elements of persuasive research design. To what extent is the literature within each of the three design categories characterized by the following (the more characteristics the better):
 - a. A theory or cause-and-effect conceptualization that is well-supported by a variety of other research?
 - b. Large Ns and limited drop out?
 - c. Pre-program as well as post-program assessments to demonstrate change within individuals?
 - d. Assessment of program implementation that demonstrates the program was faithfully and completely implemented in the program group and less so or not at all in the comparison group?
 - e. Monitoring and assessment of participant exposure to program elements (e.g., number of service hours experienced, number of participants completing the entire program) and immediate effects of program on participants (e.g., training produced learning gains; program experience improved participants self-esteem) that are hypothesized to produce immediate or long-term outcomes?
 - f. Various participant characteristics otherwise thought to influence outcome are assessed, covaried, or examined as moderators in all groups?
3. Replication. To what extent has the program been shown to be effective in each of the three design categories in two or more different studies (separate groups of participants), especially (in order of increasing persuasiveness) if they were:
 - a. Conducted by the same investigators and providers but on different participants?
 - b. Conducted by the same investigators but using different providers and participants?
 - c. Conducted by different investigators using different providers and participants?

Table 1 cont.

4. Effect size. To what extent in each of the three design categories are the effect sizes associated with the program, measured in terms of odds ratios (see Scott, Mason, & Chapman, 1999) and cost/benefit ratios, sufficient to justify replication?
5. Generality. To what extent has the program been shown to be effective within each of the three design categories using as many of the following:
 - a. *Participants* varying in gender, age, education, income, race and ethnicity, severity of risk or problem condition, and other characteristics thought to be related to outcome and program effectiveness?
 - b. *Providers* that vary in education, training, and experience in general and specifically related to this program?
 - c. *Program characteristics* that vary, such as budgets, facilities, and tangible and personnel resources and other aspects (e.g., size of program, number of participants per provider, hours of service per participant) that may influence outcome?
6. External validity specific to the replication circumstances. To what extent has the program been shown to be effective within each of the three design categories using program characteristics as described in #5 above that are similar to those likely to characterize the replicated projects?
7. Feasibility of replication. Is the program and its implementation process described comprehensively and packaged in a way that providers new to the program can faithfully replicate the program?

Characteristics of successful programs. Does the literature support one or more characteristics that may be important to producing beneficial outcomes, because they are common components of programs that were successful, especially if they are not common components of programs that were not successful (such characteristics may be participant, provider, and program characteristics [see # 5 above] or the characteristics listed in Table 2)? Is there a strong theory of change and evidence on how or why the program is effective?

Table 2. Characteristics of Successful Collaborations among Diverse Professionals

A common purpose. A collaboration is beneficial and likely to succeed if it is created to accomplish a purpose that each of the participants needs or wants but that none of the participants alone or in smaller groups can attain as well or at all. The common goal is necessary so that each participant focuses on a single set of criteria, and the mutual dependency keeps the coalition together and lays the ground work for mutual respect.

Clear, concrete, achievable, specific goals. A collaboration must identify common, clear, achievable, and specific goals, both short-term process goals and long-term outcomes, plus a plan with concrete and realistic steps and a time schedule for achieving them.

Selection of participants. The collaboration should bring together at the outset all individuals and/or organizations that are necessary to accomplish the goals, even though the contribution of some may not be necessary until the final stage. Collaborations may be small, consisting of a scholar and a service agency (e.g., school, hospital, early childhood center), or much larger, involving several organizations of the same or different kinds (e.g., family support centers, early childhood centers, schools, public and private funders, elected officials, media representatives). Early and continuous involvement of key players (e.g., policy makers) creates feelings of ownership and loyalty that may be needed later.

Team players. Participants must be selected for the resources and influence they may bring as well as for several personal characteristics, including the ability to get along with diverse people and groups, attend every meeting, represent their organization and commit its resources to the project, listen and understand divergent points of view as well as communicate clearly and honestly their own perspective, and accept a group decision even if it conflicts with their own self-interest.

Diversity. The diversity of participants that is needed can present challenges. Some participants may hold stereotypic attitudes towards others, but such beliefs do not always apply to the particular individuals and organizations represented in a collaboration. Ultimately, the members of each group need to learn the skills that each participant can contribute regardless of stereotypic roles.

Strong, balanced, sensitive leadership. Although all of the participants of a collaboration are necessary and should share in the rights, responsibilities, and credit for the collaboration's activities and products, strong leadership is essential for any collaboration to function smoothly, efficiently, and productively. The leader must be neither dictatorial nor a benign chairperson; instead, he or she must be strong enough to structure the process and activities of the collaboration, move the group along the appropriate direction, encourage and discourage participants to be appropriate to the group's process and goals, be sensitive to the needs and characteristics of each member without unnecessary accommodation, recognize and respect the contributions of each participant, and deal in a balanced and fair manner with disagreements and even conflicts within the group. It is a tall order; few people, especially scholars, excel in these characteristics; and occasionally collaborations must change leadership if necessary to preserve the collaboration and to achieve their goals.

Based upon Butterfoss, Goodman, & Wandersman, 1996; Grobe, Curnan, Melchior, & The Center for Human Resources, 1993; Kegerise, 1999; Mattessich & Monsey, 1992; Wandersman & Goodman, 1993. Table originally published in Groark and McCall (2005); reproduced with permission of the publisher.