

ASPE RESEARCH BRIEF

OFFICE OF THE ASSISTANT SECRETARY FOR PLANNING AND EVALUATION
OFFICE OF HUMAN SERVICES POLICY - U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES

CORE INTERVENTION COMPONENTS: IDENTIFYING AND OPERATIONALIZING WHAT MAKES PROGRAMS WORK

Executive Summary

Rather than being based on hunches and best guesses, intervention programs are increasingly expected to be evidence-based. However, when evidence-based programs are replicated or scaled up, it is critical not only to know whether a program works, but which program elements are essential in making the program successful. To date, though, few programs have had hard data about which program features are critical “core components” and which features can be adapted without jeopardizing outcomes.

What information is needed to select and implement programs that address the needs of identified populations? “Core components” include the functions or principles and related activities necessary to achieve outcomes. Strategies for a well-operationalized program include a clear description of: the context of the program; the core components; the active ingredients to operationally define the core components so they can be taught and learned and can be implemented in typical settings; and a practical strategy for assessing the behaviors and practices that reflect the program’s values and principles, as well as the program’s active ingredients and activities. Also, when outcomes are not achieved, an understanding of core components and whether they were implemented correctly is essential to understanding whether a program is ineffective, or alternatively, whether it was not implemented well.

Key Take-Away Messages

- Usability testing research that identifies, measures, and tests the efficacy of program core components or “active ingredients” can improve our understanding of which program elements are essential for evidence-based

ABOUT THIS RESEARCH BRIEF

This research brief was written by Karen Blase, PhD, and Dean Fixsen, PhD, of the National Implementation Research Network and the Frank Porter Graham Child Development Institute at the University of North Carolina at Chapel Hill.

In 2010, ASPE awarded Child Trends a contract for the project “Emphasizing Evidence-Based Programs for Children and Youth: An Examination of Policy Issues and Practice Dilemmas Across Federal Initiatives”. This contract was designed to assemble the latest thinking and knowledge on implementing evidence-based programs and developing evidence-informed approaches. This project has explored the challenges confronting stakeholders involved in the replication and scale-up of evidence-based programs and the issues around implementing evidence-informed strategies. Staff from ASPE’s Division of Children and Youth Policy oversaw the project.

As part of this contract, three research briefs have been developed that focus on critical implementation considerations.

Office of the Assistant Secretary for
Planning and Evaluation

Office of Human
Services Policy

US Department of Health
and Human Services

Washington, DC 20201



programs and practices to produce desired outcomes.

- Program funders should consider including requirements to specify the core components of interventions as deliverables at the end of a demonstration or pilot phase to facilitate replication and scalability.
- Decision-makers seeking to select and validate intervention might ask program developers for a description of intervention's core components, the rationales underlying each core component, fidelity measures, and measures of processes and outcomes.
- Policymakers should require that evidence-based program implementations include plans for defining, operationalizing and validating core components to ensure alignment with desired outcomes, and ongoing assessments of fidelity in delivering the core components to maintain and improve outcomes over time.
- Program developers should consider monitoring the potential social and participant-level costs when core components are missing or not clearly articulated to understand why developing core components is a sound, efficient, and strategic approach to achieving positive outcomes.

Since issues related to the core components of interventions are relevant to producing new knowledge about what works and for moving science to practice in socially significant ways, this brief is relevant for a range of professionals and stakeholders, including program developers, researchers, implementers, and policy makers.

CORE INTERVENTION COMPONENTS: IDENTIFYING AND OPERATIONALIZING WHAT MAKES PROGRAMS WORK

Purpose

This brief is part of a series that explores key implementation considerations. It focuses on the importance of identifying, operationalizing, and implementing the “core components” of evidence-based and evidence-informed interventions that likely are critical to producing positive outcomes. The brief offers a definition of “core components”, discusses challenges and processes related to identifying and validating them, highlights rationales for the importance of operationalizing core components, and explores implications for selecting, funding, implementing, scaling up, and evaluating programs. Since the issues related to core components of interventions are relevant to producing new knowledge about what works and for moving science to service in socially significant ways, this brief is relevant for a range of professionals and stakeholders, including program developers, researchers, implementers, and policy makers.

Background and Introduction

Increasingly, agencies, communities, and funders are driven to make a difference by using the best information that social science has to offer. Also, service recipients and communities are becoming increasingly savvy about asking for the data that demonstrate that programs or practices are likely to result in positive outcomes. Given the effort, time, and expense required to establish and sustain services and interventions, the return on this investment matters deeply for all stakeholders.

But what information is needed to select and implement programs that address the needs of identified populations? What data matter most? Can outcome data tell the whole story? Increasingly, researchers, evaluators, and program developers are discussing the importance of identifying the core components of complex interventions. Those who use data to make decisions (e.g., grant makers, foundations, policy makers, agency directors, and intermediary organizations) are interested in understanding which program or practice elements are “essential” and which ones can be modified without jeopardizing outcomes.

EMPHASIS ON EVIDENCE

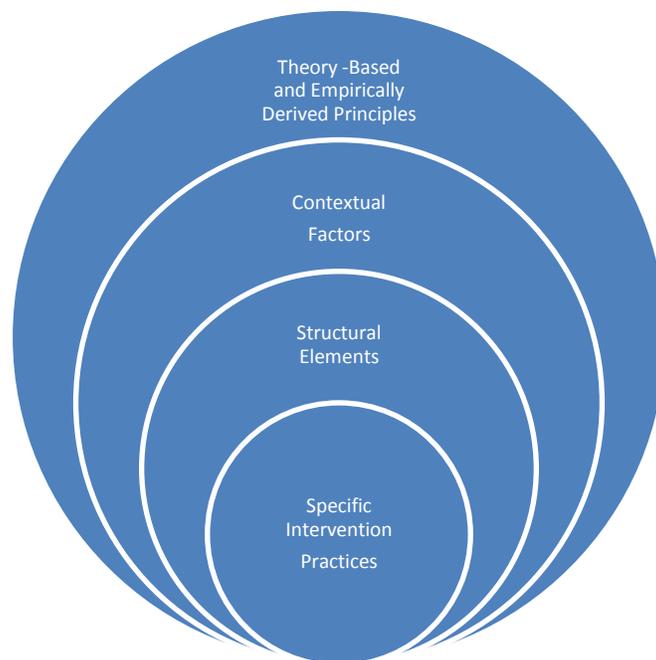
The Federal government has made a strong commitment to supporting evidence-based and evidence-informed programs, particularly for children and youth. Recent examples include: the Maternal, Infant, and Early Childhood Home Visiting Program, the Teen Pregnancy Prevention Initiative, the Permanency Innovations Initiative, the Social Innovation Fund, and the Investing in Innovation (i3) Fund.

What Do We Mean By “Core Components”?

For the purposes of this brief, we use the term core components to refer to the essential functions or principles, and associated elements and intervention activities (e.g., active ingredients, behavioral kernels; Embry, 2004) that are judged necessary to produce desired outcomes. Core components are directly related to a program’s theory of change, which proposes the mechanisms by which an intervention or program works. The core components are intended to, or have been demonstrated through research to, positively impact the proximal outcomes that address the identified needs and that increase the likelihood that longer-term outcomes will be achieved. In short, the core components are the features that define an effective program.

Core components can be cast as theory-driven, empirically derived principles and then further operationalized as the contextual factors, structural elements, and intervention practices that are aligned with these principles. For example Multi-Systemic Therapy details nine such principles, such as, “Interventions should be present-focused and action-oriented, targeting specific and well-defined problems” (Henggeler, Schoenwald, Liao, Letourneau & Edwards, 2002, p. 157). Multidimensional Treatment Foster Care articulates four such principles, such as, “providing the youth with a consistent reinforcing environment where he or she is mentored and encouraged” (Chamberlain, P., 2003, p. 304). Incredible Years posits social learning principles as core elements of the various school and parent training programs (e.g., Webster-Stratton & Herman, 2010). Sexton and Alexander (2002), surveyed the qualitative and meta-analytic reviews of research related to family-based interventions in the context of Principles of Empirically Supported Interventions (PESI), describing how such empirically derived principles can aid in identifying and developing effective treatment approaches as well as research agendas. Core components, cast as principles, inform the specification of contextual aspects of the interventions (e.g., interventions occur in schools or communities, parent and community involvement, interventions occur in families’ homes), structural elements (e.g., a low adult/child ratio, the required number and sequence of sessions), and specific intervention practices (e.g., teaching problem-solving and communication skills, practicing social skills, reinforcing appropriate behavior).

Figure 1. Core Components – From Principles to Practices



Challenges in Identifying and Validating Core Components

The core components may be developed over time by experimentally testing a theory of change (e.g., what are the mechanisms by which we expect change to occur) and by developing and

validating fidelity measures (e.g., was the intervention done as intended) that reflect the core components. Core components can be identified through causal research designs (e.g., randomized control trials, quasi-experimental designs, single-subject designs) that test the degree to which core components produce positive outcomes, as compared to results that occur in the absence of these core components. Research that demonstrates a positive correlation between high fidelity and better outcomes also increases our confidence in and understanding of the core components (e.g., higher fidelity is associated with better outcomes). However, causality cannot be inferred from such correlational research.

Core components are often equated with measures of fidelity; but such measures do not necessarily tell the whole story about what is required for effective use of an intervention in typical service settings. Moreover, identifying and validating core components through the creation of valid, reliable, and practical measures of fidelity is not a simple task. It requires research over time and across replications. Efforts to create, test, and refine fidelity measures have been conducted for programs for children and families (Schoenwald, Chapman, Sheidow & Carter, 2009; Henggeler, Pickrel, & Brondino, 1999; Bruns, Burchard, Suter, Force, & Leverentz-Brady, 2004; Forgatch, Patterson, & DeGarmo, 2005) and for programs serving adults (Bond, Salyers, Rollins, Rapp, & Zippel, 2004; Propst, 1992; Lucca, 2000; Mowbray, Holter, Teague, & Bybee, 2003; McGrew & Griss, 2005). These studies chronicle the challenges of creating fidelity measures that not only reflect the core components, but also are practical to use in typical service settings and are good predictors of socially important outcomes. Concerted effort over time by teams of researchers seems to be required to produce valid and serviceable assessments of fidelity.

While teams of researchers have successfully taken on the task of better articulating and validating the core components of some programs (Henggeler et al., 2002; Chamberlain, 2003; Forgatch et al., 2005; Webster-Stratton & Herman, 2009), on the whole, there are few adequately defined programs in the research literature that clearly detail the core components with recommendations on the dosage, strength, and adherence required to produce positive outcomes. The source of this problem has been documented by Dane and Schneider (1998). These authors summarized reviews of over 1,200 outcome studies and found that investigators assessed the presence or strength of the independent variables (the core intervention components) in only about 20 percent of the studies, and only about 5 percent of the studies used those assessments in their analyses of the outcome data. A review by Durlak and DuPre (2008) drew similar conclusions. The challenge is further exacerbated by the lack of commonly accepted definitions or criteria related to the verification of the presence or validation of the independent variables (the core components that define the program) in gold standard, randomized control studies. This means that the published research literature is likely a poor source of information about the functional core components of interventions, evidence-based, or otherwise.

One reason that very few program evaluations are able to actually research which components of the program are most strongly related to positive outcomes is that, in demonstration projects, extra efforts are made to insure that the program is implemented with fidelity, thus eliminating variations. An exception to this occurred in the early research on the Teen Outreach Program (TOP) where some variations in program implementation did occur because the program did not yet have “minimum standards” and site facilitators took liberties with the curriculum and

volunteer service components of the program (Allen, Kuperminc, Philliber and Herre, 1994, Allen et al., 1990). Variations in facilitator “style” also occurred, and data were collected on how interactions with facilitators and others were perceived by students.

The Teen Outreach research found that presence of volunteer community service was related to positive outcomes including less failure of school courses, lower rates of teen pregnancy, and lower rates of school suspension. On the other hand, variations in the amount of classroom time and exact fidelity to the curriculum were not related to these outcomes. This research also found that, when students said they had a great deal of input in selecting the volunteer work they would do and that this work was truly important, they had more positive outcomes (Allen, Philliber, Herrling, and Kuperminc, 1997). After this research was completed, the TOP adopted minimum standards for replication of TOP including 20 hours of community service and choice regarding volunteer work viewed as important by the teen. TOP also requires 25 curriculum sessions, but program facilitators can use any of the curriculum sessions they choose. In communities where teaching about sex is prohibited or restricted, this left facilitators free to leave out those lessons since their inclusion had not been shown to affect outcomes. In addition, training for TOP facilitators stresses that this is a curriculum to truly be facilitated rather than taught, and that, at the end of the program, young people should report that they did most of the talking. Fidelity data for TOP currently include measures of each of these important core components derived from examining variations in program practices and protocols as related to outcomes.

With such exceptions, there is, as noted by Dane and Schneider (1998) and Michie, Fixsen, Grimshaw, and Eccles (2009), little empirical evidence to support assertions that the components named by an evidence-based program developer are, in fact, the functional, or only functional, core components necessary for producing the outcomes. In their examination of intervention research studies, Jensen and his colleagues (Jensen, Weersing, Hoagwood, & Goldman, 2005) concluded, “when positive effects were found, few studies systematically explored whether the presumed active therapeutic ingredients actually accounted for the degree of change, nor did they often address plausible alternative explanations, such as nonspecific therapeutic factors of positive expectancies, therapeutic alliance, or attention” (p 53). This may mean that the mention or failure to mention certain components by a program developer or researcher should not be confused with their function or their lack of function in the producing hoped for outcomes in the intervention settings.

Thus, the current literature regarding evidence-based programs heavily focuses on the quality and quantity of the “evidence” of impacts. And the vetting of research design, rigor, number of studies, and outcomes has resulted in rosters of evidence-based programs such as SAMHSA’s National Registry of Evidence-Based Programs and Practices (<http://nrepp.samhsa.gov>), Blueprints for Violence Prevention (<http://www.colorado.edu/cspv/blueprints/>) with various criteria and rankings (e.g., evidence-based, evidence-informed, and promising) based on reviews of the research literature. A resource from “What Works Wisconsin” (Huser, Cooney, Small, O’Connor, & Mather, 2009) provides brief descriptions of 14 such registries (<http://whatworks.uwex.edu/attachment/EBRegistriesAug-2009.pdf>) covering a range of areas including substance abuse and violence prevention as well as the promotion of positive outcomes such as school success and emotional and social competence. The identification of programs and practices that “work” and assessing the quality and quantity of the evidence are important for

building confidence about outcomes. We need to understand the rigor and the outcomes of the research because we need to invest in “what works”. But we also need to define and understand the core components that make the “what” work.

Operationalizing Programs and Their Core Components

Defining a program and its core components matters because practitioners do not use “experimental rigor” in their interactions with those they serve; they use programs. Thus, the lack of adequately defined programs with well-operationalized core components is an impediment to implementation with good outcomes (Hall & Hord, 2006). Since the research literature, with the predominant focus on rigor and outcomes, is not yet a good source for defining programs and the attendant core components, what processes can help? And what defines a well-operationalized program?

To be useful in a real-world service setting, any new program, intervention, or innovation, whether evidence-based or evidence-informed, should meet the criteria below. When the researcher and/or program developer has not specified these elements, then funders, policy makers, and implementing agencies, with the guidance of researchers and program developers, will need to work together to do so. This means allowing the time and allocating the resources for this important work to occur before and during initial implementation of the innovation as it moves from research trials into typical service settings.

With the use of evidence-based and evidence-informed innovations in mind, we propose that the following elements comprise a well-operationalized program including the core components:

- Clear description of the context for the program.
 - This means that the philosophical principles and values that undergird the program are clearly articulated. Such principles and values (e.g., families are the experts about their children, children with disabilities have a right to participate in community and school life, culture matters, all families have strengths) provide guidance for intervention decisions, for program development decisions, and for evaluation plans. If they are a “lived” set of principles and values, they promote consistency and integrity; and they serve as a decision-making guide when the ‘next right steps’ with a child or family are complex or unclear, even when the core components are well-operationalized.
 - The context of the program also includes a clear definition of the population for whom the program is intended. Without clear inclusion and exclusion criteria and the willingness to apply these criteria, the core components will be applied inappropriately or will not even be applicable.
- Clear description of the core components. These are the essential functions and principles that define the program and are judged as being necessary to produce outcomes in a typical service setting (e.g., use of modeling, practice, and feedback to acquire parenting skills, acquisition of social skills, and participation in positive recreation and community activities with non-deviant peers).
- Description of the active ingredients that further operationally define the core components.

- One format and process for specifying the active ingredients associated with each core component involves the development of practice profiles. Practice profiles are referred to as innovation configurations in the field of education (Hall & Hord, 2011). In the context of a practice profile, the active ingredients are specified well enough to allow them to be teachable, learnable, and doable in typical service settings. Well-written practice profiles help promote consistent expectations across staff.
- A practical assessment of the performance of the practitioners who are delivering the program and its associated core components.
 - The performance assessment relates to identifying behaviors and practices that reflect the program philosophy, values, and principles embodied in the core components, as well as the active ingredients/activities associated with each core component and specified in the practice profiles. Assessments are practical and can be done routinely in the context of typical service systems as a measure of how robustly the core components are being utilized.
 - A useful performance assessment may comprise some or all of the fidelity assessment process, and across practitioners should be highly correlated with intended outcomes. Over time the researchers, evaluators, and program developers can correlate these performance assessment measures with outcomes to determine how reliable and valid they are. When higher fidelity is associated with better outcomes, there is growing evidence that the program is more effective when used as intended and that the assessment may indeed be tapping into core components.

Usability Testing to Operationalize and Validate Core Components

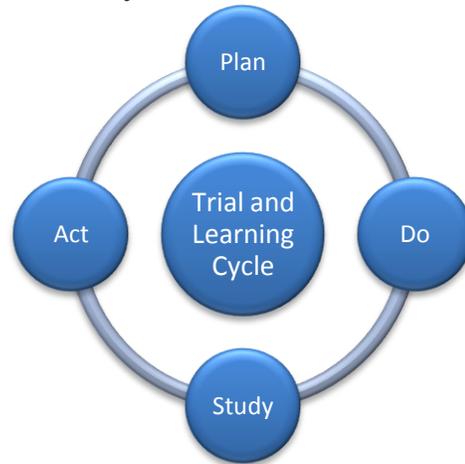
Researchers and program developers should provide information to enable agencies to support practitioners to implement a program with fidelity. The vast majority of programs (evidence-based or evidence-informed), as noted earlier, do not meet these criteria. For the evidence-based and evidence-informed interventions that have not been well-operationalized (Hall & Hord, 2011), there is a need to employ usability testing to verify or elaborate on the program's core components and the active ingredients associated with each core component before proceeding with broader scale implementation.

What is usability testing? Usability testing is an efficient and effective method for gaining the experience and information needed to better operationalize a program and its core components. Usability testing methods were developed by computer scientists as a way to de-bug and improve complex software programs or websites. Usability testing (e.g., Nielsen, 2005) employs a small number of participants for the first trial, assesses results immediately, makes corrections based on those results, and plans and executes the next, hopefully improved, version of the core component and its associated active ingredients. This cyclical process is repeated (say, 5 times with 4 participants in each iteration for a total $N = 20$) until the program is producing credible proximal or short-term outcomes related to the tested core components and the associated active ingredients.

Usability testing is an example of the Plan, Do, Study, Act (PDSA) cycle (e.g., Shewhart, 1931; Deming, 1986). The benefits of the PDSA cycle in highly interactive environments have been verified through evaluations across many domains including manufacturing, health, and

substance abuse treatment. This “trial and learning” approach allows developers of complex programs and those charged with implementing them to identify the core components and active ingredients of a program and further evaluate, improve, or discard non-essential components. Usability testing often is done in partnership with the program developers, researchers, and early implementers of the program.

Figure 2. Plan, Act, Do and Study



An example of usability testing may provide more clarity about the utility of such an approach. A home-based intervention for parents whose children have just been removed due to child welfare concerns might include a small test of the degree to which the core component of ‘engagement’ and the associated active ingredients during the initial visit (e.g., the therapist expresses empathy, asks parents to identify family and child strengths, allows the parents to tell their story) are associated with parents’ willingness to engage with the therapist.

Measures of engagement might include the number of times the family is at home at the scheduled time for visits and the number of sessions in which the parents agree to participate in parent training activities guided by the therapist. Such information can be collected very efficiently from supervisors and/or therapists. The results from the first cohort of trained therapists as they interact with families might then be assessed after three visits are scheduled and therapeutic interventions are attempted during each visit. The data, both process and outcome, are then reviewed and, if the a priori criteria are met (e.g., 75 percent of families allow the therapists into their homes for all three visits; 80 percent of families participate in the parent training activities), the same engagement processes are continued by new therapists with new families and by current therapists with subsequent families. Or, if results are not favorable, improvements in engagement strategies are made and operationally defined. Changes are made in the protocol and the process begins again. That is, new and current therapists receive additional material, training, and coaching regarding revised engagement strategies to be used during initial interactions. The revised engagement process is then tried with a second cohort, again including proximal measures of engagement. Such usability testing may occur throughout the implementation of a new program. Some program components may not occur until later in the course of the intervention (e.g., procedures related to reintegration of children into their families).

Because this is a new way of work, there can be concerns related to the cost and feasibility of usability testing. While effort and organization are required, this is not a research project but a ‘testing’ event that can be managed efficiently and effectively. The costs of maintaining program elements that are not feasible or are not yielding reasonable proximal results can be far greater than costs associated with making time to target key elements for usability testing. And, while not all core components are amenable to such testing (e.g., use of x# of sessions), this “trial and learning” process does create the opportunity to efficiently refine and improve important elements of the “program” and/or its core components and active ingredients with each iteration. Each small group is testing a new and, hopefully, improved version. Each iteration results in incremental improvements and specificity until the outcomes indicate that the program or the tested set of core components is ready to be used more systematically or on a larger scale and is ready to undergo more formal fidelity assessment and validation.

Why Is It Important to Identify Core Components?

The lack of description and specification of the core components of programs presents challenges when it comes to assessing whether or not a given program has been or can be successfully implemented, effectively evaluated, improved over time, and subsequently scaled up if results are promising. This means that, when agencies and funders promote or require the use of evidence-based programs that are not well-operationalized, and agencies and practitioners are recruited to engage in new ways of work, there can be a great deal of discussion and confusion about just what the “it” is, that must be implemented to produce the hoped for outcomes.

Benefits of increased attention to the definition and measurement of core components and their associated active ingredients include an:

- Increased ability to focus often scarce implementation resources and supports (e.g., resources for staff recruitment and selection, training, coaching, fidelity monitoring) on the right variables (e.g., the core components) to make a difference.
- Increased likelihood of accurately interpreting outcomes and then engaging in effective program improvement strategies that address the “right” challenges.
- Increased ability to make adaptations that improve fit and community acceptance, without moving into the “zone of drastic mutation” (Bauman, Stein, & Ireys, 1991).
- Increased ability to engage in replication and scale-up while avoiding program ‘drift’ that can lead to poor outcomes.
- Increased ability to build coherent theory and practice as common core components emerge that are associated with positive outcomes across diverse settings and/or programs.

These benefits are elaborated below.

Application of implementation supports to ensure and improve the use of core components.

When core components are more clearly defined, implementation supports can be targeted to ensure that the core components and their active ingredients come to life as they are used in everyday service settings. As noted in the in-home services example above, the usability testing

approach not only allows for repeated assessments and improvements in the intervention, but it also creates opportunities for improving the implementation supports – the “execution” part of usability testing. That is, each round of improvement allows for adjustments in implementation supports such as training, coaching, and the performance assessment process itself (e.g., did we execute these activities as intended?), as well as serving as fodder for further defining and the core components and active ingredients themselves.

As noted above, usability testing is a variant of the Plan, Do, Study, Act process (PDSA). PSDA cycles and implementation supports are typically rapid cycle processes to ensure that you are getting proximal outcomes. When applying a usability testing process to an incompletely operationalized evidence-based program or to an evidence-informed program, the “plan” can be to test a segment of the program or test one or more core components as intended to be used in practice. To carry out the “do” part of the PDSA cycle, the “plan” needs to be operationalized and grounded in best evidence. That is, who will say or do what activities, with whom, under what conditions to enact the plan? And to what degree are these core components and/or active ingredients supported by evaluation and research findings? This attention to the “plan” compels attention to the core components and active ingredients.

The “do” part of the PDSA cycle provides an opportunity to specify the implementation supports required to enact the plan. How will the confidence and competence of practitioners to “do” the plan be ensured? This requires attention to the implementation supports; such as the recruitment and selection criteria for staff, as well as training and coaching processes (e.g., who is most likely to be able to engage in these activities; what skill-based training is needed to “do” the “plan”; how will coaching be provided to improve practitioners’ skills and judgment as they execute the “plan?”). And the “study” portion of the PDSA cycle requires creating an assessment of performance (e.g., did practitioners “do” the plan? were our implementation supports sufficient to change practitioner behavior?), as well as the collection of proximal or near-term outcomes (e.g., were parents at home? were parents willing to engage in practice sessions with the therapist?).

As three or four newly trained staff begin providing the new services, the budding performance assessment measures can be used to interpret the immediate outcomes in the “study” part of the PDSA cycle (e.g., did we do what we intended?; if so, did doing what we intended to do result in desired proximal outcomes?). Without proximal outcomes, distal outcomes are much less likely. Once the results from the “study” segment of the cycle are known (e.g., from performance assessment data and outcomes for participants), work can commence to “act” on this information by making adjustments to segments of the program, the core components, and/or to particular active ingredients. And further action can be taken as implementation supports are adjusted for the next group of staff as the usability testing cycle begins (e.g., Fixsen, Blase, Timbers, & Wolf, 2001; Wolf, Kirigin, Fixsen, Blase, & Braukmann, 1995). The ‘act’ portion of the cycle defines the improvements to be made and initiates a new PDSA cycle related to the improvements related to training, coaching, and feedback systems to improve practitioner competence and confidence and adherence to the new, revised processes.

These brief descriptions of usability testing and implementation supports have focused on identifying and developing the core components for an initial effective working example of an

evidence-informed innovation or to develop an improved definition of an evidence-based program that has not been well operationalized. But even a well-operationalized intervention will continue to evolve as new situations are encountered and more lessons are learned about how to better operationalize core components, improve implementation supports, improve fidelity, and improve outcomes. The goal is not to “do the same thing” no matter what, just for the sake of “doing the program”. The goal is to reliably produce significant benefits, with better outcomes over time and to clearly identify, understand and skillfully employ the core components that are associated with better outcomes.

Interpreting Outcomes and Improving Programs. Identifying the core components that help to create positive outcomes, and knowing whether or not they were implemented with fidelity, greatly improves the ability to interpret outcomes (Fixsen, Naoom, Blase, Friedman, & Wallace, 2005). In addition, it reduces the likelihood of “throwing the baby (the new program) out with the bath water (poor implementation)”. Without understanding and monitoring the use of the core components, it is difficult to tell the difference between an implementation problem and an effectiveness problem. This is particularly problematic when positive outcomes are not achieved or when outcomes were not as beneficial as expected. Because strategies for improving effectiveness are different from strategies for improving fidelity and the implementation of the core components, it is important to be able to assess whether the program does not work or whether the implementation of the program was flawed. The following table illustrates the types of improvement strategies or subsequent actions that may be useful depending on the where one may “land” with respect to fidelity and outcomes.

Table 1. Analyzing data related to both fidelity assessments and outcomes helps to identify the actions needed to improve outcomes.

	High Fidelity	Low Fidelity
Satisfactory Outcomes	Continue to monitor fidelity and outcomes Consider scale-up	Re-examine the intervention Modify the fidelity assessment
Unsatisfactory Outcomes	Select a different intervention Modify the current intervention	Improve implementation supports to boost fidelity

Obviously, we want our efforts to “land” in the quadrants that involve achieving satisfactory outcomes. When there are satisfactory outcomes and fidelity is high, then continued monitoring of fidelity and outcomes helps to ensure that the core components are continuing to be used with good effect. And it may indicate that the program should be reviewed for scalability or increased reach since the core components appear to be well-operationalized and the implementation supports seem to be effective in producing high fidelity.

When satisfactory proximal and/or distal outcomes are being achieved but fidelity is low or lower than expected, it may require re-examining the intervention to determine if there are additional core components or active ingredients that have not been specified. This requires qualitative and quantitative data collection and analysis of the strategies used by practitioners

who are positive outliers (e.g., achieving good results but with low fidelity). Or it may be that the context for the program has changed. For example, there has been a change in the population (e.g., different presenting problems, different age-range), leading to the need to provide very different program strategies to meet the needs of the population. Or it may be that there are core components and active ingredients that are well operationalized and are being trained and coached, but are currently not included in the fidelity assessment process. In such cases, revising and re-testing the fidelity assessment may be called for. In any event, discovering the source of this discrepancy will be important if the program is to be sustained over time or scaled-up.

The combination of high fidelity but unsatisfactory outcomes may indicate that the selected intervention or prevention program is not appropriate for the population or does not address critical needs of the population. Since the purpose of using programs is to achieve positive results, the achievement of high fidelity with poor outcomes produces no value to the population in need. Such findings may help build theory and set future research agendas and hopefully would result in communities choosing to invest resources differently. Once unmet needs are identified through data gathering and analysis, it may be possible to modify the intervention and add in core components that have theory and evidence to support their impact on the unmet needs. Or it may be that the selection process for the intervention was flawed or that the population being served is different from the population identified during the original needs assessment. In any event, the search for programs with core components that address the needs of the population may need to be re-initiated.

If outcomes are unsatisfactory and fidelity is low, then the first approach is to improve or modify implementation supports (e.g., more targeted staff recruitment and selection, increased skill-based training, and frequency and type of coaching) in order to improve fidelity (Webster-Stratton, Reinke, Herman, & Newcomer, in press). Or it may be necessary to review the organizational factors (e.g., time allocated for coaching, access to equipment needed for the intervention) and systemic constraints (e.g., licensure requirements that limit recruitment, billing constraints, inappropriate referrals) that may be making it difficult to achieve high fidelity. Making changes to address organizational and/or systems issues (e.g., funding, licensure, billing) often require considerable time and effort. Therefore, the implementation supports of selection, training, and coaching may need to ‘compensate’ for the organizational and systems barriers (Fixsen et al., 2005). For example, it may take time to address the funding constraints that make it difficult to fund coaching of staff. While attempts to fund the coaching are being pursued, it may be necessary to use more rigorous selection criteria to recruit more experienced staff or provide increased training to ‘compensate’ for the impact of funding on the provision of coaching.

In summary, this table brings home the point that both knowledge and measurement of the presence and strength of the core components (e.g., through fidelity and other measures), are required to interpret and respond to outcome data.

Making adaptations to improve “fit” and community acceptance. There may be a variety of reasons that adaptations are considered by communities and agencies as they implement programs and innovations. There may be a perceived or documented need to attend to cultural or linguistic appropriateness or community values (Backer, 2001; Castro, Barrera, & Martinez,

2004). Or there may be resource or contextual constraints that result in decisions to adapt the program or practices. Perhaps the workforce available to implement the program influences making programmatic adaptations that are perceived to be better aligned with the background, experience, and competencies of the workforce.

Adapting evidence-based programs and evidence-informed innovations may make it more likely that communities will make the decision to adopt such programs and innovations (Rogers, 1995). However, improving the likelihood of the decision to adopt through adaptation does not necessarily mean that those adaptations in the service settings will help to produce positive outcomes. While some initial adaptations are logical (e.g., translation to the language of the population, use of culturally appropriate metaphors), recommendations by some program developers are to first do the program as intended and assess both fidelity and outcomes. Then, based on data, work with program developers and researchers to make functional adaptations. Functional adaptations are those that a) reduce “burden” and “cost” without decreasing benefits, or b) improve cultural fit, community acceptability, or practitioner acceptance while maintaining or improving outcomes. Adaptations are much more likely to be functional when the core components and the associated active ingredients are known. In addition, those engaged in adapting programs and practices must understand the underlying theory base, principles, and the functions of the core components so that adaptations do not undermine the effectiveness of the program. Finally, process and outcome data must be collected to validate that the adaptations meet the criteria for “functional”. It then stands to reason that adaptations are most likely to be functional when: the core components are well-operationalized, when implementation supports are able to reliably create competent and confident use of the intervention, and when adaptations are made in partnership with the original program developer and researcher to avoid moving into the “zone of drastic mutation” (Bauman, Stein, & Ireys, 1991) and destroying the effectiveness of the intervention; and when data verify that the changes have not undermined the effectiveness of the program or practice (Lau, 2006).

As program developers and researchers work with diverse communities, cultures, and populations to make adaptations, they may look for ways to change “form” (e.g., time, place, language, metaphors used) to improve appropriateness and acceptability while preserving the “function” (e.g. the processes that relate to effectiveness) of the core components. Collecting data to analyze the impact of making cultural adaptations is key to determining when such adaptations are functional since reducing the dosage of the core components or altering them can result in adaptations that reduce positive outcomes, as noted by Castro et al. (2004). For example, Kumpfer, Alvarado, Smith, and Bellamy (2002) describe a cultural adaptation of the Strengthening Families Program for Asian Pacific Islanders and Hispanic families that added material on cultural and family values but displaced the content related to acquiring behavioral skills – a core component. This resulted in less improvement in parental depression and parenting skills, as well as less improvement in child behavior problems than the original version, which focused only on behavioral skills.

Cultural adaptations can be made that enhance acceptability but that do not undermine the core components and active ingredients of the evidence-based program. A cultural adaptation of Parent Child Interaction Therapy (PCIT) was developed by McCabe and her colleagues (McCabe, Yeh, Garland, Lau, & Chavez, 2005). They made modifications to the core

component of engagement by including engagement protocols for immediate and extended family members to reduce the likelihood of lack of support undermining treatment. They also “tailored” the manner that certain active ingredients were framed when the results of a parent self-report questionnaire detected elements at odds with parenting beliefs. For example, for parents who expressed a commitment to strict discipline, the active ingredient of “time out” was re-framed as a punitive practice by using terms such as “punishment chair” for the time out location. Or if Mexican American parents of young children expressed concerns about the practice being too punishing for young children, the term “thinking chair” was adopted. This left the function of the time out process intact (e.g., brief removal from positive reinforcement) while tailoring or adapting the form to fit the cultural and familial norms.

Lau (2006) makes the case for selective and directed cultural adaptations that prioritize the use of data to identify communities or populations who would benefit from adaptations and are based on evidence of a poor fit. Lau makes the case for focusing on high priority adaptations that avoid fidelity drift in the name of cultural competence. In short, the process of adaptation needs to be based on empirical data and demonstrate benefits to the community or population.

In summary, modifications to core components must be done thoughtfully and in partnership with program developers and researchers, so that the underlying theory-base of the program is not inadvertently undermined. Data-based decision-making should guide modifications to core components. Linguistic adaptations aside, an implementation process that first implements the core components as intended and then analyzes results may be better positioned to make functional adaptations. Functional adaptations are those that are developed in order to improve fit, acceptability, and/or reduce burden or cost while improving outcomes or at least maintaining positive outcomes while avoiding negative impact.

Improving the success of replication and scale-up efforts . As David Olds (2002) noted, “Even when communities choose to develop programs based on models with good scientific evidence, such programs run the risk of being watered down in the process of being scaled up” (p. 168). Of course, understanding whether or not a program has been “watered down” requires an understanding of the core components and their relationship to achieving desired outcomes. Michie et al. (2009) noted that clear definitions of the required core components increase the likelihood that programs and practices can be successfully introduced in communities and scaled-up over time. However, it takes time and a number of closely controlled and monitored replication efforts by the developers to first stabilize the intervention before making the decision to attempt to more broadly scale-up the program. From the business arena, Winter and Szulanski (2001) note that, “The formula or business model, far from being a quantum of information that is revealed in a flash, is typically a complex set of interdependent routines that is discovered, adjusted, and fine-tuned by ‘doing’” (p. 371). Such fine-tuning can be done through usability testing, evaluation, and research. Scaling up too soon can lead to a lost opportunity to adequately develop, specify, and reliably produce the core components that lead to effectiveness.

Successful replication and scale-up are significantly enhanced when the core components are well specified and when effective implementation supports are in place to promote the competency and confidence of practitioners, and when organizational and systems change occurs to support the new way of work. Effectiveness and efficiency of replication and scale-up also

may be improved when there is greater clarity about the non-core components that can be adapted to fit the local circumstances including culture, language, workforce, and economic or political realities. And, as noted above, efficiency is enhanced when resources for implementation supports (e.g., training, coaching, data systems, fidelity measurement and reporting) are targeted to impact core components.

Implications

Given the importance of identifying core components, what are the implications for research agendas, program development, funding of service initiatives, and policy making, as well as for implementation in typical service settings. Research that focuses on operationalizing, measuring, and testing the efficacy of the independent variables (e.g., the core components) would improve our understanding of ‘what works’ and what is necessary for evidence-based programs and practices to produce outcomes. At present, research standards, publication constraints, and journal requirements for publication do not significantly support or encourage such detailed attention to core components. Michie et al. (2009) argue that, “If a more explicitly theoretical approach to deciding how to design and report interventions were taken, it may be that more effects may be revealed and more understanding of their functional mechanisms gleaned...promoting the understanding of causal mechanisms that both enrich theory and facilitate the development of more effective interventions”. They also argue that the use of the web for publishing allows for the publication of detailed intervention protocols, which would further improve the identification, operationalization, and testing of core components.

Funders of demonstration programs and pilots can further support the development of and attention to core components by including requirements to specify the underlying theoretical bases and the definition of interventions as deliverables at the end of the demonstration or pilot phase. Such attention then might see demonstrations and pilots as the launching pad for replications and scalability and the first in a series of development steps rather than islands of excellence that come and go.

As communities, agencies, and government entities turn to evidence-based programs and practices and evidence-informed innovations to address specific needs, they, too, can promote increased attention to the importance of well-defined core components. By asking program developers about fidelity measures, research related to fidelity measures, the rationales for core components, and the description of intervention core components, they can discern which programs and practices are more likely to be ready for use in their communities. In addition, they serve notice to program developers who intend to be purveyors (Fixsen et al., 2005) that such information may be an important deciding factor when communities and agencies select interventions or prevention programs.

Similarly, policy makers need to be aware that providing funding for evidence-based programs and practices needs to be coupled with attention to the degree to which such programs’ core components are defined, operationalized, and validated. Failure to have both identified program models and well-specified core components can lead to significant implementation and sustainability challenges. If the intervention is poorly specified and performance assessment (fidelity) measures do not target functional core components, then achieving outcomes may not be realistic. Similarly, policy makers need to support and require continued attention to fidelity

and outcome assessments in order to maintain and improve service outcomes over time and across practitioners and leadership changes. This requires funding for the infrastructure needed to collect and use data. Resources to collect, analyze and interpret data are as important as the skills of the practitioner for achieving, interpreting, and improving outcomes. At present, core components and research done related to them are not targeted by registries, or clearly cited; and in some cases, you can even find research studies related to components, but not in a systematic way or highlighted as such. This representation gap requires systematic and sustained attention.

In summary, defining, operationalizing and measuring the presence and strength of core components are important if we are to improve our knowledge about “what works” and understand how to implement with benefits in everyday service settings.

References

- Allen, JP, Kuperminc, G., Philliber S., & Herre, K (1994). Programmatic prevention of adolescent problem behaviors: The role of autonomy, relatedness and volunteer service in the Teen Outreach Program. *American Journal of Community Psychology*, 22, 617-638.
- Allen, JP, Philliber, S., Herrling, S., and Kuperminc, GP. Preventing teen pregnancy and academic failure: Experimental evaluation of a developmentally based approach. *Child Development*, 64, 729-742.
- Backer, T. E. (2001). Finding the balance—Program fidelity and adaptation in substance abuse prevention: A state-of-the art review. Center for Substance Abuse Prevention, Rockville, MD.
- Bauman, L. J., Stein, R. E. K., & Ireys, H. T. (1991). Reinventing fidelity: The transfer of social technology among settings. *American Journal of Community Psychology*, 19, 619-639.
- Bond, G.R., Salyers, M.P., Rollins, A.L., Rapp, C.A., & Zippel, A.M. (2004). How evidence-based practices contribute to community integration. *Community Mental Health Journal*, 40, 569-588.
- Bruns, E. J., Burchard, J. D., Suter, J. C., Force, M. D., & Leverentz-Brady, K. (2004). Assessing fidelity to a community-based treatment for youth: The Wraparound Fidelity Index. *Journal of Emotional and Behavioral Disorders*, 12, 69-79.
- Castro, F.G., Barrera, M., & Martinez, C.R. (2004). The cultural adaptation of prevention interventions: Resolving tensions between fidelity and fit. *Prevention Science*, 5, 41-45.
- Chamberlain, P. (2003). The Oregon Multidimensional Treatment Foster Care Model: Features, outcomes, and progress in dissemination. *Cognitive and Behavioral Practice*, 10, 303-312.
- Dane, A. V., & Schneider, B. H. (1998). Program integrity in primary and early secondary prevention: Are implementation effects out of control? *Clinical Psychology Review*, 18(1), 23-45.
- Deming, W. E. (1986). *Out of the crisis*. Cambridge, MA: MIT Press.
- Durlak, J. A., & DuPre, E. P. (2008). Implementation matters: A review of research on the influence of implementation on program outcomes and the factors affecting implementation. *American Journal of Community Psychology*, 41, 327-350. doi: 10.1007/s10464-008-9165-0
- Embry, D.D. (2004). Community-based prevention using simple, low-cost, evidence-based kernels and behavior vaccines. *Journal of Community Psychology*, 32, 575-591.
- Fixsen, D. L., Blase, K., Duda, M., Naoom, S., & Van Dyke, M. (2010). Implementation of evidence-based treatments for children and adolescents: Research findings and their implications for the future. In J. Weisz & A. Kazdin (Eds.), *Implementation and dissemination: Extending treatments to new populations and new settings* (2nd ed., pp. 435-450). New York: Guilford Press.

- Fixsen, D. L., Naoom, S. F., Blase, K. A., Friedman, R. M., & Wallace, F. (2005). *Implementation research: A synthesis of the literature*. Tampa, FL: University of South Florida, Louis de la Parte Florida Mental Health Institute, National Implementation Research Network. (FMHI Publication No. 231).
- Forgatch, M. S., Patterson, G. R., & DeGarmo, D. S. (2005). Evaluating fidelity: Predictive validity for a measure of competent adherence to the Oregon model of parent management training (PMTO). *Behavior Therapy*, 36, 3-13.
- Hall, G. E., & Hord, S. M. (2006). *Implementing change: Patterns, principles and potholes* (2nd ed.). Boston: Allyn and Bacon.
- Henggeler, S. W., Melton, G. B., Brondino, M. J., Scherer, D. G., & Hanley, J. H. (1997). Multisystemic therapy with violent and chronic juvenile offenders and their families: The role of treatment fidelity in successful dissemination. *Journal of Consulting and Clinical Psychology*, 65(5), 821-833.
- Henggeler, S. W., Pickrel, S. G., & Brondino, M. J. (1999). Multisystemic treatment of substance-abusing and -dependent delinquents: Outcomes, treatment fidelity, and transportability. *Mental Health Services Research*, 1(3), 171-184.
- Henggeler, S.W., Schoenwald, S.K., Liao, J.G., Letourneau, E.J. & Edwards, D.L. (2002). Transporting efficacious treatments to field settings: The link between supervisory practices and therapist fidelity in MST programs, *Journal of Clinical Child & Adolescent Psychology*, 31:2, 155-167
- Huser, M., Cooney, S., Small, S., O'Connor, C. & Mather, R. (2009). *Evidence-based program registries. What Works, Wisconsin Research to Practice Series*. Madison, WI: University of Wisconsin–Madison/Extension.
- Jensen, P. S., Weersing, R., Hoagwood, K. E., & Goldman, E. (2005). What is the evidence for evidence-based treatments? A hard look at our soft underbelly. *Mental Health Services Research*, 7(1), 53-74.
- Kumpfer, K. L., Alvarado, R., Smith, P., & Bellamy, N. (2002). Cultural sensitivity and adaptation in family-based prevention interventions. *Prevention Science*, 3, 241-246.
- Lau, A. S. (2006). Making the case for selective and directed cultural adaptations of evidence-based treatments: Examples from parent training. *Clinical Psychology: Science and Practice*, 13, 4, 295 – 310.
- Lucca, A. M. (2000). A clubhouse fidelity index: Preliminary reliability and validity results. *Mental Health Services Research*, 2 (2), 89 – 94.
- McCabe, K. M., Yeh, M., Garland, A. F., Lau, A. S., & Chavez, G. (2005). The GANA program: A tailoring approach to adapting parent-child interaction therapy for Mexican Americans. *Education and Treatment of Children*, 28, 111-129.

McGrew, J. H., Bond, G. R., Dietzen, L., & Salyers, M. (1994). Measuring the fidelity of implementation of a mental health program model. *Journal of Consulting & Clinical Psychology*, 62(4), 670-678.

McGrew, J.H., & Griss, M.E. (2005). Concurrent and predictive validity of two scales to assess the fidelity of implementation of supported employment. *Psychiatric Rehabilitation Journal*, 29, 41-47.

Michie, S., Fixsen, D., Grimshaw, J., & Eccles, M. (2009). Specifying and reporting complex behaviour change interventions: the need for a scientific method. *Implementation Science*, 4(1), 40.

Mowbray, C. T., Holter, M. C., Teague, G. B., & Bybee, D. (2003). Fidelity criteria: Development, measurement, and validation. *American Journal of Evaluation*, 24, 315 – 340.

Nielsen, J. (2005). Usability for the masses. *Journal of Usability Studies*, 1(1), 2-3.

Olds D. L. (2002). Prenatal and infancy home visiting by nurses: From randomized trials to community replication. *Prevention Science*, 3:153-17

Propst, R.N. (1992). Standards for clubhouse programs: Why and how they were developed. *Psychosocial Rehabilitation Journal*, 16, 25-30.

Rogers, E.M. (1995). *Diffusion of innovations* (5th ed.). New York: The Free Press.

Schoenwald, S.K., Chapman, J.E., Sheidow, A.J., & Carter, R.E., (2009). Long-term youth criminal outcomes in MST transport: The impact of therapist adherence and organizational climate and structure. *Journal of Clinical Child & Adolescent Psychology* 38:1, pages 91-105.

Sexton, T.L. & Alexander, J.F. (2002). Family-based empirically supported interventions, *The Counseling Psychologist*, Vol. 30, No.2, 238-261.

Shewhart, W. A. (1931). *Economic control of quality of manufactured product*. New York: D. Van Nostrand Co.

Webster-Stratton, C. & Herman K.C., (2010). Disseminating Incredible Years Series Early Intervention Programs: Integrating and sustaining services between school and home. *Psychology in the Schools*, 47, 36-54.

Webster-Stratton, C., Reinke, W.M., Herman, K.C., & Newcomer, L. (in press). The Incredible Years Teacher Classroom Management Training: The methods and principles that support fidelity of training delivery. *School Psychology Review*.

Winter, S. G., & Szulanski, G. (2001). Replication as strategy. *Organization Science*, 12(6), 730-743.

Wolf, M. M., Kirigin, K. A., Fixsen, D. L., Blase, K. A., & Braukmann, C. J. (1995). The Teaching-Family Model: A case study in data-based program development and refinement (and dragon wrestling). *Journal of Organizational Behavior Management*, 15, 11-68.

**DEPARTMENT OF HEALTH
& HUMAN SERVICES**

Office of the Secretary
Washington, DC

OFFICIAL BUSINESS
Penalty for Private Use \$300

