

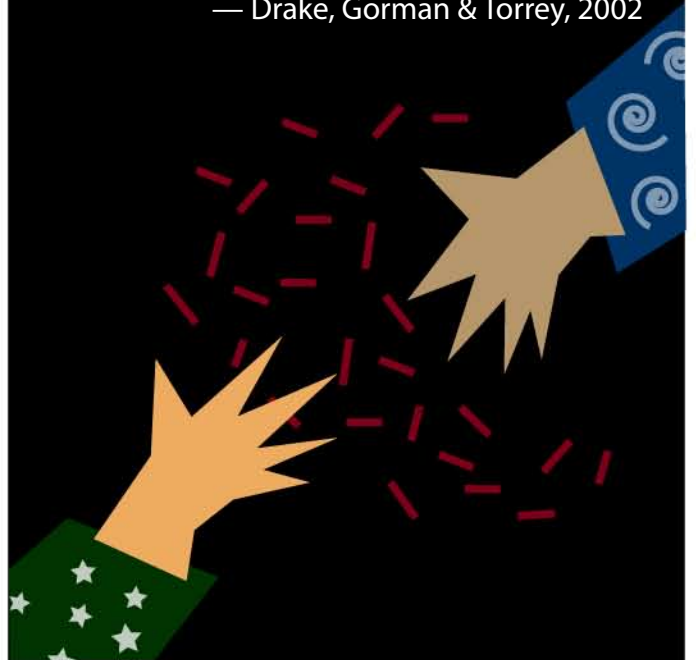
Chapter 5

Research on Core Implementation Components

- Staff Selection
- Staff Training
- Staff Coaching
- Evaluation and Fidelity

“We are faced with the paradox of non-evidence-based implementation of evidence-based programs.”

— Drake, Gorman & Torrey, 2002





Research on Core Implementation Components

Selection of staff is important to having effective practitioners, excellent trainers, effective coaches, skilled evaluators, facilitative administrators, or effective purveyors.

Logically, desired changes in important consumer outcomes in human services can only be achieved by changing the behavior of practitioners.

To be sure, enabling policies are important, appropriate funding sources are important, organizational structures and cultures are important, facilitative administrative supports are important, and capable trainers, coaches, evaluators, and administrators are important. However, in the end, all of these important factors exert their influence on consumers indirectly, through practitioners. Practitioners who competently use core *intervention* components in their interactions with consumers can have the positive effects that are the promise of evidence-based practices and programs. Thus, critical functions of implementation consist of practitioner training, coaching the practitioner on the job, regularly assessing fidelity, and using that information to improve the performance of practitioners who are carefully selected for the position. With these core *implementation* components in place (and functioning at a high level of competence themselves), practitioner behavior can be routinely changed and improved to assure competent performance of evidence-based practices and programs.

This chapter summarizes the empirical foundations for these core implementation components. Each section contains an introduction to the component, an overview of the experimental research (functional analyses using rigorous designs) concerning the component, a brief discussion of factors that may affect implementation of the component (where this information is available), an overview of other research and literature reviews regarding the component, and a brief summary section. For the interest of some readers, the well-designed experimental studies also have been summarized in Appendix C.

Staff Selection

Staff selection has been proposed as an implementation driver although it is not discussed often and rarely evaluated in human service programs. Nevertheless, selection may be a key ingredient of implementation at every level:

- selection of practitioners,
- selection of organization staff (trainers, coaches, evaluators, administrators), and
- selection of staff for purveyor groups.

Selection of staff is important to having effective practitioners, excellent trainers, effective coaches, skilled evaluators, facilitative administrators, or effective purveyors. Not everyone is suited to each role. People who are outgoing and decisive may make good practitioners or purveyors. People who are methodical and comfortable making judgments based on specified criteria may make better evaluators. People who are more comfortable with public speaking and “performing” might make better trainers. With respect to given evidence-based practices or programs, the extent of knowledge and direct experience in the specific program or practice might be more critical for some positions than others.

Experimental Research on Selection

The factors involved in staff selection interviewing were the subject of a meta-analysis of research in business (McDaniel, Whetzel, Schmidt, & Maurer, 1994). The authors found that education and background, exchange of information, and role play/behavior vignettes were effective interview techniques that related to later work outcomes for employees. An analysis of education and background as a selection criterion for the Nurse-Family Partnership prevention program was conducted by Olds et al., (2002). In this study, training, consumer: practitioner ratios, etc. were the same for two groups of practitioners. Group 1 consisted of nurses (the standard for the Nurse-Family Partnership program).

Group 2 consisted of paraprofessionals who had a high school diploma (and no further education) and strong people skills. Group 2 (paraprofessionals) also had greater access to coaching with 2 supervisors for every 10 practitioners compared to 1 for 10 in Group 1 (nurses). The results showed that pregnant women and their newborn children benefited more from Group 1 practitioners (nurses), confirming the need for candidates to have a nursing degree and background to be successful practitioners within the Nurse-Family Partnership program. Given the expense of using nurses vs. paraprofessionals, it is unfortunate there were no fidelity measures and no indications of the variability of outcomes within each group of practitioners. Without fidelity measures, there are no clues regarding the functional ways in which the two groups differed and, therefore, no clues for how to direct future efforts at implementation program development (see the analysis of FFT outcomes by the Washington State Institute for Public Policy, 2002).

These experimental studies suggest that the methods and the criteria for selecting practitioners may be important to achieving eventual intervention outcomes.

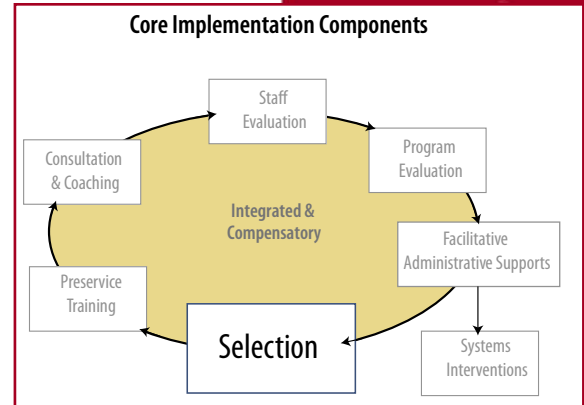
Practitioner Selection: Additional Evidence

Selection of practitioners is essential since it is at this level that evidence-based practices and programs are actually carried out (or not). Ager & O'May (2001) reviewed 103 intervention outcome studies and 42 staff "capacity to deliver" studies with regard to challenging behavior in persons with intellectual disability and acquired brain injury. They found that 25% of the primary deliverers of interventions were "researchers with...no formal or enduring relationship with the service setting," raising significant concerns about the likely success of any attempt at broad scale implementation. This finding points out that it is important to understand who is being employed to deliver evidence-based practices and programs, as reported in supporting research documents as well as during implementation (Diamond et al., 2002; Paine et al., 1984). Information on practitioners could include clear descriptions of inclusion-exclusion hiring criteria, candidate referral sources and interview procedures, exposure to skill building and professional development resources, and participation rates along with the more common

descriptions of academic and demographic characteristics of staff.

Descriptions of procedures and some data have been collected regarding practitioner selection. A manual was developed (Fixsen & Blase, 1996) to codify the process for selecting Teaching-Family homebased treatment

specialists including initial telephone contact to give information about the position and the difficulties of working in people's homes and neighborhoods, basic interview questions to get information about the candidate's capabilities, responses to behavioral vignettes, responses to role play situations, and responses to mini-training that requires behavior change and repractice in another role play. Variations of this selection process are used widely in national implementations of the Teaching-Family program (Blase et al., 1984; Fixsen et al., 1978; Maloney, Timbers, and Blase, 1977). One study analyzed the relationship between selection factors and later job performance for married couples that had applied to be Teaching-Parents in Teaching-Family treatment group homes. Maloney, et al. (1983) collected background information (years married, education, previous work experience) and measured interview behavior (responses to 10 behavioral vignettes, social interaction skills, receptivity to training, and overall interview performance) during the recruitment and hiring process. Compared with couples that were not hired, couples that were hired scored significantly higher on responses to the behavioral vignettes, receptivity to training, and overall interview performance (the social interaction skills factor was not a significant discriminator). Based on an evaluation of their performance on the job (3 - 5 months), the couples that were hired were divided into two groups: above or below the median performance for the group. The better performers were found to have differed significantly on their responses to the behavioral vignettes during the interview (but not the other three interview components). The better performers also had a higher GPA but the other background measures were not significantly different. The higher rated couples also stayed significantly longer on the job.



Research on the selection of staff to do large-scale implementation of evidence-based practices and programs is the next logical requirement in the overall scheme of things.

Another example of a behavior-based, structured selection process was reported by Reiter-Lavery (2004) for selecting therapists for MST programs nationally. Applications are screened for degree (master's preferred) and relevance of training and experience (family therapy, cognitive-behavioral approaches). Applicants then are interviewed. The first interview asks more general questions designed to get to know the candidate's style of interacting with others and solving problems and "fit" with MST ways of working. Those who successfully complete the first interview are invited to a second interview. In this interview, the details of the job are explained (e.g., how work is done, flexible hours) and the candidate's work and life experiences are explored in more detail. The final part of the interview is a series of role-play scenarios where a situation is presented then acted out with the candidate in the role of a family therapist. The candidate's responses are rated along several dimensions including collaborative and strength focused, efforts to overcome barriers, ability to use behavioral language, uses logical thinking, and is open to feedback.

Fisher & Chamberlain (2000) described the core implementation components of the Multidimensional Treatment Foster Care Program (MDTFC) including the methods to select new treatment foster parents. Advertising in various forms led to candidates who were screened for basic eligibility (adequate space in the home, no criminal history) before asking them to complete an application form. Program staff then made a home visit to meet the family, assess the family atmosphere, give detailed information about the program, and explain the training, supervision, and certification requirements. During the home visit they looked for empathy, knowledge of child development, a healthy sense of humor, willingness to take an active role in treatment, and ability to work within a structured program.

Huber et al., (2003) conducted a case study at one large hospital and commented on the fundamental importance of attracting, selecting, developing, and engaging staff in clinical settings to improve care, reduce turnover, and improve morale. In that system they did recruitment and prescreening for basic qualifications and personality characteristics, conducted interviews that consisted of giving information about the goals,

philosophy, and functions of the hospital as well as getting useful information about work experience and interaction styles of the candidates. Again, no data were reported on the outcomes or importance of the processes described. Wanberg & Banas (2000) studied practitioner characteristics in the context of organizational change at HUD and found that resilience, increased information, and self-efficacy were associated with greater acceptance of change in the workplace. These may be important selection characteristics for staff in organizations that are about to undergo changes as part of the implementation of an evidence-based program or practice.

Organization Staff Selection: Additional Evidence

Selection also is said to be important at the organizational level. Blase et al., (1984) and Fixsen & Blase (1993) described the process of selecting trainers, coaches, evaluators, and administrators to carry out the organizational change and development processes at new Teaching-Family implementation sites. Ogden et al., (in press) reported a similar process for selecting staff at the organizational levels in the national implementation of the parent management training Oregon model (PMTO) in Norway. Marks & Gersten (1998) studied the process of coaching with teachers across schools. Coaches were selected based on recommendations by district administrators who assessed candidate's ability to communicate information in a collegial style and ability to effectively teach students with learning disabilities in the regular classroom.

None of these descriptions included any data on the selection processes or criteria. However, a common theme was that the organizational staff needed to have a high level of understanding of the practices being implemented in the organization. For example, trainers or coaches in the Teaching-Family Model, PMTO program, Multidimensional Treatment Foster Care program, and the schools studied by Marks & Gersten (1998) all were required to have been practitioners in the program. In that way, they already had strong experiential knowledge and a detailed understanding of the intervention technology and only had to learn the new skills associated with being a trainer or coach.

Purveyor Staff Selection: Additional Evidence

Selection of staff at the purveyor level has been discussed by Bierman et al., (2002) and Blase et al. (1984) although no criteria or processes were noted. Havelock & Havelock (1974) described a curriculum for training “change agents” that provides a template for approaching training at the purveyor staff level although little information was provided regarding selection of candidates. Research on the selection of staff to do large-scale implementation of evidence-based practices and programs is the next logical requirement in the overall scheme of things.

Staff Selection Summary

Staff selection is a neglected area of implementation research. As implementation of evidence-based practices and programs becomes more of a national phenomenon, workforce issues likely will become much more important. There is increasing recognition of workforce development issues in behavioral health and groups such as the Annapolis Coalition on Behavioral Health Workforce Education (O’Connell et al., 2004) are discussing how to incorporate best practices in teaching methods, content, training sites, and student and instructor characteristics. Others (Morris & Stuart, 2002) are attempting to distill the generic skills needed by front-line practitioners in the behavioral health field (e.g., assessment skills, family and support system involvement, social and cultural engagement skills, treatment skills, methods to optimize recovery and empowerment, consumer relationship skills, and community resource management and coordination skills).

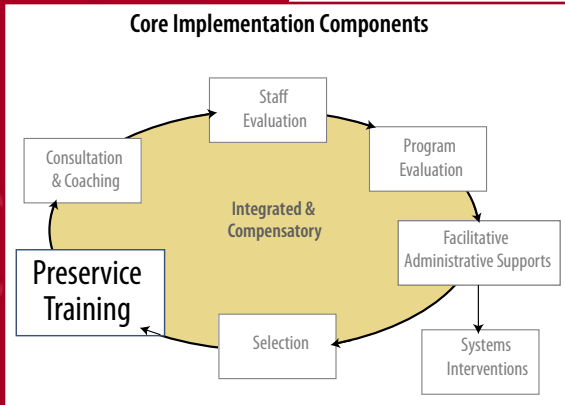
Research is needed to provide guidance to colleges and universities as they redesign their curricula to expose students to the basic theory and functions of evidence-based practices and programs, theories and methods of organizational and systems change, and a variety of evidence-based approaches in human service systems. Research on specific staff selection variables also will help promote success of implementations at each implementation site. Best practices for staff selection (core staff selection components) are not known although background, GPA, and direct observation and assessment of skills in behavioral vignettes may be important aspects of an inter-

view process for evidence-based practices and programs (Maloney et al., 1983; McDaniel et al., 1994; Olds et al., 2002).

Staff Training

Training appears to be a core implementation component for practitioners, agency staff, and purveyor staff. Rubenstein, Mittman, Yano, & Mulrow (2000) noted that, “Clinical services are delivered to patients through actions of health care providers, and the extent to which these actions mirror effective clinical practices determines quality of care. Effective interventions to improve health care reflect an understanding of health care provider behavior, the influences that shape it, and the methods that can be used to change it” (p. I-129). The content of training will vary considerably depending upon the evidence-based practice or program, clinical practice guideline, or management strategy that is being implemented. The methods of training seem to be less variable. There seem to be common approaches to imparting knowledge, skills, and abilities in programs to train practitioners (e.g., Bedlington, Booth, Fixsen, & Leavitt, 1996; Joyce & Showers, 2002; Schoenwald et al, 2000), trainers (e.g., Braukmann & Blase, 1979; Ogden et al., in press), coaches (e.g., Smart, Blase, et al., 1979; Joyce & Showers, 2003), fidelity evaluators (Davis, Warfel, Maloney, Blase, & Fixsen, 1979; Wineman, et al., 1979), and administrators (Baron, Watson, Coughlin, Fixsen, & Phillips, 1979; Atherton, Mbekem, & Nyalusi, 1999). During training, information about history, theory, philosophy, and rationales for program components and practices can be conveyed in lecture and discussion formats geared to knowledge acquisition and understanding. Skills and abilities related to carrying out the program components and practices can be demonstrated (live or on tape) then followed by behavior rehearsal to practice the skills and receive feedback on the practice (Blase et al., 1984; Joyce & Showers, 2002; Kealey, Peterson, Gaul, & Dinh, 2000).

Some programs have developed manuals for training practitioners (e.g., Bedlington et al., 1996; Braukmann & Blase, 1979; Schoenwald et al, 2003; VanDenBerg & Grealish, 1998), training trainers (Dreisbach & Smart, 1980),



The “train-and-hope” approach (Stokes & Baer, 1977) to implementation does not appear to work.

and training behavior rehearsal leaders and confederates (Dreisbach, Luger, Ritter, & Smart, 1979; Luger, Dreisbach, Smart, & Smart, 1979). The authors of these manuals point out a difference between role play (“pretend you are someone else and try this”) and

behavior rehearsal (“you are in your position as a practitioner and you are confronted with the following”). Role plays might sharpen a practitioner’s understanding or empathy. Behavior rehearsals are direct preparation for the real thing and are meant to be as much like the clinical setting as possible.

Factors that Impact Training

Several factors are thought to impact training (again, supporting data are lacking). Buston, Wight, Hart, & Scott (2002), evaluated the implementation of a sex education curriculum in Scottish schools. In the process of doing the study, they found that it was difficult to secure release time for teachers to participate in training, absences and turnover negatively impacted availability for training, and role play was difficult for teachers. Joyce & Showers (2002) noted that new learning that is outside the experience of the trainee or new learning that requires a more complex repertoire of skills is more difficult for trainees to learn and master and demands greater planning and precision from the trainers and coaches.

Joyce & Showers also emphasize that the content of training must be useful and, ultimately, beneficial to consumers. They examined the teacher training content for one state and found that, even if implemented completely, only 5% of the content being taught to teachers was different enough from common practice to have any possible benefit to children. Evidence-based practices and programs that have well defined core intervention components should be able to meet this criterion of potential benefit.

Experimental Research on Training Outcomes

As has been shown in a variety of settings, the “train-and-hope” approach (Stokes & Baer, 1977) to implementation does not appear to work. In the Schectman, et al. (2003) study discussed in Chapter 3, physicians randomly assigned to the physician education and feedback on usage group were not different from the control group with regard to adherence to clinical guidelines. Kelly et al., (2000) randomly assigned HIV service organizations to one of three groups: technical assistance manuals only, manuals plus a 2-day training workshop, or manuals plus training plus follow-up consultation. The addition of training produced a modest gain compared to the manuals-only group but the largest increase in reported adoptions of the HIV service guidelines occurred when consultation was added to training.

Smeele, et al. (1999) randomly assigned physicians to a non-intervention control group or a group that received an intensive small group education and peer review program. The results showed that the physicians in the experimental group demonstrated increased knowledge of the clinical guidelines pertaining to asthma and chronic obstructive pulmonary disease but patient care was not changed. McCormick et al., (1995) randomly assigned 22 school districts to experimental or control conditions. Health teachers and administrators in each group were provided with curriculum materials. In addition, the experimental group health teachers and administrators were given in-depth training on the use of a tobacco use prevention curriculum. The results indicated that teachers who had been trained were more likely to use more of the curriculum compared to control teachers. However, for health teachers in both groups combined, only 23% of the teachers initially used at least 90% of the curriculum and only 14% continued to use the curriculum for one year. Joyce & Showers’ (2002) meta-analysis of research on training and coaching in education was reviewed in detail in Chapter 3. Those results showed little change in classroom performance as a result of teacher training by itself or in combination with feedback on performance. A meta-analysis (Davis, 1995) found similar results in medicine. Davis concluded that, “formal CME conferences and activities, without enabling or practice reinforcing strategies, had little impact” (p. 700).

These experimental studies suggest that train-

ing by itself does not result in positive implementation outcomes (changes in practitioner behavior in the clinical setting) or intervention outcomes (benefits to consumers).

Experimental Research on Training Methods

While they are not effective by themselves for producing changes in clinical settings, training workshops are an efficient way to impart important information to practitioners and, when coupled with coaching, can contribute to important outcomes (e.g., Joyce & Showers, 2002). Some of the core components of training have been identified in experimental studies.

A series of studies was carried out by researchers attempting to implement the Teaching-Family Model. Kirigin et al., (1975) conducted an experimental analysis of the effects of training for Teaching-Parents (married couples who staff Teaching-Family group home programs). Training consisted of a 5-day workshop with presentations and discussion of history, theory, and philosophy; descriptions and demonstrations of skills; and behavior rehearsal of skills to criteria for mastery. Using a multiple-baseline design across participants, the authors found training produced significant improvements in key aspects of the “teaching interaction,” a core component of the Teaching-Family Model. A systematic replication was conducted by Maloney et al., (1975) with similar results: instructions plus practice plus feedback on practice were most effective in teaching skills important to the operation of a Teaching-Family group home.

Additional research on practitioner training was conducted by Dancer et al., (1978). As part of a 6 day, 50 hour preservice training workshop, one section (2 hrs) was for teaching “observing and describing behavior,” a foundation skill for other skills integral to the Teaching-Family Model (e.g., teaching social, academic, and self-care skills; providing feedback to youths regarding their ongoing behavior; and working with teachers and parents). Material was presented using brief lectures, discussions, live and video modeling, behavioral rehearsal to criterion, and constructive feedback. Using a multiple baseline design across groups, measures of observing and describing skills improved substantially after training. Another component of the Teaching-Family treatment program is provid-

ing personal rationales to youths (descriptions of natural and explicit consequences that may result from a youth’s behavior). Braukmann, Kirigin Ramp, Braukmann, Willner, & Wolf (1983) used a multiple baseline design to assess the effects of training consisting of a self-instruction manual, lecture/discussion, and behavior rehearsal on the use of rationales. Social validity was assessed via ratings by girls referred for delinquency issues in a Teaching-Family Model group home. Training produced large changes in the use of rationales and social validity ratings indicated that the girls preferred interactions that included rationales.

These experimental studies combined with the meta-analysis of research studies carried out by Joyce & Showers (2002) indicate that effective training workshops appear to consist of presenting information (knowledge), providing demonstrations (live or taped) of the important aspects of the practice or program, and assuring opportunities to practice key skills in the training setting (behavior rehearsal).

Training Practitioners: Additional Evidence

While there is wide agreement about the need for training as an important part of the implementation process, there are fewer studies that directly assess the impact of training on participants’ implementation in work settings. Dixon et al., (1999) compared implementation in 4 agencies where staff received a standard didactic presentation (lecture and discussion of the model and supporting data) with implementation in 5 agencies where staff received the standard presentation plus intensive training (information, discussion, demonstrations, role play). None of the standard presentation sites changed their approach to family services while 3 of the 5 agencies whose staff received intensive training did enhance their family services to some degree.

Kealey et al., (2000) reported the results of training about 500 teachers in 20 school districts in a smoking prevention program. Training included presentation of theory, description of skills, modeling of new skills and methods, and practice with feedback. A modest level of coaching was provided after the workshop. Ratings of the training workshops were high and teachers reported feeling prepared and confident upon completion of the workshops. Over 85% of the teachers were

Effective training workshops appear to consist of presenting information (knowledge), providing demonstrations (live or taped) of the important aspects of the practice or program, and assuring opportunities to practice key skills in the training setting (behavior rehearsal).

Recommendations for Training:

- Emphasize practice and use feedback on practice to teach the finer points.
- Use practice sessions to help trainees integrate thinking and doing.
- Provide guidance with respect to the boundaries of using the technique, describing when it may be useful and when it may not be useful.
- Provide guidance on the flexible use of the core components.
- Encourage peer and administrative support.

observed during the first year after training and teachers delivered 89% of the smoking prevention lessons according to the protocol. They concluded that practitioners must be motivated to adopt new practices, know what actions constitute the practices, have the tools to perform those actions, and have the ability and confidence to perform those actions (self-efficacy).

Ross, Luepker, Nelson, Saavedra, & Hubbard (1991) evaluated training for health education teachers in experimental schools that adopted the Teenage Health Teaching Modules (THTM). The modules originally were designed to be used without any training for health education teachers because training was thought to be too costly and therefore would hinder use of the curriculum. To test this assumption, training was offered to a group of randomly selected teachers and not offered to another group of randomly selected teachers. The analysis focused on 45 teachers who were trained and 25 who were not trained. Training consisted on 20 hours of orientation to the modules, practice on brainstorming and role-play-instructional methods, and discussion followed by telephone consultation. The results indicated that the trained teachers completed significantly more of the activities required in the modules and modified fewer of them compared to untrained teachers. In addition, students in the classes taught by trained teachers made significant gains in health knowledge and attitude scores while the students in classes taught by untrained teachers were no different pre to post. Thus, while teacher training added to the cost, effective use of the curriculum did not occur without the benefits of the training experience.

Dansereau & Dees (2002) discuss the development and evolution of a process for training counselors to use cognitive mapping, a method for spatially organizing and relating ideas, feelings, and actions with a consumer. Effective training processes were developed through an iterative process of defining the basic components of mapping, teaching those components to a group of counselors, coaching the counselors as they attempted to use mapping, then evaluating how the counselors did with respect to fidelity, competence, and comfort in using the mapping procedures. The coaching and evaluation experiences led to modifications in the next training session and the whole process was repeated until an effective training and coaching system resulted. Based on their

experiences in developing the training process, the authors made five recommendations for training:

- **Emphasize practice and use feedback on practice to teach the finer points of mapping.** Overemphasis on “rules” or drilling trainees on details before having a chance to practice can overwhelm some trainees and put off others who view it as inflexible.
- **Use practice sessions to help trainees integrate thinking and doing.** Didactic training tends to be linear while practice is multidimensional and dynamic. Practice and discussions of practice help integrate “what” and “why” starting with simpler examples and working to the more complex.
- **Provide guidance with respect to the boundaries of using the technique, describing when it may be useful and when it may not be useful.** Coaching is important to helping trainees find appropriate opportunities to use (and practice using) mapping.
- **Provide guidance on the flexible use of the core components of mapping.** Coaching is important to helping trainees adapt mapping to fit their own clinical style while retaining the essential components of the technique.
- **Encourage peer and administrative support** to build a culture of acceptance and support for effective use of mapping with consumers.

The authors caution that, “technologies designed to enhance counselors’ skills ... present a different set of problems. Movement from initial exposure to adoption and long-term practice depends heavily on the counselor’s confidence in executing the skills and a vision of how such skills can be integrated into ongoing activities. In addition to initial training, substantial hands-on coaching and practice may be necessary before a counselor feels comfortable with this new strategy” (p. 226).

Based on a review of teacher training programs, Gingiss (1992) noted that learning generally progresses from orientation and new learning to mechanical use, routine use, refinement, integration, and innovation as new knowledge, skills, and abilities become fully developed.

Training Organizational Staff: Additional Evidence

There is little research information concerning training staff at an *organizational* level. Fixsen & Blase (1993) reported data on attempts to implement the Teaching-Family Model in whole organizations. After developing methods to systematically train site staff (trainers, coaches, fidelity evaluators, administrators), more attempted organizational implementations were successful (30% pre to 80% post). In another study, Fixsen et al. (2001) showed that over 85% of the treatment programs associated with Teaching-Family sites with systematically-trained site staff were sustained over many years compared to about 15% of the treatment programs that operated independently of a site.

Palsha & Wesley (1998) developed a program to train consultants for early childhood education (ECE) centers. They found that 62% of the consultant trainees completed training and pre-post tests showed significant improvements in the quality of ECE provided to children 0 - 5 years old for those centers that were the subject of their consultation.

A top down, “cascade model” of training trainers was used in Zimbabwe to provide AIDS education nationally (O’Donoghue, 2002, reported in the previous chapter). The evaluation indicated poor results for implementation and for intervention. A similar method is being used in Norway to implement the Parent Management Training Oregon model (Ogden et al., in press). However, in Norway a bottom up approach is being used with the second and third groups of trainers being chosen from the ranks of practitioners who learned the program first hand. Wells, Sherbourne et al., (2000) trained trainers and nurse specialists in 46 primary care clinics to provide clinician and patient training for patients with depression. Leader training, staff training, and monitoring were provided according to the protocol in 100% of the clinics. However, less than 40% of the patients received treatment in keeping with the protocol.

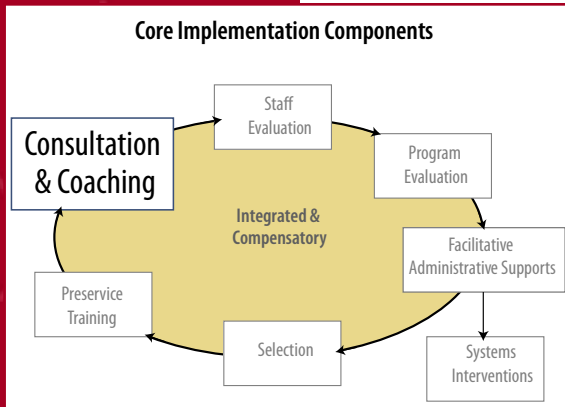
While these studies are encouraging and help to define some relevant aspects of training, none demonstrated a functional relationship between organizational staff training and implementation outcomes at the consumer level.

Staff Training Summary

The essence of implementation is behavior change. Training by itself seems to be an ineffective approach to implementation. However, it appears that the functional components of staff training are knowledge of the program and practices, demonstrations of key skills, and practice to criterion of key skills. Training for trainers and special training for behavior rehearsal leaders and confederates may be required to maximize learning for the trainees. The essential aspects of training may be similar for imparting knowledge and skills to key organizational staff (trainers, coaches, evaluators, administrators) and purveyors as well as practitioners.

Research is needed to assess the most effective and efficient conditions for training practitioners and for training organizational staff. Analyzing staff selection and training interaction effects may be especially useful as implementers of evidence-based practices and programs have to decide the relative merits of working with current staff of agencies (“conscripted staff”) or recruiting and training new staff for important roles (practitioner, trainer, coach, administrator, etc).

The essence of implementation is behavior change.



Staff Coaching

In their review of operations of ministries of health for the World Health Organization, Unger, Macq, Bredo, & Boelaert (2000) stated that systems reform (such as implementation) depends upon “training of field staff, on-the-spot expert

coaching, and promotion of a new organizational structure.” Spouse (2001) noted that formal knowledge (“episteme”) needs to be supplemented with craft knowledge (“phronesis”) so practitioners can learn to see the relevance of what they have learned to the situations at hand. Coaching needs to be work based, opportunistic, readily available, and reflective (e.g., debriefing discussions). Spouse (2001) described four main roles of a coach:

- Supervision
- Teaching while engaged in practice activities
- Assessment and feedback
- Provision of emotional support

After a few decades of research on training teachers, Joyce & Showers (2002) began to think of training and coaching as one continuous set of operations designed to produce actual changes in the classroom behavior of teachers. One without the other is insufficient. Behavior change is difficult for most people (for example, some people hire personal coaches to help them exercise more or change their eating behavior or stop smoking). With newly learned behavior there are several simultaneous problems that must be faced:

Newly-learned behavior is crude compared to performance by a master practitioner. Training usually is designed to introduce the learner to the essential elements of a new set of skills. For example, there are nine components of a “teaching interaction” (Phillips et al., 1974) and these components are taught to and rehearsed by practitioners in a preservice training workshop until they reach mastery criteria (Kirigin et al., 1975). However, there are uncounted nuances of when and how to use the components in various com-

binations in proactive teaching, reactive teaching, conceptual teaching, effective praise, proactive prompting, and so on given the treatment plans for and immediate behavior of particular children, families, or adults. This functional and adaptable set of skills is developed in practice with the help of a consultant/coach who shares craft knowledge as he or she observes, describes, and tutors the practitioner (Smart et al., 1979). With experience and effective coaching, a practitioner develops a personal style that is comfortable for the practitioner while still incorporating the core intervention components of the evidence-based practice.

Newly-learned behavior is fragile and needs to be supported in the face of reactions from consumers and others in the service setting. Behavior change directly impacts others in the environment. For example, when a teacher makes a significant change in his or her behavior in the classroom, 20 to 30 children and their families react to that change. When Nurse-Family Partners make a significant change in their behavior, 25 families and a variety of stakeholders react to that change. Joyce & Showers (2002) recommend having discussions with students and their parents to prepare them for the new ways of teaching that are about to be implemented. Although we could find no data on the topic, this probably is a good idea.

When practitioners change their behavior the reactions from consumers and stakeholders initially may not be positive, effectively punishing the practitioner for making a change. For fragile, new behavior the negative reaction may be enough to discourage the practitioner from persisting. One role of a coach is to prepare the practitioner for potential reactions and support the practitioner through the early stages of implementation until the new behavior is more skillfully embedded in the clinical environment (Joyce & Showers, 2002). Bierman et al., (2002) describe this as a counter-control function of a coach. That is, to help the practitioner engage in the new behavior even though they are not yet proficient and despite the negative reactions to using the new behavior (sometimes poorly).

Newly-learned behavior is incomplete and will need to be shaped to be most functional in a service setting. When designing workshop training experiences, there is only so much that can be accomplished effectively within the time avail-

able. Preservice workshop training can be used to develop entry-level knowledge and skills. Then, coaching can help practitioners put the segmented basic knowledge and skills into the whole clinical context. Coaches can help practitioners see how their personal beliefs and attitudes can be integrated with the skills, knowledge, philosophy, values, and principles of the program as well as other aspects of the clinical context (Smart et al., 1979).

In addition to helping to establish new behavior in the clinical environment, emotional and personal support is another role for a coach (Spouse, 2001). In human services, *practitioners are the intervention*. Evidence-based practices and programs inform when and how they interact with consumers and stakeholders but it is the person (the practitioner) who delivers the intervention through his or her words and actions. In the transactional interplay between practitioner and consumer, each affects the other in complex ways (for example, Fixsen & Blase, (1993) pointed out that each dependent variable is also an independent variable in a treatment environment; in this case, the consumer is “treating” the practitioner as well as being treated by the practitioner). In clinical work, practitioners often come face to face with their own issues and sensitivities as they work with consumers and stakeholders. A coach can help support a practitioner during times of stress or discomfort (Spouse, 2001). However, an overemphasis on emotional support may be counterproductive (Schoenwald et al., 2004).

Factors that Impact Coaching

The amount of time devoted to coaching often is not reported, but seems to vary widely. Diamond et al., (2002) provided 2 hours of coaching per week for therapists using drug treatment models. Supervision in Australian mental health settings typically occurred monthly for about 2 hours (Kavanagh et al., 2003). Coaching of teachers in special education classrooms occurred twice a week for an hour or so (Marks & Gersten, 1998). In Multisystemic Therapy for children and their families in the delinquency system, group coaching (primarily based on practitioner reports) occurs once or twice a week for about 90 minutes for each group of 3 to 4 therapists and the coaches themselves receive individual consultation once a

week for about an hour (Schoenwald et al., 2000). For the Teaching-Family Model, consultation occurs weekly (more often for new practitioners, less often for certified practitioners) with several hours devoted to on-site direct observation of the practitioner while he or she is providing direct services, feedback after the observation, and skill development in keeping with a professional development plan for each practitioner coupled with more frequent telephone consultation and coaching (Smart et al., 1979).

Denton, Vaughn, & Fletcher (2003) reviewed attempts to implement reading programs for students with reading and learning disabilities. While noting that effective coaching was the most critical factor in successful implementation, they cautioned that effective coaching depended upon the availability of coaches who are expert in the content, techniques, and rationales of the program. It is said that good mentors are encouraging, supportive, committed, sensitive, flexible, respectful, enthusiastic, diplomatic, patient, and willing to share information, credit, and recognition (McCormick & Brennan, 2001). In their survey in Kentucky, McCormick & Brennan (2001) found that coaching was impacted by time allotted to do the work, reluctance to seek information from the mentor, role confusion due to the dual role of supervisor and coach, feelings of inadequacy on the part of the mentors, poor match between the coach and practitioner, and lack of availability of coaches in rural areas.

Joyce & Showers (2002) pointed out that leadership, organizational culture, labor-relations, scheduling, interpersonal relationships, and engagement in participatory planning all impact the availability and effectiveness of coaching. In addition, coaches need to be trained and coached to provide specialized coaching functions for teachers, and that requires more organizational leadership and more resources (Marks & Gersten, 1998). Kavanagh et al., (2003) found that high caseloads and inadequately trained supervisors were major impediments to adequate supervision. Bond et al. (2001) noted that coaching sometimes suffered due to lack of information and skills, lack of time, inadequate staff resources, and a focus on paperwork instead of outcomes.

Showers & Joyce (1996) described the evolution of coaching and recommended that coaching

**In human services,
practitioners are the
intervention.**

Coaching relationships should start during training so parts of the training experience (practice new skills, receive feedback, re-practice) can facilitate the development of the coaching relationship.

relationships should start during training so parts of the training experience (practice new skills, receive feedback, re-practice) can facilitate the development of the coaching relationship (a strategy also recommended by Smart et al., 1979).

Experimental Research on Coaching

The value of on-the-job coaching repeatedly appeared in the overall implementation evaluation literature. In Chapter 4 the results of the Joyce & Showers meta-analysis were presented showing that implementation in educational settings occurred primarily when training was combined with coaching in the classroom. A similar result was obtained in a mental health setting (Kelly et al., 2000) and a medical setting (Fine et al., 2003) as reviewed earlier in this chapter.

Van den Hombergh, Grol, Van den Hoogen, & Van den Bosch (1999) compared two different types of coaches in a randomized group design. One group of physicians was assigned to a coach who was a “peer physician” and the other group was assigned to a coach who was a “practice assistant” (not a physician). In each case, the coaches followed a standard protocol to assess practice management and organization (issues not related to direct patient care). The results indicated that both groups improved on many of the 33 measures of practice management but peer-visited physicians showed significantly greater improvement on several practice dimensions. Joyce & Showers (2002) also recommended the use of peer coaches although they did not have experimental data to support their conclusion.

While these studies point to the importance of coaching in any attempt to implement a practice or program, we did not find any experimental analyses of the functional components of coaching. Thus, at this point, we know that coaching is important but we do not know (experimentally) what a coach should do or say with a practitioner to be most effective.

Additional Evidence for Coaching

Some non-experimental data do provide some clues to what may be the functional components of coaching. Kavanagh et al. (2003) conducted a telephone survey of nearly 300 mental health practitioners in Australia. They found that constructive feedback and praise were common components of supervision but there was very little direct observation of clinical practice by the supervisors (median = 0; also see Walker, Koroloff, & Schutte (2002) who found a similar result for persons supervising treatment planning teams). Four factors accounted for 62% of the variance of the perceived impact of supervision on practice: supervisor taught new skills, strengthened confidence, offered safety in sessions, and devoted time to discipline-specific skills (as opposed to generic skills).

Ager & O’May (2001) conducted a literature review of “best practice” for intervention for challenging behavior in persons with intellectual disability and acquired brain injury and found 42 papers that directly addressed the issue of the capacity of direct care for the delivery of interventions. They found that staff training has little impact on staff performance in clinical settings without additional help from a coach. The use of consultants (for feedback, supervision, and support) was found to be necessary for changes in staff performance. Schoenwald et al., (2004) evaluated a Consultant Adherence Measure (CAM) developed to measure clinical consultation in the multisystemic treatment (MST) program. They found that items related to perceived consultant competence (knowledgeable, skilled in MST, able to teach MST) were related to higher Therapist Adherence Measures (TAMS) and better youth outcomes. Items related to MST procedures (use of MST-specific assessment, intervention, and analytic techniques) were not related to TAMS scores for therapists and had mixed results for youth outcomes. Items related to alliance (attentive and supportive of therapists) were associated with lower TAMS scores and poorer youth outcomes. Although the results are not conclusive, this study represents an important step forward in finding ways to measure the interaction of core implementation components, core intervention components, and outcomes for consumers.

Looking at data from the first 17 years of development and implementation of the Teaching-Family Model, Fixsen & Blase (1993) analyzed the success of implementation attempts before and after systematic consultation and supports were provided to Teaching-Parents in Teaching-Family group homes. Only 24% of the attempted group home implementations lasted 6 years or more before and 84% were sustained for 6 years or more after systematic consultation and supports were provided.

Harchik, Sherman, Sheldon, & Strouse (1992) examined the effects of consultation on staff in a community group home for adults with severe mental retardation. They found that consultation had a positive impact on staff members' appropriate use of the token reinforcement system, constructive teaching interactions, and engagement and participation in activities. Kelly et al., (2000) compared technical assistance manuals on how to implement HIV prevention interventions with manuals plus staff training plus consultation on how to conduct implementations of the program. The addition of the consultation component produced a significant improvement in the number of implementations of the prevention program (about 60% adoption rate compared to about 35% for the manuals-only group).

Staff Coaching Summary

As stated earlier, implementation of evidence-based practices and programs cannot occur unless the practitioner is well-prepared to deliver the required practices in his or her interactions with a consumer. Coaching makes clear contributions to the preparation of practitioners, both in the experimental and other research literature. The core coaching components seem to be teaching and reinforcing evidence-based skill development and adaptations of skills and craft knowledge to fit the personal styles of the practitioners (changing form, not function). Support during stressful times was mentioned as a key ingredient by several sources but that function may not be supported empirically (Schoenwald et al., 2004). Given the key interpolative role of coaching between staff selection and training on the one hand and staff performance assessments on the other hand, research is needed that evaluates the relative contributions of selection, training, and coaching and

(especially) the interaction effects among the three factors (e.g., see Schoenwald et al., 2004). One purported aspect of the implementation driver framework is that the components are integrated and compensatory. Thus, the interaction effects may provide very useful information to inform the practice and theory of implementation.

Evaluation and Fidelity

In the reviews of staff evaluation and fidelity, two functions quickly became apparent. First, 46% of the articles reviewed in this section used measures at the practitioners performance level in order to help improve performance in the context of an organizational environment. The performance improvement function usually was embedded in organizations as an essential part of the treatment program. However, the majority (54%) of the articles used measures of staff performance in order to evaluate adherence to research protocols. The protocol adherence function usually is conducted outside the service organization and has utility only for the duration of the evaluation project.

Second, a subset of the articles described fidelity measures at the organizational level. Interestingly, about 2/3 of the articles regarding assessments of staff performance that were part of the treatment programs (including all of those at the organizational level) concerned Fountain House clubhouses, Assertive Community Treatment, or the Teaching-Family Model, three evidence-based treatment programs that have been involved in national implementation since the 1970s. It appears that more mature programs have learned the value of a similar set of practitioner-level and organizational-level performance measures that must be built into any organization using their program.

Staff evaluation and fidelity seem to consist of some combination of measures of context, compliance, and competence (Waltz, Addis, Koerner, & Jacobson, 1993; Forgatch, Patterson, & DeGarmo, in press). With respect to these measures:

- **Context** refers to the prerequisites that must be in place for a program or practice to operate (e.g., staffing qualifications or numbers, practitioner-consumer ratio, supervisor-practitioner ratio, location of service provision, prior completion of training).

Given the key interpolative role of coaching between staff selection and training on the one hand and staff performance assessments on the other hand, research is needed that evaluates the relative contributions of selection, training, and coaching and (especially) the interaction effects among the three factors.

In a highly functional systems, staff evaluation is part of a sequence of supports designed to have good people well prepared to do an effective job.

- **Compliance** refers to the extent to which the practitioner uses the core intervention components prescribed by the evidence-based program or practice and avoids those proscribed by the program or practice.
- **Competence** refers to the level of skill shown by the therapist in using the core intervention components as prescribed while delivering the treatment to a consumer (e.g., appropriate responses to contextual factors and consumer variables, recognizing the key aspects of the presenting problems, understanding the consumer's individual life situation, sensitivity of timing, recognizing and acting on opportunities to intervene).

Table 2 provides some examples of these forms of fidelity measures. How fidelity is measured is described briefly in the left hand column. Specific questions asked to measure context, compliance, or competence are then provided in the other three columns. For example, supported employment programs measure context fidelity by asking about caseload size and consumer eligibility for participation (among others). These are prerequisite conditions for providing supported employment as defined by the researchers. Assertive Community Treatment programs measure compliance fidelity by asking where the work is done (“in the community rather than office”) and who does supervision (among others). These examples of compliance measures tell program managers and others whether or not a procedure or process is in place. The Parent Management Training Oregon Model measures competence fidelity by directly observing the performance of practitioners as recorded on videotaped sessions. Their measures (among others) assess the occurrence of clinical episodes (“therapist sets up role

play”) and how well the practitioner performed when those episodes occurred (“capitalizes on opportunities,” “balances verbal teaching and active teaching”). These examples of competence measures tell practitioners,

coaches, managers, and others how well the practitioner is performing the core intervention components of an evidence-based program or practice.

Staff Evaluation for Performance Improvement

Huber et al., (2003) described highly effective hospital management systems that included recruitment and prescreening for basic qualifications and personality characteristics; interview procedures designed to give information about the goals, philosophy, and functions of the hospital as well as obtaining information about work experience and style; post-hiring orientation to the workplace and the specific role of the person; ongoing training and education focusing on specific skills needed, cross training on related roles, and in-services and monthly dinners for discussion; performance evaluations based on direct observation to assess practice knowledge, communication skills, and use of time with prompt verbal feedback followed by a write up with recommendations; and quality improvement information systems to keep the system on track (see Core Implementation Components).

In a highly functional systems, staff evaluation is part of a sequence of supports designed to have good people well prepared to do an effective job. In these cases, assessments of performance are well integrated with what has been taught and coached and there are no surprises for the practitioner.

The feedback from the more formalized assessment provides information for the coaching process (Phillips et al., 1974; Davis, Warfel, Fixsen, Maloney, & Blase, 1978; Smart et al., 1979; Schoenwald et al., 2000) and is an outcome measure for the quality of coaching (Blase et al., 1984; Schoenwald et al., 2004).

In the Teaching-Family Model practitioners are selected, trained, coached, and then evaluated at 6 months, 12 months, and annually thereafter with respect to their performance, the satisfaction of the consumers they have treated, and the satisfaction of the stakeholders with whom they have contact (Phillips et al., 1974; Wineman & Fixsen, 1979). Performance is evaluated by two trained evaluators who directly observe a practitioner for 2 to 3 hours as he or she provides treatment (Davis et al., 1978). A standard form is used to make detailed comments on the practitioner's performance and provide a rating for each of several areas that

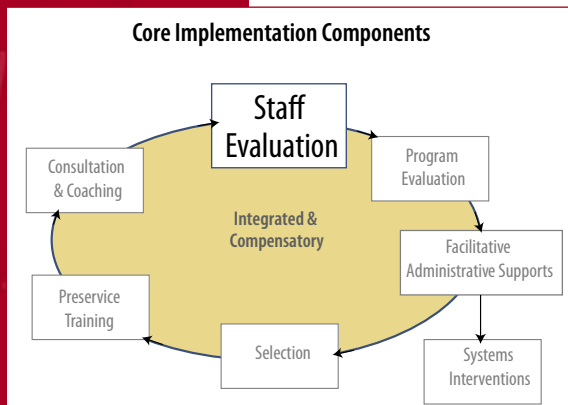


Table 2
Examples of Different Types of Fidelity Measures Across Programs

Program	Type of Fidelity Measurement		
	Context Measures	Compliance Measures	Competence Measure
Supported employment (Bond, et al., 1997) — 15 items, 5-pt. scale, interview a knowledgeable staff person	Employment specialists manage caseloads of up to 25 clients No eligibility requirements for consumer participation	“Employment specialists provide only vocational services” “Job search occurs rapidly after program entry”	
Assertive Community Treatment (Teague, et al., 1998) — 26 items, interview knowledgeable staff or managers, record review	Vocational specialist, nurse, psychiatrist, substance abuse specialist on staff Client-provider ratio of 10:1 Responsible for crisis services	“Staff monitors status and develops skills in the community rather than office” “Supervisor provides direct services as well” “High number of service contacts, high amount of time”	
Teaching-Family Model homebased treatment (Fixsen, et al., 1992) — survey of children, parents, and stakeholders; direct observation of performance by trained evaluator	Family Specialist-family ratio of 1:2 Completion of 60-hour preservice workshop and access to at least weekly inservice coaching and consultation	Treatment plan for each family Paperwork done promptly	“Family Specialist engages family members in the treatment process (building partnerships with parents, active participation by family members, family investment of time and energy)” “Family Specialist provides conceptual feedback (useful strength and improvement concepts, adequate specific examples, convincing rationales)”
Multisystemic Therapy (Henggeler, et al., 1992) — 27 items, 5-point scale, interview parent	Completion of 5-day training workshop On-going involvement by consultants at MST Services Inc.	“The therapist recommended that family members do specific things to solve our problems” “Family members and the therapist agreed upon the goals of the sessions”	“The therapist tried to understand how my family’s problems all fit together.” “There were awkward silences and pauses during the session”
Parent Management Training Oregon Model (Forgatch, et al., in press) — direct observation of video-taped sessions		Follows an agenda Includes appropriate sections	“Therapist sets up role play and capitalizes on opportunities” “Therapist balances verbal teaching and active teaching while engaging the family and providing rationales”

have been demonstrated to be core intervention components in the Teaching-Family Model (e.g., relationship development, teaching, self-determination, use of motivation systems). The individual consumer interview asks about the fairness, helpfulness, and concern of the practitioners. In addition, another set of questions asks each consumer about staff practices that may be unethical or illegal to help assure the safety of consumers (especially in residential treatment settings like group homes or foster homes). Finally, a brief set of questions is mailed to stakeholders who are asked to rate and provide comments concerning the practitioners

performance (an 80%+ response rate is typical). The specific stakeholder questions were derived from interviews with consumers and practitioners and from the overall mission and goals of the program (e.g., cooperation, communication, respect for opinions, effectiveness, helpfulness, concern). Detailed verbal and written reports of the findings, conclusions, and recommendations are promptly provided to the practitioner, coach, and manager. Staff evaluations are conducted by evaluators in a Certified Organization and are reviewed in detail as part of the Organizational Certification process (described below).

The discriminant validity of the practitioner fidelity measures was tested by comparing Teaching-Family treatment homes with other group homes and with a state detention center. Teaching-Family practitioners scored higher on ratings by school teachers and administrators, parents, and youths. There were no differences in ratings by juvenile court personnel, social services personnel, or members of the Boards of Directors (Kirigin & Fixsen, 1974; Kirigin, Fixsen, & Wolf, 1974). Predictive validity was tested by Kirigin, Braukmann, Atwater, & Wolf (1982) who correlated staff evaluation and fidelity measures with eventual youth delinquency outcomes and found that higher fidelity was associated with greater reductions in delinquency. Kirigin et al., (1982) also found that overall consumer and stakeholder ratings discriminated between Teaching-Family and control group homes with significant differences for youth and school teacher/administrator ratings. There were no differences in ratings by parents, juvenile court personnel, social services personnel, or members of the Boards of Directors. These findings were extended by Solnick, Braukmann, Bedlington, Kirigin, & Wolf (1981) who correlated a measure of one core intervention component (the “teaching interaction”) with self-reported delinquency and found a high level of correspondence between more teaching and less delinquency and between more teaching and higher satisfaction ratings by the youths in Teaching-Family group homes.

The multisystemic treatment (MST) program has monthly assessments of practitioner adherence to the 9 principles that are the foundation of the program (Schoenwald et al., 2000). Monthly fidelity assessments (called the TAM: Therapist Adherence Measure) occur via a telephone call (or other contact) with a parent who is asked to rate the practitioner on 27 items. After practitioners are selected and trained in a 5-day workshop, they begin work with youths and families with the support of a local supervisor. The web-based fidelity data are collected by MST Services, Inc. and the information is used to inform a chain of consultants including those employed by MST Services, Inc. to consult with area or organization-based MST consultants who consult with team supervisors who consult with practitioners.

At the practitioner level, Henggeler, Melton, Brondino, Scherer, & Hanley, (1997) found that higher fidelity scores during treatment were associated with better delinquency outcomes for youths. Schoenwald, Halliday-Boykins, & Henggeler (2003) conducted an interesting study that related fidelity to characteristics of the youths served. They found that practitioner fidelity was lower when working with youths who were referred for a combination of criminal offenses and substance abuse. In addition, practitioner fidelity was lower when working with youths who had more pretreatment arrests and school suspensions. Practitioner fidelity measures were higher when working with youths with educational disadvantage and higher when there was an ethnic match between practitioner and parent. Recently, the Consultant Adherence Measure (CAM) has been developed and tested to assess adherence to the MST consultant protocol. In an important study that linked the TAM, CAM, and youth outcomes, Schoenwald et al., (2004) found that higher consultant fidelity was associated with higher practitioner fidelity, and higher practitioner fidelity was associated with better youth outcomes. In another study, Schoenwald et al., (2003) found that practitioner fidelity was associated with better outcomes for youths but fidelity was not associated with measures of organizational climate. Organizational climate was presumed to be a mediation variable for adherence but this hypothesis was not borne out by the data.

Fidelity measures for the highly individualized Wraparound process (J. D. Burchard, S. N. Burchard, Sewell, & VanDenBerg, 1993) are being developed and tested (Bruns, Burchard, Suter, Force, & Leverentz-Brady, 2004; Bruns, Henggeler, & Burchard, 2000; Bruns, Suter, Burchard, Leverentz-Brady, & Force, in press; Bruns et al., 2004; Epstein, Jayanthi, McKelvey, Frankenberry, Hary, Potter, et al., 1998). The Wraparound Fidelity Index (WFI) consists of asking wraparound team facilitators, parents, and youths to rate 11 dimensions of the services for a family (voice and choice, youth and family team, community-based supports, cultural competence, individualized, strength-based, use of natural supports, continuity of care, collaboration, use of flexible resources, outcome based). When high fidelity implementations were compared to those

with low fidelity as measured by the WFI, high fidelity implementations resulted in improved social and academic functioning for children, lower restrictiveness of placements, and higher levels of satisfaction (Bruns et al., in press). High fidelity implementations were associated with training, coaching, and supervision for providers and the consistent use of data collection systems to inform the overall process.

The Washington State Institute for Public Policy (2002) evaluated the statewide implementation of the Functional Family Therapy (FFT) program for juvenile offenders (Alexander, Pugh, Parsons, & Sexton, 2000). The results showed that youths and families treated by therapists with high fidelity scores had significantly better outcomes. FFT, Inc. (the purveyor of FFT) conducted therapist fidelity measures and found that 19 (53%) of the 36 therapists were rated as competent or highly competent, and those therapists treated a total of 48% of the families in the study. When compared to the control group, youth with a highly competent or competent therapist had a lower 12-month felony recidivism rate. However, within this group of highly competent or competent therapists, the recidivism rates varied considerably. The authors lamented the lack of fidelity measures at the organizational level and speculated that variations in the amount or quality of training, supervision, or organizational support may have been important to therapist fidelity and youth outcomes. They also noted that measures of FFT fidelity built into local organizations might be more useful as a tool to guide the implementation process compared to having this function performed centrally by FFT, Inc.

Organization-Level Fidelity Assessments

The International Center for Clubhouse Development (ICCD) developed a measure of fidelity for Fountain House clubhouse organizations (Macias, Propst, Rodican, & Boyd, 2001). The measure was developed over a period of 5 years and is well-grounded in a series of workshops, surveys, and pilot studies involving nearly all of the international clubhouse community. Criterion validity was established when results of the fidelity measure were compared to the results of a more exhaustive 3-day site visit by trained evaluators to determine ICCD Certification.

ICCD Certification was established in 1994 to certify clubhouse organizations that meet all the criteria set forth by the ICCD. The certification process has a manual, a process to select and train site evaluators, and a review board that judges the quality and makes certification recommendations. Prior to a site visit, the clubhouse prepares a detailed self-study. The 3-day visit consists of record reviews, interviews with members and staff, visits with collaborators, and direct observations of the daily activities. An in-depth report and consultation is provided at the end of the visit and a written report is prepared after the visit (20–40 pages). Site visitors remain on call for continuing consultation and to make return visits to assess implementation of any agreed-upon changes.

The Teaching-Family Association developed a two-tier system of fidelity review and certification in 1978 (Blase et al., 1984; Wolf et al., 1995). Organizational Certification (the first tier) has a process to select and train site evaluators and a national Certification Committee that assesses quality and makes certification recommendations to the Board of Directors of the Association. Certification consists of a full report of organization activities related to the program (e.g., selection, training, coaching, staff evaluation, administrative supports) and an organizational consumer evaluation (360-degree evaluations internally as well as external stakeholder evaluations by funders, referral sources, others) conducted by the national Certification Committee. The 2 to 3-day site visit involves interviews with practitioners, consumers, and members of the Board of Directors; observations of on-going treatment; and interviews with trainers, coaches, and staff evaluators regarding the technical aspects of the program. A detailed report of the findings, conclusions, and recommendations (often 100+ pages) is prepared for the organization and the Certification Committee. The second tier is Practitioner Certification (described in the previous section) conducted by trained evaluators employed by a Certified Organization.

McGrew, Bond, Dietzen, & Salyers (1994) described a process for evaluating organizations providing Assertive Community Treatment (ACT). Experts rated a pool of 73 items proposed as critical to ACT operations. Based on the expert review, a 17-item subset was used to construct a

Another approach to program evaluation has been taken in the state of Michigan where they have instituted a state-wide continuous monitoring system.

fidelity index with 3 subscales:

- Staffing — client-staff ratio, team size, psychiatrist on team, nurse on team
- Organization — team as primary therapist, location of team, shared caseloads, daily team meetings, coordinator provides direct client service, 24-hour availability, time-unlimited resources
- Service — frequency and hours of face-to-face contact, in office contact, all contact

In applications of the fidelity scale to 18 ACT programs, internal consistency of the items was acceptable and higher total scale scores and scores for the staffing and organization subscales were associated with greater reductions in days spent in psychiatric hospitals. The fidelity measure also detected program drift; scores were linearly related to successive iterations, or "program generations."

Teague, Drake, & Ackerson (1995) found similar results in a comparison of ACT with standard case management at 7 sites over a 27-month period. Teague, Bond, & Drake (1998) revised the ACT fidelity scale and applied it to four groups known to differ in their approach. Fidelity scores were highest for ACT, with an average score above 4 on a 5-point scale, demonstrating discriminant validity. In these studies, the staff and service components of the ACT fidelity scale accounted for more of the variance in outcome measures than organization subscale. It is interesting that ACT does not have a practitioner-level staff fidelity measure.

Another approach to program evaluation has been taken in the state of Michigan where they have instituted a state-wide continuous monitoring system using the Child and Adolescent Functional Assessment Scale (CAFAS) as a common measure (Hodges & Wotring, 2004). Over a period of 6 years, the monitoring system was implemented, a culture of using evidence to aid decision-making was created among clinical professionals in the state, and the overall benefits of mental health treatments for over 5,000 children with serious emotional disturbance were assessed (Hodges, Xue, & Wotring, 2004). By analyzing the characteristics of the children for whom treatments were most successful and least successful, the Michigan evaluation system was able to identify those problems that were most intractable in the "treatment as

usual" system and, therefore, good candidates for implementing evidence-based programs (Hodges, Xue, & Wotring, 2004; Xue, Hodges, & Wotring, 2004). A similar approach has been taken by the Nurse-Family Partnership program, the Functional Family Therapy program, and the Multidimensional Treatment Foster Care program where program evaluation measures have been built into the paperwork flow and outcome measures are collected via a web-based reporting system. Korfmacher et al. (1998) analyzed the program evaluation information for 228 infants and their mothers who were served by the Nurse-Family Partnership program in Tennessee. They were able to assess how closely the clinical applications of the program compared to the experimental versions and they began an analysis of the contributions of individual program components to the outcomes for infants and their mothers.

While the relationships between fidelity measures and outcome measures are consistent across programs, they are correlational. The results so far could be related to therapist enthusiasm or consumer characteristics (e.g., Schoenwald et al., 2003) or other aspects of the therapeutic situation rather than to the core intervention components. Thus, the actual relationship between fidelity to the prescribed core intervention components of evidence-based practices and programs and their outcomes must await the results of eventual experimental analyses of those relationships.

Factors that Impact Staff Evaluation for Performance Improvement

McGrew et al., (1994) noted that the development of fidelity measures is hampered by 3 factors: (1) most treatment models are not well defined conceptually, making it difficult to identify core intervention components, (2) when core intervention components have been identified, they are not operationally defined with agreed-upon criteria for implementation, and (3) only a few models have been around long enough to study planned and unplanned variations.

Staff evaluations need to be practical so they can be done routinely in an organization (Blase et al., 1984; Henggeler et al., 1997) and staff evaluators need to be prepared for their roles. Wineman & Fixsen (1979) developed a detailed procedure manual for conducting a rigorous staff evaluation in the context of a Teaching-Family treatment group home. Freeman, Fabry, & Blase (1982) developed a comprehensive program for training staff evaluators for national implementations of the Teaching-Family Model. The staff evaluator training included instruction in direct observation of practitioner behavior, conducting record review, youth, parent and stakeholder evaluations, and analysis and presentation of evaluation findings to practitioners, coaches and managers. Workshop training included practice to criterion on the critical skills and was followed by a series of “co-evaluations” at implementation sites to assess agreement and provide opportunities for coaching on staff evaluation skills (Blase et al., 1984; Fixsen & Blase, 1993).

Given the integrated nature of any organization, it is likely that administrative decisions, changes in budget, office moves, etc. can have unintended and undesirable impacts on practitioner behavior and, therefore, impact fidelity. However, no measures were found in the literature.

Experimental Research on Evaluation

The review of the general implementation evaluation literature provided many examples of the importance of staff evaluation, implementation fidelity, and program evaluation. However, no experimental analysis of staff or program evaluation methods or outcomes appeared in the review. Experimental analyses of staff and program evaluation methods seem to be warranted given the presumed importance of evaluation-driven feedback loops and the resources necessary to routinely measure practitioner and organizational performance. Experimental exploration of evaluation efforts could yield more effective and efficient methods that could be adopted by purveyors of evidence-based practices and programs.

Staff Evaluation for Performance Improvement: Additional Evidence

Most of the research makes use of the staff performance data as predictors of consumer outcomes showing that programs with higher fidelity produce better outcomes for consumers (e.g., Felner et al., 2001; Henggeler et al., 1997; Henggeler, Pickrel, & Brondino, 1999; Kirigin et al., 1982; Kutash, Duchnowski, Sumi, Rudo, & Harris, 2002; Solnick et al., 1981). An interesting case study by Hodges, Hernandez, Nesman, & Lipien (2002) demonstrated how a theory of change exercise can help programs clarify their strategies and develop fidelity measures to assess their use of those strategies. Similarly, Shern, Trochim, & LaComb (1995) used concept mapping to develop fidelity measures for an adult mental health program. In another interesting study, Forthman, Wooster, Hill, Homa-Lowry, & DesHarnais (2003) found that feedback, provided in a timely fashion (short feedback loops, recurring), and delivered personally by a respected source was most effective when accompanied by written material and attended to the motivation of the audience (e.g. interest in improving quality for patients).

Testing Validity of Fidelity Measures

- Reliability across respondents
- Internal structure of the data
- Known groups
- Convergent validity
- Predictive validity

Organization-Level Fidelity Assessments: Additional Evidence

The research cited above regarding the ACT program established the validity and reliability of the organizational fidelity measures used for that program. A General Organizational Index has been recommended for use with the adult toolkits that were developed by SAMHSA (SAMHSA's Mental Health Information Center, 2004) but no data support its use. Fixsen & Blase (1993) and Fixsen et al., (2001) used organizational fidelity as an outcome to measure organizational implementation success but did not assess the measure itself.

Staff Evaluation to Measure Adherence to Research Protocols

The majority of articles that measured adherence to a research protocol simply reported the outcomes of having done so. In a review of 34 programs deemed to be effective by the Prevention Research Center (Domitrovich & Greenberg, 2000), 59% included some rating of fidelity and adherence in their implementation data but only 32% used the implementation measures as a source of variance in their data analysis. Gresham, Gansle, & Noell (1993) reviewed 158 articles in the *Journal of Applied Behavior Analysis* (1980-1990) to see how many assessed "implementation of the independent variable." After the articles were coded, the results showed that 34% of the studies provided an operational definition of the independent variables and 16% reported levels of treatment integrity. In a broader review of the literature, Moncher & Prinz (1991) reviewed 359 outcome studies (1980-1988). A detailed assessment of the studies showed that 32% used a treatment manual, 22% supervised the treatment agents, and 18% measured adherence to the protocol. Only 6% did all three (manual + supervision + adherence) while 55% did none of the three. They also found that 26% of the studies reported training the practitioners and only 13% of those assessed practitioner competence in using the protocol.

Factors that Impact Staff Evaluation to Measure Adherence to Research Protocols

None were found in the literature reviewed. Again, most measures of adherence to research protocols simply reported the measures and results. Well-funded research efforts may have fewer issues with measures of adherence compared to those that are built into organizational routines and consume a variety of organizational resources. Nevertheless, given the importance of measuring the degree of implementation of independent variables, it may be useful for researchers to report the factors that enable or compromise such measures.

Staff Evaluation to Measure Adherence to Research Protocols: Additional Evidence

Bond, Becker, Drake, & Vogler (1997) developed a fidelity scale (questions regarding staffing, organization, service) for the Individual Placement and Support (IPS) model of helping consumers find employment. They tested the scale with 9 IPS programs, 11 other supported employment programs, and 7 other vocational rehabilitation programs. The majority had been in existence for at least one year. The results showed the scale distinguished between the programs that were utilizing the IPS model and those that were not. As expected, the IPS programs had greater consistency with the IPS model scale than other supported programs. However, other supported employment programs were more "partially consistent" with the IPS model than the non-supported employment (other vocational rehabilitation) programs. Thus, the scale showed discriminant validity. Brekke & Test (1992) constructed a fidelity scale for the Assertive Community Treatment (ACT) program. They used questions related to client characteristics, location of services, mode of delivering services, frequency and duration of contact with consumers, staffing patterns, and continuity of care. Nearly all of the data were collected from record reviews, a time consuming process. The results demonstrated the ability of the fidelity measure to discriminate among intensive community programs.

Mowbray et al., (2003) point out that fidelity is important to internal validity and can enhance statistical power by explaining more of the variance. Fidelity can assess whether the program (independent variable) is really there in the experimental

condition and not there in the control condition, if it is really there in multi-site studies, and if it is really there across studies in a meta-analysis. After describing a developmental process (similar to that used by McGrew et al., 1994), the authors recommended testing several forms of validity:

- **Reliability across respondents** (various measures of agreement)
- **Internal structure of the data** (factor analysis, cluster analysis, internal consistency reliability)
- **Known groups** (apply measures to groups that are known to differ in ways important to the program)
- **Convergent validity** (correlating various measures from different sources with the fidelity measure)
- **Predictive validity** (relate fidelity scores with important outcome measures)

Another approach to developing a fidelity scale was taken by Paulsell, Kisker, Love, & Raikes (2002). When developing a scale to assess implementation in early Head Start programs, they based the items on the Head Start Program Performance Standards published by the government. The scale included items related to services (assessments, frequency, individualized, parent involvement), partnerships (family, community), and management supports (staff training, supervision, compensation, retention, morale). In 1997, after about 1 year of operation, 6 (35%) of the 17 programs had reached full implementation. By 1999, 12 (70%) had reached full implementation. The biggest improvements were in community partnerships (from 8 to 15 fully implemented) and management systems and procedures (from 7 to 14 fully implemented). The smallest gains were in the areas of child development (from 8 to 9 fully implemented) and family partnership (from 9 to 12 fully implemented). Early implementers started with a strong child development focus, had low staff turnover, and consistent leadership. Later implementers responded promptly to feedback from early site reviews, shifted from family support to a child development focus, and had early changes in leadership. Incomplete implementers had trouble responding to feedback from site visits, had trouble shifting to a child development focus, had higher staff turnover, had turnover in leadership, and had difficulties in community partnerships.

Forgatch et al., (in press) are developing an extensive fidelity measure for the Parent Management Training Oregon (PMTO) model, a clinical program being implemented in parts of the US and nationally in Norway. The fidelity measure consists of detailed coding and analyses of videotapes of treatment sessions. Trained observers use a 9-point scale to rate 5 dimensions of practitioner performance during the session: knowledge of PMTO, use of PMTO structures, teaching, clinical process, and overall quality. Their study found a significant positive relationship between practitioner fidelity and improvements in the parenting behaviors of mothers and stepfathers in the families being treated.

Evaluation and Fidelity Summary

The most effective intervention will not produce positive effects if it is not implemented. Thus, assessments of performance are a critical component of implementation. Context fidelity measures describe the necessary precursors to high-level performance (e.g., completion of training, acceptable practitioner-coach ratio, acceptable caseload, availability of colleagues with special skills, availability of certain resources) for a particular program or practice. Compliance fidelity measures provide an outline of the core intervention components and their use by the practitioner. Competence fidelity measures are essential for determining the extent to which the core intervention components were delivered with skill and attention to the craft when interacting with consumers. The results of fidelity measures and staff evaluations seem to have many practical uses. Coaches can use the information to sharpen their professional development agendas with practitioners. Administrators can use the information to assess the quality of training and coaching. Purveyors can use the information as a guide for implementation at the practice and program development levels. And, researchers can use the information as an outcome measure for some studies and as an independent variable in others.

The most effective intervention will not produce positive effects if it is not implemented. Thus, assessments of performance are a critical component of implementation.

