Challenges in Measuring the Fidelity of a Child Welfare Service Intervention

AMY STUCZYNSKI and MADELEINE H. KIMMICH
Human Services Research Institute, Tualatin, OR, USA

This article is a case study of an effort to implement and measure the fidelity of a child welfare intervention, family team meetings (FTMs). It describes the challenges practitioners faced in implementing FTMs. It discusses the evaluation challenges encountered in specifying, measuring, and reporting on model fidelity. It presents the findings of an interim fidelity assessment and the implications for next steps in the program evaluation. The authors reach a 3-fold conclusion: (a) the use of fidelity assessment can aid practitioners to focus on the specifics of their implementation; (b) qualitative factors related to program implementation and local context are critically important to evaluation efforts; and (c) fidelity criteria can provide the basis for targeted outcome measurement.

KEYWORDS fidelity, evaluation, child welfare, implementation

The current movement towards evidence-based practice (EBP) in the social services challenges practitioners and evaluators to pay increasing attention to interventions that have been shown to improve client outcomes. In the child welfare field, federal waivers to Title IV-E of the Social Security Act enable state child welfare programs to pursue systems reform by allowing them to redirect federal funds from foster care to alternative services for children suffering abuse or neglect. Consistent with the EBP movement, the Waiver program requires that the agencies’ efforts be evaluated to show how reform was achieved and whether it resulted in improved child and family outcomes.

Before evaluating the impact of services on child outcomes, it is critical to assess fidelity, in order to measure the extent to which the program was...
implemented as intended and thus would be expected to reach the expected outcomes (Crea, Usher & Wildfire, 2009). Acceptable levels of fidelity need to be established before an outcome evaluation is undertaken (Gilliam, Ripple, Zigler, & Leiter, 2000). In addition, fidelity assessment helps practitioners to make sense of the efforts they have undergone with children and families and provides a framework for evaluators to identify and acknowledge the constraints and uncertainty surrounding their findings.

This article is a case study of an effort to implement and measure the fidelity of a child welfare intervention, Family team meetings (FTMs), being used as the core strategy under Ohio’s Title IV-E waiver demonstration project. It describes the challenges practitioners faced in implementing FTMs and how this influenced the measurement of fidelity. It discusses the evaluation challenges encountered in specifying, measuring, and reporting on model fidelity. It also presents the findings of an interim fidelity assessment and the implications for next steps in the program evaluation. The article reaches a three-fold conclusion: (1) the use of fidelity assessment can aid practitioners to focus on the specifics of their implementation, (2) qualitative factors related to program implementation and local context are critically important to evaluation efforts, and (3) fidelity criteria can provide the basis for targeted outcome measurement.

CONTEXT: OHIO’S TITLE IV-E WAIVER DEMONSTRATION PROJECT AND EVIDENCE-BASED PRACTICE

In October 1997, the Ohio Department of Job and Family Services (ODJFS) began implementation of ProtectOhio, a demonstration project authorized under Title IV-E of the Social Security Act. Ohio’s Title IV-E Waiver focuses on overall reform of the child welfare system, by using flexible federal funds to support a wide array of services and supports for children and families impacted by child abuse and neglect. The Waiver offers Ohio an opportunity to alter the core fiscal dynamics of its county-administered child welfare system. By allowing foster care dollars to be used on services other than foster care, counties can expand supports to children in their own homes, reducing the need for out-of-home placement without any loss of federal revenue. This flexibility raises the question: what services do counties provide as an alternative to placement, and do the children receiving those alternative services avoid or delay out-of-home placement and remain safe?

Of Ohio’s county-administered public child-serving agencies (PCSAs) Fourteen opted to participate in ProtectOhio. (Because one county provided insufficient data and another county was temporarily removed from the Waiver, this article presents case-level data from 12 counties.) The participating counties were geographically spread throughout the state, ranging
in population density from urban to rural. Since 1998, Ohio has contracted
with an evaluation team led by the Human Services Research Institute (HSRI)
to conduct a rigorous evaluation of the ProtectOhio demonstration. The
evaluation design includes a comparison group consisting of 14 additional
counties with similar demographics to the demonstration counties.

Each county had its own ideas about how best to utilize the flexibility—
and how to handle the risk—of limited Title IV-E funds. Under the first 5-
year Waiver authorization in 1997, the focus was on “flexibility” and locally
developed strategies to reduce placement costs; the counties were given no
particular directives in how to use their IV-E funds. After 5 years, the external
evaluation reported a modest overall impact on child outcomes, but it could
not be attributed to any specific practice change.

During this time, increasing attention was being paid in the social ser-
vices field to identifying and implementing EBP, narrowly defined as using
empirically supported interventions. It is within this context that the federal
Children’s Bureau sought to establish better linkages between services and
outcomes. Thus the second Waiver authorization in 2004 mandated that the
counties focus on particular service interventions, including the requirement
that all counties participate in one core strategy.

The Ohio Waiver counties selected FTM as their common strategy be-
cause they were already experimenting with various forms of family meetings
under the first Waiver, staff were familiar with the philosophy and practice,
and they believed it to be a potent strategy. While this approach represents a
common way in which interventions are chosen, it does not reflect a consid-
eration of research evidence. Gambrill (2006) points out that, according to its
originators, EBP is more than simply using empirically supported interven-
tions, but “the integration of best research evidence with clinical expertise
and [client] values” (Sackett, Straus, Richardson, Rosenberg & Haynes, 2000,
as cited by Gambrill, 2006). As explained by Gibbs and Gambrill (2002), the
EBP approach involves line workers and clients in formulating research ques-
tions regarding decisions in their practice, searching electronically for the
answer, critically appraising what they find, carefully considering whether the
findings apply to a particular client, and then, taking into account the client’s
values and expectations, selecting an option to try and evaluating the results.

Family Team Meetings Practice and Research

FTM is a method for engaging family members and other people who can
support the family for shared case planning and decision-making. It is charac-
terized by regularly scheduled meetings; facilitated by a trained professional;
and bringing together family, friends, service providers and advocates. The
goal of FTMs is to design creative and effective solutions to case challenges,
ultimately reducing the need for foster care placement and improving per-
manency outcomes.
The child welfare field is increasingly embracing practices such as FTMs with the expectation that they will result in more appropriate and timely services to families. Different models of FTMs such as team decision-making (TDM) and family group decision-making (FGDM) have been promoted by the Annie E. Casey Foundation and American Humane Association, respectively, as promising practices. Counties throughout Ohio have experimented with these and other models of FTMs.

A significant amount of research on the process of FTMs has clarified how the models work and what participants think makes them work well (Crampton, 2007). Research has demonstrated that family members come to FGDM meetings when they are given an opportunity; they participate appropriately and develop plans that create stability for children; both family members and child welfare professionals are satisfied with the process; and children involved in meetings are more likely to be placed with members of their extended families (Merkel-Holguin, Nixon, & Burford, 2003).

To date, only limited evaluation has been done of FTMs models. The first full-scale external evaluation of TDM is scheduled for completion in late 2009 (Crea et al., 2009). Reviews of the limited research on the outcomes of FGDM has shown positive or neutral effects, but many of the studies suffer from small sample sizes or a lack of adequate comparison groups (Berzin, 2006; Crampton, 2007; Sundell & Vinnerljung, 2004). Furthermore, the studies have not addressed fidelity, which makes their findings difficult to interpret (see for example Gunderson, Cahn, & Wirth, 2003; Litchfield, Gatowski, & Dobbin, 2003; Penell & Burford, 2000). Thus, while FTM practice is popular and has been implemented around the world, there remain many questions about the effectiveness of FTMs. While FGDM and TDM may be considered promising practices, work on implementation fidelity is critically needed if FTMs are to have a solid evidence base.

Within this context, and without a proven model, the Ohio counties built on their existing practice and defined a common model that they could all agree to implement. This model had four specific components:

- Meetings are held over the entire period of ongoing services (sometimes called protective services), including at a minimum (i) within 30 days of case opening to ongoing services, and (ii) at least quarterly after the case plan is completed.
- A range of people attend the meetings: Participants may include the birth parents, primary caregivers and other family members, foster parents, support people, and professionals.
- Facilitators are trained and their responsibilities are to: arrange the meetings, help assure that participants attend and know what to expect (provide some orientation for potential participants), and support the family in the meetings and in preparing for them.
Regarding the population served, the counties agreed that FTM targets all children in newly opened ongoing/protective services cases where the initial case plan goal is to reunify or to maintain the child in the home. They cease FTM with the family when the case plan goal changes from reunification or maintain in home to something else, and when the child moves to permanent custody, planned permanent living arrangement, or legal custody to kin.

In addition, the counties articulated two other aspects of their model:

- The FTM process includes at least these components: agenda, introduction, information sharing, planning, and decision process.
- FTMIs are held at other critical events in the case, such as placement or custody changes.

While these pieces are important to the ProtectOhio practice, they have not been included as part of the fidelity assessment, partly because they have been difficult for the counties to know how to handle appropriately, and difficult for the evaluators to know what happened, as explained further below.

County leaders and FTM facilitators created a logic model to illustrate how model components were expected to lead to outcomes (Table 1). Specifically, they believe that families who receive FTM, characterized by frequent meetings that include a wide range of people, will be linked to more appropriate and timely services (for which the Waiver may help to provide funding), reducing the need for foster care placement and improving permanency outcomes.

Implementation Process and Challenges

Research has shown that not only is it important that EBP be used in selecting an intervention, but also that the intervention is put into practice using a specified set of planned activities. A recent review of the research on implementation concluded, “implementation appears most successful when:

- Carefully selected practitioners receive coordinated training, coaching and frequent performance assessments;
- Organizations provide the infrastructure necessary for timely training, skillful supervision and coaching, and regular process and outcome evaluations;
- Communities and consumers are fully involved in the selection and evaluation of programs and practices; and
- State and federal funding avenues, policies, and regulations create a hospitable environment for implementation and program operations.” (Fixsen, Naoom, Blase, Friedman & Wallace, 2005, p. vi).
TABLE 1 Logic Model for Family Team Meetings (FTMs)

<table>
<thead>
<tr>
<th>Inputs</th>
<th>Activities</th>
<th>Outputs</th>
<th>Outcomes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training</td>
<td>For children with case plan goal of reunification or maintain in home:</td>
<td>More appropriate and timely services</td>
<td>Avoid initial placements</td>
</tr>
<tr>
<td>Buy-in</td>
<td>1. FTMs over the entire period of ongoing services, including at a minimum</td>
<td>More clarity in case plans</td>
<td>Shorter time in placement</td>
</tr>
<tr>
<td>Child's age, previous</td>
<td>• Within 30 days of case opening to ongoing services,</td>
<td>Stronger family relationships, more natural supports</td>
<td>Of children who are placed, more children are placed with kin</td>
</tr>
<tr>
<td>child welfare</td>
<td>• At critical events in the case,</td>
<td>More consistent agency practice in deciding whether to place</td>
<td>More reunification</td>
</tr>
<tr>
<td>history</td>
<td>• At least quarterly.</td>
<td></td>
<td>Quicker permanence</td>
</tr>
<tr>
<td></td>
<td>2. Participants may include the birth parents, primary caregiver and other</td>
<td></td>
<td>Less re-entry to substitute care</td>
</tr>
<tr>
<td></td>
<td>family members, foster parent (if child goes to placement), support</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>people, and professionals.</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3. Facilitator responsibilities include: arrange meetings, help assure</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>that participants attend and know what to expect (provide some</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>orientation for potential participants), and support the family in the</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>meetings and in preparing for them.</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>4. The FTM process includes at least these components: agenda,</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>introduction, information sharing, planning, and decision process.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The Ohio demonstration counties conducted a set of implementation activities that resemble the previously noted list. Once the FTM model had been defined, counties independently began their implementation processes: they hired facilitators, developed internal policies and procedures for their FTM programs, and provided training or orientation for caseworkers. Most counties offered some sort of orientation to FTM practice for community partners. Because all counties operated under the ProtectOhio Waiver, all had the same degree of flexibility in using Title IV-E funds.

Nonetheless, the demonstration counties faced a number of challenges in establishing FTM. These challenges reveal both weaknesses in their implementation processes and weaknesses in the model definition. These implementation challenges have substantial impact on the sites’ ability to adhere to the defined FTM model, i.e., to have high model fidelity. These findings about the implementation process strongly underscore the central role of local contextual factors in fostering or impeding program fidelity.
The implementation challenges included:

- Working together to implement common service interventions: Public child welfare services in Ohio are county-administered and thus practice normally varies somewhat among any particular group of counties as they are accustomed to operating independently. Related to this independence, there was no single state-level leader or strategic champion for the initiative.
- Gaining worker, line supervisor, and sometimes management commitment (or buy-in) to the ProtectOhio FTM model: Even when staff were supportive, the agency sometimes lacked leadership commitment to a philosophy change and needed technical assistance and training in the new service intervention.
- Recruiting, training, and retaining qualified and effective FTM facilitators: Turnover among facilitators, arising from reassignment within the agency or resignations, made recruitment and training a necessity throughout the waiver period.
- Meeting the evaluation mandate to serve all cases transferring from intake to ongoing services: Despite agencies’ general desire to learn what works for whom, this requirement did not reflect what the agencies would have chosen to do, absent the evaluation.
- Incorporating new data collection processes into existing systems and building shared understanding of the correct usage of the data elements: Workers and facilitators often struggled to adapt to new data collection demands.
- Integrating the new practice into the larger patterns of agency practice: This required clarifying staff roles, incorporating legal requirements for case plan development and review, and maintaining confidentiality. Insofar as child welfare agencies routinely hold many case planning and review meetings, sometimes it was unclear how FTMs differed from other meetings held.
- Managing limited resources: Families requiring FTM services accumulate over time; unless the agency is immediately successful at shortening the time cases are open or otherwise streamlining the workload, they need to develop more and more FTM capacity.
- Inviting and getting the attendance of all relevant people at the meeting: If the primary concern in a case is parental drug abuse, it would be important to have the parent’s drug treatment counselor attend. However, third-party providers often found it difficult to attend meetings because of their own workloads or defined boundaries (e.g., they may not be able to bill time they spend in an FTM). For parents, the challenges included their own physical or mental health, nervousness about attending the meeting, and work or other constraints on their time.
- Gaining the cooperation of other parties: When the agencies were already holding so many meetings, it was difficult to maintain a schedule of
quarterly FTMs unless they could merge or integrate some of the existing meetings, which might have somewhat distinct stakeholders and agendas. In addition, the juvenile court’s receptivity to FTM impacted whether the decisions made in the meetings were supported and able to be implemented, or were overridden by the judge.

- Keeping focused on the intervention: Other major initiatives, such as the federally-mandated transition to the State Automated Child Welfare Information System, have demanded time and effort from PCSA staff.

These wide-ranging challenges made it very difficult for facilitators to demonstrate consistent fidelity to the FTM practice model. However, recognizing these issues helped the facilitators work as a group to improve the consistency of their practice. To assist them in their implementation, the facilitators began meeting together on a quarterly basis. These meetings provided an opportunity for the facilitators to clarify aspects of the practice model, review evaluation issues and data, and discuss other implementation challenges. Melde, Esbensen, and Tusinski (2006) note the importance of continued communication among program stakeholders as an important element in maintaining some level of cross-site consistency.

Measuring Fidelity

Building on the program implementation process is the process of measuring fidelity to the intended practice model. Fidelity can be seen in many ways (Dusenbury, Brannigan, Falco & Hansen, 2003). This study defines fidelity as whether the services as implemented conform to the intended model.

Researchers have consistently noted the importance of measuring fidelity and its necessity in determining program effectiveness (Melde et al., 2006; Zvoch, Letourneau & Parker, 2007). Fidelity measurement helps explain why outcomes are or are not met by identifying the mediating variables in a program (i.e., what has been changed and how those changes impact outcomes) (Dusenbury et al., 2003). When fidelity is poor, the effectiveness of a practice cannot be determined (Sosna, 2007).

When dealing with models that do not have a prior research base, “the appropriate use of fidelity criteria can assist program evaluation designs, with or without RCTs [randomized clinical trials], to contribute to establishing the evidence base for any program” (Mowbray, Holter, Teague, & Bybee, 2003, p. 316). In the early implementation phase, fidelity assessment can help establish when a model has been well implemented and is ready for an outcome evaluation (Crea et al., 2009).

Another reason for collecting data on program fidelity is to document problems in delivering programs in real world settings (Melde et al., 2006). When fidelity information is made available, it enhances the ability of program developers and those replicating programs to improve upon future
work, rather than repeating similar mistakes time and again (Melde et al., 2006). In evaluations involving multiple sites, fidelity information is particularly important, otherwise the evaluator must assume that the treatment was delivered to all recipients in the same manner across all sites (Zvoch et al., 2007). Fidelity measures document deviations from an intended model, both across time within sites and across sites (also called program drift) (Crea et al., 2009; Melde et al., 2006; Mowbray et al., 2003). By studying fidelity of implementation, one can assess the feasibility of an intervention, or how likely it is that the intervention can be implemented with fidelity (Dusenbury et al., 2003).

Fidelity measurement reveals not only problems with the program, but problems with the implementation process. Fidelity has been shown to be impacted by therapists’ beliefs and confidence (Symes, Remington, Brown & Hastings, 2006); intervention agents’ skill, training and supervision (Dumas, Lynch, Laughlin, Smith & Prinz, 2001); and shifting state and local policies which may change the treatment as usual or comparison group (Holden, Rousseau O’Connell, Connor, Brannan, Foster, Blau & Panciera, 2002).

But fidelity measurement is still relatively rare in program evaluation studies (Zvoch et al., 2007). The literature often lacks detailed descriptions of the development and use of fidelity criteria (Dusenbury et al., 2003; Mowbray et al., 2003). Measurement is often poor, commonly relying on self-reports or based on single observations (Dusenbury et al., 2003). Melde et al. (2006) stress the importance of using multiple methods when assessing fidelity, given evidence that the method of documenting fidelity may impact the results.

Constructing fidelity criteria “involves making many choices, especially when there is no existing program model with established efficacy on which fidelity criteria can be based.” (Mowbray et al., 2003, p. 333). Of primary importance in measuring fidelity is that specific components of the intervention are identified and well defined (Mowbray et al., 2003; Scheirer & Rezmovic, 1983). Mowbray et al. (2003) note that the methods of developing fidelity criteria should be described. While it is preferable that program manuals and literature reviews are used, it may be necessary to use experts or other qualitative methods when dealing with programs that are not tested in a thorough manner.

Mowbray et al. (2003) suggest that fidelity criteria should include aspects of structure (the framework for service delivery) and process (the way in which services are delivered). The authors note that when programs are initially being evaluated and fidelity criteria are first developed, an emphasis on structure over process items may be appropriate. Fixsen et al. (2005) point to two factors when measuring fidelity. First, it is necessary to explore both staff performance measures and organizational performance measures; successful implementation is not possible unless the organization has the capacity to adopt a new treatment model and staff are adequately trained.
and supported in administering the new model. Second, efforts to measure fidelity should focus on three areas: context (“prerequisites that must be in place for a program or practice to operate”), compliance (“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”), and competence (“level of skill shown by the therapist in using the core intervention components as prescribed while delivering the treatment to the consumer”) (pp. 47–48).

These factors are all key components necessary for successful fidelity exploration. At this stage, the interim fidelity assessment focuses on structure more than process, and on context and compliance more than competence. Limitations and plans for future work will be further addressed in text.

Family Team Meeting Fidelity Measurement and Results

The counties are seeking to develop a model that works and to use fidelity assessment to test the combination of elements that they have put together. In order to establish fidelity criteria the evaluation team worked closely with county staff to identify and define measures and set standards of performance. The evaluation team, in consultation with key stakeholders from the counties, determined that the model components would be measured at various levels using a combination of (1) semi-structured staff interviews conducted by members of the evaluation team and (2) meeting data recorded by the facilitators after each FTM in a database created specifically for the demonstration (Table 2). In addition, on-site observations by the evaluation team provide an opportunity to examine what occurs during the meetings.

<table>
<thead>
<tr>
<th>Model component</th>
<th>Data source</th>
<th>Variable level</th>
</tr>
</thead>
<tbody>
<tr>
<td>Meetings are held over the entire period of ongoing services</td>
<td>Demonstration database (administrative data provide additional information for analysis such as case opening dates)</td>
<td>Child</td>
</tr>
<tr>
<td>Participants may include birth parents, primary caregivers and other family members, foster parents, support people, and professionals</td>
<td>Demonstration database, observations</td>
<td>Child-level meeting</td>
</tr>
<tr>
<td>Facilitators are trained and their responsibilities include: arrange meetings, help assure that participants attend and know what to expect, and support family in meetings and in preparing for them</td>
<td>Staff interviews, observations</td>
<td>County</td>
</tr>
</tbody>
</table>
and see how it coincides with what the interviews and facilitator-reported data show. Meeting data is being analyzed at the child level because that is the level at which we will analyze outcomes. The final analysis will be able to take into consideration the number of children per meeting and may be able to consider a multi-level approach. The facilitator’s quarterly meetings provided a forum to review the model definition, articulating acceptable degrees of variation and reducing ambiguity. In particular, the evaluation’s focus on fidelity assessment provided some impetus for facilitators to specify the model, encourage conformance to it, and reconcile the model with day-to-day practice. From these conversations, performance benchmarks were established that reflected both current practice and expectations of what best practice would be (Table 3). The performance benchmarks set a high standard, with the idea that they would be used throughout the evaluation to assess change over time for each demonstration county. In addition, reaching the benchmarks would convincingly show that the practice was being consistently applied, one of the conditions required in order to link the resulting outcomes to the intervention.

To pilot the performance benchmarks, and to provide an interim assessment to aid the counties in refining their practices, the performance benchmarks were applied to the meeting data from the first seven to twelve months of implementation of the ProtectOhio FTM model in each county. Subsequent measurement will look at data covering as much as four years of implementation. The findings are reported in Table 4 and further described below (Human Services Research Institute [HSRI], 2007). They show wide variation among counties on each component, pointing to the importance of

<table>
<thead>
<tr>
<th>Component</th>
<th>Definition and benchmark</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial FTM within 30 days</td>
<td>At least 90% of children who had FTM had their initial FTM within 35 days before or after the case opening date.¹</td>
</tr>
<tr>
<td>Second FTM within 90 days</td>
<td>At least 90% of children who had FTM had their second FTM within 100 days of the previous FTM.</td>
</tr>
<tr>
<td>Range of attendees at meetings</td>
<td>At least 75% of meetings included the minimum participant grouping: at least one parent or primary caregiver, at least one staff person, and at least one other type of person.</td>
</tr>
<tr>
<td>Facilitators are trained and help organize meetings</td>
<td>Based on information from county interviews, the county must have used a trained facilitator for the entire study period, and that facilitator must play at least some role in helping to organize meetings.</td>
</tr>
</tbody>
</table>

Note. ¹Evaluation team chose to use a 35-day and 100-day window when making judgments about the timing of initial and quarterly meetings to avoid issues where weekends or holidays might impact when a meeting was held.
TABLE 4  Fidelity Assessment Findings

<table>
<thead>
<tr>
<th>Component</th>
<th>Overall rate</th>
<th>Range among counties</th>
<th>Counties (12) that met fidelity benchmark</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial FTM within 30 days of case opening to ongoing services</td>
<td>76%¹</td>
<td>20% to 100%</td>
<td>3</td>
</tr>
<tr>
<td>Subsequent meetings at least quarterly</td>
<td>81%²</td>
<td>12% to 98%</td>
<td>4</td>
</tr>
<tr>
<td>Meeting participants</td>
<td>50%³</td>
<td>19% to 91%</td>
<td>4</td>
</tr>
<tr>
<td>Facilitators are trained and help organize the meetings</td>
<td>NA</td>
<td>NA</td>
<td>4</td>
</tr>
</tbody>
</table>

Note. ¹n = 2,165 children who received at least one FTM; ²n = 1,429 children who were eligible for a subsequent FTM (i.e., child's case was still open 100 days after the preceding FTM); ³n = 4,173 child-level meetings (this figure represents the count of meetings held for each child, summed across all 2,165 children who had one or more FTMs).

local contextual factors through which program implementation and fidelity is filtered.

INITIAL FAMILY TEAM MEETING WITHIN 30 DAYS OF CASE OPENING TO ONGOING SERVICES

Through the quarterly facilitators’ meetings, measurement of this model component was clarified to encompass initial meetings held either before or after the transfer to ongoing services, thus respecting practice differences in the various counties. Across all 12 counties, 76% of the 2,165 children receiving FTM had their first meeting within 35 days of the case opening to ongoing services (a 35-day measurement window was used to avoid issues where weekends or holidays might impact when a meeting was held). At the individual county level the rate ranged from 20% to 100%. Three counties met the fidelity benchmark, having at least 90% of the eligible children receive their first meeting within 35 days.

SUBSEQUENT MEETINGS AT LEAST QUARTERLY

Through discussions with FTM facilitators and staff, the evaluation team clarified the requirement for quarterly FTMs to mean that if a meeting is not held for some other reason (i.e., a critical event), the ProtectOhio model calls for meetings to be held at least quarterly (measured as within 100 days of the previous FTM). Across all 12 counties, 81% of the 1,429 children who were eligible for a second FTM (that is, the child’s case was open long enough that we would expect to see another FTM) met the criterion—they had their second FTM within 100 days of the previous FTM. At the individual county level the rate ranged from 12% to 98%. Four counties met the fidelity benchmark, with at least 90% of children having their second FTM within
100 days. (At the time of this fidelity assessment, only a small number of children were eligible for a third FTM. Therefore, the data were not included in this fidelity assessment. In subsequent analyses, sufficient data should be available to generate results for third- and higher-order meetings.)

**Meeting participants**

The ProtectOhio model does not specify what participant grouping is the minimum standard for a meeting, merely stating that meeting participants may include the birth parents, primary caregivers and other family members, foster parents, support people, and professionals. Across all 12 counties, only 50% of the 4,173 child-level FTMs recorded were attended by what the evaluation team judged to be an absolute minimum mix of people: at least one parent or primary caregiver, at least one caseworker or other agency staff, and at least one other type of person. Four counties met the fidelity benchmark that at least 75% of meetings included this minimum participant grouping. At the individual county level, the number of meetings that included this minimum participant grouping varied from 19% to 91%. These results are prompting a further look into what kind of participant grouping is necessary for a productive FTM.

**Facilitators trained and help organize meetings**

In many projects, fidelity is enhanced through the use of uniform training (Dumas et al., 2001). A three-day training session was provided for facilitators at the beginning of the initiative, with two shorter trainings taking place subsequently, targeted to new hires. Evaluation team members attended parts of these trainings and documented what was discussed. However the evaluation is limited in that it did not measure participants’ reactions or learning. While the lack of uniformity across training events is a limitation to making a judgment about fidelity, the model definition and fidelity benchmark merely required that all facilitators had received some sort of training in FTM before facilitating meetings. This benchmark was met in nine of 12 counties.

In addition, the fidelity benchmark required that the facilitator take responsibility for some tasks—arranging the meetings, helping assure that participants attend and know what to expect, and supporting the family in the meetings and in preparing for them. In fact, interviews with county supervisors and FTM facilitators found that caseworkers played a large part in arranging the meetings and inviting participants, and that facilitators had some role in arranging FTMs in only six of 12 counties. Four counties met both this and the training criteria.

This raises the question of whether the counties still believe that trained FTM facilitators might do a better or more consistent job of arranging and preparing participants for meetings. Based on our interviews and observations, it appears that caseworkers tended to take on these responsibilities
due to three main reasons: (1) The facilitator’s workload is often too large for them to take on these tasks; (2) the caseworker has already established a relationship with the family so, at least for the first FTM, the familiarity is an asset; and (3) in striving to maintain an independent perspective in the meeting, facilitators prefer not having any extended contact with the family or caseworker prior to the first meeting. Often, subsequent FTMs are scheduled at the prior FTM, making the arrangements much simpler to do.

COUNTY-LEVEL FIDELITY

The demonstration counties varied considerably in the extent they implemented the FTM strategy in strict accordance with the model. These variations stem from multiple implementation challenges, as described in detail above, and reflect only 1 year of FTM practice. Again, the purposes of this initial fidelity assessment were to pilot the fidelity measures and to provide some interim feedback that would help the counties to focus on the model components. Future fidelity assessments will be conducted using 4 years of FTM data, allowing the evaluation team to measure the extent to which full implementation is reached after additional time.

Table 5 displays county-by-county scores on each of the four FTM fidelity criteria. At 1 year after strategy implementation, the overall performance of the counties is modest: only two counties showed high fidelity on at least three of the four criteria. The counties tended to do well in holding the first meeting on time but not as well in having timely subsequent meetings. It is interesting to note that counties that met the meeting timing criteria tended to do less well on the range of attendees component, and,

<table>
<thead>
<tr>
<th>TABLE 5 Fidelity to the ProtectOhio Model: Selected Components</th>
</tr>
</thead>
<tbody>
<tr>
<td>County</td>
</tr>
<tr>
<td>--------</td>
</tr>
<tr>
<td>Target</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>3</td>
</tr>
<tr>
<td>4</td>
</tr>
<tr>
<td>5</td>
</tr>
<tr>
<td>6</td>
</tr>
<tr>
<td>7</td>
</tr>
<tr>
<td>8</td>
</tr>
<tr>
<td>9</td>
</tr>
<tr>
<td>10</td>
</tr>
<tr>
<td>11</td>
</tr>
<tr>
<td>12</td>
</tr>
<tr>
<td>Total</td>
</tr>
</tbody>
</table>
conversely, counties that met the range of attendees criterion tended to score lower on the meeting timing elements. Perhaps pressing to meet certain timelines caused meetings to be held at times which were not convenient to all involved parties, a finding that was also suggested by Crea et al. (2009). At this point it is not clear whether or not certain components are critical to children’s outcomes, but it is something that could be rigorously examined if stronger fidelity is reached.

Evaluation Challenges

To better contextualize the findings reported above, it is necessary to acknowledge and discuss the associated limitations of the evaluation, especially as related to fidelity assessment. Three of the most critical evaluation challenges include model definition, fidelity measurement, and ongoing monitoring and reporting.

Challenges with the Model Definition

True to the independent nature of county child welfare services in Ohio, each county developed its own implementation steps and schedules, often affected by each county’s prior experience with some form of FTM. As counties began implementing the model, less precise aspects of the model definition came to light and became subject to interpretation. Given the limited availability of technical assistance and leadership to push adherence to the model, county practice gradually diverged. When this combined with the difficulties encountered in implementing the new model, the pressure to change or re-interpret the model was great. The counties and evaluation team agreed to make certain modifications to the model after it was initially defined, which allowed greater flexibility in implementation (e.g., considering an initial FTM to occur within the required timeframe if it occurred either 35 days before or after the case transfer date). This is consistent with suggestions that evaluators be flexible and responsive to the organizational realities and logistical obstacles practitioners identify (Rubin, 1997). Sensitivity to the local context is an important factor in encouraging decision makers to take evaluation to heart (Weiss, Murphy-Graham, Petrosino, & Gandhi, 2008). Reinvention or adaptation may be necessary to tailor programs to the needs of a particular setting and to promote ownership (Dusenbury et al., 2003).

As is often the case, this model experienced a tension between a research need for specificity and a practice demand for something that is feasible. While the model was broad from its inception, a stricter model may not have been desirable and would likely have had less support from the counties, leading to even poorer implementation. One thing is clear: the act of assessing fidelity led to a much more intense focus on the part of
practitioners on what exactly was being implemented. Having well-defined model components and performance benchmarks gave practitioners something to latch onto and quickly prompted detailed questions about implementation. While initial scores on the fidelity criteria varied considerably among counties and were very low in a few sites, we believe that focusing on fidelity brought more consistency to the cross-county effort, facilitating efforts to implement the ProtectOhio FTM model and to begin to evaluate its effectiveness.

Looking at the big picture, Mowbray et al. (2003) describe the need to balance structural (what is done) and process (how it is done) fidelity criteria. Using this conceptualization, the ProtectOhio FTM model components provide a clear structure that can be evaluated, in terms of frequency of meetings and who attends them, but describe little of the process. In so doing, the model may have missed the heart of what makes the intervention work. On the other hand, it may more closely reflect the nature of social services, allowing some room for professional judgment on the part of program providers (Babor, Steinberg, McRee, Vendetti, & Carroll, 2002). Similarly, following the construct of Fixsen et al. (2005), this fidelity assessment examines factors related to context and compliance, but has little to offer in terms of measuring competence. In order to address this, the evaluation team has increased the number of meeting observations it is conducting, which has enabled examination of the meeting process and roles and skills of facilitators and caseworkers. We expect to observe approximately 50 meetings across all the demonstration and comparison sites (using a convenience sample), and the findings will be reported as part of the process evaluation.

CHALLENGES WITH MEASUREMENT

Fidelity criteria and measurement methods need to be viable for the long haul if one expects to apply them over time and to connect them to outcomes. Due in part to the nature of the model that was defined, the evaluation chose to limit the counties’ data collection to concrete items such as dates of meetings and meeting attendees. Items that would measure the way services are delivered (i.e., the content of the meeting) are reserved for observations, focus groups, and interviews conducted by the evaluation team. This arrangement balances the need for specific data with the need to ensure that data is not too onerous to collect. In this manner the evaluation has managed to collect child-level data on FTM dates and meeting attendees for more than two years. Rather than being an isolated scientific exercise, data collection becomes relevant to counties’ own information needs, making it more likely that they will continue measurement after the end of the demonstration.

Unfortunately, data imprecision limits the ability to comprehensively assess fidelity. Certain aspects of the model have proven very difficult to
accurately measure. For example, it is difficult to accurately establish whether an FTM is being held due to a critical case event. The data do not specify when a critical event is avoided due to an FTM (because no placement or move occurs). The model component speaking to the content of the FTM process (i.e., that the meeting process includes presentation of the agenda, introductions, information-sharing, etc.) attempts to clarify what happens in meetings, although it is broadly defined in order to be acceptable to all counties. Thus far, these aspects of the FTM model are being explored in interviews and meeting observations, while the interim fidelity assessment focused on aspects of the model that were more readily measurable.

For those measures that have been examined, another continuing tension is how to define high fidelity. Applying a high standard would enable the evaluation to more clearly attribute model effectiveness, but the range of measurement challenges and the external influences that shaped implementation may argue for a different approach. While it is useful to keep some consistent fidelity measures, there may be additional, intermediate measures that would provide a more nuanced understanding. For example, there is more that can be learned about the meeting participant mix, beyond whether or not a minimum grouping was present.

Challenges with Monitoring and Reporting

The evaluation team has consistently faced the tension between answering practitioners’ questions about what they should be doing, thereby promoting consistent understanding of the model, and simply observing their implementation process. The near-absence of leadership for the initiative meant there was little outside pressure to promote adoption of the specific FTM model. In addition, the inadequacy of training and technical assistance resources meant there were limited external resources that could be consulted, impeding the quality and consistency of the model.

Given the difficulties faced in implementation and model definition, the evaluation team found it useful to frequently share data with the counties. After a few months of implementation, evaluators held individual meetings with each county to review its data and to discuss how well it appeared to be meeting the fidelity criteria. The evaluators continue to share data at quarterly facilitators’ meetings and bi-monthly meetings of the county leaders, finding that this preliminary feedback helps the evaluators to better understand the counties’ practices, thereby measuring fidelity in a more meaningful way, as well as helping the counties to reflect on their practices and the need to make changes. It is important that evaluators’ discussions with the county agencies target both directors and practitioners, as communication within the agencies about the demonstration is often overlooked in the face of heavy day-to-day responsibilities.
IMPLICATIONS OF IMPLEMENTATION FINDINGS AND FIDELITY RESULTS FOR FUTURE EVALUATION AND PRACTICE

The implementation findings and the fidelity results point to the mediating influence of the local context: the variations across the demonstration sites in leadership, practice philosophy, staff motivation and workload, as well as fiscal and administrative resources, all play a strong role in the process of adopting an innovative child welfare intervention. The lack of any state countervailing pressure—to foster innovation despite local barriers—further burdens the reform initiative. For Ohio to seriously embrace FTM as a promising practice and to commit to testing its efficacy in a rigorous fashion will take much more time and resources than have been heretofore applied; and this is an unlikely prospect, given the challenging state economy and weakened infrastructure.

While expectations for FTM practice fidelity may be low for the near future, it remains the case that both sponsors and practitioners want to know how the initiative is proceeding—how have we done, despite the myriad challenges? The fidelity assessment conducted on the initial 4,173 child-level FTMs has provided participant counties with preliminary feedback to encourage continuing implementation and evaluation, and has offered insights about practice feasibility. As the demonstration nears completion, the following issues will need to be addressed in order to provide a better understanding of fidelity and a more meaningful contribution to practice.

UNDERSTANDING REASONS FOR POOR PERFORMANCE IN FIDELITY ASSESSMENT

Poor fidelity may be the result of a poor program (weaknesses in the model design) or poor implementation (due to political and other contextual factors). It may be more acceptable politically for practitioners to say that a model is not feasible, rather than to say “we don’t want to do this.” For this reason, the evaluation team is particularly interested in examining two issues: the steps counties took to build buy-in and the identification of any factors influencing implementation that appear to be immutable. As Melde et al. (2006) state, it is necessary to question the extent to which program failure is an artifact of poor programs or poorly implemented programs, and for this reason evaluators need to know the extent to which a program is implemented before speaking to the issue of effectiveness.

LEARNING FROM THE VARIED APPLICATIONS OF A SINGLE-MODEL INTERVENTION

Though there may not ever be strong fidelity to the ProtectOhio FTM model across all sites, the field can nonetheless benefit from knowledge of how individual sites accomplished implementation and under what conditions.
Supplementing the fidelity assessment work with case studies will help identify aspects of the individual county environment that impact implementation, fidelity, and outcomes, leading to a better description of the practice changes that occurred. Information on how the practice was carried out is still needed and important, and often generates the most interest from practitioners (Weiss et al., 2008).

Measuring and Interpreting Outcomes

If the county agencies are not consistently following the model, it will be difficult for hypothesized changes in child outcomes to materialize. Either no changes in outcomes will occur, because of an inconsistent model within a county or because the various sites cancel each other out; or any observed changes in outcomes will not be able to be attributed to the intervention. This makes fidelity assessment even more important. Fidelity criteria can provide a basis for analyzing separately data from sites that deviate too far from the model (Mowbray et al., 2003). Thus, children in high-fidelity programs, or cases with strong fidelity on a particular component, can be examined separately. In the outcomes analysis planned for the end of the second waiver period, we will examine groupings of children according to their level of fidelity to the FTM model. This will help clarify whether close adherence to the FTM model is related to improvements in outcomes. Unfortunately, as is often the case, the difficulties faced in implementing the model will make it difficult to have a single, strong model to evaluate.

This case study has illustrated the widespread challenges encountered in the child welfare field when policy makers seek to introduce an innovative practice model without the requisite political, financial and technical support. This situation creates a tension with evaluation demands, because the lack of support leads to an uneven implementation process and low levels of fidelity to the model intervention. Child welfare professionals are pursuing various forms of FTMs with great enthusiasm and high expectations, and program evaluation that includes fidelity assessment is essential to show that it can achieve positive child outcomes. While the study described in this article is not able to yield the rigorous outcomes findings desired by policy makers, it has nonetheless revealed the importance of FTMs having a well-defined structure. Practice can thus be measured against the model, and the relationship between the practice model and changes in child outcomes can be fully examined.

Model fidelity is the essential starting point for EBP research. It reveals problems with the implementation process and clarifies what was implemented so that an examination of outcomes is meaningful. If the child welfare field is to become more evidence-based in its clinical practices, more attention must be given to the process of fidelity measurement. Evaluators need to take fidelity assessment into account beginning with their evaluation designs.
Methods of measuring fidelity that take local contextual factors into account and leave room for practitioner judgment need to be developed. Researchers and practitioners need to look beyond studies that simply produce outcomes without reporting on contextual factors, implementation processes, and program fidelity.

REFERENCES


**CONTRIBUTORS**

**Amy Stuczynski** is Research Associate, Human Services Research Institute, Tualatin, OR.

**Madeleine H. Kimmich** is Senior Research Fellow, Human Services Research Institute, Tualatin, OR.