Published in final edited form as:

Child Youth Serv Rev. 2014 April ; 39: 153–159. doi:10.1016/j.chilyouth.2013.10.001.

From Practice to Evidence in Child Welfare: Model Specification and Fidelity Measurement of Team Decisionmaking

Sarah Kate Bearman,

Ferkauf Graduate School of Psychology, Yeshiva University

Ann F. Garland, and

University of San Diego, Department of School, Family, & Mental Health Professions

Sonja K. Schoenwald

Medical University of South Carolina, Department of Psychiatry and Behavioral Sciences

Abstract

Fidelity measurement methods have traditionally been used to develop and evaluate the effects of psychosocial treatments and, more recently, their implementation in practice. The fidelity measurement process can also be used to operationally define and specify components of emerging but untested practices outside the realm of conventional treatment. Achieving optimal fidelity measurement effectiveness (scientific validity and reliability) and efficiency (feasibility and relevance in routine care contexts) is challenging. The purpose of this paper is to identify strategies to address these challenges in child welfare system practices. To illustrate the challenges, and operational steps to address them, we present a case example using the “Team Decisionmaking” (TDM; Annie E. Casey Foundation) intervention. This intervention has potential utility for decreasing initial entry into and time spent in foster care and increasing rates of reunification and relative care. While promising, the model requires rigorous research to refine knowledge regarding the relationship between intervention components and outcomes—research that requires fidelity measurement. The intent of this paper is to illustrate how potentially generalizable steps for developing effective and efficient fidelity measurement methods can be used to more clearly define and test the effects of child welfare system practices.

Keywords

Fidelity measurement; Team Decisionmaking

The quest to ensure an individual with a particular health problem received effective treatment for the problem—regardless of the individual’s demographic characteristics, geographic location, payer plans, and the practice preferences of local physicians—catalyzed research on the nature and implementation of “evidence-based medicine” (see, e.g., Grimshaw et al., 2001; Grol & Grimshaw, 1999). That quest extended relatively

quickly to the realm of mental health care, and, more recently, to the range of human services provided by child welfare service systems.

Research on the implementation and outcomes of evidence-based psychosocial interventions and other practice innovations by child welfare systems has been relatively sparse (Aarons, Hurlburt, & Horwitz, 2011). This may be due in part to features unique to the child welfare sector. For example, child welfare, but not mental health, agencies are mandated to execute several of the following distinct objectives: investigation of abuse or neglect for the purposes of legal prosecution; child protection from abuse or neglect; provision of health and education for children removed into protective custody; termination of parental rights; facilitation of adoption; family reunification; and, in some systems, provision of treatment to parents in the service of safe reunification. There is inherent indeterminacy and variability in human services designed to address such objectives, and the services are often loosely specified (Glisson, 1992). In efforts to bring clarity to areas in which services are loosely specified, child welfare organizations may emphasize rules, conformity, and adherence to organizational procedures and authority. Such attempts may be “a misguided effort to inject certainty into what is an inherently uncertain technology” (Glisson, 2002, p. 237).

More promising approaches are guided by program theories or conceptual models which posit that a particular objective can be met by taking specific actions; however, the extent to which the actions are implemented, and objectives are met, are often unknown, or are evaluated via uncontrolled or qualitative evaluations. Efforts are underway to more clearly and systematically define the unique objectives of child welfare systems (e.g., to decrease out-of-home placement without increasing the incidence of abuse and neglect), specify strategies to meet these specific objectives, and evaluate the implementation and effects of these strategies (see, e.g., Kaye & Osteen, 2011; Stuczynski & Kimmich, 2010). The challenges that characterize such efforts closely mirror those confronted in research on the development, effectiveness, and implementation of any effective health and mental health care intervention: to define the intervention (or “program” model), develop adequate (reliable, valid, and feasible) indicators of intervention fidelity, and measure the implementation and effects of the intervention on desired outcomes.

To the extent the goals of the child welfare and other health sectors differ, the strategies used in each sector to meet those goals could logically be expected to differ (Chen, 1990); to the extent the goals of each sector are similar, one might logically expect similar strategies to be implemented, and with similar effects. To improve the quality and effectiveness of services provided by and through child welfare systems, discernment is needed regarding the objectives and functions unique to that system, as well as those also executed by other service sectors. Evidence is needed regarding the extent to which strategies found effective for meeting a goal in one service sector effectively implemented—in another. For objectives unique to the child welfare sector, unique strategies may need to be developed and tested; or, previously tested strategies may need to be adapted for use by child welfare systems and their effects evaluated.

In this paper we focus on the benefits and challenges of developing adequate indicators of intervention fidelity; that is, the extent to which the intervention is delivered as intended

(i.e., specified procedures are implemented, proscribed procedures are not). The development of intervention fidelity indicators drives the need for clarity in specification and operational definition of essential practice components. Thus, fidelity measurement methods are not just tools to assess the implementation and effects of a treatment or practice that is already evidence-based, but also can be used to build an evidence base for a newly emerging practice, or for a practice that is already in use, but untested.

This paper illustrates, using a case example, how the process of developing fidelity measurement methods that are both effective (characterized by evidence of valid and reliable use of scores) and efficient (feasible and relevant in routine care) (Schoenwald et al., 2011) can be used to more clearly define practices developed by child welfare systems and assess their implementation and effects. The “Team Decisionmaking” (TDM; Annie E. Casey Foundation) intervention is an example of such a practice, currently underway. We briefly describe the TDM model, rationale related to its potential utility for decreasing foster care utilization and increasing rates of reunification and relative care, and the need to measure fidelity to the model. Then, we recap a framework to guide development of effective and efficient fidelity measurement methods (Schoenwald et al., 2011) and consider the operational steps in the development of TDM fidelity measurement methods in light of this framework. We also briefly describe the empirical test of the TDM fidelity measurement procedures that is currently underway.

Family Engagement in Placement Decisionmaking

Background and Impetus for Implementation Research

In recent years, particular attention has been focused on the need for increased family engagement in what has historically been the adversarial process of decision-making regarding child removal and out-of-home placement in response to safety concerns (Berzin, Cohen, Thomas, & Dawson, 2008). In theory, collaborative efforts among agencies tasked with child protection and the families, community support members, and youth most impacted by agency involvement should yield more creative and acceptable solutions to case challenges, and contribute to reductions in the use of residential foster-care, as well as increased permanency outcomes. A number of family engagement models for child removal and placement have been promoted in recent years (American Humane Organization), including Family Group Decision Making (FGDM), Family Group Conferencing (FGC), Family Unity Meetings (FUM) and Team Decisionmaking (TDM). All of these approaches are characterized by at least one formally scheduled meeting, facilitated by a trained professional, and attended by family, friends, service providers and advocates (Stuczynski & Kimmich, 2010).

Research examining these family engagement models has provided some information about how families referred for services are selected for meetings (Crampton, 2007), scope of site implementation following agency initiation of family team meetings (Crea, Crampton, Abramson-Madden, & Usher, 2008), and participant perceptions of meetings (Rauktis, Huefner, & Cahalane, 2011). Although it is widely believed that these practices show “promising outcomes” (Pennell & Anderson, 2005, p. 4), a randomized clinical trial of one variant, FGDM, did not show statistically significant positive outcomes—placement

changes, family stabilization, or length of time to reunification—for youth receiving the intervention compared to those receiving traditional services (Berzin et al., 2008). Unfortunately, consistent with much of the research in this area, there was no measure of intervention adherence or differentiation across conditions in this trial, thus prohibiting conclusions regarding the effectiveness of the FGDM model. For example, the researchers noted that in each agency where the experimental intervention was implemented, caseworkers were exposed to the philosophies and principles of FGDM, which may have led to contamination of the comparison group—thereby decreasing the noticeable differences between the two groups. In this instance, systematic measurement of the extent to which both groups used the prescribed elements of FGDM—i.e., FGDM fidelity—would have clarified whether the two interventions were indeed different. Likewise, the researchers posited that there may have been site (agency) differences in FGDM implementation that attenuated its effects. Here, too, fidelity measures assessing the integrity of the intervention to the theoretical model at each site would have produced data to discern the potential efficacy of the FGDM model from problems with its implementation (Berzin et al., 2008).

In general, studies assessing the impact of family engagement practices on outcomes of interest within the child welfare sector have not adequately addressed practice fidelity, obscuring our understanding of the results (see, e.g., Gunderson, Cahn, & Wirth, 2003; Litchfield, Gatowski, & Dobbin, 2003; Pennell & Burford, 2000). Without measurement of implementation fidelity, it is impossible to know the extent to which an intervention is delivered as intended. When fidelity to the intervention is documented, the association between the intervention and the outcome of interest can be interpreted with greater confidence. Moreover, without fidelity measurement, it is challenging for agencies to pinpoint areas where practice needs refinement or improvement, and it is impossible for policymakers to differentiate between potentially effective programs that were implemented poorly and those that are not efficacious (Bellg et al., 2004; Calsyn, 2000). Finally, evaluation of associations between fidelity to the components of a theoretical model of the intervention in question, and outcomes of interest, is needed to infer that observed outcomes are attributable to intervention effects. Such evaluation facilitates discernment of the essential, non-essential, and innocuous—or, worse—detrimental, components of the theoretical model of the intervention. To identify and differentiate components that are essential and non-essential, and positively or negatively related to the desired outcome, we need accurate and acceptable measurement strategies. Thus, while family engagement strategies are popular and have been applied widely, assessing the fidelity of their implementation is a necessary step towards establishing the effectiveness and public health utility of these strategies.

The Team Decisionmaking (TDM) Model

Background

Team Decisionmaking is among the most evaluated of family engagement strategies, and is a core component of Family to Family (F2F), a national child welfare reform initiative sponsored by the Annie E. Casey Foundation. TDM meetings invite key agency personnel, birth family, community support people and extended family to convene for any state-

proposed youth removals or placement changes in response to safety concerns (Crampton, Usher, Wildfire, Webster, & Cuccaro-Alamin, 2011). Using specified strategies (DeMuro & Rideout, 2002) and skillfully facilitated group process, TDM aims to effectively involve families and their natural support systems in proposing recommendations for youth placement decisions.

The growing number of youth in care of the child welfare system, coupled with a decline in the number of foster families nationally, has led to the increased placement of youth in group or residential settings, or with family members that are not able to provide appropriate care. The F2F initiative grew out of an attempt to generate alternatives to these sub-optimal placements, while acknowledging that fewer than 50% of youth requiring removal could be placed in foster care (DeMuro & Rideout, 2002). F2F focused, therefore, on maximizing families' natural resources and strengths in order to promote safe reunification, finding and maintaining local supports to address families' needs or appropriate temporary kinship placement, engaging family members in all placement decisions, and using outcome data to evaluate practice. The Annie E. Casey Foundation has selected TDM as a stand-alone component of F2F to be studied. TDM is based on the assumption that families treated with respect can better identify the needs of their family and children. Further, when families are actively involved in the decisionmaking process, their investment in the services and tasks prescribed will be enhanced (DeMuro & Rideout, 2002). TDM has eight "essential elements" that were distilled from the practice experiences of multiple agencies; these are: Teamwork, Consensus, Active Family Involvement, Skillful Facilitation, Safety Planning, Strength-Based Assessment, Needs-Driven Services, and Involvement of the Community into Long-Term Support Networks.

There are proximal and distal intended outcomes of TDM. Proximal outcomes include increasing client satisfaction and engagement, and increasing the family's connection to services within its community. More distal outcomes include reducing the number of youth who are placed out of home overall, and—of those placed—increasing the number of youth placed in less restrictive placements, with relatives, and in the same placement as siblings. Other outcomes of interest for TDM include reductions in change of placement, placement reentry after exit, reduced moves to more restrictive placements, and reduced use of long-term foster care. These outcomes logically follow the guiding assumptions of TDM: a group can be more effective in decisionmaking than an individual; families are experts on themselves; members of the community add value to the process by serving as natural allies to the family and experts on the community's resources; and, as noted previously, when families are respectfully included in the decisionmaking process, they can identify and participate in meeting their own needs (<http://www.aecf.org/MajorInitiatives/FamilytoFamily/Resources>).

TDM has been implemented by child protective agencies in 60 cities across 17 states in the last decade (Crea, Wildfire, & Usher, 2009). Despite its widespread use, formal evaluation of TDM has been limited to largely descriptive research regarding the implementation process. To date, there is limited controlled research examining the prospective, predictive impact of TDM on outcomes such as permanency, placement disruption, or repeated child welfare involvement, relative to other services. Where TDM has been evaluated, it has

shown promising associations with reunification and relative care (Crampton et al., 2011); however, this was not a controlled trial involving random assignment.

It has been suggested by some researchers that traditional methods for evaluating the effectiveness of TDM may not be appropriate—specifically, Usher and Wildfire (2003) have suggested that randomized designs, often considered the “gold standard” for measuring intervention effectiveness (Kessler, Gira, & Poertner, 2005), may not be able to manage interventions that are necessarily systemic and contextual. Similar perspectives have historically been proffered with respect to a range of human service and social welfare programs; one consequence of which has been the proliferation of programs absent evidence of their effectiveness (Adams, 1994; Bickman, 1996; Weiss, 1972). Accordingly, efforts are underway to fully operationalize the TDM model, identify critical components of intervention fidelity and develop reliable and valid measures of this fidelity, and ultimately, to examine the efficacy of TDM in a randomized trial.

Development of TDM Intervention Fidelity Measurement Methods

Table 1 lists the broad goals and steps for developing and evaluating effective and efficient fidelity measurement methods, adapted from the framework of Schoenwald and colleagues (2011). That framework identifies attributes of measurement methods and treatment contexts to be considered conjointly in the quest to develop instruments that are both psychometrically valid and usable in routine care settings. The first column of the table identifies general steps in the development of effective and efficient fidelity measurement methods. The second column in Table 1 provides examples of how the general step is being operationalized for the TDM intervention; each of the four steps in the measurement development process is described in the text that follows. The two columns at the right side of Table 1 reference examples of how the TDM measurement development process reflects attention to criteria for both effective (scientifically valid and reliable) and efficient (feasible and relevant for use in routine care contexts) measurement.

1. Convene task force to identify purpose(s) of measurement—A task force dedicated to developing and testing a fidelity measurement system for TDM was established in summer 2011, sponsored by the Child Welfare Strategy Group of the Annie E. Casey Foundation. Throughout the process, the TDM task force has tried to maintain a dual focus on (a) rigorous research methodology (i.e., effective measurement focus), and (b) constructing a system that would have clear clinical and administrative utility and feasibility within the agencies for which TDM is already routine practice (i.e., efficient measurement focus). Selection of group members thus reflects complementary expertise regarding these dual objectives, including stakeholders with TDM practice expertise, a former child welfare administrator who oversaw the implementation of TDM in a large state system, the director and associate director of the AECF Child Welfare Strategy Group, and the director of a Department of Children and Family Services where TDM is currently practiced, who also developed the guidelines for TDM. The workgroup also included researchers who had previously studied the implementation of TDM, as well as a researcher with expertise in the development of effective fidelity measurement for evidence-based practices.

Given that different measurement purposes require differential emphasis on effective vs. efficient measurement, one of the most important steps in measurement development is to identify how fidelity measurement will be used (Schoenwald et al., 2011). Members of the task force agreed on two primary purposes of fidelity measurement. First, the measures would be used to assess implementation fidelity for future research comparing TDM to usual services for removal and placement change decisions. Second, the fidelity measurement would be used in an ongoing fashion to provide practical feedback on the implementation of TDM by individual facilitators, or to examine the fidelity standards of entities such as agencies or states. These purposes provided some useful parameters that guided the development of the measurement system. For use in a rigorous research trial, the taskforce wanted to develop measures that were very sound psychometrically (i.e., effective); for ongoing practical utility, the taskforce recognized the need for measures that were user-friendly, cost-efficient, and would be acceptable to front-line stakeholders (i.e., efficient). Researchers on the taskforce had expertise in measurement methods informed by Classical Test Theory, and the TDM fidelity measurement development process is grounded in that theory.

2. Specify model: Develop intervention model and translate into

operationalized constructs—As noted in the introduction, specification of an intervention or practice (Schoenwald et al., 2011) drives fidelity measurement, and the effort to develop fidelity measurement tools can further clarify and operationally define intervention or practice components. Key elements of TDM were identified by the task force, including *structural* elements such as: meeting composition—ideally defined as the presence of parent(s), multiple child welfare agency staff, community partners invited by the public child welfare agency, friends and extended family invited by the family, service providers from other agencies and organizations, and a non-case-carrying, trained meeting facilitator; meeting location—specified as a community location away from the public child welfare agency; and, meeting timing—TDM meetings should be prompted by any safety concern where a placement change may be considered, occur prior to any placement moves, or in cases of imminent risk, by the next working day, and should always occur before the initial court hearing in cases of removal (Crea et al., 2008). Additionally, TDM has clearly delineated *process* variables, including both prescribed and proscribed meeting strategies, and procedures to be followed upon arriving at the live recommendations generated by the meeting.

Previously articulated definitions of TDM were examined in order to refine the practice model, starting with the existing model of the more comprehensive F2F program. Next, we examined the materials developed for training and coaching facilitators to lead TDM meetings, and consolidated the prior research on TDM to create an updated schematic of the aspects of TDM thought to be central to the theory of change. The TDM model that emerged had two main components: Meeting Structure (Meeting Composition, Meeting Location, Facilitation by a trained TDM facilitator, Meeting Triggers, Meeting Date) and Meeting Process (what actually happened during and immediately following the meeting). These designations map onto Fixsen's (Fixsen, Naoom, Blase, Friedman, & Wallace, 2005) characterization of context ("prerequisites that must be in place for a program or practice to

operate,” such as facilitator training, the presence of appropriate participants at the meeting, and the meeting occurrence in response to a safety concern) and 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,” such as listing family strengths during the meeting, or asking each participant to suggest options that would result in youth safety) as elements of fidelity that should be measured. Additionally, competence, or the level of skill exhibited in the intervention delivery, is also recommended as an element of fidelity measurement.

Once the conceptual model of TDM had been refined, members of the task force considered how each of the theoretical components might result in specific, measurable actions or events. In particular, each component of the model was described in terms of (a) the criteria for ideal practice, and (b) the data source that could be measured as indicators of fidelity. TDM *structural* elements have previously been studied (Crea, Usher, & Wildfire, 2009; Crea et al., 2008); therefore, a methodology that is acceptable to practitioners of TDM, and provides reliable data for research purposes, has already been developed for examining these indicators. The workgroup identified that meeting *process* had not previously been operationalized or systematically evaluated, and would require novel measurement development. In this discussion, the workgroup expressed specific interest in measuring whether TDM meeting facilitators were compliant with the prescribed intervention components of TDM, as defined in the training and coaching materials. Accordingly, within the *process* domain of the TDM intervention, the indicator of fidelity prioritized by the group is meeting facilitator behavior. Thus, other aspects of meeting process (the extent to which other team members participate, given the opportunity by the facilitator) were judged as lower priority and are not the focus of the current project.

TDM meetings are discrete events, typically lasting about 90 minutes. Prescribed facilitator behaviors are described fully in the *Team Decisionmaking Implementation Manual* (Annie E. Casey Foundation), and standards for assessing these skills in practice are defined by the *TDM Facilitator Development and Coaching Tool*, created for in-house evaluation by other TDM experts and already used for TDM implementation in a number of sites. The operationalizing of TDM meeting process occurred in several iterations. First, all possible events described by the TDM manual and coaching tool were defined and listed. Next, practice experts removed events from the list that they believed to be common among routine removal and placement change meetings, and not unique to TDM. The remaining constructs, considered essential to TDM, were then defined in behavioral terms, vetted with a group of researchers, and with the developers of the TDM manual and practice guides, and revised accordingly. The vetting process was designed to yield the most valid (theoretically consistent) and specific components of the intervention, as well as the most practical description of how the intervention components are delivered in community contexts, thus balancing aims of measurement effectiveness and efficiency in routine care.

Members of the task force also identified a number of variables that could affect the success of TDM, but were not prescribed elements of the intervention. Among these variables are institutional support of the endeavor, individual facilitator factors, strategies designed to support fidelity such as supervision and the manualization of procedures, and short-term

outcomes such as family engagement and team cohesion. These variables are not indicators of fidelity, although they may influence fidelity.

3. Determine how to measure operationalized intervention constructs—Once interventions have been well specified and operationally defined, the next task is to determine how these intervention constructs will be measured, in terms of data source (e.g., informant self-report, observational coding, etc.), specific informant, rating or coding method, timeframe, and sampling. The TDM taskforce contemplated several potential data sources and data collection methods for assessing the targeted intervention elements of the decisionmaking meeting, and speculated that the best-trained indigenous source, in terms of knowledge regarding the expectations of facilitator behavior, is the facilitator of the meeting. However, a number of concerns exist regarding the validity of sole reliance on the facilitator as a reporter. The facilitator is directly involved in the implementation of TDM, and may have a personal interest in reporting high fidelity in an evaluative scenario. In other behavioral health implementation studies, adherence measures completed by the interventionist have been incongruent with third-party observational ratings (Hurlburt, Garland, Nguyen, & Brookman-Frazee, 2010).

Other meeting participants, such as the youth, birth parents, state social worker, and other family/community support members are less likely to be biased due to evaluative concerns, but are also less informed about the specific expectations of the TDM meeting and what “fidelity” to this practice actually entails. As described elsewhere (Schoenwald et al., 2011), little is known about the capacity of individuals not trained in an intervention to identify what happens in the intervention. Furthermore, the high stakes of meeting recommendations (i.e., placement decisions) may heavily influence participant perspectives on the meeting process. For example, it may be very hard for the parent of a youth for whom immediate removal from the home has been recommended, to objectively report on meeting proceedings. Despite these concerns about potential limitations in the validity of participants’ self-report, use of observational coding methodology, while absent the concerns of the other sources, was deemed prohibitively expensive and cumbersome for the ultimate wide-scale purposes of the fidelity measurement (thus not meeting efficiency criteria).

Ultimately, the task force decided that self-report questionnaires, completed by the meeting participants and the TDM facilitator, might provide a reasonable balance of effective and efficient measurement, and would be the most parsimonious method of collecting information regarding meeting content. Whereas the decision to rely on self-report questionnaires to measure intervention fidelity reflected consideration of measurement efficiency, the task force also wanted to ensure that the questionnaires were psychometrically sound, so that conclusions regarding the intervention integrity can be made with reasonable confidence. As a means of evaluating the validity of the questionnaires, the task force decided to conduct a pilot study using observational coding data, and trained raters for a sample of 100 audio recorded TDM meetings in a site where TDM is a part of routine practice. During these meetings, the newly developed self-report questionnaires would likewise be administered.

The use of recordings to document fidelity—in ongoing practice or even in the pilot trial—is illustrative of the tension that can arise between efficiency and effectiveness. Even for the discrete number of meetings in the pilot trial, video recording was deemed overly intrusive and risky given the vulnerable nature of the child welfare population. Live observation was considered; however, the on-site research workforce was limited to complete this task in full. Audio recording appeared to be the best option because it was considered less intrusive and did not require the presence of research staff at the meetings. The task force concluded that the benefit of observational data was worth the added expense and burden of two months' data collection during the pilot test to validate the self-report measures. Whereas the most effective methodology might involve ongoing recording, this compromise reflects the need to consider the realities of routine practice.

Although the observational pilot study represents an additional step in the development of the fidelity measurement system, the information that can be gleaned from it may result in future time and cost savings. As an example, comparing the responses of multiple reporters and coded observational data may determine who is the most reliable reporter of meeting content, and thus eliminate the need for multiple respondents in the future (i.e., increased efficiency). The pilot study will also allow the workgroup to examine the extent to which the different reporters can adequately assess the occurrence or non-occurrence of an event or strategy, and whether or not they can further differentiate between different levels of implementation quality (e.g., poorly done to well done). Analysis of the internal consistency of questionnaire items and test-retest reliability will also be conducted, and the results used to improve the instruments psychometrically for use in future research (i.e., increased effectiveness).

From a practical perspective, completing this small pilot trial will allow for a “test-run” of the extent to which the questionnaires, intended for wide-scale use, can be used given practice context constraints. How well does this form of information gathering fit into current practice? Are participants comfortable completing these forms at the end of the TDM meeting? How much time does this form completion add to the meeting procedures? In addition, since prior TDM research has not required the informed consent of team participants, participant responses to consent procedures and data collection are unknown. The proportion of youth, caregivers, agency staff, and other supporting participants who agree to participate in the study may inform the design of future research endeavors, especially with regards to participant recruitment and meeting sampling approaches (e.g., if not all individuals agree to participate, how many meetings of how many participants are needed to get an adequate assessment of fidelity?).

4. Pilot test measurement methods for validity, internal consistency, reliability, feasibility and utility—The pilot test is being conducted in partnership with the Cuyahoga County Department of Children and Family Services (CCDCFS), a large agency that adopted TDM as part of its routine process for arriving at removal and placement change decisions in 1994. There are currently 15 trained TDM facilitators within the agency, and a number of TDM meetings occur each day. Prior research has been conducted at this site, examining the fidelity of TDM meetings with regard to composition and timing (Crea, Usher, & Wildfire, 2009). In the pilot test, a project research assistant will

invite meeting participants to be a part of a research study that seeks to better understand the process of TDM. Those who consent will complete the questionnaires following the meeting, which will be audio recorded. A subset will receive the questionnaires in the mail several days later, and be asked to complete them a second time to provide measurement of test-retest reliability.

Using the identified and agreed-upon components of TDM meeting process, questionnaires have been developed with one item to assess each of the behavioral indicators of the model components (for example, “Provided an agenda for the meeting structure.”). The director and deputy director of the CCDFCS reviewed the questionnaires for face validity, and more recently the questionnaires were distributed to the directors of the administrative division of CCDFCS that provides oversight of TDM. At each step, item refinement occurs in response to feedback. Next steps include seeking the feedback of TDM facilitators, state caseworkers, and other TDM meeting participants (youth, family members) regarding the readability of the questionnaires. Once the questionnaires are finalized for use in the pilot study, a companion coding measure, and accompanying coding protocol manual, must be developed, and coders trained to acceptable levels of reliability. Results of the pilot study will inform which meeting participant(s) will be expected to complete questionnaires in future research and practice, as well as what response format can be used to yield the most reliable data (e.g., presence/absence versus measures of extensiveness).

Addressing Challenges in the TDM Fidelity Development Process

The current project utilized a number of strategies to advance research on the implementation of family engagement practices in child welfare placement decisions. Such research is needed to build an empirically informed knowledge base about the nature of these practices and their effects on desired outcomes. Strategies used in the current research project include the following: (1) The task force was, by design, composed of individuals with both practice expertise and research knowledge, (2) deliberate consideration was given to the proposed uses of the fidelity measurement before the development began, (3) much of the work focused on developing a measurement method to document the process (as opposed to structure) of the meeting, an aspect for which systematic and methodologically sound data collection has largely been absent in prior research, (4) in developing the meeting process fidelity questionnaires, the task force has incorporated feedback from local experts on TDM practice to heighten face validity and acceptability, and (5) validation of the measures will include observational coding methods generally considered exemplary practice for this type of research (Schoenwald et al., 2011). In these ways, the current project has sought equipoise between effective and efficient measurement.

Despite these efforts, a number of challenges, both logistical and philosophical, required careful thought and creative solutions. At CCDFCS, TDM has been a part of routine practice since 1994, pre-dating the formal development of the critical elements considered today to define TDM. Some facilitators are part of an originally trained cohort, while others have received initial training more recently. Although facilitators receive ongoing review and booster training, a number of supervisory staff has suggested that there is considerable variability regarding adherence to TDM—in other words, some fidelity drift has occurred.

Although researchers have explicitly stated that such variability is beneficial for testing the sensitivity of the questionnaires and coding system, the very act of implementing systematic observational procedures may result in increased attention to the procedural details due to pressure within the agency. To address this possibility, the task force proposed that data on individual facilitator fidelity would not be available to the agency from this pilot test, and would rather be presented in aggregate form. While this strategy may alleviate some anxiety and result in a better test of the measurement system, it also somewhat compromises the value of the study for the agency, which had hoped to use these data to provide individualized feedback to meeting facilitators. Although this was identified as a key purpose to the measurement in the initial meetings, it was not well-understood by the practice site that the fidelity measurement instrument must first be proven reliable and valid before it could be used for the purpose of feedback for practice improvement. In this instance, the need to maintain rigorous research methodology (effectiveness) was somewhat at odds with the agency's practical interest in making immediate use of information they had agreed to provide (efficiency).

Another challenge to carrying out the research is practical: because of the sensitive nature of the meeting content, it is possible that some participants may not wish to audio record the meetings. The number of participants who may attend a given meeting increases the chances that someone will object to the recording, and all must agree. To manage this challenge, the task force has decided to train a research assistant to complete the coding measures live, during the meeting, in the event that this option may be preferable for some participants who object to a recording.

Finally, the process of this type of research is often incremental, with each step providing information that is essential to the next step. Before we can confidently test the impact of an intervention, we need to operationalize the components of the intervention, document that these components can be characterized and measured, and then apply measures of fidelity to ensure we are evaluating the true intervention of interest. If some components of a theoretical model predict the optimal outcomes and others do not, the theoretical model—and thus the indices of fidelity to that model—is revised iteratively. This iterative process makes good sense from a research perspective; from a practical perspective it may seem laborious and slow. In a site where the intervention being studied is already in place and is relied upon as a crucial aspect of service provision, the tension between the methodical approach of research and the urgency of practice must be confronted.

In the case of TDM, it has been necessary to engage the staff of the CCDCFS as partners in the data collection process, and to provide as much transparency as possible regarding the overall aims of the research, as well as the theoretical and scientific considerations. For example, a brief didactic presentation to program leaders and staff about the uses of fidelity measurement appears to have mitigated much of the confusion around the need for the audio recording process. Furthermore, those on the front-lines have been very eager to provide feedback on the materials, and have volunteered to let researchers observe the TDM meetings throughout the development process. By inviting service providers to be a more direct part of the research process, it has been possible to gain greater levels of commitment and cooperation, and for the research to benefit from practical expertise.

A number of issues in developing the TDM implementation fidelity measurement system remain. Whereas the criteria for acceptable fidelity are well-defined for the structural components of TDM, compliance with TDM meeting process is less clear and operationalized specifics of the intended process are still being refined. Is the presence of all prescribed items required for a meeting to be deemed adherent? How is competence, in addition to compliance, being addressed—and can non-expert raters be expected to differentiate between different levels of skillfulness in implementation? Whereas for research purposes, fidelity to TDM may be examined as a continuous variable (more fidelity presumed to be better), for performance feedback, some criteria regarding exemplary versus satisfactory—and versus unsatisfactory—TDM implementation would be valuable—yet what is the best approach to defining these criteria, without data-guided demarcations? To answer these questions, to document the extent to which the practice of TDM reflects the theoretical model, and to understand the impact of TDM on child placement decisions and child outcomes, we need fidelity measurement strategies that are both effective and efficient.

Conclusion

The child welfare sector is complex, its service provision is vast and diverse, and the outcomes of interest—child safety and wellbeing—are multi-determined. The challenges of defining promising practices, measuring their implementation, and evaluating their effectiveness are not, however, fundamentally different in child welfare relative to other service sectors, such as mental health, education, or medicine. In the same vein, the strategies applied to address these challenges in other service sectors provide a blueprint for collaborative, community-based research to help address them in child welfare. Unexpected hurdles may require creative solutions, but it is unlikely that they are insurmountable. In the end, the results of defining, developing, and measuring an intervention—and its outcome—will be greater effectiveness of the services provided, regardless of the individual characteristics of those served. In their consideration of evidence-based practice in the child welfare context, Whiting Blome and Steib (2004) note evidence-based change will require more than the implementation of a promising, but untested practice; it necessitates careful measurement of both process and outcomes—work that “is not easy—it is just the only way to be assured that children and families get the best at a time they need it the most” (Whiting Blome & Steib, 2004, p. 614). The TDM project described here illustrates a systematic process to define a child welfare intervention, and to design and test a measure to assess fidelity to that intervention, that has the potential to be both effective and efficient, thereby paving the way for tests of the effectiveness of the intervention. Our account of the project offers both a specific example, and, a potentially generalizable process by which the “not easy” steps can be undertaken to build an evidence base for interventions deployed in the child welfare sector.

Acknowledgments

Completion of this research was supported in part by a Career Award (R00 MH083887) and an Advanced Center award (P30 MH074778, J. Landsverk, PI) from the National Institute of Mental Health, and a grant from the Annie E. Casey Foundation awarded to Sarah Kate Bearman. We thank them for their support and acknowledge that the findings and conclusions presented in this article are those of the authors alone, and do not necessarily reflect the opinions of the Annie E. Casey Foundation.

References

- Aarons GA, Hurlburt M, Horwitz SM. Advancing a conceptual model of evidence-based practice implementation in public service sectors. *Administration and Policy in Mental Health and Mental Health Services Research*. 2011; 38:4–23.10.1007/s10488-010-0327-7 [PubMed: 21197565]
- Adams P. Marketing social change: The case of family preservation. *Children and Youth Services Review*. 1994; 16:417–431.
- Bellg AJ, Borrelli B, Resnick B, Hecht J, Minicucci DS, Ory M, Czajkowski S. Enhancing treatment fidelity in health behavior change studies: Best practices and recommendations from NIH Behavior Change Consortium. *Health Psychology*. 2004; 23:443–451. [PubMed: 15367063]
- Berzin SC, Cohen E, Thomas K, Dawson WC. Does family group decision making affect child welfare outcomes? Findings from a randomized control study. *Child Welfare*. 2008; 87:35–54. [PubMed: 19391466]
- Bickman, L., editor. *The Journal of Mental Health Administration*. 1996. Special Issue: The Fort Bragg Experiment; p. 23
- Calsyn RJ. A checklist for critiquing treatment fidelity studies. *Mental Health Services Research*. 2000; 2:107–113. [PubMed: 11256717]
- Chen, HT. *Theory-driven evaluations*. Newbury Park: Sage; 1990.
- Crampton DS. Research Review: Family group decision-making: a promising practice in need of more programme theory and research. *Child and Family Social Work*. 2007; 12:202–209.
- Crampton DS, Usher CL, Wildfire JB, Webster D, Cuccaro-Alamin S. Does community and family engagement enhance permanency for children in foster care? Findings from an evaluation of the family-to-family initiative. *Child Welfare*. 2011; 90:61–77. [PubMed: 22413380]
- Crea TM, Crampton DS, Abramson-Madden A, Usher CL. Variability in the implementation of Team Decisionmaking (TDM): Scope and compliance with the Family to Family practice model. *Children and Youth Services Review*. 2008; 30:1221–1232.
- Crea TM, Usher CL, Wildfire JB. Implementation fidelity of Team Decisionmaking. *Children and Youth Services Review*. 2009; 31:119–124.
- Crea TM, Wildfire JB, Usher CL. The association of team composition and meeting characteristics with foster care placement recommendations. *Journal of Social Service Research*. 2009; 35:297–310.
- DeMuro, P.; Rideout, P. *Team Decisionmaking: Involving the family and community in child welfare decisions*. Baltimore, MD: The Annie E. Casey Foundation; 2002.
- Fixsen, DL.; Naoom, SF.; Blase, KA.; Friedman, RM.; Wallace, F. *Implementation Research: A synthesis of the literature*. Tampa, FL: University of South Florida, Louis de la Parte Florida Mental Health Institute, The National Implementation Research Network; 2005. (FMHI Publication #231)
- Glisson, C. Structure and technology in human services organizations. In: Hasenfeld, Y., editor. *Human services as complex organizations*. Beverly Hills: Sage; 1992. p. 184–202.
- Glisson C. The organizational context of children's mental health services. *Clinical Child & Family Psychology Review*. 2002; 5(4):233–253. [PubMed: 12495268]
- Grimshaw JM, Shirran L, Thomas R, Mowatt G, Fraser C, Bero L, O'Brien MA. Changing provider behavior: An overview of systematic reviews of interventions. *Medical Care*. 2001; 39 Suppl 2(8):II-2–II-45. [PubMed: 11583120]
- Grol R, Grimshaw J. Evidence-based implementation of evidence-based medicine. *Journal on Quality Improvement*. 1999; 25:503–513.
- Gunderson K, Cahn K, Wirth J. The Washington State long-term outcome study. *Protecting Children*. 2003; 18:42–47.
- Hurlburt MS, Garland AF, Nguyen K, Brookman-Frazee L. Child and family therapy process: Concordance of therapist and observational perspectives. *The Journal of Administration and Policy in Mental Health*. 2010; 37:230–244.
- Kaye S, Osteen PJ. Developing and validating measures for child welfare agencies to self-monitor fidelity to a child safety intervention. *Children and Youth Services Review*. 2011; 33:2146–2151.

- Kessler ML, Gira E, Poertner J. Moving best practice to evidence-based practice in child welfare. *Families in Society*. 2005; 86:244–250.
- Litchfield MM, Gatowski SI, Dobbin SA. Improving outcomes for families: Results of an evaluation of Miami's Family Decision Making program. *Protecting Children*. 2003; 18:48–51.
- Pennell J, Burford G. Family group decision making: Protecting children and women. *Child Welfare: Journal of Policy, Practice, and Program*. 2000; 79:131–158.
- Pennell, J.; Anderson, G. Widening the circle: The practice and evaluation of family group conferencing with children, youths, and their families. Washington, DC: NASW Press; 2005.
- Rauktis ME, Huefner J, Cahalane H. Perceptions of fidelity to family group decision making principles: Examining the impact of race, gender, and relationship. *Child Welfare*. 2011; 90:41–59. [PubMed: 22413379]
- Schoenwald SK, Garland AF, Chapman JE, Frazier SL, Sheidow AJ, Southam-Gerow MA. Toward the effective and efficient measurement of implementation fidelity. *Administration and Policy in Mental Health and Mental Health Services Research*. 2011; 38:32–43.10.1007/s10488-010-0321-0 [PubMed: 20957425]
- Stuczynski A, Kimmich M. Challenges in measuring the fidelity of a Child Welfare Service intervention. *Journal of Public Child Welfare*. 2010; 4:406–426.
- Usher CL, Wildfire JB. Evidence-based practice in community-based child welfare systems. *Child Welfare*. 2003; 82:597–614. [PubMed: 14524427]
- Weiss, CH. Evaluating educational and social action programs: A “treeful of owls”. In: Weiss, CH., editor. *Evaluating Action Programs: Readings in Social Action and Education*. Boston: Allyn & Bacon; 1972.
- Whiting Blome W, Steib S. Whatever the problem, the answer is “evidence-based practice”—Or is it? *Child Welfare League of America*. 2004; LXXXIII:611–615.

Table 1

Applying a Fidelity Measurement Development Framework to Child Welfare Interventions: General Steps, TDM Exemplars, and Effectiveness & Efficiency Considerations by Step

General Steps ^a	TDM Measurement Development Examples	Examples of attention to:	
		Effective Measurement	Efficient Measurement
Identify measurement purpose and theory	1. Convene task force of diverse stakeholders with complementary expertise to identify purpose(s) of measurement system	Applied Classical Test Theory Prioritized psychometric soundness	Diverse task force stakeholders identified multiple measurement purposes and prioritized utility and feasibility
Define adherence	2. a) Specify intervention model b) Define each intervention construct c) Operationally define intervention components for each construct	Characterized unique prescribed components of intervention including structure, process, content	Vetted intervention model, constructs and components with stakeholders for practice relevance
Choose data collection methods	3. Determine <i>how</i> to measure operationalized constructs (e.g., data source, informant, rating method, timing, sampling, etc.)	Critiqued validity of potential data source(s)	Emphasized parsimonious data collection methods
Determine and test scoring system	4. Pilot test methods for validity, feasibility, internal consistency, reliability, and utility	Tested validity of self-report; Consider continuous scores for research purposes	Tested feasibility of methods; Considered cut-off scores for ease of interpretation in practice

^a = Adapted from Schoenwald et al. (2011).