ORIGINAL ARTICLE

Reliability of Therapist Self-Report on Treatment Targets and Focus in Family-Based Intervention

Aaron Hogue · Sarah Dauber · Craig E. Henderson · Howard A. Liddle

Abstract Reliable therapist-report methods appear to be an essential component of quality assurance procedures to support adoption of evidence-based practices in usual care, but studies have found weak correspondence between therapist and observer ratings of treatment techniques. This study examined therapist reliability and accuracy in rating intervention *target* (i.e., session participants) and *focus* (i.e., session content) in a manual-guided, family-based preventive intervention implemented with 50 inner-city adolescents at risk for substance use. A total of 106 sessions selected from three phases of treatment were rated via post-session self-report by the participating therapist and also via videotape by nonparticipant coders. Both groups estimated the amount of session time devoted to model-prescribed treatment targets (adolescent, parent, conjoint) and foci (family, school, peer, prosocial, drugs). Therapists demonstrated excellent reliability with coders for treatment targets and moderate to high reliability for treatment foci across the sample and within each phase. Also, therapists did not consistently overestimate their degree of activity with targets or foci. Implications of study findings for fidelity assessment in routine settings are discussed.

A. Hogue (✉) · S. Dauber
Treatment Research Division, National Center on Addiction and Substance Abuse (CASA) at Columbia University, 633 Third Avenue, 19th floor, New York, NY 10017, USA
e-mail: athogue@aol.com

C. E. Henderson
Sam Houston State University, Huntsville, TX, USA

H. A. Liddle
Center for Treatment Research on Adolescent Drug Abuse, University of Miami Miller School of Medicine, Miami, FL, USA

Introduction

Need for Validated Therapist Self-Report Measures of Fidelity to Manualized Behavioral Interventions in Routine Care

Whereas observational assessment of treatment implementation of manualized behavioral interventions remains the gold standard for scientific research on treatment fidelity (Hogue et al. 1996; Garland et al. 2010a, b), it is critical to develop reliable complements or even alternatives to observational methods that are cost-effective and easy to use by non-researchers in clinical practice. The most promising method is therapist self-report measures, which offer several advantages over observational ratings (Carroll et al. 1998; Weersing et al. 2002): they are quick, inexpensive, and non-intrusive; they can be completed throughout treatment, which facilitates evaluation of infrequent but clinically meaningful interventions; and they can assess therapist intentions as well as observed behaviors.

Reliable therapist-report methods may also be essential for improved adoption of evidence-based practices (EBPs) specified in manualized treatments (Garland et al. 2010a, b). Transporting manualized treatments into usual care invariably requires expensive and time-consuming quality assurance "superstructures" consisting of intensive on-site training of line staff by model experts, followed by continuous monitoring via booster trainings and remote consultation (Hogue et al. 2013). Thus, resource-efficient

quality assurance tools that monitor EBP implementation in field settings with precision and clinical sophistication are urgently needed (Kelley et al. 2010; Schoenwald et al. 2011). Valid therapist-report measures could support technology transfer in several ways: as a self-check by therapists to mark their own progress in treating individual cases; as a supervision aid for trainers and agency supervisors to monitor treatment fidelity; and as administrative data for stakeholders and external reviewers to evaluate therapist- and agency-level clinical performance (Bearsley-Smith et al. 2008; Carroll et al. 1998; Garland et al. 2010a, b). Therapist-report measures could eventually develop standardized "red line" scores to serve as benchmarks for determining if a given session was faithful to an EBP (Dobson and Shaw 1993).

Extant Research on the Reliability of Therapist-Reported Fidelity to EBPs

To date only a handful of studies have tested a therapist-report measure of EBP implementation (e.g., Weersing et al. 2002). Fewer still have attempted to confirm the validity of therapist self-reports by comparing them to observational ratings. Pioneering work with EBPs for adult substance abuse have found modest to weak correspondence between nonparticipant raters and therapist reports of fidelity to manualized intervention techniques (Carroll et al. 1998; Martino et al. 2009). In the youth treatment arena, Hurlburt et al. (2010) found that observational coders reported substantially less occurrence and lower intensity of EBPs compared to therapist report in front-line mental health care. Therapist reports of adherence to multisystemic therapy (MST) for antisocial youth have been linked to various client, therapist, and organizational factors (e.g., Schoenwald et al. 2005), including ongoing expert consultation (Schoenwald et al. 2004). However, the MST therapist-report fidelity scale assesses adherence to basic intervention principles that guide the MST model specifically—limiting its utility for measuring other EBPs—and has not yet been validated with non-participant observational coding.

Methodological Innovations of the Current Study

The current study advances research on therapist ratings of EBPs in three ways. First, it examined therapist reliability and accuracy in rating intervention *target* (session participants, i.e., to whom interventions are addressed) and *focus* (session content, i.e., domains of functioning addressed). Therapist-report studies have paid close attention to treatment principles and specific techniques but have not

addressed the "contours" of implementation (Schoenwald et al. 2011) defined by the parameters of a given treatment (i.e., service delivery aspects of implementation: to whom, where, and how often); nor have they examined prescribed session content (Garland et al. 2010a, b; Hogue et al. 2004). To make headway in developing efficient fidelity measures for real-world application, it is critical to explore the feasibility of various methods for assessing various dimensions of implementation (Schoenwald 2011). Asking therapists to judge the (more) readily defined targets and foci of their interventions, rather than treatment techniques that are often multifaceted and interwoven, sets the fidelity measurement bar a notch lower, which might engender improved reliability. Along these lines Kelley et al. (2010) developed a brief therapist-report measure of session focus that showed acceptable internal consistency and distinguished between clinician versus client influences on session content, though it has not been validated with observational data.

Fidelity evaluation of treatment targets and foci is especially germane to family-based intervention, one of the most common treatment modalities in child mental health services (Hoagwood 2005) and one that places strong emphasis on tracking treatment activity with various family members and social systems (e.g., school, social services) that serve as key developmental contexts for youth (Glisson et al. 2010; Schoenwald et al. 2008). Whereas adherence to signature family therapy techniques has been linked to in-session changes in parenting and family interactions (Diamond and Liddle 1996, 1999; Schmidt et al. 1996) and to posttreatment outcomes (Huey et al. 2000; Hogue et al. 2008a, b; Robbins et al. 2011), research on benefits associated with faithful attention to treatment parameters and content has rarely surfaced. One exception is a study by Hogue et al. (2006) that incorporated observational ratings of treatment foci in showing that both family-focused and adolescent-focused interventions in family therapy predicted improvement in adolescent symptoms and family processes up to one year posttreatment. Looking ahead, as manualized family therapies progress in efforts to adapt to new settings and clinical populations (e.g., Hogue et al. 2002a, b; Liddle et al. 2006) and to ensure fidelity standards during large-scale dissemination initiatives (e.g., Barnoski 2003; Zazzali et al. 2008), adherence to model tenets about what persons and domains of functioning to address in what stages of treatment will be prominently featured in quality assurance procedures.

A second key feature of this study is inclusion of research-funded therapists rather than community practitioners (see also Carroll et al. 1998). Although the ultimate goal of fidelity measurement is real-world utility

(Schoenwald 2011), at this nascent stage in the development of therapist-report instruments there are two practical advantages to using research therapists. First, study therapists were made fluent and then continuously supervised in the intervention components being rated, and they participated in designing the measure itself. In this way they received training in the coding scheme that was equivalent to training received by the observational coders (see "Therapist and Observational Coder Training" section). Second, study therapists were operating in a de facto "low stakes" context, wherein the self-report data they provided were tallied for scientific and clinical purposes only. This stands in contrast to "high stakes" fidelity evaluation that may obtain in front-line settings, wherein practitioner self-reports of model implementation can have direct implications for performance ratings, immediate rewards or punitive action, and even job security (Schoenwald et al. 2011). Thus, utilizing research therapists created optimal conditions in which to measure their fundamental capacity for accurate reporting, uncontaminated by potential negative consequences for subpar fidelity. Once baseline benchmarks for self-report accuracy are established in low stakes contexts such as this, it will be possible to move forward with the critical task of determining which training and motivational factors influence (compromise—or perhaps enhance?) accuracy in higher stakes settings.

Third, whereas most EBP fidelity measures assess intervention extensiveness (i.e., the thoroughness and/or frequency with which an intervention is delivered; Hogue et al. 1996), this study examined therapist ratings of the number of minutes devoted to each treatment target and focus. This self-rating task aligns with standard quality assurance procedures in everyday practice, wherein it is important to account accurately for billable clinical hours spent with various members of the client system as well as time devoted to model- or client-specific clinical topics (Fixsen et al. 2005; Schoenwald et al. 2011).

Study Hypotheses

The main study hypothesis was that research therapist ratings of session time devoted to various treatment targets (adolescent alone, parent(s) alone, adolescent-parent conjointly) and developmental foci (family, peer, school, prosocial activities, substance use) would be reliable with ratings made by observational coders, and that strong reliability would be consistent across treatment phases (early, middle, late). A secondary hypothesis was that therapists would overestimate the average amount of therapeutic activity in these modules and domains (Hurlburt et al. 2010). The observational coder data used in the current study have already been examined in a previous study with this sample (Hogue et al. 2005) to demonstrate

that study therapists adhered to manual-specified guidelines regarding session activity in treatment modules and domains for a manualized family-based intervention, multidimensional family prevention (MDFP; Hogue et al. 1999, 2002a, b). Of interest in the current study was whether therapists were able to make consistently reliable judgments about their own adherence to these MDFP intervention components.

Method

Participants

Clients

Clients ($N = 50$) were adolescents and families who participated in a randomized trial of the MDFP model (Hogue et al. 2002b). Clients were recruited from a community youth enrichment program after the target adolescents were identified as indicated risk for developing substance use problems based on individualized risk screening (see Hogue et al. 1999). The mean age of the adolescents was 12.5 years (SD = 0.79, range 11–14). There were 24 boys (48 %) and 26 girls (52 %), and 98 % identified as African American. Families were headed by single biological parent (54 %), grandparent(s) (16 %), one biological and one stepparent (14 %), two biological parents (12 %), and other (4 %). A total of 60 % reported annual family income under $15,000, and 62 % received some form of public assistance.

Therapists

Four male therapists completed post-session self-reports of MDFP implementation: two African American, one European American, and one Asian American (M age: 31 years). Three had a master's in counseling and one a doctorate in Clinical Psychology, and they averaged two years of experience as family counselors.

Observational Coders

Observational coders were nine undergraduates and one Psychology graduate student: six European American females, two Asian American females, one European American male, one Asian American male. Undergraduates were considered qualified for this task because they were not required to make clinically sophisticated judgments about thoroughness, quality, or appropriateness of interventions within the treatment model (Hogue et al. 1996, 2008a, b).
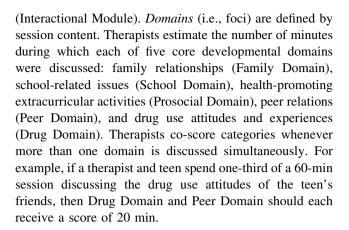
Intervention Model and Fidelity Measure

Multidimensional Family Prevention (MDFP)

MDFP (Hogue and Liddle 1999) is a family-based preventive intervention for indicated-risk adolescents. MDFP is a prevention version of multidimensional family therapy, an empirically supported treatment for adolescent substance use (as nominated in Becker and Curry 2008; Waldron and Turner 2008). MDFP has shown efficacy in preventing problem behaviors in multiple developmental domains (Hogue et al. 2002b). MDFP session composition varies on a case-by-case and session-by-session basis, and therapists regularly spend time working individually with family members to accomplish family-wide goals. A total of 15–25 sessions are held over a 3–4 months period. Initial sessions are dedicated to assessment of adolescent and family functioning in several risk and protection domains, particularly family relations, school performance, extracurricular prosocial activities, peer relations, and drug use attitudes and experiences. Therapist and family then review the risk profile and construct a treatment agenda for addressing the most significant themes within three primary modules. The *Adolescent Module* focuses on the teen's developmental milestones, coping and problem-solving skills, investment in prosocial institutions, and risky behaviors associated with drug use and delinquency. The *Parent Module* fosters parenting competency by supporting consistency in limit-setting and discipline, teaching age-appropriate behavioral management techniques, and encouraging monitoring of school performance and other behavior outside the home. The *Interactional Module* builds relationship skills and fosters autonomy and relatedness in the parent–teen relationship. In-session conversation among family members is shaped to increase family cohesion, problem-solving, and clarity of communication and roles. As an individually tailored model, MDFP makes few prescriptions about specific modules or domains being implemented in specific sessions.

Therapist Self-Report Checklist (TSRC)

The TSRC is a therapist-report measure of the amount of time devoted in session to predefined treatment targets and domains of adolescent functioning. TSRC items are consistent with the MDFP model but also with general implementation principles that guide the family-based approach for adolescent behavior problems. *Modules* (i.e., targets) are defined by who attends the session. Module ratings require a therapist to estimate the number of minutes during which s/he was (a) with the adolescent only (Adolescent Module), (b) with parent(s) only (Parent Module), or (c) with parent(s) and adolescent conjointly (Interactional Module). *Domains* (i.e., foci) are defined by session content. Therapists estimate the number of minutes during which each of five core developmental domains were discussed: family relationships (Family Domain), school-related issues (School Domain), health-promoting extracurricular activities (Prosocial Domain), peer relations (Peer Domain), and drug use attitudes and experiences (Drug Domain). Therapists co-score categories whenever more than one domain is discussed simultaneously. For example, if a therapist and teen spend one-third of a 60-min session discussing the drug use attitudes of the teen's friends, then Drug Domain and Peer Domain should each receive a score of 20 min.

Procedures

Sampling Design

MDFP therapists videotaped every session (as possible) of every case and also completed a TSRC after every session. For observational coding purposes, one session was randomly chosen for videotape review from each available treatment phase for every case in order to sample representatively across the duration of treatment. Treatment phases were defined as: Phase 1 (sessions 1–5), Phase 2 (sessions 6–12), and Phase 3 (sessions 13 and higher). Thus every study case had between 1 and 3 sessions selected, depending on how long the given case was retained in treatment, and these sessions were rated by both the MDFP therapist and observational coders. Of the 106 sessions selected, 44 (42 %) were from Phase 1, 38 (36 %) from Phase 2, and 24 (23 %) from Phase 3. Due to logistical barriers and client preferences it was not possible to videotape a Phase 1 session for six of the 50 study cases. Successively fewer sessions were available for coding in Phases 2 and 3 due to treatment attrition during the trial. Of the 65 families who participated in the MDFP trial (Hogue et al. 2002b), 15 were not included in the current study for the following reasons: four did not complete a posttreatment assessment, one attended no treatment sessions at all, and ten refused to be videotaped. There were no significant differences between the original trial sample and the current study sample on any demographic variables (Hogue et al. 2005).

Therapist and Observational Coder Training

Therapists were introduced to the structure and content of the TSRC during a 1-h training session. Then, for 4 weeks they dedicated a portion of their routine clinical supervision to reviewing their self-reported ratings of module and domain activity; questions regarding TSRC fidelity ratings were resolved at these meetings. Observational coders

trained in a group format for ninety minutes per week over a 4-months period to reach adequate pre-study reliability (Intraclass correlation coefficient$_{(1,2)}$ >0.65 for most study items). Training consisted of didactic instruction and discussion of the rating manual, trainer and peer review of practice scales using pilot cases, and coding exercises designed to test and expand understanding of each scale item. Once rating of study tapes commenced, coders reconvened on a weekly basis for supportive training and to prevent rater drift.

Ratings

Therapists completed the TSRC after every session. Observational coders completed TSRC items after viewing entire videotaped sessions (average length: 60 min). Coders were kept unaware of the intent of the study, instructed that family involvement and session composition would vary according to the contingencies of each case, and informed that each scale item could arise in every session. Two coders rated every session; coders were randomly assigned to sessions following balanced incomplete block design procedures (Fleiss 1981). Reliability and fidelity of these observational data were described in a previous study (Hogue et al. 2005). Intraclass correlation coefficients (ICCs; Shrout and Fleiss 1979) were adequate for Module items: Adolescent Module ICC$_{(1,2)}$ = 0.62; Parent = 0.91, and Interactional = 0.85. ICCs for Domain items were adequate as well: Family Domain ICC$_{(1,2)}$ = 0.80, School = 0.76, Prosocial = 0.50, Peer = 0.72, and Drug = 0.82. The final observational rating scores for each item were calculated by averaging the time estimates provided by each coder.

The study was conducted under approval by the governing Institutional Review Board. Active consent from caregivers and assent from adolescents were collected. Therapists provided active consent for sessions to be judged for MDFP model adherence.

Plan of Analysis

Therapist reliability in rating the amount of session time devoted to MDFP modules and domains was calculated using the one-way random ICC (Shrout and Fleiss 1979). Therapist accuracy was tested by comparing therapist versus coder mean scores using paired samples t tests. Cohen's d effect size was calculated for mean differences of $p < 0.10$; according to Cohen (1988), $d = 0.20$ is a small effect, 0.50 is medium, and 0.80 is large. Analyses were conducted first on all 106 study sessions and then separately by phase (1 = early, 2 = middle, 3 = later) to examine consistency over the course of intervention. Variance components analysis was then used to determine the proportion of variance in both therapist and observer scores attributable to client and therapist effects.

Results

According to Cicchetti (1994), ICC > 0.70 is acceptable reliability and >0.80 strong reliability for molar behavioral ratings. Results for the full sample (Table 1) shows strong reliability for modules: Adolescent ICC$_{(1,2)}$ = 0.96; Parent ICC = 0.95; Interactional ICC = 0.86. Therapists reported more time spent working in the Parent ($t(105) = -2.45$, $p < 0.05$, $d = 0.48$) and Interactional ($t(105) = -2.21$, $p < 0.05$, $d = 0.43$) modules than did observers. Domain ratings showed moderate to high reliability: ICC$_{(1,2)}$ was above 0.80 for Drug and above 0.70 for Family, Peer, and School; one exception was a low ICC = 0.59 for Prosocial. Mean comparisons revealed a trend for observers to give higher scores in the Family Domain ($t(105) = 1.80$, $p < 0.10$, $d = 0.35$) but no differences in any other domain.

Results for each treatment phase are presented in Table 2. Phase-specific results for modules mirror those found for the overall sample, with ICCs above 0.80 in all three phases and relatively stable across phase for each module. Mean comparisons revealed trend-level effects for higher therapist scores in Parent Module Phase 2 ($t(37) = -1.98$, $p < 0.10$, $d = 0.65$) and in Interactional Module Phase 1 ($t(43) = -1.79$, $p < 0.10$, $d = 0.55$) and

Table 1 Interrater reliability and mean scores for therapist self-report and observational ratings of treatment modules and domains for the full sample ($N = 106$ sessions)

	Therapist-observer ICC	Therapist score (# min) M (SD)	Observer score (# min) M (SD)	Paired-sample t- and cohen's d coefficients
Adolescent module	0.96	12.6 (18.0)	12.7 (16.5)	Ns
Parent module	0.95	13.8 (20.2)	11.8 (17.2)	$t = -2.45$**, $d = 0.48$
Interactional module	0.86	20.3 (21.9)	17.3 (19.6)	$t = -2.21$**, $d = 0.43$
Family domain	0.71	29.2 (15.4)	32.1 (19.4)	$t = 1.80$*, $d = 0.35$
Peer domain	0.70	5.2 (6.3)	5.0 (5.8)	Ns
School domain	0.74	7.3 (7.0)	6.9 (8.0)	Ns
Prosocial domain	0.59	5.9 (5.8)	5.1 (5.5)	Ns
Drug domain	0.87	1.9 (4.8)	1.9 (5.2)	Ns

** $p < 0.05$, * $p < 0.10$

Table 2 Interrater reliability and mean scores for therapist self-report and observational ratings of treatment modules and domains by treatment phase

	Phase 1 (n = 44)			Phase 2 (n = 38)			Phase 3 (n = 24)		
	Therapist-observer ICC	Therapist score M (SD)	Observer score M (SD)	Therapist-observer ICC	Therapist score M (SD)	Observer score M (SD)	Therapist-observer ICC	Therapist score M (SD)	Observer score M (SD)
Adolesent module	0.95	13.0 (18.3)	12.9 (16.0)	0.97	14.7 (19.9)	14.8 (17.6)	0.95	8.7 (14.1)	9.1 (15.4)
Parent module	0.95	12.6 (19.6)	10.8 (17.2)	0.93	14.3 (21.2)	11.4 (16.7)*	0.97	15.2 (20.3)	14.6 (18.5)
Interactional module	0.92	21.7 (21.6)	18.7 (19.0)*	0.83	18.0 (24.2)	16.2 (22.6)	0.83	21.7 (19.0)	16.5 (15.8)*
Family domain	0.65	27.1 (13.9)	29.2 (19.9)	0.75	29.4 (15.1)	34.1 (17.9)*	0.73	32.6 (18.3)	34.0 (20.7)
Peer domain	0.69	6.3 (7.6)	5.8 (6.2)	0.73	4.3 (5.1)	5.4 (6.6)	0.61	4.5 (5.1)	2.9 (2.8)*
School domain	0.87	8.2 (7.3)	8.2 (7.3)	0.67	6.2 (7.0)	6.1 (7.0)	0.62	7.5 (6.7)	5.7 (10.2)
Prosocial domain	0.61	5.8 (5.5)	5.2 (6.6)	0.57	5.4 (5.3)	4.7 (4.4)	0.58	6.8 (6.9)	5.5 (5.2)
Drug domain	0.93	2.1 (5.1)	1.9 (6.2)	0.63	1.5 (3.3)	2.2 (4.6)	0.92	2.2 (6.3)	1.7 (4.2)

* $p < 0.10$

Phase 3 ($t(23) = -2.00$, $p < 0.10$, $d = 0.83$). Though not statistically significant, Table 2 suggests a trend for Adolescent interventions to decrease as therapy enters the later phases, whereas Parent and Interactional work remain constant or increase. ICCs for domains were also relatively stable across time, with the exception of the School and Drug Domains, which both declined from Phase 1 to Phase 2, and then increased in Phase 3 for Drug Domain. Trend-level differences between reporters were found for Family Domain Phase 2 ($t(37) = 2.02$, $p < 0.10$, $d = 0.66$), with observers giving higher scores, and Peer Domain Phase 3 ($t(23) = -1.83$, $p < 0.10$, $d = 0.66$), with therapists giving higher scores.

Variance components analyses were conducted on all sessions using a restricted maximum-likelihood estimation method for three terms: Therapist, Client (nested within Therapist), and Error (Note: There were not enough observations to calculate the multiply nested Phase and Coder terms). Each term was entered as a random effect in the analysis, and estimates of variance for each term were transformed into proportions of variance based on the estimates of total variance across terms. For observational ratings of treatment modules, Therapist accounted for 8 % (Adolescent), 2 % (Parent), and 0 % (Interactional) of respective total score variance, whereas Client accounted for 14 % (Adolescent), 28 % (Parent), and 0 % (Interactional). For therapist ratings of modules, Therapist accounted for 5 % (Adolescent), 0 % (Parent), and 0 % (Interactional), whereas Client accounted for 23 % (Adolescent), 20 % (Parent), and 0 % (Interactional), respectively. Therapist and

Client component variances were less than 5 % for each domain score for both observer and therapist ratings, with three exceptions: Client term for therapist ratings in the Prosocial Domain (24 %), Client term for therapist ratings in Family (17 %), and Therapist term for observer ratings in Family (11 %). These data show that therapists and coders concurred that session time spent with adolescents and parents varied somewhat across clients but hardly at all across therapists. Informants also agreed that time spent in the Interactional Module, and in most domains, was highly consistent across therapists and clients.

Discussion

Main Study Findings

This study found that therapists were quite reliable in judging their own activity in treatment modules and domains: Correlations with nonparticipant observers ranged from ICC = 0.57–0.97. This level of reliability is substantially higher overall than found in previous research on therapist self-report of treatment techniques (Carroll et al. 1998; Hurlburt et al. 2010; Martino et al. 2009). Contrary to hypotheses, therapists were not prone to overestimating their degree of activity in modules and domains, especially when compared to strong positive biases (i.e., over-reporting) documented for techniques (Carroll et al. 1998; Hurlburt et al. 2010). Therapists and observers also agreed that the amount of time spent

working in treatment modules and (to a lesser extent) domains varied somewhat across clients but not across therapists. Pending replication with other EBPs and client populations, results suggest that therapists can be much more reliable and accurate in judging their implementation of treatment targets and foci than specific techniques.

The comparatively strong reliability documented in this study may be attributable to any of the following reasons: use of research-hired therapists (vs. Martino et al. 2009), thorough training and monitoring of therapists in the coding task (vs. Hurlburt et al. 2010), or the measurement specificity associated with rating treatment targets and foci rather than techniques (vs. Carroll et al. 1998). Clearly, the shallow pool of studies on therapist-report fidelity needs to be considerably deepened in order to identify which person, task, and context factors most influence therapist reliability, including whether and how self-report fidelity will be affected by high-stakes performance incentives enacted in front-line settings (Schoenwald et al. 2011). The current study advances this area of work by establishing reliability benchmarks that are achievable under optimal conditions for accurate reporting on treatment parameters and content.

The pattern of ICCs indicates that the highly specifiable Module items—defined primarily by who attends the session—garnered stronger therapist-observer concordance than did the Domain items, which are more broadly defined by the content of participant speech in session. The Prosocial Domain, which was more broadly defined than other Domain items (and rarely benefitted from the word "prosocial" being uttered in session), demonstrated the weakest correlations overall. Also, results from variance components analyses suggest that therapists can accurately discern that they spend time discriminately with adolescents and parents—devoting more or less time in these Modules depending on the presenting needs of the case—but remain constant across cases in the amount of time devoted to interactional work, in keeping with MDFP principles (Hogue et al. 2002a). These results mirror those generated by the observational coders for MDFP therapist adherence to core family therapy techniques for these same cases (Hogue et al. 2005).

Study Limitations

The study sample was too small to test reliably for fidelity-outcome correlations. An important next step is to determine whether levels of prescribed activity with treatment targets and foci predict client outcomes for EBPs of all kinds. Establishing fidelity-outcome links, or the absence thereof, is a high priority for implementation science. There remains little knowledge regarding which aspects of EBPs directly influence outcomes, and virtually no knowledge about which aspects are, and are not, essential for producing key effects. As it stands, model-specific treatment techniques, which have received the most attention to date, have not consistently predicted outcome (Perepletchikova and Kazdin 2005), and when they have, effect sizes are typically small (Webb et al. 2010). However, it is possible that technique-outcome links are stronger in family-based treatment for adolescent behavior problems, as several studies have reported at least some adherence effects on outcome (Hogue et al. 2008a, b; Huey et al. 2000; Robbins et al. 2011). Investigating whether fidelity to the contours of family therapy also predicts outcome is a logical next step. More broadly, advancing research on multiple dimensions of fidelity-outcome links involving the techniques, content, and service delivery parameters of treatment is perhaps the most promising method for uncovering mechanisms of change in EBPs (McLeod et al. 2013).

Another limitation is that this study focused on only aspect of treatment fidelity—treatment adherence (i.e., quantity)—and did not attempt to measure therapist competence (i.e., quality) in implementing MDFP modules and domains. Although virtually every manualized treatment offers guidelines for skillful implementation—how to deliver the appropriate interventions at the appropriate time—it has proven exceedingly difficult to assess competence reliably (Barber et al. 2007; Hogue et al. 2008a, b). Moreover, even reliable measures of competence do not reliably predict outcomes, a counterintuitive but persistent finding (Barber et al. 2007; Hogue et al. 2008a, b).

Other limitations include the small sample size for therapists, the limited number and range of treatment domains measured by the study's fidelity instrument, and inclusion of male therapists only. Study clients were enrolled in prevention counseling and may therefore differ in meaningful ways from families that typically seek outpatient behavioral treatment; however, the central task of accurately self-reporting on model implementation appears little affected by this potential difference. Also, although there was some variability in the strength of correlations across treatment phases, the small samples within each phase, combined with the modest variability in mean scores reported within-phase for each Module and Domain item, make it premature to draw conclusions about phase-specific differences in reporter accuracy for these aspects of treatment.

Implications for Mental Health Services

This study was not designed to validate a specific fidelity instrument or promulgate treatment elements specific to MDFP, but instead, to test a question broadly applicable to fidelity measures and quality assurance procedures of all

kinds: Can therapists estimate their own activity with regard to the targets and foci of treatment with moderate to high reliability and acceptable accuracy? Study results confirm that they can, and this is welcome news for service providers and evaluators invested in quality assurance methods for EBPs. Many evidence-based models have explicit guidelines for whom to treat and what themes to address. Specifications for activity with treatment targets and foci are often contained in quality assurance procedures designed to promote fidelity during the training and sustainability phases of EBP dissemination (McHugh and Barlow 2010). Such specifications are especially prominent in dissemination toolkits for family-based models (Henggeler and Sheidow 2012) and are primary indices of fidelity success in family therapy transportability research (e.g., Liddle et al. 2006), marking them as central to effective implementation in routine care.

In addition to straightforward fidelity monitoring, therapist-report procedures can supply front-line implementation data that establish concrete benchmarks for therapist performance in real-world conditions (McLeod et al. 2013). Such fidelity benchmarks should prove to be essential for calibrating EBP training efforts and assessing the adequacy and consistency of programmatic EBP delivery across clients, therapists, and organizational levels. Indeed, it may be difficult to achieve meaningful improvements in EBP delivery, and subsequent client outcomes, without compiling fidelity benchmarks for treatment parameters and techniques that have been carefully vetted in routine care (Hogue et al. 2013).

For all these reasons, quality assurance procedures for EBP dissemination and implementation will be greatly enhanced by validated methods for therapist self-report of treatment targets and foci, particularly methods that can be easily incorporated into therapist training guidelines, supervision and fidelity monitoring procedures, and clinical feedback systems to improve client care (Garland et al. 2010a, b; Hogue et al. 2013). Moreover, methods for training therapists to rate their own fidelity to treatment contours might "lift the whole boat" by generalizing to ratings of treatment techniques and socializing therapists to become more reflective and programmatic about their interventions. The methods described in this study represent a generic assessment approach that can be readily adapted to fit specifications for all varieties of models—including prescriptions as appropriate for how much time and activity should be devoted to each—and then incorporated as evaluation items into companion fidelity toolkits.

References

Barber, J. P., Sharpless, B., Klostermann, S., & McCarthy, K. S. (2007). Assessing intervention competence and its relation to therapy outcome: A selected review derived from the outcome literature. *Professional Psychology, 38*, 493–500.

Barnoski, R. (2003). *Outcome evaluation of Washington State's research-based programs for juvenile offenders.* Washington State Institute for Public Policy (wsipp.wa.gov).

Bearsley-Smith, C., Sellick, K., Chesters, J., & Francis, K. (2008). Treatment content in child and adolescent mental health services: Development of the treatment recording sheet. *Administration and Policy in Mental Health and mental Health Services Research, 35*, 423–435.

Becker, S. J., & Curry, J. F. (2008). Outpatient interventions for adolescent substance abuse: A quality of evidence review. *Journal of Consulting and Clinical Psychology, 76*, 531–543.

Carroll, K. M., Nich, C., & Rounsaville, B. J. (1998). Utility of therapist session checklists to monitor delivery of coping skills treatment for cocaine abusers. *Psychotherapy Research, 8*, 307–320.

Cicchetti, D. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment, 6*, 284–290.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Diamond, G. S., & Liddle, H. A. (1996). Resolving a therapeutic impasse between parents and adolescents in multidimensional family therapy. *Journal of Consulting and Clinical Psychology, 64*, 481–488.

Diamond, G. S., & Liddle, H. A. (1999). Transforming negative parent-adolescent interactions: From impasse to dialogue. *Family Process, 38*, 5–26.

Dobson, K. S., & Shaw, B. F. (1993). The training of cognitive therapists: What have we learned from treatment manuals? *Psychotherapy, 30*, 573–577.

Fixsen, D.L., Naoom, S.F., Blasé, K.A., Friedman, R.M., & Wallace, F. (2005). *Implementation research: A synthesis of the literature.* Tampa, FL: University of South Florida, Louis de la parte Florida Mental Health Institute, The National Implementation Research Network. (Report No.: FMHI Publication #231).

Fleiss, J. L. (1981). Balanced incomplete block designs for inter-rater reliability studies. *Applied Psychological Measurement, 5*, 105–112.

Garland, A. F., Bickman, L., & Chorpita, B. F. (2010a). Change what? Identifying quality improvement targets by investigating usual mental health care. *Administration and Policy in Mental Health and Mental Health Services Research, 37*, 15–26.

Garland, A. F., Hurlburt, M. S., Brookman-Frazee, L., Taylor, R. M., & Accurso, E. C. (2010b). Methodological challenges of characterizing usual care psychotherapeutic practice. *Administration and Policy in Mental Health and Mental Health Services Research, 37*, 208–220.

Glisson, C., Schoenwald, S. K., Hemmelgarn, A., Green, P., Dukes, D., Armstrong, K. S., et al. (2010). Randomized trial of MST and ARC in a two-level evidence-based treatment implementation strategy. *Journal of Consulting and Clinical Psychology, 78*, 537–550.

Henggeler, S. W., & Sheidow, A. J. (2012). Empirically supported family-based treatments for conduct disorder and delinquency in adolescents. *Journal of Marital and Family Therapy, 38*, 30–58.

Hoagwood, K. E. (2005). Family-based services in children's mental health: A research review and synthesis. *Journal of Child Psychology and Psychiatry, 46*, 690–713.

Hogue, A., Dauber, S., Chinchilla, P., Fried, A., Henderson, C. E., Inclan, J., et al. (2008a). Assessing fidelity in individual and family therapy for adolescent substance abuse. *Journal of Substance Abuse Treatment, 35*, 137–147.

Hogue, A., Dauber, S., Samuolis, J., & Liddle, H. A. (2006). Treatment techniques and outcomes in multidimensional family therapy for adolescent behavior problems. *Journal of Family Psychology, 20*, 535–543.

Hogue, A., Henderson, C. E., Dauber, S., Barajas, P. C., Fried, A., & Liddle, H. A. (2008b). Treatment adherence, competence, and outcome in individual and family therapy for adolescent behavior problems. *Journal of Consulting and Clinical Psychology, 76*, 544–555.

Hogue, A., Johnson-Leckrone, J., & Liddle, H. A. (1999). Recruiting high-risk families into family-based prevention and prevention research. *Journal of Mental Health Counseling, 21*, 337–351.

Hogue, A., & Liddle, H. A. (1999). Family-based preventive intervention: An approach to preventing substance use and antisocial behavior. *American Journal of Orthopsychiatry, 69*, 278–293.

Hogue, A., Liddle, H. A., & Becker, D. (2002a). Multidimensional family prevention for at-risk adolescents. In T. Patterson (Ed.), *Comprehensive handbook of psychotherapy, Vol. II Cognitive-behavioral approaches* (pp. 141–166). New York: Wiley.

Hogue, A., Liddle, H. A., Becker, D., & Johnson-Leckrone, J. (2002b). Family-based prevention counseling for high-risk young adolescents: Immediate outcomes. *Journal of Community Psychology, 30*, 1–22.

Hogue, A., Liddle, H. A., Dauber, S., & Samuolis, J. (2004). Linking session focus to treatment outcome in evidence-based treatments for adolescent substance abuse. *Psychotherapy, 41*, 83–96.

Hogue, A., Liddle, H. A., & Rowe, C. (1996). Treatment adherence process research in family therapy: A rationale and some practical guidelines. *Psychotherapy, 33*, 332–345.

Hogue, A., Liddle, H. A., Singer, A., & Leckrone, J. (2005). Intervention fidelity in family-based prevention counseling for adolescent problem behaviors. *Journal of Community Psychology, 33*, 191–211.

Hogue, A., Ozechowski, T. J., Robbins, M. S., & Waldron, H. B. (2013). Making fidelity an intramural game: Localizing quality assurance procedures to promote sustainability of evidence-based practices in usual care. *Clinical Psychology, 20*, 60–77.

Huey, S., Henggeler, S. W., Brondino, M. J., & Pickrel, S. G. (2000). Mechanisms of change in multisystemic therapy: Reducing delinquent behavior through therapist adherence and improved family and peer functioning. *Journal of Consulting & Clinical Psychology, 68*, 451–467.

Hurlburt, M. S., Garland, A. F., Nguyen, K., & Brookman-Frazee, L. (2010). Child and family therapy process: Concordance of therapist and observational perspectives. *Administration and Policy in Mental Health and Mental Health Services Research, 37*, 230–244.

Kelley, S. D., Vides de Andrade, A. R., Sheffer, E., & Bickman, L. (2010). Exploring the black box: Measuring youth treatment process and progress in usual care. *Administration and Policy in Mental Health and Mental Health Services Research, 37*, 287–300.

Liddle, H. A., Rowe, C. L., Gonzalez, A., Henderson, C. E., Dakof, G. A., & Greenbaum, P. E. (2006). Changing provider practices, program environment, and improving outcomes by transporting multidimensional family therapy to an adolescent drug treatment setting. *The American Journal on Addictions, 15*, 102–112.

Martino, S., Ball, S., Nich, C., Frankforter, T. L., & Carroll, K. M. (2009). Correspondence of motivational enhancement treatment integrity ratings among therapists, supervisors, and observers. *Psychotherapy Research, 19*, 181–193.

McHugh, R. K., & Barlow, D. H. (2010). The dissemination and implementation of evidence-based psychological treatments. *American Psychologist, 65*, 73–84.

McLeod, B. D., Southam-Gerow, M. A., Tully, C. B., Rodriguez, A., & Smith, M. M. (2013). Making a case for treatment integrity as a psychosocial treatment quality indicator for youth mental health care. *Clinical Psychology, 20*, 14–32.

Perepletchikova, F., & Kazdin, A. E. (2005). Treatment integrity and therapeutic change: Issues and research recommendations. *Clinical Psychology, 12*, 365–383.

Robbins, M. S., Feaster, D. J., Horigian, V. E., Puccinelli, M. J., Henderson, C. E., & Szapocznik, J. (2011). Therapist adherence in brief strategic family therapy for adolescent drug abusers. *Journal of Consulting & Clinical Psychology, 79*, 43–53.

Schmidt, S. E., Liddle, H. A., & Dakof, G. A. (1996). Changes in parenting practices and adolescent drug abuse during multidimensional family therapy. *Journal of Family Psychology, 10*, 12–27.

Schoenwald, S. K. (2011). It's a bird, it's a plane, it's…fidelity measurement in the real world. *Clinical Psychology, 18*, 142–147.

Schoenwald, S. K., Chapman, J. E., Kelleher, K. E., Hoagwood, K. E., Landsverk, J., Stevens, J., et al. (2008). A survey of the infrastructure for children's mental health services: Implications for the implementation of empirically supported treatments (ESTs). *Administration and Policy in Mental Health and Mental Health Services Research, 35*, 84–97.

Schoenwald, S. K., Garland, A. F., Chapman, J. E., Frazier, S. L., Sheidow, A. J., & Southam-Gerow, M. A. (2011). Toward the effective and efficient measurement of implementation fidelity. *Administration and Policy in Mental Health and Mental Health Services Research, 38*, 32–43.

Schoenwald, S. K., Letourneau, E. J., & Halliday-Boykins, C. (2005). Predicting therapist adherence to a transported family-based treatment for youth. *Journal of Clinical Child and Adolescent Psychology, 34*, 658–670.

Schoenwald, S. K., Sheidow, A., & Letourneau, E. J. (2004). Toward effective quality assurance in evidence-based practice: Links between expert consultation, therapist fidelity, and child outcomes. *Journal of Clinical Child and Adolescent Psychology, 33*, 94–104.

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin, 86*, 420–428.

Waldron, H. B., & Turner, C. W. (2008). Evidence-based psychosocial treatments for adolescent substance abuse. *Journal of Clinical Child and Adolescent Psychology, 37*, 238–261.

Webb, C. A., DeRubeis, R. J., & Barber, J. P. (2010). Therapist adherence/competence and treatment outcome: A meta-analytic review. *Journal of Consulting and Clinical Psychology, 78*, 200–211.

Weersing, R. V., Weisz, J. R., & Donenberg, G. R. (2002). Development of the therapy procedures checklist: A therapist-report measure of technique use in child and adolescent treatment. *Journal of Clinical Child Psychology, 31*, 168–180.

Zazzali, J. L., Sherbourne, C., Hoagwood, K. E., Greene, D., Bigley, M. F., & Sexton, T. (2008). The adoption and implementation of an evidence based practice in child and family mental health services organizations: A pilot study of functional family therapy in New York State. *Administration and Policy in Mental Health and Mental Health Services Research, 35*, 38–49.