

Adherence Measurement in Treatments for Disruptive Behavior Disorders: Pursuing Clear Vision Through Varied Lenses

Sonja K. Schoenwald, Medical University of South Carolina

Ann F. Garland, University of California, San Diego and Child and Adolescent Services Research Center, Rady Children's Hospital

Michael A. Southam-Gerow, Virginia Commonwealth University

Bruce F. Chorpita, University of California, Los Angeles

Jason E. Chapman, Medical University of South Carolina

To extend the reach, transparency, and accountability for the implementation and outcomes of effective treatments in routine care, more clarity is needed about what happens in treatment. We attempt to (a) clarify terminology to describe and measure psychological treatment and (b) consider what treatment adherence instruments can tell us about what happens in treatment. We reviewed the content of 11 adherence instruments for 14 evidence-based treatments for disruptive behavior problems in youth identified in an ongoing review of adherence measurement methods used in psychosocial treatment studies from 1980 to 2008. Item number, content, and level of detail varied widely. Implications are considered for the definition of effective treatments and design and testing of strategies to measure and monitor their delivery.

Key words: evidence-based treatment, fidelity measurement, treatment specification. [*Clin Psychol Sci Prac* 18: 331–341, 2011]

Address correspondence to Sonja K. Schoenwald, Family Services Research Center, Department of Psychiatry & Behavioral Sciences, MUSC, 67 President Street, STE MC406, MSC 861, Charleston, SC 29425. E-mail: schoensk@musc.edu.

The use of evidence-based practices has been identified as critical to improve the outcomes of mental health care. Within the child treatment and services research community, several investigative lenses are being used to study the processes involved in taking evidence-based interventions to scale and to leverage the knowledge base about these interventions and their implementation to improve the quality and effectiveness of care obtained by larger numbers of children and their families. One set of lenses has focused primarily on understanding the status quo or routine care: (a) services received currently by youth and families and paid for by individual, private, and government payers; (b) factors influencing service access, duration, and effects; and (c) psychotherapy practices as executed by clinicians on the front lines. Another set of lenses has focused primarily on investigating the implementation and outcomes of one or more specific, evidence-based treatments in routine care settings.

In both types of research, an explicit focus has been on identifying potential levers for high quality and sustained uptake of the evidence-based programs and practices. This focus requires grappling with the conceptual and methodological challenges of reliably characterizing what actually happens in treatment, and the extent to which that which happens resembles what is supposed

to happen, given the client, nature of the client's problems, and evidence regarding treatment effects for those problems. Some progress in this regard has been made via three main lines of research. First, some researchers have investigated the therapeutic strategies that characterize distinct treatment orientations (see, e.g., Kennard et al., 2010; McLeod & Weisz, 2010; Weersing, Weisz, & Donenberg, 2002). A second line of research has focused on the role of common factors in psychotherapy (e.g., alliance, empathy, expectations, satisfaction; see, e.g., Nock & Kazdin, 2001; Shirk & Karver, 2003) and the extent to which the topography of common factors can or cannot be understood apart from the types of intervention used (Malik, Beutler, Alimohamed, Gallagher-Thompson, & Thompson, 2003; Sexton, Ridley, & Kleiner, 2004). A third approach has been to use data analytic methods such as meta-analyses (Dieckmann, Malle, & Bodner, 2009; Kaminski, Valle, Filene, & Boyle, 2008; McCarty & Weisz, 2007) and distillation and matching (Chorpita & Daleiden, 2009) to identify those practice elements that appear in multiple evidence-based treatments for a particular clinical problem or class of problems, and to evaluate the extent to which what clinicians do in practice resembles any of these elements (Garland et al., 2010).

The scientific harvest generated by practice and intervention implementation research includes some consistent findings related to the detection of what happens in treatment. For example, studies of routine care practice find therapists report using more practice elements than is objectively observed (see, e.g., Hurlburt, Garland, Nguyen, & Brookman-Frazee, 2010). Similarly, therapists in several multisite trials of evidence-based substance abuse treatment report using elements of evidence-based treatment more frequently than was observed (Carroll, Martino, & Rounsaville, 2010). Such convergence of findings may signal the ubiquity of some human tendencies—social desirability, for example—to influence practitioners' assessment of their deployment of more evidence-based treatments, or elements of EBTs, in new contexts. The findings likely also reflect some of the challenges specific to the status of treatment definition, specification, and measurement. Indeed, the lack of a common language to define and measure what happens in treatment may slow efforts to build a coherent literature on the

dissemination and implementation of effective psychosocial treatments.

There is significant variation in the treatment theory and specification of treatment models and programs that have empirical support and limited consensus on optimal methods to measure the extent to which specific evidence-based treatment models or programs are implemented. Accordingly, the first objective of the current study is to clarify treatment-related terms often used for disparate referents and purposes. The second objective is to see what the lens of adherence measurement can teach us about distinct treatment programs and protocols focused on attenuating a class of clinical problems. As detailed elsewhere (Schoenwald et al., 2010), fidelity (including adherence) measurement methods have been developed and validated for relatively few evidence-based psychosocial interventions; to develop adherence instruments that are both effective (valid and reliable) and efficient (feasibly used in routine practice) requires the navigation of both conceptual and methodological challenges.

To begin to address our second objective, we reviewed the content of adherence instruments for evidence-based treatments for youth with disruptive behavior disorders and their families. These treatments and adherence instruments were identified as part of a larger project evaluating the adherence measurement methods used with evidence-based psychosocial treatments in studies published between 1980 and 2008. The objectives and parameters of the larger review are described elsewhere (Schoenwald et al., 2010). Of the 14 treatments we identified for disruptive behavior disorders, we were able to obtain adherence instruments for 11 of them. Using those 11 instruments, we developed methods to identify the commonalities and distinctions among these treatments, described shortly.

WORKING DEFINITIONS

We begin by clarifying what we mean by terms such as *treatment models*, *programs*, *protocols*, and *practice elements*, as these terms will be used throughout the article.

Terms Describing Treatment

First, we use the term *model* to refer to a broad, theoretically driven psychological treatment approach. As

an example, parent training would be a *model*. Second, we use the term *program* to refer to a clearly defined psychological treatment from a particular model, as specified in a manual. As an example, Parent-Child Interaction Therapy (PCIT) is a program. Third, we use the term *protocol* to refer to a specific instantiation of a program. A single program like PCIT may have multiple protocols associated with it (e.g., early vs. later versions of the manual, variations of the same manual as adapted for a specific population). Note that treatment models and treatment programs are informed by specific treatment theories regarding hypothesized causes of the problems at hand and mechanisms of therapeutic action required to attenuate those problems, and this treatment theory drives treatment specification (content, process, structure; Lipsey, 1988). The level of detail used to specify treatment models and programs varies considerably. Finally, we use the term *practice elements* to refer to specific techniques or components derived from a program or protocol. As an example, parental praising of a child's positive behavior is one practice element contained in the PCIT program (and many other parent training programs).

Treatment Integrity Measurement

To draw valid inferences from clinical trials, treatments must have integrity or fidelity; in other words, the treatment must be well specified, well tested, and carried out as intended (Kazdin, 1994). *Treatment integrity* refers to the degree to which a treatment was delivered as intended and is composed of three components: *treatment adherence*, *treatment differentiation*, and *therapist competence* (Perepletchikova, Treat, & Kazdin, 2007; Waltz, Addis, Koerner, & Jacobson, 1993). *Treatment adherence* refers to the extent to which the delivered treatment includes techniques prescribed by the particular model or manual and excludes or avoids proscribed techniques (Waltz et al., 1993; Yeaton & Sechrest, 1981). To the extent that a treatment manual or protocol specifies the sequencing and amount of treatment operations to be delivered, adherence may also reflect dosage (e.g., McLeod, Southam-Gerow, & Weisz, 2009; Perepletchikova et al., 2007), although dosage is often identified as distinct from adherence in treatment trials and implementation research (Mihalic, 2004).

Treatment differentiation refers to the extent to which treatments under study differ along appropriate lines defined by the treatment manual or protocol; here, the appropriate concept is treatment purity. Differentiation is the extent to which a treatment contains the required elements, does *not* contain prohibited elements, and can be reliably distinguished from other treatment protocols. Differentiation measures like the Treatment Process Observation Coding System (TPOCS; McLeod & Weisz, 2010) have also been used to gauge the contents of usual care, much like a spectrometer does for light (Brookman-Frazee, Haine, Baker-Ericzen, Zoffness, & Garland, 2010; Garland et al., 2010).

Finally, *competence* refers to the level of skill and degree of responsiveness demonstrated by the therapist when delivering the technical and relational elements of treatment. Some have argued that competence can be subdivided into at least two separate components: (a) competence with a particular model and (b) competence at what have been termed common factors, including relational factors (e.g., McLeod et al., 2009). Adherence, differentiation, and competence capture unique aspects of treatment integrity that together, and/or in isolation, may be responsible for therapeutic change or lack thereof (Perepletchikova et al., 2007).

USING THE LENS OF ADHERENCE TO CHARACTERIZE TREATMENT MODELS, PROGRAMS, AND PROTOCOLS

The focus of the present study is treatment adherence, that is, the extent to which therapists adhere to a particular treatment model, program, protocol, or practice element. The importance of adherence (and other aspects of treatment integrity) goes beyond its function as an independent variable manipulation check. As multisite effectiveness and implementation trials become more common, there is an increasing need to ensure that the treatment model, program, protocol, or element of interest is implemented in a similar manner across multiple sites. Absent valid and reliable adherence data, for example, the observation of significant site effects on client outcomes could reasonably be interpreted as an indicator the treatment worked in some sites but not others. Evidence of site differences in adherence, however, would suggest an alternative interpretation, namely that outcome differences reflect

decrements in treatment implementation, rather than treatment failure.

Despite the importance of establishing treatment integrity, the definition and measurement of all three components of treatment integrity are underdeveloped in child therapy (Perepletchikova et al., 2007; Weisz, Doss, & Hawley, 2005). In a review of 226 child therapy studies published from 1962 to 2002, Weisz et al. (2005) found only 32.2% used any form of integrity check. Perepletchikova et al. (2007) reviewed 147 child and adult therapy studies published from 2000 to 2004 and found only 3.5% *adequately* measured treatment integrity. Emerging interest in treatment adherence also comes from outside of intervention and implementation science. As evidence-based programs become more commonplace across a variety of mental health contexts, payers and other stakeholders in the system are asking for methods of knowing if the treatment program they have paid for is the one being delivered. Here, the adherence measure can serve as an accountability tool (Schoenwald et al., 2010; Yeaton & Sechrest, 1981).

Methods to characterize the variability and commonality in treatment theory and specification across treatment models, programs, and protocols are needed to advance research designed to increase the reach and adequate implementation of effective psychosocial treatment for particular target populations. As noted previously, quantitative methods such as meta-analysis and distillation and matching offer some promise in this regard. These approaches aim to illuminate the elements of treatments used across specific treatment programs and systematic variation in characteristics of client populations with which these elements were used. So far, however, these approaches do not purport to address the combination, sequencing, dosage, and decision-making processes needed to deploy the elements for a particular class of problems or for an individual client, although identification of common processes, algorithms, etc., can be pursued using the same methodology (Chorpita, Daleiden, & Weisz, 2005). An alternative strategy is to deploy rational (rather than empirically identified or mined) clinical decision-making algorithms that incorporate client response and comorbid conditions when implementing common elements (e.g., Chorpita & Weisz, 2009).

Finally, theory-driven approaches emerging in the adult and child psychotherapy literature, such as transdiagnostic treatment models, aim to address the common mechanisms of action known to contribute to distinct disorders (i.e., cognitive distortions in anxiety and depression) with common mechanisms of therapeutic action (Ehrenreich, Goldstein, Wright, & Barlow, 2009; McHugh, Murray, & Barlow, 2009).

It would also seem reasonable to expect that one could derive, either empirically or rationally, some information about the variation and commonality in critical contents, processes, and structures of treatment models, programs, and protocols from the instruments used to index adherence to them. This was the hypothesis that catalyzed the current exploratory study. We attempted a rational approach to cataloging similarities and differences in treatment programs for a common clinical problem as specified in the instruments used to index adherence to these treatments.

A STUDY OF ADHERENCE INSTRUMENT CONTENT

The current study was designed to evaluate the extent to which fidelity measurement instruments for evidence-based child and family treatments for a particular class of youth problems—disruptive behavior disorders—might illuminate a pathway to the development of a “yardstick of therapeutic accountability” (Yeaton & Sechrest, 1981). Such a “yardstick” could be used in routine care settings to “provide individual-level metrics” (Miranda, Azocar, & Burnam, 2010, p. 206) of clinician behavior that is consistent with effective treatments for disruptive behavior disorders. Likewise, this could ultimately be evaluated with respect to relations with youth outcomes. Could existing adherence measures for evidence-based treatments for this particular class of clinical problem illuminate critical contours of these treatments to which fidelity of implementation could be indexed (reliably and validly) by a single set of common indicators?

Although prior research has identified several practice elements that characterize multiple individual treatment programs for this clinical problem area (Chorpita & Daleiden, 2009; Garland, Hawley, Brookman-Frazee, & Hurlburt, 2008), the extent to which their adherence measurement methods share commonality is unknown. Thus, the aims of this study were to

compile all the items from the existing adherence instruments for evidence-based treatments for children with disruptive behavior disorders and develop a classification scheme to evaluate the extent to which identical or functionally similar items appeared across instruments.

METHOD

Identification and Procurement of Adherence Instruments

The data for the current manuscript were obtained from an ongoing investigation of the Implementation Methods Research Group (IMRG; see, e.g., Schoenwald et al., 2010). That investigation is evaluating the adherence measurement methods reported in research on evidence-based psychosocial treatments for children and adults published in peer-reviewed journals between 1980 and 2008. In that ongoing investigation, *evidence-based* was defined as treatments demonstrating efficacy or effectiveness in randomized trials. Article inclusion criteria included presence of the terms *fidelity* or *adherence* in the abstracts, keywords, or Method section of the articles. A total of 342 articles were retained for the larger study. Four trained coders used a manual to code articles with respect to the details of a wide range of study-level (e.g., clinical setting, clinician type, demographics, clinical problem treated, and client demographics), treatment-level (e.g., treatment type, specific treatment program, or protocol name), and adherence measurement method (e.g., type, sampling method and procedures, psychometric properties including distinct types of reliability and validity, documentation of relations between adherence and outcomes) variables. Following the establishment of adequate inter-rater reliability in a sample of 40 articles using Krippendorff's alpha (Hayes & Krippendorff, 2007), 10% of all articles were rated by multiple coders throughout the coding process.

From this pool of 342 articles, 53 articles reporting on 46 studies of 14 treatment programs and protocols for youth with disruptive behavior disorders were identified. The treatments are listed in Table 1. Copies of the adherence measurement instruments for these treatments were obtained from the articles, treatment program Web sites, and the developers of the treatment programs and protocols. Using these procurement methods, adherence instruments were obtained for 11

of the 14 treatments. Reasons instruments were not obtained were as follows: (a) Checklists identified as indicators of treatment implementation in published studies were not considered by the developers to be adherence instruments and had not been validated ($n = 1$); (b) The instrument indexed something other than adherence (consumer satisfaction) ($n = 1$); and (c) The instruments were neither publicly available nor provided by investigators using them ($n = 2$). For two treatment protocols, insufficient information was presented in the pertinent articles to code adherence measurement.

Analyses

Before conducting the item content review, descriptive analyses were conducted to evaluate the reliability statistics (e.g., Cohen's kappa, intraclass correlations, Kendall's coefficient of concordance, Cronbach's alpha, split-half reliability, test-retest reliability reported as Spearman's rho, kappa, or other methods) and validity statistics (construct, concurrent, discriminant, predictive) of the instruments. Then, two authors reviewed all instruments to identify literal and functional similarities of the instrument items. The remaining authors reviewed subsets of items and, using an iterative process, worked toward consensus regarding a classification scheme for characterizing item content across instruments. One candidate scheme, for example, included the dimensions of structure, content, process, techniques, and participants. The iterative process failed, however, to generate a classification scheme with sufficiently few categories and sufficiently large numbers of items within a category to render it meaningful. The Results section presents attributes of the instruments believed to contribute to the failure of our item classification efforts.

RESULTS

A total of 3,521 items appeared across the instruments used to index adherence for the 11 treatments. There was a large range (4 to 1,733) in the number of items on the instruments. Three treatment programs—Parent-Child Interaction Therapy (PCIT; Zisser & Eyberg, 2010), Incredible Years (IY; Webster-Stratton & Reid, 2010), and Coping Power (CP; Lochman, Boxmeyer, Powell, Barry, & Pardini, 2010)—used

Table 1. Treatments for youth with disruptive behavior disorders identified in literature review as having adherence measures

Treatment Program or Model

Brief Strategic Family Therapy (BSFT)
Cognitive Behavioral Therapy (CBT) with Interactional Group Therapy
CBT with Problem Solving Skills Training (CBT + PSST)
Coping Power
Functional Family Therapy (FFT)
Incredible Years (IY)
Multidimensional Family Therapy (MDFT) plus CBT
Multidimensional Treatment Foster Care (MTFC)
Multisystemic Therapy (MST) and Multisystemic Therapy—Contingency Management (MST-CM)
Parent–Child Interaction Training (PCIT)
Parenting Through Change
Parent Management Training (PMT)
Parent Training Program
Triple P

session-by-session checklists, and the number of items on these checklists (3,328) accounted for 94.5% of the total number of items (3,521) observed across instruments. These were not 3,521 uniquely worded items in that the same item (e.g., “*Review goal sheets and main points from previous session*”) could be included in more than one session checklist, but they are counted as separate items because they are used uniquely in the context of different sessions. In addition, two treatment programs included different versions of the instrument used with each of several participants in treatment: the clinician, the parent, and the youth.

The extent to which adherence was assessed and adherence instruments were validated also varied. Adherence was assessed in 67.4% (31) of the 46 published studies. The psychometric properties of adherence measures were reported in 58.1% (18) of this subsample of 31 studies in which adherence was assessed. Use of an adherence instrument to distinguish between the treatment and another type of treatment was reported in 6.5% (2) of these 31 studies, and adherence–outcomes relations were reported for 19.4% (6) of the 31 studies. For one treatment program, an observational coding instrument had been validated, whereas the clinician report version had not.

Across the adherence instruments, items differed dramatically across numerous dimensions, including most notably the level of detail used to specify the construct for assessment (e.g., a macro-level principle vs. a more micro-level-specific detailed element). Items also differed dramatically on what was being assessed (e.g., an environmental or contextual feature, materials or

content provided, provider behavior, client behavior, or some interaction between these). Items also differed markedly on how adherence was assessed (e.g., presence/absence of an element, amount/duration of activity, or quality of intervention activity) and by whom it was assessed (e.g., clinician self-report, observer rating, client report).

To illustrate just some of this variability, Table 2 presents selected examples of different types of adherence items along two dimensions: level of specificity and what is being assessed. This table is not meant to reflect the comprehensive range of items, but rather just to highlight some aspects of the diversity. The dimensions that appear in this table are two of the most salient. There are, however, many other ways in which items differ. Many of the items could not be classified in accordance with the dimensions on the table, nor in accordance with several multidimensional frameworks generated for the purposes of item categorization.

DISCUSSION

Our review of the contents of the instruments used to index adherence to empirically supported treatment programs and models for youth with disruptive behavior disorders revealed tremendous variation in the number and details of items. The variation was so great that we were unable to identify dimensions (e.g., content, process, structure, techniques, participants) along which the items could be classified for the purposes of further evaluation. On the one hand, this variability is to be expected, given the distinctions in treatment

Table 2. Examples of variation in adherence measurement items by level of specificity assessment target

Level of Specificity	What Is Being Assessed	
	Environment/Context	Provider Behavior
Micro	Generic: Had item X available to accomplish goal Z 1. Was a co-leader present for this session? 2. Was equipment set up and workable?	Generic: Used item X to teach client skill Y to accomplish goal Z 1. Have puppets say goodbye 2. Ask each child to identify one positive thing about himself/herself
Intermediate	Generic: Materials available to accomplish goal Z 1. Write the agenda on the board 2. Set up a friendship station in the classroom	Generic: Taught skill Y 1. Have children practice the actual behaviors being taught through puppet plays, role plays, live role plays, and peer-coached play 2. Discuss difficulty expressing feelings and normalize feelings
Macro	Generic: Session supported working on goal Z 1. Did the atmosphere support humor? 2. Create a feeling of safety in the group	Generic: Worked to accomplish goal Z 1. Were positive aspects of the youth's behavior noted? 2. Joins with children/adolescents

theories and specification that characterize empirically supported treatment models, programs, and protocols. That specification is typically informed by an accumulation of years of scientific data, clinical trial and error, and aims to maximize therapeutic gain via the synergistic sequencing of treatment content. On the other hand, this high level of variability did not allow us to evaluate the extent to which adherence to any operations embodied in the treatment programs could be understood using a common frame of reference and common terminology. Developing such a frame of reference for treatment specification, and for adherence measurement methods, could help to more rapidly advance research on strategies to extend the reach of effective psychosocial treatment for children and families in diverse practice contexts. The remainder of this section identifies strategies to begin to address this research agenda.

Treatment Definition and Specification

We propose that advancing the reach of effective treatment in routine care may require some rational agreement among treatment development, research, and implementation experts regarding the optimal *definition* of at least those practice elements commonly found across a number of distinct models and programs within models as well as the adequate specification of those elements. In short, consensus is needed regarding the answer to the question, “What level of detail is needed to adequately define which treatment opera-

tions?” The common elements identified in distillation and matching, meta-analytic, and practice coding studies provide a promising starting place for the development of such consensus. For example, a set of common elements of treatments for a class of disorders derived using one of these methods could be compared for evidence of overlap, and specification detail, with the content of the manuals for distinct evidence-based treatments.

Adherence Measurement Methods

Consensus regarding the definition and specification of common treatment operations would also assist in the development of fidelity *measurement methods*. Here, the issue of measurement *purpose* looms large. Fidelity measures have typically been developed to serve a distinct purpose, and that purpose drives the many decisions involved in measurement development and testing. However, the desired uses of a fidelity measure may vary across contexts and stakeholders (Schoenwald et al., 2010). For example, fidelity measures designed to ensure the precise delivery of the independent variable in a treatment trial might require a higher level of specificity (i.e., a greater number, or more detailed items or both) than fidelity measures designed for the purposes of meta-analytic knowledge accumulation. The latter measures would need to specify treatment operations sufficiently broadly to allow for aggregation across different protocols (e.g., time-out procedures from two different protocols with slightly different

procedures could be considered the same procedure in a meta-analytic summary).

To accurately and ethically serve “high stakes” decisions in routine care (i.e., regarding program continuation or termination, reimbursement for psychotherapy sessions, hiring and firing staff), more detailed definition of treatment operations may be needed in adherence items and in the metrics used to detect adherence (presence/absence, amount, etc.). Broader treatment specification and measurement methods may be appropriate for quality management (clinician training to particular skills or competencies, monitoring and improvement of these as needed) in contexts such as community mental health centers, and where such centers are part of a specific service system, as one indicator of quality at a system level. In each of these scenarios, assurance is needed that the data generated by the measurement can be legitimately used to answer the question(s) at hand. Obtaining such assurance involves both scientific and practical considerations, as detailed elsewhere (Schoenwald et al., 2010). Just as sound measures that are impractical will not likely be used, so too practicality is not sufficient to justify use. A highly practical measure may not be suitable for making valid and reliable “high stakes” decisions.

Are Variable View Strategies Viable?

This potential diversity in the purposes and levels of specification and measurement raises the question: to what extent can treatment operations and the fidelity measurement instruments designed to index their implementation be specified in ways that allow for aggregation and dissection to accommodate distinct purposes? This is an empirical question. Ideally, the development and evaluation of fidelity measures would proceed according to the following steps: (a) determination of the purpose of those measures (e.g., independent variable delivery in randomized treatment trials, meta-analysis, verification of model, program, or protocol delivery in practice settings, quality management), (b) rational agreement on an appropriate level of specificity for that purpose, (c) design of the measures, (d) validation of measures, and (e) application of the measures in their appropriate context.

Elaboration is warranted about the substantive implications of step b (selecting a preferred level of

specificity) in this sequence. Rational treatment specification is more than simply a matter of choosing the level of detail. As explicated by others (Lipsey, 1988; Yeaton & Sechrest, 1981), it also forces the articulation of aspects considered to be the *essence* of that treatment. For example, some treatments could look quite different from one episode to another, from case to case, and yet would be considered the same treatment by the developer because they employ the same underlying logic or principles. This possibility illuminates the need in the field for greater clarity regarding the mechanisms of action by which treatment models, programs, or protocols are hypothesized to affect client progress and outcomes and empirical evaluation of these hypotheses. Ultimately, empirical evidence is needed to demonstrate the extent to which aspects of treatment considered to be its essence are indeed putative. Such aspects likely include not only specific treatment content but also the logic used to select and sequence content delivery (see, e.g., Chorpita & Weisz, 2009; Henggeler, Schoenwald, Borduin, Rowland, & Cunningham, 2009). Thus, regardless of the level of detail that characterizes the specification of a treatment model, program, or practice element, if the logic used to choose and apply the content is considered (and, ultimately, demonstrated to be) critical to the treatment and its outcomes, then indices of fidelity would also need to capture this logic and sequencing.

It is possible that an instrument designed to detect independent variable delivery in research settings can also be effectively and efficiently used for purposes such as quality management. Alternatively, the stepwise sequence of measurement development may have to begin anew if the primary purpose of measurement is quality assurance (i.e., if use of the original measure is infeasible). Another possibility is that the adaptation and application of existing measures for new purposes could take place in the final step (i.e., application of the measures in their appropriate context). It is this possibility we hoped to exploit in the current study. We were perhaps overly ambitious and also naïve with respect to the scope (14 treatments for youth with disruptive behavior disorders) and methods required to identify dimensions along which extant adherence instruments could be characterized. Methodologies such as semantic differential tasks, concept mapping,

and item coding are among those that may hold promise for similar, future efforts.

A study recently funded by the NIMH takes an initial step in this direction by classifying different types of treatment operations reflected in two different evidence-based treatment programs for child anxiety evaluated across three separate randomized controlled trials (Southam-Gerow & McLeod, R01-MH086529). The investigators will develop, on the basis of this classification scheme, an adherence instrument that will be tested across the three randomized controlled trials, two of which are effectiveness trials including a usual care control condition. The study is the first attempt to our knowledge to build on an existing adherence instrument developed for the primary purpose of independent variable verification (and characterized by the “gold standard” of observational coding) to generate a valid, reliable, and lower-burden fidelity instrument that could be used to detect in routine care the implementation of both a treatment model (cognitive behavioral therapy for individual youth) and distinct treatment elements.

CONCLUSION

In conclusion, our results signal the importance of clarifying the level of specificity of treatment contents, process, sequencing, and logic to advancing the effectiveness and utility of treatment integrity measurement methods used in effectiveness, implementation, and dissemination research. The field is ripe for the development of a clearly delineated, hierarchical structure that captures meaningful differences in the level of specification with which treatment operations are defined and, in turn, can be used to inform advances in adherence measurement. Sometimes consumers, referral agents, and payers will want to know if IY is being purchased and delivered. At other times, the goal will be known if a parent training was delivered (regardless of the specific program). Or, the goal may be to know whether or not a behaviorally based (or other model) therapy was delivered. Measures can be designed to detect any of these levels, but a single measure cannot likely be constructed that detects all of them. Accordingly, we do not think it likely that a single, common “yardstick of accountability” will emerge for treatments found to be effective for a particular class of disorders or target population.

The notion that variations in the operational definitions of a phenomenon can be hierarchically organized to facilitate the valid and reliable aggregation or disaggregation of data about that phenomenon for a particular purpose is not unique to treatment research, but characterizes other sciences as well, as does the search for methodologies to accomplish this task (Chorpita et al., 2005). Although not an example from the social sciences, the popular Internet-based product Google Earth illustrates the potential utility of such an approach. Google Earth has the capacity to generate views that range from macro-planetary outlines to large geographic parcels of land to specific streets and even to the homes on such streets. Similarly, a hierarchically organized framework of treatment operations could aid in the organization of existing adherence instruments and in the development of new instruments from which some data could be used to generate variable views.

First steps in the development of such a hierarchically organized framework include the development of greater conceptual consensus on the level of detail necessary and sufficient to characterize key treatment operations (content, process, structure, participants) for particular classes of disorders or types of clinical problems. Validation will be needed of instruments to assess the extent to which operations thus defined occur. Finally, empirical evaluation will be needed of the extent to which (a) these operations as indexed by the instruments relate to one another within and across treatment protocols, programs, and models; and (b) data from the instruments can be aggregated and disaggregated to yield valid and reliable indicators for the desired purposes, including assessment of relations with client outcomes.

ACKNOWLEDGMENT

The primary support for this manuscript was provided by NIMH research grant P30 MH074778 (J. Landsverk, PI). Sonja K. Schoenwald is a board member and stockholder in MST Services, LLC, which has the exclusive licensing agreement through the Medical University of South Carolina for the dissemination of MST technology. Bruce F. Chorpita is the president of PracticeWise, LLC, a private behavioral health consulting partnership. The authors thank Drs. John Lochman, Jose Szapocznik, and Ron Prinz and their colleagues

for providing some of the adherence instruments and information about their uses.

REFERENCES

- Brookman-Frazee, L., Haine, R. A., Baker-Ericzen, M., Zoffness, R., & Garland, A. F. (2010). Factors associated with use of evidence-based practice strategies in usual care youth psychotherapy. *Administration and Policy in Mental Health and Mental Health Services Research, 37*, 254–269.
- Carroll, K. M., Martino, S., & Rounsaville, B. J. (2010). No train, no gain? *Clinical Psychology: Science and Practice, 17*, 36–38.
- Chorpita, B. F., & Daleiden, E. L. (2009). Mapping evidence-based treatments for children and adolescents: Application of the distillation and matching model to 615 treatments from 322 randomized trials. *Journal of Consulting and Clinical Psychology, 77*, 566–579.
- Chorpita, B. F., Daleiden, E. L., & Weisz, J. R. (2005). Identifying and selecting the common elements of evidence-based interventions: A distillation and matching model. *Mental Health Services Research, 7*, 5–20.
- Chorpita, B. F., & Weisz, J. R. (2009). *MATCH-ADTC: Modular approach to therapy for children with anxiety, depression, trauma, or conduct problems*. Satellite Beach, FL: PracticeWise, LLC.
- Dieckmann, N. F., Malle, B. F., & Bodner, T. E. (2009). An empirical assessment of meta-analytic practice. *Review of General Psychology, 13*, 101–115.
- Ehrenreich, J. T., Goldstein, C. R., Wright, L. R., & Barlow, D. H. (2009). Development of a unified protocol for the treatment of emotional disorders in youth. *Child & Family Behavior Therapy, 31*, 20–37.
- Garland, A. F., Brookman-Frazee, L., Hurlburt, M. S., Accurso, E. C., Zoffness, R., Haine, R. A., et al. (2010). Mental health care for children with disruptive behavior problems: A view inside therapists' offices. *Psychiatric Services, 61*, 788–795.
- Garland, A. F., Hawley, K. M., Brookman-Frazee, L. B., & Hurlburt, M. S. (2008). Identifying common elements of evidence-based psychosocial treatments for children's disruptive behavior problems. *Journal of the American Academy of Child and Adolescent Psychiatry, 47*, 505–514.
- Hayes, A. G., & Krippendorff, K. (2007). Answering the call for a standard reliability measure for coding data. *Communication Methods and Measures, 1*, 77–89.
- Henggeler, S. W., Schoenwald, S. K., Borduin, C. M., Rowland, M. D., & Cunningham, P. B. (2009). *Multisystemic therapy for antisocial behavior in children and adolescents* (2nd ed.). New York: Guilford Press.
- 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.
- Kaminski, J. W., Valle, L. A., Filene, J. H., & Boyle, C. L. (2008). A meta-analytic review of components associated with parent training program effectiveness. *Journal of Abnormal Child Psychology, 36*, 567–589.
- Kazdin, A. E. (1994). Methodology, design, and evaluation in psychotherapy research. In A. E. Bergin & S. L. Garfield (Eds.), *Handbook of psychotherapy and behavior change* (pp. 19–71). New York: John Wiley & Sons.
- Kennard, B. D., Clarke, G. N., Weersing, V. R., Asarnow, J. R., Shamseddeen, W., Porta, G., et al. (2010). Effective components of TORDIA cognitive-behavioral therapy for adolescent depression: Preliminary findings. *Journal of Consulting and Clinical Psychology, 77*, 1033–1041.
- Lipsey, M. W. (1988). Juvenile delinquency intervention. In H. S. Bloom, D. S. Cordray, & R. J. Light (Eds.), *Lessons from selected program and policy areas: New directions for program evaluation* (pp. 63–84). San Francisco, CA: Jossey-Bass.
- Lochman, J. E., Boxmeyer, C. L., Powell, N. P., Barry, T. D., & Pardini, D. A. (2010). Anger control training for aggressive youths. In J. R. Weisz & A. E. Kazdin (Eds.), *Evidence-based psychotherapies for children and adolescents* (pp. 227–242). New York: Guilford Press.
- Malik, L., Beutler, L. E., Alimohamed, S., Gallagher-Thompson, D., & Thompson, L. (2003). Are all cognitive therapies alike? A comparison of cognitive and noncognitive therapy process and implications for the application of empirically supported treatments. *Journal of Consulting and Clinical Psychology, 71*, 150–158.
- McCarty, C. A., & Weisz, J. R. (2007). Effects of psychotherapy in children and adolescents: What we can (and can't) learn from meta-analysis and component profiling. *Journal of the American Academy of Child and Adolescent Psychiatry, 46*, 879–886.
- McHugh, R. K., Murray, H. W., & Barlow, D. H. (2009). Balancing fidelity and adaptation in the dissemination of empirically-supported treatments: The promise of transdiagnostic interventions. *Behavior Research and Therapy, 47*, 946–953.
- McLeod, B. D., Southam-Gerow, M. A., & Weisz, J. R. (2009). Conceptual and methodological issues in treatment integrity measurement. *School Psychology Review, 38*, 541–546.
- McLeod, B. D., & Weisz, J. R. (2010). The Therapy Process Observational Coding System for Child Psychotherapy

- Strategies Scale. *Journal of Clinical Child and Adolescent Psychology*, 39, 436–443.
- Mihalic, S. (2004). The importance of implementation fidelity. *Emotional and Behavioral Disorders in Youth*, 4, 83–86, 99–105.
- Miranda, J., Azocar, F., & Burnam, M. A. (2010). Assessment of evidence-based psychotherapy practices in usual care: Challenges, promising approaches, and future directions. *Administration and Policy in Mental Health and Mental Health Services Research*, 37, 205–207.
- Nock, M. K., & Kazdin, A. E. (2001). Parent expectancies for child therapy: Assessment and relation to participation in treatment. *Journal of Child and Family Studies*, 10, 155–180.
- Perepletchikova, F., Treat, T. A., & Kazdin, A. E. (2007). Treatment integrity in psychotherapy research: Analysis of the studies and examination of the associated factors. *Journal of Consulting and Clinical Psychology*, 75, 829–841.
- Schoenwald, S. K., Garland, A. F., Chapman, J. E., Frazier, S. L., Sheidow, A. J., & Southam-Gerow, M. A. (2010). Toward the effective and efficient measurement of implementation fidelity. *Administration and Policy in Mental Health and Mental Health Services Research*, 38, 32–43. doi: 10.1007/s10488-010-0321-0
- Sexton, T. L., Ridley, C. B., & Kleiner, A. J. (2004). Beyond common factors: Multilevel-process models of therapeutic change in marriage and family therapy. *Journal of Marital and Family Therapy*, 30, 1–12.
- Shirk, S. R., & Karver, M. (2003). Prediction of treatment outcome from relationship variables in child and adolescent therapy: A meta-analytic review. *Journal of Consulting and Clinical Psychology*, 71, 452–464.
- Waltz, J., Addis, M. E., Koerner, K., & Jacobson, N. E. (1993). Testing the integrity of a psychotherapy protocol: Assessment of adherence and competence. *Journal of Consulting and Clinical Psychology*, 61, 620–630.
- Webster-Stratton, C., & Reid, M. J. (2010). The Incredible Years parents, teachers, and children training series: A multifaceted treatment approach for young children with conduct disorders. In J. R. Weisz & A. E. Kazdin (Eds.), *Evidence-based psychotherapies for children and adolescents* (pp. 194–210). New York: Guilford Press.
- Weersing, V. R., 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.
- Weisz, J. R., Doss, A. J., & Hawley, K. M. (2005). Youth psychotherapy outcome research: A review and critique of the evidence base. *Annual Review of Psychology*, 56, 337–363.
- Yeaton, W. H., & Sechrest, L. (1981). Critical dimensions in the choice and maintenance of successful treatments: Strength, integrity, and effectiveness. *Journal of Consulting and Clinical Psychology*, 49, 156–167.
- Zisser, A., & Eyberg, S. M. (2010). Parent-child interaction therapy and the treatment of disruptive behavior disorders. In J. R. Weisz & A. E. Kazdin (Eds.), *Evidence-based psychotherapies for children and adolescents* (pp. 179–193). New York: Guilford Press.

Received September 18, 2011; accepted September 18, 2011.