Journal of Clinical Child & Adolescent Psychology, 42(1), 44–55, 2013

ISSN: 1537-4416 print/1537-4424 online
DOI: 10.1080/15374416.2012.700505

Routledge
Taylor & Francis Group

EVIDENCE-BASED PRACTICE AND DISSEMINATION ISSUES

Tracking Evidence Based Practice with Youth: Validity of the MATCH and Standard Manual Consultation Records

Alyssa M. Ward, Jennifer Regan, Bruce F. Chorpita, Nicole Starace, Adriana Rodriguez, and Kelsie Okamura
Department of Psychology, University of California, Los Angeles

Eric L. Daleiden
Kismetrics, LLC

Sarah Kate Bearman and John R. Weisz
Department of Psychology, Judge Baker Children's Center, Harvard University

The Research Network on Youth Mental Health

This study sought to evaluate the agreement between therapist report and coder observation of therapy practices. The study sampled session data from a community-based, randomized trial of treatment for youth ages 7 to 13. We used therapist report of session content and coverage gathered using formal Consultation Records and developed complimentary records for coders to use when watching or listening to therapy tape. We established initial reliability between coders and then conducted a random, stratified, and comprehensive sample of sessions across youth ($N = 121$), therapists ($N = 57$), conditions (MATCH and Standard Manuals), and study sites (Honolulu and Boston) to code and compare with therapist record reports. Intraclass correlation coefficients (ICCs) representing coder versus therapist agreement on manual content delivered ranged from .42 to 1.0 across conditions and problem areas. Analyses revealed marked variability in agreement regarding whether behavioral rehearsals took place (ICCs from −.01 to 1.0) but strong agreement on client comprehension of therapy content and homework assignments. Overall, the findings indicate that therapists can be accurate reporters of the therapeutic practices they deliver, although they may need more support in reporting subtle but valuable aspects of implementation such as types of behavioral rehearsals. Developing means to support accurate reporting is important to developing

The Research Network on Youth Mental Health: Charles Glisson, Children's Mental Health Services Research Center, University of Tennessee at Knoxville; Evelyn Polk Green, Attention Deficit Disorder Association, Wilmington, Delaware; Kimberly Eaton Hoagwood, Department of Mental Health Services and Policy Research, New York State Psychiatric Institute, Columbia University, New York; Kelly Kelleher, Research Institute of Nationwide Children's Hospital, and Ohio State University, Columbus; John Landsverk, Child and Adolescent Services Research Center, University of California, San Diego; and Stephen Mayberg, California Department of Mental Health, Sacramento.

Financial Disclosure: Modular Approach to Treatment of Children with Anxiety, Depression, or Conduct Problems was the precursor to a manual for which Drs Chorpita, Daleiden, and Weisz may receive income. Funding/Support: The study was supported by the MacArthur Foundation, the Norlien Foundation, and the National Institute of Mental Health. Role of the Sponsors: The funders/sponsors did not participate in the design and conduct of the study; in the collection, analysis, and interpretation of the data; or in the preparation, review, or approval of the manuscript.

Correspondence should be addressed to Alyssa M. Ward, Department of Psychology, University of California, 1285 Franz Hall, Box 951563, Los Angeles, CA 90095. E-mail: alymargaret@gmail.com

future clinical feedback methodology applicable to the implementation of evidence-based treatments in the real world.

INTRODUCTION

The past half century has seen a proliferation of randomized controlled trials to understand what psychotherapeutic treatment approaches work best for youth (Chorpita et al., 2011; Kazdin, Bass, Ayers, & Rodgers, 1990; Weisz, Doss, & Hawley, 2005), and the last few decades of this work have witnessed increased emphasis on the more subtle aspects of why treatments do or do not succeed as they are tested in increasingly diverse contexts (Kazdin, 2000; Southam-Gerow et al., 2010; Weisz et al., 2009). One aspect of this deeper understanding of the ''why'' or ''how'' treatments perform is treatment integrity, that is, the degree to which treatments are implemented as intended and the quality with which they are delivered (Perepletchikova & Kazdin, 2005; Waltz, Addis, Koerner, & Jacobson, 1993).

This need to put a lens on the treatment delivery process becomes increasingly relevant when considering testing or implementing evidence-based treatments (EBTs) in everyday service settings. For example, real-world therapists have voiced numerous concerns about the fit of EBTs into regular clinical practice (Addis & Krasnow, 2000; Addis, Wade, & Hatgis, 1999), and reports indicate important differences between client populations in research versus regular clinical care (Baker-Ericzen, Hurlburt, Brookman-Frazee, Jenkins, & Hough, 2009; Southam-Gerow, Chorpita, Miller, & Gleacher, 2008). Given the diversity of clients, workforces, system goals, and organizational contexts, everyday settings are likely to elicit local adaptations and modifications to treatments (Rogers, 2003), and without some metrics for monitoring practice there is no way to distinguish which adaptations are ''drift'' and which are enhancements. Such inferences are important to both the research and the service enterprise.

In the field of psychotherapy research, the collective findings are somewhat equivocal regarding the relation of treatment adaptation to clinical outcomes (McHugh, Murray, & Barlow, 2009), and the complexities of even interpreting integrity findings have been well documented (e.g., Feeley, DeRubeis, & Gelfand, 1999; Webb, DeRubeis, & Barber, 2010). On one hand, some literature suggests that fidelity is an important factor in assuring positive clinical outcomes when treatments are implemented in community contexts (e.g., Schoenwald, Carter, Chapman, & Sheidow, 2008). On the other hand, others have argued that flexibility can be beneficial (Chorpita, Taylor, Francis, Moffitt, & Austin, 2004; Chu & Kendall, 2009; Jacobson et al., 1989), especially in complex settings (Chorpita & Daleiden, 2010). Some of this issue is perhaps a false dichotomy, in that some protocols are inherently flexible by design and, hence, treatment adaptation can be consonant with treatment fidelity. Further complicating matters is that there is no consensus on what is meant by flexibility or fidelity, with some arguing that fidelity requires that adaptation be limited to within-session customization of protocol content and others arguing that varying such things as content, order, and dose are within the boundaries of treatment fidelity. Amidst this controversy, a recent randomized effectiveness trial examining this very question compared therapists who used EBT procedures for anxiety, depression, and conduct problems in a standard fashion with those who used those procedures as part of a modular treatment system that allowed systematic adaptation (Chorpita & Weisz, 2009; Weisz et al., 2011). The findings suggested that controlled adaptation was associated with the most improvement on child and parent reported outcomes, with the modular condition showing superior outcomes to the standard EBTs. Further work is necessary to explore the complex and perhaps curvilinear relationship between fidelity and treatment outcome, and instrumentation around measuring therapy practices may be useful in this pursuit.

With the premise that adaptation might be important as EBTs are engineered for use in complex settings, a critical question becomes how to guide that adaptation. Bickman (2008b) emphasized the use of feedback and monitoring systems that track client progress, such that feedback indicating a poor response to treatment would identify opportunities for changing the treatment approach (cf. Daleiden & Chorpita, 2005). Not surprisingly, feedback systems designed to provide written/graphic information about clinical progress have been shown to improve outcomes in adult mental health (e.g., Hawkins, Lambert, Vermeersch, Slade, & Tuttle, 2004; Lambert et al., 2001; Whipple et al., 2003). More recently, Chorpita, Bernstein, Daleiden, and the Research Network on Youth Mental Health (2008) outlined the advantages of measuring client progress not only as a guide but also as a record of practice history. Thus, therapists can review in a single feedback display the client progress and its temporal association with a history of practices delivered. Others such as Garland, Bickman, and Chorpita (2010) have similarly argued that understanding the composite practices of care delivered in community settings is an important aspect of quality improvement.

The Child System & Treatment Enhancement Projects (Child STEPs) study (Weisz et al., 2011) employed a feedback system to allow consultants (clinical ''supervisors'') to keep track of the treatment practices employed and

clinical progress of treated youth in order to provide guidance and support to therapists. The measure used to collect the treatment practice data and thus inform this part of the Child STEPs feedback system is the subject of this article. Feedback systems have the potential to be complicated and expensive. Developing practical, feasible metrics of implementation are important as EBTs move into regular, wide-spread clinical use (Bickman, 2008a).

Measuring practices, however desirable, comes at a cost (Schoenwald et al., 2011). Existing methodologies for adherence measurement in clinical trials have included assessment by supervisors, coding by trained observers, survey of consumers/participants in treatment, and reporting by therapists (Martino, Ball, Nich, Frankforter, & Carroll, 2009). Adherence monitoring in these forms is a difficult enterprise and presents pragmatic challenges. For example, clinical supervisors in large-scale clinical trials may find it difficult to review multiple hours of therapy recordings for each client enrolled in an active treatment group. Training coders to review and code therapy session recordings is a time-consuming and expensive process.

Reports of session content directly from therapists is obviously a more efficient method requiring fewer resources (e.g., Carroll, Nich, & Rounsaville, 1998), but it remains unclear whether this approach yields valid information. For example, Martino and colleagues (2009) examined the relationship between supervisor, therapist, and third-party observer ratings of adherence to a motivational enhancement therapy protocol and found that although all three reporters had reasonable agreement on presence or absence of intervention components, both therapists and supervisors tended to report more extensive coverage of material than was reported by observers. Further evaluation of the accuracy of therapist and supervisor reports may help inform efforts to assess integrity to evidence-based models. It is important to note that establishing methods that promote accuracy in therapist report of session content may help decrease dependency on resource-exhausting efforts to code recordings in the assessment of treatment content and dose.

The current study is an effort to evaluate the accuracy of therapist report on a semistructured interview instrument of session content in a large-scale, community-based clinical trial. As previously mentioned, the Child STEPs Clinic Treatment Project (Weisz et al., 2011) was a multisite study conducted between 2005 and 2009 based in Boston, Massachusetts, and Honolulu, Hawaii, that compared the effectiveness of standard manual treatments, a modular manual treatment (Modular Approach to Therapy for Children [MATCH]), and usual care practice for youth presenting with clinical levels of anxiety, depression, and/or disruptive behavior. This study seeks to evaluate the performance of checklist interview tools (MATCH Consultation Record and Standard Consultation Record) used by Child STEPs consultants to obtain therapist report of session content.

METHODS

Therapist Participant Sample

Eighty-four participating therapists were recruited through 10 community-based mental health clinics. Therapists were 80% female and 56% Caucasian, 23% Asian American, 6% African American, and 6% Pacific Islander; they averaged 40.6 years of age and 7.6 years of clinical experience, and 40% were social workers, 24% psychologists, and 36% other (e.g., licensed mental health counselor, licensed marriage and family therapists). Clinicians were randomized to therapy condition upon entry into the study, and there were no significant differences across condition on any of the therapist characteristics (Weisz et al., 2011). The current study focuses on an examination of therapist reports from the Standard and MATCH conditions.

Therapists in the Standard and MATCH conditions participated in 6 days of training in the treatments corresponding to their condition by experts in the field. MATCH condition therapists were trained in the MATCH Program (Chorpita & Weisz, 2005) and Standard therapists were trained in the Defiant Children Parent Management Training protocol (Barkley, 1997), Coping Cat (Kendall, Kane, Howard, & Siqueland, 1990) and Primary and Secondary Control Enhancement Training (Weisz, Thurber, Sweeney, Proffitt, & LeGagnoux, 1997). In addition to these initial trainings, clinicians received weekly individual consultation from doctoral-level study staff. The vast majority of these consultation meetings were held in person, with a small percentage held over the phone. During these meetings, study consultants discussed the events of the last therapy session and collaborated with therapists to complete a standardized record specific to each condition (MATCH Consultation Record; Standard Manual Consultation Record). The two Consultation Records were then used to populate an Internet-based ''clinical dashboard'' (feedback tool) that allowed consultants to track treatment practices alongside weekly assessments of clinical progress collected as a part of the study (Chorpita et al., 2008).

Youth Participant Sample

As described in Weisz et al. (2011), the study involved 174 youth between the ages of 7 and 13 years old ($M = 10.59$ years, $SD = 1.76$). Of these youth, 70% were

boys ($n = 121$); 45% were Caucasian, 32% multiethnic, 9% African American, 6% Latino/a, 4% Asian American/Pacific Islander, and 2% other. Annual family income was below $40 K for 55% of the sample, $40 to $79 K for 28%, $80 to $119 K for 12%, and $120 K or higher for 6%; 53% lived in single-parent households. Youth in the study had either *Diagnostic and Statistical Manual of Mental Disorders* (4th ed. [*DSM–IV*]; American Psychiatric Association, 1994) disorders or clinically elevated problem levels in the areas of anxiety, depression, and/or disruptive conduct. The mean number of *DSM–IV* disorders was 2.74 ($SD = 1.52$) per youth. See Weisz et al. (2011) for additional details. For the current project, we were interested only in youth from the MATCH and Standard conditions and thus the sample employed here involved 121 youth and 57 therapists.

Measures

Consultation record (MATCH and Standard versions). Project consultants completed a Consultation Record in collaboration with clinicians for each session discussed during weekly consultation meetings. This measure was created for use in the Child STEPs trial and is composed of a matrix of checkboxes, with rows representing specific treatment practices corresponding to session content (for the Standard manuals) or modules (for MATCH). The rows were grouped into clusters representing different treatments (Standard) or treatment foci (MATCH). For example, the first section of the matrix corresponded to the focus area of anxiety, and the first row in that section corresponded to ''Getting Acquainted,'' which was the first treatment session for anxiety in both conditions. Because there was not an exact one-to-one match between the names of Standard sessions and the MATCH treatment modules, there were 35 rows for the MATCH Consultation Record and 47 rows for the Standard Consultation Record.

The columns of the record indicated activities or procedures related to each session. For example, three columns contained headings related to whether a youth had fully, partially, or not completed homework corresponding to sessions or modules. Thus, checking the box in the matrix in the Homework Completed Full column in the Cognitive BLUE row would indicate that homework material related to the Cognitive BLUE session had been assigned in the prior session and was brought completed to the current session. Other columns contained headings related to the coverage of the session or module content (full or partial coverage), the presence of role play or in vivo exercises, the clinician's assessment as to the session attendee's comprehension of the session/module content (full, partial, or no comprehension), and whether homework was assigned. Finally, the rightmost columns contained headings related to whether the therapist and consultant had reviewed therapy recordings, discussed, modeled, or role-played or made plans for the next session corresponding to the different sessions/modules.

The consultation records also contain a checkbox to indicate the presence of a ''crisis of the week,'' defined as a major event in the life of youth or family (e.g., illness, injury, accident, natural disaster, severe conflict, or other trauma). A corresponding series of checkboxes allows for an indication as to whether the therapist was able to transition from discussion of the crisis to the intended session /module content ''fully,'' ''partial,'' or ''not at all.''

The consultation records were completed as a semi-structured interview in each consultation meeting. The consultant posed questions corresponding to the consultation record to obtain the information necessary to complete the record over the course of the regular meeting. Many of the items could be completed based on information gleaned from the natural process of case discussion, but in any cases where the therapist's report was unclear, the consultant would refer back to the options from the record for clarification (e.g., ''You said that you all practiced relaxation in your session—did you role play this skill or were you doing an in-vivo rehearsal?'') Consultants were trained on use of the measure and definitions of the terms. Consultants were trained to assess for crisis of the week and to probe to determine the nature of the crisis and whether handling of this event prevented transition to the intended psychotherapeutic content.

Consultation record validation coding sheet. Coders completed a Consultation Record Validation Coding Sheet for each session randomly assigned to them (see next for randomization procedure; see Appendix A and B for examples). This measure was created for use in this validation study and contains the same matrix with identical items as the Client Consultation Record except the ''in-supervision'' category was disregarded (as recording of supervision was not available to code). Coders indicated items based on what was present in the audio- or video-recording of the therapy sessions. A coding manual was developed by study staff with extensive experience conducting the original consultation in collaboration with study Principal Investigators and the coding team. The manual set forth definitions of codes (e.g., when to endorse ''full'' vs. ''partial'' coverage of session/module content, identifying ''crisis of the week,'' etc.) based on criteria identical to those that were used by the original consultants when they filled out the consultation records.

Procedure

Data procedures. After obtaining consent from children and their parents, therapists made audio- (Honolulu and Boston) or video- (some of Boston) recordings of each one of their sessions. Recorded data were not available for every session held in the trial as therapists sometimes forgot to record sessions and recorded files were sometimes corrupted or of poor quality. Thus, recorded data were not available for every session held in the trial. Coders for the current study logged all existing recordings into a database.

The sampling strategy for this project involved selecting a random sample from the available recording data with constraints introduced to ensure a comprehensive representation of all possible session/module content areas, so that all items on the Consultation Record would have the opportunity for validation. Further, a stratified sampling strategy was employed to ensure that an equal number of recordings were chosen from the two clinical trial sites (Boston and Honolulu), thus to minimize any site-based, nonrandom error. The sampling procedure began by establishing the rules for recording selection. Overall, we sought to pick five recordings representing each possible session/module type for the Standard and MATCH conditions (45 total session/module content types for Standard, 31 for MATCH). On the Consultation Record, each session/module content type is marked as either ''covered full'' or ''covered part''; we intended to randomly select at least two ''covered full'' recordings for each session/module type from each site (four total), and then to randomly select at least one ''covered part'' recording from the whole sample of recordings (cross-site). Because sessions sometimes included more than one session/module type, we were able to sample fewer recordings overall than 3 times the number of session/module types while still ensuring that all types were represented as intended. On average, four sessions were selected from each case to meet these criteria. The number of sessions selected ranged from one to 10 tapes depending on availability of tapes per area of content across cases. Similarly, items not associated with specific session/module types on the Consultation Record (e.g., Role-Play, In Vivo, Understood Full, Understood Part, Understood No, HW Done Full, HW Done Part, HW Not Done, HW Assigned) were typically represented at least five times within the larger sample of records. Then, in cases where this criterion was not met, an additional record was randomly selected from those records whose Consultation Records indicated that therapists reported the recording would contain the item of interest.

Coding procedures. The coding team included three graduate students; seven postbaccalaureate research assistants; and two postdoctoral supervisors, both of whom had participated as consultants in the clinical trial. Members of the coding team began training by reading each treatment manual (Coping Cat, Primary and Secondary Control Enhancement Training, Defiant Children, and MATCH-ADC) and participating in coding group discussions. Once the members of the coding team endorsed comprehension of the treatments, they coded the three training cases as a group and discussed what items would be appropriate to code from these records. During coding, the team followed along in each corresponding manual and paused recordings to discuss when criteria were reached for the items on the coding sheet. Following the training recordings, three prereliability cases were then assigned to the coding team, and each member independently coded all three sessions before meeting to compare and discuss responses. The group discussed coder rationales contributing to discrepancies between codes, areas of particular confusion (e.g., role play versus in vivo coding was a common theme), and idiosyncrasies in therapist delivery of session/module content that led

TABLE 1
Intraclass Correlation Coefficients Representing Average Agreement Across Conditions (MATCH & Standard) and Problem Areas (Anxiety, Disruptive, & Depression)

	Practice	*Rehearsal*	*Receipt*	*HW Assigned*	*HW Review*
MATCH Overall	.74	.42	.71	.60	.38
MATCH Anxiety Codes	.66	.09	.70	.53	.10
MATCH Disruptive Codes	.74	.74	.63	.47	.56
MATCH Depression Codes	.79	.32	.81	.75	.04
STANDARD Overall	.71	.44	.72	.73	.47
Standard Anxiety Codes	.79	.45	.80	.73	.49
Standard Disruptive Codes	.56	.59	.57	.64	.40
Standard Depression Codes	.77	.28	.79	.83	.53

Note: MATCH = Modular Approach to Therapy for Children; HW = Homework.

TABLE 2
Intraclass Correlation Coefficients Representing Agreement Between Coder Records and Consultation Records for MATCH Therapy Practices

	Practice	*Rehearsal*	*Receipt*	*HW Assigned*	*HW Review*
Getting Acquainted	.75	.00	.72	.39	n/a
Fear Ladder	.74	.00	.73	n/a	−.00
Psychoeducation – C	.85	.00	.86	.69	.50
Psychoeducation – P	.71	n/a	.81	.66	n/a
Practicing	.42	.54	.42	.31	.00
Maintenance	.53	n/a	.59	−.01	.00
Cognitive STOP	.45	−.00	.62	.66	.00
Wrap Up	.80	−.00	.83	1.00	n/a
Problem Solving	.81	.66	.77	.85	.36
Activity Selection	.80	.40	.84	.80	.70
Relaxation	.64	.66	.78	.70	.81
Secret Calming	.61	.39	.66	.86	.45
Talents and Skills	.85	.35	.88	.76	.09
Positive Self	.87	.61	.78	.86	1.00
Cognitive BLUE	.99	.80	.99	.75	.33
Cognitive FUN	.97	−.00	1.00	.89	.16
Three-Step Plan	.63	−.00	.55	.56	−.02
Engaging Parents	.79	n/a	.76	.00	n/a
One-on-One Time	.65	.62	.58	.65	.56
Praise	.67	.28	.55	.74	.73
Active Ignoring	.89	.71	.94	.79	.61
Effective Instructions	.84	1.00	.86	.72	.29
Rewards	.55	.80	.50	.65	.59
Time Out	.89	1.00	.73	.66	n/a
Anticipating Problems	.59	n/a	.49	−.01	n/a
Daily Report Card	n/a	n/a	.00	n/a	n/a
Future Problems	.82	n/a	.84	.00	n/a

TABLE 3
Intraclass Correlation Coefficients Representing Agreement Between Coder Records and Consultation Records for Standard Therapy Depression Practices

	Practice	*Rehearsal*	*Receipt*	*HW Assigned*	*HW Review*
Formulation & Orientation – P	.95	—	.92	—	—
Home Visit	.98	—	1.00	—	—
Getting Acquainted – C	.91	.00	.85	1.00	—
Psychoeducation – C	.90	.00	.87	1.00	—
STEPS	.70	.42	.68	1.00	—
Activity Selection	.94	.72	.91	.92	.96
Relaxation	.90	.72	.88	.72	.86
Secret Calming	.85	.86	.85	.80	1.00
Positive Self	.86	.49	.87	.79	.89
Talents & Skills	1.00	.40	.99	.85	.58
Formulation Review – Parent	.71	—	.71	—	—
Think Positive	.77	.15	.78	.71	.52
H-I-N	.69	−.01	.69	.75	.63
Keep Thinking[a]	.96	−.02	.96	See below	See below
Three-Step Plan[a]	.73	.39	.73	See below	See below
Best Fit ACT & THINK[a]	.25	−.02	.15	See below	See below
Keep Thinking + Three-Step Plan + Best Fit[a]	—	.18	—	1.0	.81
ACT & THINK in Real Life	.57	-.03	.45	.42	.07
Final Summation – Parent	.85	—	.85	—	—
Review, Commercial, Party	.95	.27	.90	—	—

Note: Items for which intraclass correlation coefficients were not provided come from sessions wherein those types of content were not included. For example, there was neither opportunity for "rehearsal" nor any homework to assign in the Psychoeducation – P module.

[a]Keep Thinking, Three-Step Plan, and Best Fit ACT & THINK were all prescribed by the manual to occur in a single session, though were distinct interventions. We examined their level of coverage, rehearsal, and receipt separately, but because final rehearsal and homework for this session represented all three interventions, we coded them together.

TABLE 4
Intraclass Correlation Coefficients Representing Agreement Between Records and Supervision Records for Standard Therapy Anxiety Practices

	Practice	*Rehearsal*	*Receipt*	*HW Assigned*	*HW Review*
Engagement – P	.87	—	1.00	—	—
Engagement – C	.92	—	.85	.88	—
Psychoeducation – C	.83	1.00	.82	.88	.83
Psychoeducation – P	.78	—	.75	—	—
Relaxation	.97	.49	.99	.64	.62
Coping Thoughts	.87	.66	.77	.75	.07
Problem Solving	.72	.27	.79	.85	.45
Results & Rewards	.68	.00	.68	.71	.17
FEAR Plan	.59	.28	.54	.36	.62
Parent Involvement	.70	—	.70	—	—
Practice	.76	.46	.86	.74	.68
Review, Commercial, Party	.82	—	.81	—	—

Note: Items for which intraclass correlation coefficients were not provided come from sessions wherein those types of content were not included. For example, there was neither opportunity for "rehearsal" nor any homework to assign in the Psychoeducation – P module.

to coding confusion. These discussions were then used to update and clarify procedures in the coding manual.

During coding, raters selected whether treatment practices were covered "none," "partial," or "full" during the session. They indicated presence or absence for "role-play" and "in vivo practice" of skills. They rated client comprehension or "understanding" of the material presented in the session as "full," "partial," or "none." As these were the most subjective codes, the coding manual instructed coders to default to a "full" rating unless they detected evidence to the contrary (e.g., inability to perform behavioral rehearsal, numerous questions that indicated lack of comprehension and were not resolved, etc.). Finally, they indicated presence or absence of homework assignment by the therapist and homework completion as reported by the client. For the purposes of analysis and reporting, some of these codes from the Consultation Record were combined. Treatment practice content coverage is represented in Tables 1 to 4 under the Practice column. In Vivo and Role-Play items were combined into a single item titled Rehearsal, and therapist ratings of client comprehension/understanding of the content are referred to as Receipt in Tables 1 to 4.

RESULTS

For all analyses, the reliability of therapist and coder reports of session content were calculated using intraclass correlation coefficients (ICC; Shrout & Fleiss, 1979). There are six possible forms of ICC from which to choose; this choice is dependent on (a) whether all raters of interest rate all of the targets (i.e., records), a random sample of all possible raters rates all of the targets, or a different random set of raters rates each of the targets, and (b) whether reliability estimates correspond to individual raters versus mean ratings of a set of k raters (McGraw & Wong, 1996). The reliability coefficients for this project represent the model ICC (A, 1), based on a two-way random effects model. This model is most appropriate because our raters were intended to represent a random sample of possible raters (i.e., their reliability could be generalized to others) and because each target was rated by a different random subset of raters. Our ICC model employed a single rater model because we wanted to estimate the reliability of a single therapist reporting on practices performed, which is ultimately the predictor of interest. Following Cicchetti and Sparrow (1981), ICCs values below .40 reflect "poor" agreement, ICCs from .40 to .59 reflect "fair" agreement, ICCs from .60 to .74 reflect "good" agreement, and ICCs .75 and higher reflect "excellent" agreement.

Reliability of Record Codes

We established initial reliability for the coders by assigning the first seven randomly selected recorded sessions to every member of the coding team and then calculating ICCs comparing all coders to an index coder (the first author) to determine adequate interrater reliability. The index coder provided consultation on the therapy protocols in the clinical trial, was substantially involved in the code development, and had prior experience coding therapy records from other projects (e.g., McLeod & Weisz, 2010). The overall reliability across all items (Practice, Rehearsal, Understanding, Homework Assignment, and Homework Completed) was found to be acceptable as defined by Cicchetti and Sparrow (1981), with average ICCs between the index and other coders ranging from .70 to .75. ICCs within the session/module coverage codes (those that specified whether the coverage of prescribed content was "full"

or "partial") were more variable (range =.25 – 1.0) but were, for the most part, acceptable. Agreement on role plays compared with other session categories was somewhat lower (range =.33 – .50) and was thus addressed during the coding meetings directly following this initial analysis for establishing initial coder reliability.

Over the course of the subsequent 12-month coding process, 24 recordings were randomly selected for double coding across raters (at least one coder and the index rater were assigned to the record) to assess for coder drift. The ICCs across these double-coded sessions ranged between .50 and 1.0, with an average of .80. Therefore, reliability between coders appeared to be consistently high over time.

Therapist Report and Coder Agreement

ICCs were calculated for the full sample across all items of the Standard and MATCH versions of the Consultation Validation Coding Sheet (see Tables 1–5). ICCs representing agreement between therapist report and coder observation of coverage ("full" vs. "partial") for session/module types delivered in each record ("practice covered" in Tables 1–5) ranged from .42 to 1.0 across conditions (Standard and MATCH) and problem areas (Disruptive, Depression, and Anxiety), and all were in the acceptable range ("fair" or higher by Cicchetti & Sparrow, 1981, standards). Average agreement for coverage of session/modules types was similar between MATCH (*M* ICC =.71) and Standard (*M* ICC =.74) conditions as well as across problem areas (anxiety, depression, disruptive; see Table 1).

Therapists and coders demonstrated greater variability in their agreement in terms of whether rehearsal of session/module type (e.g., conducting relaxation exercise) took place in session, with ICCs ranging from −.01 to 1.0. As was true for ratings of session/module type coverage, rater agreements were similar for Standard (*M* ICC =.44) and MATCH (*M* ICC =.42) conditions. Although these agreements still fell in an acceptable range by Ciccetti and Sparrow's (1981) standards, they were lower overall than the practice coverage agreements. A closer look at these rehearsal agreements indicated that those involving cognitive skill rehearsals were some of the lowest across both Standard (e.g., Think Positive, ICC =.15; HIN cognitive coping, ICC = − .01; Keep Thinking, ICC = − .02), and MATCH conditions (e.g., Cognitive FUN, ICC = .00; Cognitive STOP, ICC = .00). Closer examination of these low scores for cognitive rehearsals indicated a consistent trend for each wherein therapists reported rehearsals that coders did not identify for that session.

Therapist and coder agreement on the client's comprehension of the session/module content is represented in the Receipt category and was rated on a 3-point scale (*fully understood, partially understood, no understanding*).

TABLE 5
Intraclass Correlation Coefficients Representing Agreement Between Records and Supervision Records for Standard Therapy Conduct Practices

	Practice	*Rehearsal*	*Receipt*	*HW Assigned*	*HW Review*
Assessment & Engagement – P	.84	—	.76	.90	.51
Why Children Misbehave	.77	—	.77	.49	−.04
Pay Attention	.87	.59	.76	.66	.68
Attending to Compliance[a]	.79	.00	.72	See below	See below
Independent Play[a]	.49	.33	.49	See below	See below
Attending to Compliance & Independent Play Combined[a]	—	.27	—	.61	.31
Effective Instructions	.74	.74	.65	See below	See below
Compliance Training	.19	.00	.20	See below	See below
Effective Instructions & Compliance Training Combined	—	.64	—	.66	.00
Home Point System	.32	.00	.32	See below	See below
Response Cost	.53	.00	.57	See below	See below
Home Point System & Response Cost Combined	—	.00	—	.64	.75
Time Out (Noncompliance)	.42	1.00	.24	.30	.30
Time Out (House Rules)	.60	.49	.76	.74	.37
Time Out Combined	.61	.49	.68	.64	.49
Anticipating Problems	.90	—	.82	.79	.67
Daily Report Card	1.00	—	1.0	—	—
Future Problems	.98	—	1.0	—	—

Note: Items for which intraclass correlation coefficients were not provided come from sessions wherein those types of content were not included. For example, there was neither opportunity for "rehearsal" nor any homework to assign in the Psychoeducation – P module.

[a]Attending to Compliance and Independent Play were prescribed by the manual to occur in a single session, though were distinct interventions. We examined their level of coverage, rehearsal, and receipt separately, but because final rehearsal and homework for this session represented all three interventions, we coded them together. The same was done for Effective Instructions and Compliance Training as well as the pair of Home Point System and Response Cost.

Agreement on receipt ranged from the "poor" to "excellent" range, though two thirds of the ratings (37 of 42 possible practice areas) across Standard and MATCH for this area fell in at least the "good" range.

Overall, agreement between coders and therapists on homework assignment was high across the Standard condition (*M* ICC =.73) and MATCH (*M* ICC =.60). Agreement for homework review was lower for both conditions (Standard *M* ICC =.47; MATCH *M* ICC =.38).

DISCUSSION

This study marks one of the first efforts to evaluate the agreement between therapist report and coder observation of session content and activity across a broad range of cognitive-behavioral and parent management training practices tested in a community-based randomized controlled trial. This study aimed to evaluate a checklist that may be an effective supervision and consultation tool to obtain therapist report of clinical practice and monitor patterns of treatment delivery relevant to integrity assessment or quality improvement efforts. There were several major findings within this effort.

First, therapists and coders demonstrated generally "good" to "excellent" agreement on the presence and absence of session/module content that occurred in sessions, supporting previous findings along these lines by Martino and colleagues (2009). There were a few areas within practice that fell in the "fair" range. Low agreement for Practicing (exposure for anxiety) in the MATCH condition is interesting given the "excellent" range agreement found in its Standard condition version (Practice). Examination of the data revealed that in the MATCH cases, there were numerous incidents where therapists reported having covered this content, but it was not identified by coder observation of those sessions. Similar findings occurred within cognitive session/modules for anxiety, as the use of "Cognitive STOP" in MATCH demonstrated low agreement while again a very similar session in the Standard condition (Coping Thoughts) showed excellent agreement. In both these cases, it may have been that therapists in the Standard condition followed the manual in a more strict fashion that made it easier for coders to follow along with the manual and identify the coverage of content, whereas MATCH therapists may have shown a more flexible or adapted approach (e.g., using more sessions to cover the content) thereby making it more difficult for coders to identify content. These discrepancies could also be an artifact of other variables related to therapist reporting (e.g., years of experience, expertise with CBT, time passed since the session being reported on, number of clients on their caseload, etc.) to be explored in a forthcoming analysis. It was also noted that two of the Parent Management Training practices in the Standard condition showed low agreement (Home Point System =.32, Compliance Training =.19). These practices were prescribed to take place paired with other interventions in sessions (e.g., Home Point System along with Response Cost) but the coverage of each type of content was coded as distinct. Low coverage codes here may represent that competent coverage of this material may have required more than one session to cover, creating confusion in reporting on content or inability for the coder to see all that was presented on the topic by the therapist across sessions. More intensive coding of the extensiveness of practice implementation would be helpful in understanding these differences. However, the generally strong agreement between therapists and coders on session content indicates that therapists can be quite accurate reporters of the presence (partial vs. full) or absence of treatment practices in evidence-based treatment sessions.

Second, therapists and coders demonstrated acceptable but only "fair" agreement on reports of rehearsals of therapeutic practices in session. That is, they did not generally show strong agreement on whether or not role play or in vivo applications of skills occurred. Agreements were low for rehearsals of cognitive skills, which may be explained by the consideration that cognitive skill rehearsals are (a) more difficult for coders to identify in session as they may happen naturalistically and without structure or prompt and (b) may be difficult for coders to differentiate from more general processing of past and current thought patterns.

Third, therapists and coders showed overall strong agreement in terms of client comprehension of the session content ("receipt") as well as whether homework was assigned in session. However, therapists and coders showed less consistent agreement on whether homework was reviewed in session. This may be in part due to homework review that could have occurred "off record" and thus not allowing coders to observe it (e.g., while walking from the waiting room to therapist office, or before therapist remembered to turn on the record), or it may have been woven into more general discussion of "how the week had gone" at the beginning of session that made it difficult for coders to discern its presence or absence.

This study has some notable limitations. Although our sampling strategy involved coding of a very large number of sessions overall (more than 500), the records that we selected are only a relatively small proportion of therapist behavior. Further, we did not have recording for every single type of session content across manuals and conditions, as there were a few of these practices that occurred at a very low frequency across the cases in the trial, and we also had instances where therapists

forgot to record sessions or encountered technological difficulties, and other times where the quality of the sound on recordings was too poor to allow for coding.

A few limitations are also relevant to the recording of therapy sessions and the nature of the data collection. First, the Honolulu site relied exclusively on audiotape, whereas the Boston site employed a mix of audio- and video-recording. Despite this difference, our analysis did not indicate troublesome differences in the average ICCs between the Boston (*M* MATCH ICC = .53, *M* Standard ICC = .62) and Honolulu (*M* MATCH ICC = .52, *M* Standard ICC = .53). Also, therapist behavior in session and their report of session activity may have been affected by the fact that their sessions were taped. In some ways this may limit how well these findings generalize to typical community care where session recording is uncommon. Further, therapist report of session content may have been affected by confirmatory bias (therapists conforming to supervisor expectations) or the cognitive challenges associated with remembering content despite large case loads and busy schedules. Reports could also have been affected by therapist level of experience with and competence in the therapeutic techniques, as greater experience/competence may have aided some in their interpretation of session events and description of activities. Finally, codes of Understanding (Receipt) are subject to bias and are likely more difficult for coders to report on than for therapists.

Another limitation is that our coding did not take into account subtle extensiveness ratings; instead most items were coded for either presence/absence (Homework Assignment, Homework Review, Rehearsal) or Partial/Full/None (Practice, Receipt). A closer look at extensiveness is important to examine more subtle aspects of treatment integrity and to better understand what particular parts of prescribed content may benefit from flexible or adapted implementation. Examination of this type is vital in order to provide feedback to treatment developers as we move lab-designed treatments into community contexts with increasingly complex cases and therapists from a greater diversity of training backgrounds. If our goal is to build and adapt treatments most appropriate and effective for the youth and therapists in our communities, we must direct our attention to the qualitative experiences of therapists implementing these treatments on the front line.

Finally, our coding did not document session content that was proscribed, or that which was ''outside'' of the intended treatment model. Our coders qualitatively observed that therapists sometimes adopted strategies that were outside of the protocols, though it is hard to say what proportion of this went on in the trial without more extensive quantification through coding. Understanding the dose of these ''proscribed'' interventions is important in the interpretation of trial outcome, as it represents part of what could have been the active treatment components.

Several implications arise from the results of this study. The first are relevant to the supervision or consultation with therapists who participate in clinical trials. Our findings that therapists are good reporters of the presence versus absence of practices but less strong in their ability to report on rehearsal of practices may imply that increased structure could support more accurate reporting of what actually goes on in therapy sessions. For example, making definitions of terms explicit (e.g., What does each practice mean? What is full vs. partial coverage of prescribed content? What is role play vs. in vivo rehearsal?) and making a template means of reporting this information is important particularly if it will inform a mechanism for tracking clinical progress and practice (Bickman, 2008a; Chorpita et al., 2008). Establishing integrity-promoting practices within supervision may be important as previous findings have indicated that adherence to supervisory principles can be related to therapist adherence to treatment principles (Schoenwald, Sheidow, & Chapman, 2009). Finding means to improve communication about treatment integrity during clinical trial testing is important so that supervisors are able to provide the most relevant support to therapists so as to protect the internal validity of the project.

Implementation of evidence-based treatments outside of clinical trials is a parallel process to trial supervision and may present even greater challenges in terms of maintenance of integrity. Therapists in community contexts are presented with significant administrative (Donohue et al., 2009) and clinical challenges (Addis et al., 1999) to making treatments ''fit'' their way of practice and the presenting problems raised by youth and families (Weisz & Addis, 2006). These challenges underscore the value of monitoring progress and practice, and thereby means of obtaining the most accurate report possible from therapists as to what adaptations are being made in session to prescribed content. Our study demonstrates that rehearsals of clinical practices (e.g., role-plays, in vivo practice) may be particularly difficult for therapists to report on during supervision. Past work has also shown that therapists in a clinical trial for youth CBT intervention tend to implement significantly more didactic than rehearsal content of material (Marder, 2007). As rehearsal or behavioral activation of skills is theorized to be an important component of therapeutic change (Weisz, 2004), finding means to document therapist enactment of prescribed rehearsal as well as support problem solving around barriers to enactment are important. Such work would wisely be integrated into a broader rubric for supervising EBT as outlined by Friedberg, Gorman, and Biedel (2009).

In sum, therapist report of session content is one efficient means of tracking clinical practice, as third-party verification is expensive and in real-world settings, probably unrealistic. Keeping tabs on the manner in which interventions are employed is important in understanding and interpreting clinical trial results, but also for individual treatment plan in real-world application of evidence-based care. This study demonstrates that—when using a semistructured interview procedure as part of supervision—therapists can be reliable reporters of the basic therapeutic practices delivered in their sessions, although they may need more support in reporting more subtle aspects of implementation such as behavioral rehearsal of concepts, particularly when the concepts are cognitive in nature. Future work aimed at better understanding therapists' ability to track clinical progress and use it to guide the intelligent application and adaptation of evidence-based treatment is an important step in continuing to connect scientific findings with clinical practice.

REFERENCES

Addis, M., & Krasnow, A. (2000). A national survey of practicing psychologists' attitudes toward psychotherapy treatment manuals. *Journal of Consulting and Clinical Psychology*, *68*, 331–339.

Addis, M., Wade, W., & Hatgis, C. (1999). Barriers to evidence-based practices: Addressing practitioners' concerns about manual-based psychotherapies. *Clinical Psychology: Science and Practice*, *6*, 30–441.

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: Author.

Baker-Erikzen, M. J., Hurlburt, M. S., Brookman-Frazee, L., Jenkins, M. M., & Hough, R. L. (2009). Comparing child, parent, and family characteristics in usual care and empirically supported treatment research samples for children with disruptive behavior disorders. *Journal of Emotional and Behavioral Disorders*, *18*, 82–99. doi:10.1177/1063426609336956

Barkley, R. A. (1997). *Defiant children: A clinician's manual for assessment and parent training* (2nd ed.). New York, NY: Guilford.

Bickman, L. (2008a). A Measurement Feedback System (MFS) is necessary to improve mental health outcomes. *Journal of the American Academy of Child & Adolescent Psychiatry*, *47*, 1114–1119.

Bickman, L. (2008b). Why don't we have effective mental health services? *Administration and Policy in Mental Health*, *35*, 437–439. doi:10.1007/s10488-008-0192-9

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

Cicchetti, D., & Sparrow, S. (1981). Developing criteria for establishing interrater reliability of specific items: Applications to assessment of adaptive behavior. *American Journal of Mental Deficiency*, *86*, 127–137.

Chorpita, B. F., Bernstein, A., Daleiden, E., & the Research Network on Youth Mental Health (2008). Driving with roadmaps and dashboards: Using information resources to structure the decision models in service organizations. *Administration and Policy in Mental Health and Mental Health Research*, *35*, 114–123.

Chorpita, B. F., & Daleiden, E. L. (2010). Building evidence-based systems in children's mental health. In A. E. Kazdin & J. R. Weisz (Eds.), *Evidence-based psychotherapies for children and adolescents* (pp. 482–499). New York, NY: Oxford University Press.

Chorpita, B. F., Daleiden, E. L., Ebesutani, C., Young, J., Becker, K., Nakamura, B. J., ... Starace, N. (2011). Evidence based treatments for children and adolescents: An updated review of indicators of efficacy and effectiveness. *Clinical Psychology Science & Practice*, *18*, 154–172.

Chorpita, B. F., Taylor, A. A., Francis, S. E., Moffitt, C., & Austin, A. A. (2004). Efficacy of modular cognitive behavior therapy for childhood anxiety disorders. *Behavior Therapy*, *35*, 263–287.

Chorpita, B. F., & Weisz, J. R. (2005). *MATCH-ADC: A Modular Approach to Treatment for Children with Anxiety, Depression, and Conduct Problems*. Honolulu, HI, and Boston, MA: University of Hawaii at Manoa and Judge Baker Children's Center, Harvard Medical School.

Chu, B., & Kendall, P. (2009). Therapist responsiveness to child engagement: Flexibility within manual-based CBT for anxious youth. *Journal of Clinical Psychology*, *65*, 736–754.

Daleiden, E., & Chorpita, B. (2005). From data to wisdom: Quality improvement strategies supporting large-scale implementation of evidence-based services. *Child and Adolescent Psychiatric Clinics of North America*, *14*, 329–349.

Donohue, B., Allen, D., Romero, V., Hill, H., Vasaeli, K., Lapota, H., ... Van Hasselt, V. (2009). Description of a standardized treatment center that utilizes evidence-based clinic operations to facilitate implementation of an evidence-based treatment. *Behavior Modification*, *33*, 411–436.

Feeley, M., DeRubeis, R., & Gelfand, L. (1999). The temporal relation of adherence and alliance to symptom change in cognitive therapy for depression. *Journal of Consulting and Clinical Psychology*, *67*, 578–582.

Friedberg, R., Gorman, A., & Biedel, D. (2009). Training psychologists for cognitive-behavioral therapy in the raw world: A rubric for supervisors. *Behavior Modification*, *33*, 104–123.

Garland, A. F., Bickman, L., & Chorpita, B. F. (2010). Change what? Identifying quality improvement targets by investigating usual mental health care. *Administration and Policy in Mental Health and Mental Health Research*, *37*, 15–26. doi:10.1007/s10488-010-0279-y.

Hawkins, E., Lambert, M., Vermeersch, D., Slade, K., & Tuttle, K. (2004). The therapeutic effects of providing patient progress information to therapists and patients. *Psychotherapy Research*, *14*, 308–327. doi:10.1093/ptr/kph027

Jacobson, N. S., Schmaling, K. B., Holtzworth-Munroe, A., Katt, J. L., Wood, L. F., & Follette, V. M. (1989). Research-structures vs. clinically flexible versions of social learning based marital therapy. *Behaviour Research & Therapy*, *27*, 173–180.

Kazdin, A. E. (2000). *Psychotherapy for children and adolescents: Directions for research and practice*. Oxford, UK: Oxford Univiversity Press.

Kazdin, A. E., Bass, D., Ayers, W. A., & Rodgers, A. (1990). Empirical and clinical focus of child and adolescent psychotherapy research. *Journal of Consulting and Clinical Psychology*, *58*, 729–740.

Kendall, P. C., Kane, M., Howard, B., & Siqueland, L. (1990). *Cognitive-behavioral therapy for anxious children: Therapist manual* (2nd ed.). Ardmore, PA: Workbook.

Lambert, M. J., Whipple, J. L., Smart, D. W., Vermeersch, D. A., Nielsen, S. L., & Hawkins, E. J. (2001). The effects of providing therapists with feedback on patient progress during psychotherapy: Are outcomes enhanced? *Psychotherapy Research*, *11*, 49–68. doi:10.1093/ptr/11.1.49

Marder, A. M. (2007). *Measuring therapist adherence to a manual-based treatment tested in a community setting: The PASCET*

Manual Adherence Scale (P-MAS). (Unpublished doctoral dissertation). Virginia Commonwealth University, Richmond, VA.

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.

McGraw, K., & Wong, S. (1996). Forming inferences about some intraclass correlation coefficients. *Psychological Methods*, *1*, 30–46.

McHugh, R. K., Murray, H. W., & Barlow, D. (2009). Balancing fidelity and adaptation in the dissemination of empirically-supported treatments: The promise of transdiagnostic interventions. *Behaviour Research and Therapy*, *47*, 946–953.

McLeod, B., & Weisz, J. (2010). The Therapy Process Observational Coding System for Child Psychotherapy strategies scale. *Journal of Clinical Child & Adolescent Psychology*, *39*, 436–443. doi:10.1080/15374411003691750

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

Rogers, E. M. (2003). *Diffusion of innovations*. New York, NY: Free Press.

Schoenwald, S., Carter, R., Chapman, J., & Sheidow, A. (2008). Therapist adherence and organizational effects on change in youth behavior problems one year after multisystemic therapy. *Administration and Policy in Mental Health and Mental Health and Mental Health Services Research*, *35*, 379–394.

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

Schoenwald, S. K., Sheidow, A. J., & Chapman, J. E. (2009). Clinical supervision in treatment transport: Effects on adherence and outcomes. *Journal of Consulting and Clinical Psychology*, *77*, 410–421.

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

Southam-Gerow, M. A., Chorpita, B. F., Miller, L., & Gleacher, A. (2008). Are children with anxiety disorders privately referred to a university clinic like those referred from the public mental health system?. *Administration and Policy in Mental Health and Mental Health Research*, *35*, 168–180. doi:10.1007/s10488-007-0154-7.

Southam-Gerow, M., Weisz, J., Chu, B., McLeod, B., Gordis, E., & Connor-Smith, J. (2010). Does cognitive behavior therapy for youth anxiety outperform usual care in community clinics? An initial effectiveness test. *Journal of the American Academy of Child & Adolescent Psychiatry*, *49*, 1043–1052. doi:10.1016/j.jaac.2010.06.009

Waltz, J., Addis, M. E., Koerner, K., & Jacobson, N. S. (1993). Testing the integrity of psychotherapy protocol: Assessment of adherence and competence. *Journal of Consulting and Clinical Psychology*, *61*, 620–630.

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

Weisz, J. R. (2004). *Psychotherapy for children and adolescents: Evidence-based treatments and case examples*. Cambridge, UK: Cambridge University Press.

Weisz, J. R., & Addis, M. E. (2006). The research-practice tango and other choreographic challenges: Using and testing evidence-based psychotherapies in clinical care settings. In C. D. Goodheart, A. E. Kazdin & R. J. Sternberg (Eds.), *Evidence-based psychotherapy: Where practice and research meet* (pp. 179–206). Washington, DC: American Psychological Association. doi:10.1037/11423-008

Weisz, J., Chorpita, B., Palinkas, L., Schoenwald, S., Miranda, J., Bearman, S. K., ... Research Network on Youth Mental Health. (2011). Testing standard and modular designs for psychotherapy with youth depression, anxiety, and conduct problems: A randomized effectiveness trial. *Archives of General Psychiatry*. Advance online publication. doi:10.1001/archgenpsychiatry.2011.147.

Weisz, J., Doss, A. J., & Hawley, K. (2005). Youth psychotherapy outcome research: A review and critique of the evidence base. *Annual Review of Psychology*, *56*, 337–363.

Weisz, J., Southam-Gerow, M., Gordis, E., Conner-Smith, J., Chu, B., Langer, D., ... Weiss, B. (2009). Cognitive-behavioral therapy versus usual clinical care for youth depression: An initial test of transportability to community clinics and clinicians. *Journal of Consulting and Clinical Psychology*, *77*, 383–396. doi:10.1037/a0013877

Weisz, J. R., Thurber, C. A., Sweeney, L., Proffitt, V. D., & LeGagnoux, G. L. (1997). Brief treatment of mild to moderate child depression using primary and secondary control enhancement training. *Journal of Consulting and Clinical Psychology*, *65*, 703–707. doi:10.1037/0022-006X.65.4.703

Whipple, J., Lambert, M., Vermeersch, D., Smart, D., Nielsen, S., & Hawkins, E. (2003). Improving the effects of psychotherapy: The use of early identification of treatment failure and problem-solving strategies in routine practice. *Journal of Counseling Psychology*, *50*, 59–68. doi:10.1037/0022-0167.50.1.59