6

ASSESSMENT OF TREATMENT INTEGRITY IN PSYCHOTHERAPY RESEARCH

FRANCHESKA PEREPLETCHIKOVA

Precision and control are vital to the methodology of empirical testing of treatment efficacy, and treatment integrity plays an integral role. *Treatment integrity* refers to the extent to which treatment was implemented as intended (Vermilyea, Barlow, & O'Brien, 1984). Treatment outcome research is full of examples of how failure to ensure that the treatment was carried out appropriately resulted in unjustified conclusions, inadequate clinical care, inappropriate

---

Some of the material used in this chapter is from the following sources:

Treatment Integrity in Treatment Outcome Research: Adequacy of Procedures, Associated Factors, Implications for Research and Practice, Guidelines and Recommendations, by F. Perepletchikova, 2009, Saarbrücken, Germany: Lambert Academic Publishing. Copyright 2009 by F. Perepletchikova. Adapted with permission.

"Treatment Integrity and Therapeutic Change: Issues and Research Recommendations" by F. Perepletchikova & A. E. Kazdin, 2005, *Clinical Psychology: Science and Practice*, 12, 365–383. Copyright 2005 by Wiley. Adapted with permission.

"Treatment Integrity in Psychotherapy Research: Analysis of the Studies and Examination of the Associated Factors" by F. Perepletchikova, T. A. Treat, & A. E. Kazdin, 2007, *Journal of Consulting and Clinical Psychology, 75*, 829–841. Copyright 2007 by the American Psychological Association.

http://dx.doi.org/10.1037/14275-008

Treatment Integrity: A Foundation for Evidence-Based Practice in Applied Psychology, L. M. H. Sanetti and T. R. Kratochwill (Editors)

recommendations, and premature dissemination of treatments into clinical practice (e.g., Krumholz et al., 1998; Lauritsen, 1992; Sechrest, White, & Brown, 1979). Failure to ensure treatment integrity poses threats to the experimental validity of a study and has serious implications for inferences drawn about the relationship between treatment and outcome (e.g., Gresham, MacMillan, Beebe-Frankenberger, & Bocian, 2000; Moncher & Prinz, 1991). Once a treatment is established as empirically supported and is disseminated into clinical practice, treatment integrity continues to play an important role in preventing deviation from and gradual alteration of the treatment manual, referred to as a *therapeutic drift* (Peterson, Homer, & Wonderlich, 1982).

The methodological necessity of treatment integrity has long been recognized (e.g., Marks & Tolsma, 1986; Morris, Turner, & Szykula, 1988). Yet, despite the critical significance of treatment integrity for testing therapeutic efficacy, only a fraction of the psychotherapy outcome studies address this topic. Literature reviews indicate that only 6% to 30% of outcome studies assess and report data on treatment integrity (e.g., Armstrong, Ehrhardt, Cool, & Poling, 1997; Borrelli et al., 2005; Gresham, Gansle, & Noell, 1993; Wiese, 1992). Of those studies that do mention treatment integrity, only 10% to 50% provide quantitative data concerning the degree to which procedures were implemented as designed (e.g., Gresham, Gansle, Noell, Cohen, & Rosenblum, 1993; Gresham et al., 2000). An examination of randomized controlled trials (RCTs) of psychosocial treatments indicated that only about 3.5% of the evaluated RCTs have adequately implemented treatment integrity procedures (Perepletchikova, Treat, & Kazdin, 2007). This may mean that treatment effects from only 3.5% of the examined RCTs can be unambiguously interpreted. Evaluation of barriers to adequately addressing treatment integrity suggests that psychotherapy researchers appreciate its importance but indicate that lack of general knowledge about treatment integrity and specific guidelines hinders adequate attention to treatment integrity (Perepletchikova, Hilt, Chereji, & Kazdin, 2009).

The main goal of this chapter is to provide recommendations on the assessment of treatment integrity. In the chapter, I discuss the importance of treatment integrity assessment, factors that affect assessment procedures, guidelines on assessment strategies, recommendations on evaluating psychometric properties of the assessment instruments, and considerations on data representativeness and accuracy, and I provide an overview of the reporting procedures.

WHY ASSESS TREATMENT INTEGRITY

Assessment of treatment integrity (a) ascertains that inferences drawn about treatment effects are justified, (b) augments therapist training and supervision procedures, (c) prevents therapeutic drift, (d) offers opportunities

for isolating active ingredients of change, and (e) may help simplify dissemination procedures (Kazdin, 2003; Miller & Binder, 2002; Moos & Finney, 1983).

Experimental Validity

Assessment of treatment integrity is critical for demonstrating that treatment was delivered as designed. Shadish, Cook, and Campbell (2002) outlined four types of experimental validity—internal, external, construct, and statistical conclusion validity. Establishing which manipulation (treatment or alternative factors) resulted in a change on dependent measures would not be possible without assessment of treatment integrity, and the *internal validity* of the study would be threatened. Indeed, the most cited reason for assessing treatment integrity is to understand ambiguous treatment outcome results or results that were not in the expected direction (Hohmann & Shear, 2002). Assessment of the degree to which a treatment was implemented as intended helps clarify whether such results are due to the failure of the treatment or the failure of the implementation. The failure of the implementation also has to be ruled out when results are in the expected direction, because the breakdown in treatment integrity can indeed enhance the effectiveness of the treatment (Gresham et al., 2000). Deviations can augment procedures and alter the protocol to better suit the treated population. Low treatment integrity does not mean that the treatment is weak, just that it is different from that which was originally intended. When results are in the expected direction but treatment integrity is low, establishing what was actually done by therapists can provide clues for developing a more promising treatment. Thus, data and materials collected as part of a treatment integrity assessment (e.g., sessions tapes, expert ratings, therapist self-reports, clinical notes, written homework assignments, data collection sheets) can be critical for further research.

Lack of treatment integrity assessment can also hinder attempts to replicate the study and evaluate its *external validity*. Generalizability of the findings cannot be established without the exact description of the treatment and how it was delivered. When a treatment is not provided as planned, the *construct validity* of the experiment is also compromised. Imprecision in treatment delivery can cause ambiguity in evaluating what the treatment was and why it produced the effect. When treatment integrity is compromised, the essence of the treatment cannot be separated from the factors that covaried with the treatment. Further, evaluation of the effectiveness of a treatment depends on multiple considerations, including the computational aspects of statistical tests. When treatment is not implemented as intended, *statistical conclusion validity* can be compromised because an unsystematic error is introduced into the data. Such "noise" increases the within-group variability, which reduces the obtained effect size and statistical power and, thus, obscures treatment effect. A treatment might fail to produce

significant change on dependent measures because the variability in treatment implementation decreased the likelihood of detecting the effect.

Therapeutic Drift

Assessment of treatment integrity procedures is essential for prevention of the therapeutic drift and can also augment therapist training and supervision methods. Therapeutic drift can result from a multitude of factors. Therapists may not be sufficiently trained in treatment delivery. Further, therapists may adjust their presentation of the prescribed procedures to fit their personality and style of treatment implementation. They might view certain aspects of the protocol as awkward or irrelevant and might alter certain parts. Therefore, failure to adequately train and monitor therapists and assess their performance might threaten the treatment integrity of a protocol. Therapeutic drift is especially common in clinical settings (e.g., Bond, Evans, Salyers, Williams, & Kim, 2000; Tobin, Banker, Weisberg, & Bowers, 2007), and ongoing evaluation of treatment delivery can alert supervisors to its occurrence.

Therapist Training Procedures

Well-trained therapists are less susceptible to deviation from specified treatment protocol (Beidas & Kendall, 2010; Sholomskas et al., 2005). Faithful rendition of the treatment is more likely with direct training procedures that include opportunities for practice and involve procedures such as role-playing, modeling, feedback, rehearsal, and periodic booster sessions (e.g., Kratochwill, Sheridan, Rotto, & Salmon, 1991; Sterling-Turner, Watson, Wildmon, Watkins, & Little, 2001). Still, insufficient opportunities to observe treatment implementation is one of the main problems in therapist training (e.g., Fairburn & Cooper, 2011; Sharpless & Barber, 2009). Video modeling has been shown to increase treatment integrity of implemented techniques (e.g., DiGennaro-Reed, Codding, Catania, & Maguire, 2010; Moore & Fisher, 2007). Over the past decade, e-learning (through electronic media) and Internet-enhanced training have also been gaining interest and recognition as valuable, clinically rich, and easily accessible tools (e.g., Fairburn & Cooper, 2011; Weingardt, Cucciare, Bellotti, & Lai, 2009). Further, training therapists in research and clinical settings to adequate treatment integrity levels can be facilitated by having therapists watch videotapes of sessions by other therapists and rate their performance with treatment integrity measures. Rating adherence and competence of other therapists and comparing these ratings to those made by experts or independent judges may help advance knowledge of treatment components, facilitate learning of required strategies, and improve identification of the adequate and inadequate

implementation of treatment procedures. Further, use of self-report measures of treatment integrity by therapists in self-monitoring of treatment integrity levels can greatly enhance training, as well as help therapists continue to adjust their performance after training is completed.

Therapists' self-reports of treatment integrity levels offer immediate access to integrity data. Such access allows ongoing adjustment of treatment delivery and suggestions for improvement via review of self-reports during supervision. Performance feedback on implementation may increase treatment integrity when low levels are detected, may prompt therapists to implement treatment with integrity, and can enhance homogeneity across therapists (Gresham, 1997; K. Jones, Wickstrom, & Friman, 1997). Therapists' self-reports can also be compared with ratings of independent observers or supervisors. Feedback on self-ratings can cue therapists to strategies and procedures that they tend to under- or overrate, overlook, or incorrectly code, as well as further their appreciation of the difference between required and auxiliary procedures and improve their understanding of the competent implementation of treatment components.

Mechanisms of Change

Careful assessment of the implementation of different treatment components can help one identify mechanisms of change. Demonstrating a causal effect between the treatment and the outcome does not establish the way the change occurred. Treatment is a package of components that can be distinguished as essential for change, sufficient ingredients, or facilitative ingredients via dismantling and constructive studies (Kazdin, 2007). Dismantling studies allow for analysis of individual components of a treatment by providing a full intervention package to one group and reduced variation to another group or other groups. Constructive studies are used to evaluate whether adding components to a treatment package enhances the effect of an intervention. For example, in a psychoanalytic treatment, mechanisms research (i.e., dismantling and constructive studies) can help one evaluate whether expressive techniques (e.g., analysis of transference and countertransference) contribute to the therapeutic effect above and beyond the uncovering (e.g., interpretations) facets. However, without an operational definition of each component, meticulous monitoring of implementation, and precise assessment of procedures, such research cannot be done.

Dissemination

Adherence to a protocol is necessary when treatment is defined by all of its representative components, but many of these components may not be

essential. Understanding which factors underpin therapeutic effectiveness may simplify the transition of treatments from research laboratories to clinical settings (e.g., Jensen, Weersing, Hoagwood, & Goldman, 2005; Weisz, Weiss, & Donenberg, 1992). Treatments may become more precise and specific once active ingredients of change are known and essential components are separated from sufficient and facilitative ingredients.

FACTORS THAT AFFECT TREATMENT INTEGRITY ASSESSMENT PROCEDURES

Conceptualization of Treatment Integrity

Conceptualization of treatment integrity determines assessment procedures. For example, does assessment encompass adherence, competence, or both? Is treatment differentiation assessed as well? Are proscribed procedures monitored? Is a client's compliance used to determine treatment integrity levels? What treatment components are assessed—those essential for change, required by protocol, and/or auxiliary? The following section elucidates these issues.

Aspects of Treatment Integrity

Treatment integrity consists of three aspects: treatment adherence, therapist competence, and treatment differentiation (e.g., Margison et al., 2000; Waltz, Addis, Koerner, & Jacobson, 1993). *Adherence* refers to the degree to which the therapist utilizes specified procedures and avoids prohibited tasks (e.g., follows the manual verbatim, performs all prescribed tasks and activities). *Competence* refers to the level of skill and judgment shown by a therapist in delivering a treatment (e.g., contingent reinforcement of behavior, provision of prompts and feedback, accurate modeling of techniques). In the context of treatment integrity, competence is conceptualized as the level of skill in performing a specific treatment as opposed to a general therapeutic competence (e.g., empathy, warmth), which is related to common factors. *Differentiation* refers to whether treatments under investigation differ from each other along critical dimensions (e.g., implementing procedures prescribed by the manual for Treatment A and avoiding procedures prescribed for Treatment B and vice versa). Adherence and treatment differentiation are closely related in the sense that a measure of adherence is sufficient to determine whether treatments are distinct (Waltz et al., 1993).

Therapist adherence cannot be substituted for competence. Adherence represents a quantitative aspect of treatment integrity (how frequently a therapist implements prescribed and avoids proscribed procedures), whereas

competence is its qualitative aspect (how well prescribed procedures are implemented). Therapists can adhere to a manual and still deliver a treatment incompetently (e.g., appropriate procedures provided in an inappropriate time; application of tasks was not sensitive to the client's needs). If competence is not evaluated, factors that contributed to an obtained effect or lack of effect cannot be identified.

Treatment integrity is sometimes extended to include participant responsiveness (e.g., Dane & Schneider, 1998; Dusenbury, Brannigan, Falco, & Hansen, 2003; H. A. Jones, Clarke, & Power, 2008). However, participant receipt and enactment considerations move treatment integrity beyond implementation aspects to include treatment outcome. Assessing participant compliance with treatment is indeed important and may moderate treatment effect. Yet, treatment integrity may not even be associated with an outcome (Perepletchikova & Kazdin, 2005). If a treatment is not effective, high treatment integrity cannot be expected to lead to better outcomes, regardless of how a client responded to the used strategies or how closely she or he followed the therapist's recommendations. Low treatment integrity, in this case, may actually improve results, as added procedures may better address a client's needs. Further, a client's responsiveness may depend on the multitude of factors not associated with a treatment. Client characteristics can play an important role here, including the client's difficulty (e.g., anger, hostility), cognitive abilities, and developmental level; the problem's severity, duration, and comorbidity; and the client's readiness for change. Treatment integrity answers a question of whether therapy was delivered as intended by a therapist and not whether it was received as intended by a client. The former is the independent variable, whereas the latter is the dependent variable.

Definition of Treatment Integrity

As Sanetti and Kratochwill (2009) pointed out, the field is still a long way from a consensus on a definition of treatment integrity and specification of its aspects. Given the similarities across different conceptual models of treatment integrity, Sanetti and Kratochwill proposed to define treatment integrity as the "extent to which essential intervention components are delivered in a comprehensive and consistent manner by an interventionist trained to deliver the intervention" (p. 448). This conceptualization is much broader and better captures the complexity of the topic. However, although the term *essential component* is frequently used in the treatment integrity literature, such use may lead to some confusion. The question of which ingredients are essential relates to treatment outcome and taps into the mechanisms of change and treatment specificity, rather than implementation. A clear-cut differentiation between essential and nonessential treatment components is

not always established. I propose describing the components as "required" instead of "essential." The term *required components* includes components that are regarded as key therapeutic ingredients, consistent with theoretical framework, that differentiate the treatment from other models. Further, the above definition does not include proscribed procedures, the specification of the target, or the framework of delivery. Thus, I propose to extend the definition to incorporate these critical aspects: *Treatment integrity is the extent to which required intervention components are delivered as designed in a competent manner while proscribed procedures are avoided by an interventionist trained to deliver the intervention in a particular setting to a particular population*.

Treatment Specificity

Treatments differ in conceptual approaches, therapeutic components, and operational definition of competent implementation. Thus, requirements for demonstrating adherence and competence and measurement procedures may differ as a function of treatment type. For example, skills-based approaches, such as parenting intervention, may assess adherence in terms of the degree to which prescribed tasks are implemented per each session content (e.g., review last week's practice assignment, discuss time-out technique, show videotaped vignette, role-play techniques; e.g., Breitenstein et al., 2010). Principal-based approaches, such as acceptance and commitment therapy, may be measuring adherence of implemented strategies that are prescribed for all sessions (e.g., identifying thoughts as thoughts and not necessarily as reality, highlighting that thoughts and feelings do not lead to action, facilitating willingness to contact and accept difficult feelings, encouraging commitment to all aspects of life) and avoidance of proscribed strategies (e.g., challenging cognitions, in-session exposure, experiential avoidance change strategies; Plumb & Vilardaga, 2010). Thus, most treatment integrity measures are treatment specific.

Treatment Manual

Assessment of treatment integrity also depends on how the treatment is defined. An operational definition of a treatment provides (a) a clear description of procedures, strategies, and activities that should be implemented and those that should be avoided; (b) specification of the length, duration, and intensity of the services; and (c) definition of the target population. Explicit description of procedures (a) ensures that active ingredients of a treatment are being delivered and proscribed procedures are being avoided; (b) reduces complexity of manipulation checks and amount of inferences required in coding; and (c) increases precision of the assessment measures and data accuracy (e.g., Elkin, Pilkonis, & Sotsky, 1988; Heimberg & Becker, 1984).

Treatment procedures can be detailed in a manual form. Manuals can be developed from conceptual frameworks, pilot studies, and consultations with implementers (Nelsen, 1985). Manuals reduce the variability in treatment implementation (Drozd & Goldfried, 1996; Rounsaville, Chevron, & Weissman, 1984) and enhance treatment integrity (Ehrhardt, Barnett, Lentz, Stollar, & Reifin, 1996; Schinke, Gilchrest, & Snow, 1985). Manuals can (a) discuss the theoretical basis of a treatment, (b) outline its structure, (c) detail required and auxiliary components, (d) specify therapist behaviors (e.g., provide verbatim statements to be made by therapists), (e) describe the sequence of the techniques, (f) give procedures for competent implementation of tasks, and (g) provide procedures for handling deviations (e.g., Dobson & Shaw, 1988; McMahon, 1987; Nelsen, 1985). Nezu and Nezu (2008) suggested developing treatment manuals with treatment integrity implementation and assessment procedures in mind, including detailing adequate and inadequate performance criteria, specifying relevance of therapists' behavior by context, and matching treatment manual and treatment integrity protocol. Further, greater consistency in treatment delivery may arise from allowing some built-in flexibility, in which required therapeutic ingredients are presented in conjunction with procedures that are optional or indicated for just some clients. This built-in flexibility must be reflected in the assessment guidelines and procedures.

Flexibility of an approach may also involve creative presentation of material. Creativity is usually seen as an important aspect of therapist competence; however, when it is used as a key therapeutic component, creativity enters into the domain of treatment adherence. For purposes of assessing adherence, treatment components require operational definition. Yet, meaningfully manualizing creative responding in a moment may be very difficult. For example, training in a humanistic therapy involves providing therapists with an understanding of humanistic philosophy and the theoretical basis of a treatment and facilitating the development of spontaneity, empathy, and genuineness. Creative responding and improvisation in the moment are valued as key ingredients of therapeutic process, whereas specific treatment protocols are viewed as counterproductive (Bohart, O'Hara, & Leitner, 1998). However, such idiosyncratic responding of therapists reduced uniformity in therapists' behavior. Problems with operationalizing treatment are inherent in any psychotherapy research (Frances, Sweeney, & Clarkin, 1985). The task may be challenging because including all potential scenarios of treatment delivery, considering various comorbid diagnoses, and accounting for a client's difficulty is not always possible or feasible. The complexity of the task can at least partially explain the finding that only about 65% of RCTs of psychosocial treatments use specific manuals (Perepletchkova et al., 2007). Yet, lack of a precise operational definition of a treatment introduces random

variation into the delivery of a treatment, impedes treatment integrity assessment, reduces statistical power, and compromises the internal and external validity of the study.

ASSESSMENT STRATEGIES

Treatment integrity can be assessed via direct, indirect, and hybrid strategies. In deciding on a strategy, one should consider the strengths and weaknesses of each approach, as detailed next.

Indirect Strategies

Indirect methods of assessment include self-reports, rating scales, interviews, and permanent products of treatment implementation (e.g., homework sheets; Gresham et al., 2000). Self-reports can be obtained directly from therapists and include rating on the degree to which procedures were implemented as intended. Likert scales can be used for such assessment. The Therapy Procedures Checklist (Weersing, Weisz, & Donenberg, 2002), a therapist self-report measure that encompasses psychodynamic, cognitive, behavioral, and family approaches, is notable for its ability to assess treatment differentiation. However, although self-ratings are convenient and easy to obtain, they can be biased and distorted by self-interest. Demand characteristics and a need for social approval can affect the accuracy of the reported adherence and competence.

Clients can be debriefed via interview or questionnaire on what was done by a therapist during treatment sessions. Clients may provide information regarding the manner in which procedures were executed and what was received by subjects (Docherty, 1984; Kazdin, 2003). Subjective recollections may be inaccurate, but responses that vary systematically among experimenters may provide important information about what was implemented. For example, clients may describe specific behaviors of a therapist or report the nature of assignments, monitoring of homework completion, rehearsal of techniques, and clarification of difficult material. Systematic endorsement of a therapist's specific tendencies may provide clues about deviations from the manual and the therapist's competence in treatment delivery.

Research that relies primarily on the indirect, subjective evaluations of treatment integrity is likely to be weak in its ability to measure treatment integrity accurately. Such methods are more likely to be reactive and to be influenced by social desirability and demand characteristics; thus, such methods are less reliable and valid. A potentially more accurate indirect method of treatment integrity assessment includes collection of the permanent products,

such as written homework assignments or data collection sheets. Each component of the treatment is referenced to the permanent product corresponding to each treatment step. Evaluation of the presence and absence of the corresponding products is a more reliable indirect method of treatment integrity assessment.

Direct Strategies

Assessments that rely on self-report are subject to various threats to measurement accuracy, as discussed above. Observations conducted by trained staff are considered the gold standard because they can provide a more objective account of the implemented procedures (Hogue, Liddle, & Rowe, 1996; Mowbray, Holter, Teague, & Bybee, 2003). Observations can be done in the treatment setting, by viewing sessions through a one-way mirror or via monitors, or by listening to sessions via review of video/audio records. Attending sessions is more likely when a treatment is provided to a large group of people, such as in a classroom. However, the presence of an observer can alter performance of a therapist and may result in higher adherence to specified procedures during observed sessions as compared to when observations are not conducted (K. Jones et al., 1997; McMahon, 1987). Differential adherence may artificially inflate estimates of treatment integrity and compromise accuracy of integrity data.

To ameliorate reactivity, observers can "spot check" treatment implementations on a variable-time schedule (Peterson et al., 1982). Viewing sessions via monitors or one-way mirror may be a feasible alternative to direct observations in the treatment setting. Such observations are unobtrusive because they can be conducted without therapist awareness, and they may provide more accurate data. Videotaping sessions with subsequent coding by trained observers is a common approach for research protocols. Usually, 20% to 40% of all sessions are observed or videotaped (Schlosser, 2002). Reactivity may be ameliorated when all treatment sessions are videotaped, with subsequent evaluation of a random subset of recordings.

An observational measure, the Therapy Process Observational Coding System for Child Psychotherapy Strategies Scale (TPOCS-S; McLeod & Weisz, 2010), was recently developed to address the limitations of the self-reported Therapy Procedures Checklist. This measure assesses the extent to which treatment procedures from several therapeutic approaches are employed from direct observations of treatment sessions. The ability to examine different therapeutic approaches is a unique strength of this measure, as most treatment integrity assessment instruments are treatment specific. Such a measure can be used to examine implementation of evidence-based approaches in clinical practice. Implementation research is central to understanding the

needs and preferences of clinical providers and can inform dissemination efforts. For example, researchers using the TPOCS-S found that usual care therapists utilize multiple approaches and favor nonbehavioral treatments, such as client-centered therapy (McLeod & Islam, 2011).

Hybrid Strategies

Indirect methods of treatment integrity assessment are usually used to supplement direct strategies. Data obtained via direct and indirect measures may be compared to clarify treatment implementation issues (Bergan & Kratochwill, 1990; Gresham, 1989). Only a fraction of sessions is usually utilized for coding of observational data; thus, therapist self-reports collected after each session may offer a more detailed assessment of treatment delivery. Further, therapist self-reports can be used to differentiate between frequently and infrequently used strategies. This information can provide clues to which procedures therapists regard as more effective for further treatment refinement (Carroll et al., 2000).

It should be noted, however, that there is low agreement between direct and indirect methods (Carroll, Nich, & Rounsaville, 1998; Gresham, 1997; Wickstrom, Jones, LaFleur, & Witt, 1998). Some reasons for such disparity include demand characteristics and need for social approval. Inherent limitations of observational ratings may also play a role. Raters are removed from the treatment process and may, therefore, (a) miss subtle strategies or those that are imbedded in reference to previous discussions, (b) misunderstand statements made by therapists, and (c) fail to see the full interaction between client and therapist (e.g., beginning and end of sessions may not always be recorded, and sound and picture quality of the videotapes or TV monitors may preclude accurate representation of a therapist's behaviors; Carroll et al., 2000). Yet, as has been noted, self-reports offer immediate access to the adherence and competence data, and they can be used to adjust treatment delivery.

Reliability of Treatment Integrity Measures

There is no conventional method of establishing reliability of treatment integrity measures. Reliability data have been reported as the percentage of agreement between raters or observers (e.g., Dusenbury, Brannigan, Hansen, Walsh, & Falco, 2005), test–retest reliability (e.g., Resnicow et al., 1998), and intraclass correlations for raters (Carroll et al., 2000). Other methods have been suggested, such as factor analysis to determine internal structure of a measure and internal consistency indices (e.g., coefficient alpha; McGrew, Bond, Dietzen, & Salyers, 1994; Sheridan, Swanger-Gagne, Welch, Kwon, &

Garbacz, 2009). However, coefficient alpha reflects a correlation of variables. As suggested by Gresham (2009), the presence of such a correlation is a dubious assumption, because treatment components cannot be expected to relate to each other. For example, in dialectical behavior therapy, therapeutic strategies assessed for adherence include problem assessment, problem solving, cognitive modification, case management, irreverence, reciprocal communication, exposure-based procedures, and crisis protocols. All of these strategies cannot be expected to correlate. Calculation of interrater reliability, the most common method of establishing measure reliability, is done by dividing the number of agreements by the number of agreements plus the number of disagreements and multiplying by 100%. Agreement can also be calculated by obtaining the kappa coefficient (Cohen, 1960). The kappa coefficient is specifically useful when no raw data are missing, and it may provide better control for chance agreement.

Validation of Treatment Integrity Measures

Validation of measures may be challenging because treatment integrity encompasses constructs—adherence and competence—that are not necessarily related. The validation process may involve examination of how measures behave in relation to one another. Factors (e.g., predictors, causal agents, moderating, mediating variables) that are known to affect constructs under investigation may be used to examine the validity of an instrument. Several factors, including client and treatment characteristics, specificity of a treatment manual, and the levels of therapist training and supervision, are known to be associated with treatment integrity (Perepletchikova & Kazdin, 2005).

Complexity of the treatment, required multiple resources and materials, number of treatment agents, time needed for treatment implementation, and rate of behavioral change are hypothesized to be inversely related to the level of treatment integrity (Gresham, 1989; Gresham et al., 2000; Noell, Gresham, & Gansle, 2002). Construct validity of the treatment adherence measure may be evaluated by examining its association with treatment characteristics. *Construct validity* refers to the relation of a measure to other measures or domains of functioning. Two treatments that differ on the characteristics that negatively affect treatment integrity (e.g., time consuming, requires multiple resources and therapists) may be compared on their effects on adherence levels. Validity may be supported by demonstrating that the characteristics of the treatments are differentially associated with levels of treatment integrity.

Further, evidence on the differential training effects on therapist adherence levels may support validity of a measure. Adherence to protocol is more

likely with direct training of therapists (e.g., opportunities for practice and feedback) than with indirect training (e.g., didactic instructions; e.g., Kratochwill et al., 1991; Sterling-Turner et al., 2001). Specificity of a treatment manual may also differentially affect treatment integrity, where higher specificity contributes to greater treatment integrity levels. Manual specificity that differs as a function of treatment may be used to examine construct validity of a treatment integrity measure.

Discriminant validity is suggested if a measure shows little or no correlation with measures with which it is not expected to correlate. Discriminant validity may be supported when the measure of adherence to treatment protocol is associated with the treatment for which the protocol was originally devised (e.g., cognitive-behavioral treatment) and is not associated with the treatment for which conditions are different (e.g., psychodynamic therapy). The same measure of adherence may be employed with both treatments, and significant difference on adherence ratings as a function of therapy type in the expected direction may be indicative of the measure's ability to discriminate between treatments. Further, supervision and monitoring of treatment delivery can help reduce therapeutic drift and may facilitate adherence to the specified treatment protocol. Thus, concurrent validity of a treatment integrity measure may be supported via its association with the levels of provided supervision (i.e., the higher the level of monitoring and feedback, the higher the adherence).

Supervision and ongoing monitoring may also enhance the competence of treatment delivery. Thus, relationship between the measures of the levels of provided supervision with the measure of therapist competence may serve to support the concurrent validity of treatment integrity measures. *Concurrent validity* refers to the association of a measure with performance on another measure at the same point in time. Concurrent validity of the competence measure can also be examined by evaluating the association of therapist competence with the measures of client characteristics, because therapist performance may vary as a function of client difficulty, hostility, high problem severity, duration, and comorbidity (e.g., Detrich, 1999; Foley, O'Malley, Rounsaville, Prusoff, & Weissman, 1987; Waltz et al., 1993).

Criterion validity of a competence measure can be suggested when a measure can distinguish between different therapeutic modalities (e.g., Barber & Crits-Christoph, 1996; Barber, Liese, & Abrams, 2003). *Criterion validity* refers to a correlation of a measure with some other specific or dichotomous criterion. Therapeutic approaches usually have specific requirements for demonstrating competence. For example, cheerleading, self-disclosure, and irreverent communication are commonly used in dialectical behavior therapy but are proscribed for psychoanalytic therapists.

Representativeness of Treatment Integrity Data

Therapists' adherence and competence may vary across subjects, tasks, and time. Representativeness depends upon the number and length of observations and collection of data across treatment phases, therapists, situations, cases, and sessions (Moncher & Prinz, 1991; Peterson et al., 1982). Sampling across these aspects of treatment delivery informs consistency in treatment integrity data (Docherty, 1984).

1. *Across treatment phases.* A treatment usually consists of several phases, such as introduction to therapy and assessment of pathology (Phase 1), skills training (Phase 2), and relapse prevention (Phase 3). Phase 1 may be most conducive to higher treatment integrity ratings, because tasks in this phase are more straightforward. Phase 2, on the other hand, may have lower treatment integrity ratings because more complex tasks, such as the actual training of skills, are administered.
2. *Across therapists.* Therapists may have high variability in their performance due to personality factors, motivation, and previous training and experience (Gresham, 1989; Miller & Binder, 2002; Weissman, Rounsaville, & Chevron, 1982). When data are overrepresentative for some therapists, treatment integrity ratings may be skewed in a particular direction.
3. *Across situations.* At times, therapists may have to deal with unexpected circumstances, such as when a client presents with a crisis. In such situations therapists may have to partially or completely deviate from a treatment protocol in order to address specific concerns. Hence, the integrity of treatment implementation may be lower.
4. *Across sessions.* The material of sessions may vary in complexity and difficulty. Therapists may find it easier to adhere to guidelines when material is more straightforward. When such sessions are overrepresented in the sample, treatment integrity ratings may be higher.
5. *Across cases.* A therapeutic relationship with a difficult (e.g., angry, hostile) client may be less reinforcing for a therapist and may require greater effort. Greater effort in the face of little success may discourage a faithful rendition of the treatment plan. Severe cases may require more direction and coaching and incorporation of additional techniques to address a client's specific concerns. So, treatment integrity may be higher with uncomplicated clients than with more difficult clients.

Further, treatment integrity data may be more informative when they encompass all three aspects involved in their specification: adherence, competence, and treatment differentiation. Adherence measures that include prescribed as well as proscribed procedures can also assess treatment differentiation. Waltz et al. (1993) recommended that adherence measures should include items pertaining to four types of therapist behaviors: (a) those that are unique and essential to the specific treatment (e.g., assigning homework in behavior therapy); (b) those that are essential but not unique to the treatment (e.g., setting treatment goals); (c) those that are compatible with the treatment; that is, not prohibited but neither unique nor essential (e.g., therapeutic self-disclosure); and (d) those that are proscribed (e.g., interpreting resistance or transference in behavior therapy). Competence measures cannot rely on the level of experience and training but should be independently verified by measuring how sensitively the treatment protocol is applied to individual clients. Within this framework, ratings of competence should consider (a) stage of therapy, in terms of information about number of sessions completed and extent of progress; (b) client difficulty, which may impact the level of therapist activity and involvement; and (c) therapist approach to the presenting problem in a manner consistent with the prescribed procedures (Waltz et al., 1993).

Data Accuracy

The accuracy of the rating of treatment integrity depends on the competence of raters, the sophistication of the measures, and the coding procedures. Raters have to be trained in all of the major and minor treatment components and subtle aspects of the treatment (Stein, Sargent, & Rafaels, 2007). Raters who are themselves skilled in the treatment delivery may be most suitable for treatment integrity rating. However, when raters are affiliated with the project, their ratings may be biased. Rater bias occurs when ratings are influenced by the subjectivity of a rater. It can be reduced by (a) using raters not associated with a study; (b) keeping raters unaware of treatment assignment, which is called *blinding;* (c) using multiple raters; and (d) performing consensus ratings and interrater reliability checks (Marcus et al., 2006; Wu, Whiteside, & Neighbors, 2007). Hoyt (2000) offered bias correction procedures to minimize its adverse effect on findings. He proposed four types of rater bias and delineated formulas to correct for attenuation and inflation of observed effect sizes.

Sophistication of the treatment integrity measure can have a significant impact on the accuracy of rating, with less sophisticated measures, such as indirect assessment methods, contributing to higher integrity rates (Miller & Binder, 2002; Robbins & Gutkin, 1994; Wickstrom et al., 1998).

Further, sensitivity and adequacy of the selected measures can affect treatment integrity data. Sensitivity of the measure can be constrained by ceiling or floor effects. Restriction in the range of scores can prevent continued increments in performance and result in a lower treatment integrity rating. A measure may not be comprehensive or specific enough to address constructs of interest. The importance of measurement adequacy was highlighted in the evaluation of the relationship between therapist competence and treatment outcome in the National Institute of Mental Health Treatment of Depression Collaborative Research Program (Shaw et al., 1999). This study showed only weak effects between these variables; however, the competence measure may have failed to tap into important aspects of cognitive-behavioral therapists' performance. The selected scale may have been useful for quality-control monitoring, but it was inadequate as a measure of therapist competence.

Procedures for coding treatment integrity can influence the coding accuracy. Coding of treatment integrity can include evaluation of the occurrence and nonoccurrence of each treatment component and the extent to which each component is performed competently and as specified. Thus, a detailed analysis of each task can increase coding accuracy. Coding videos of sessions in random order can also increase coding accuracy by controlling for observer drift (Kazdin, 1977; O'Leary & Kent, 1973).

Reporting Treatment Integrity

Treatment integrity applies not only to the overall performance of a therapist in treatment delivery but also to the implementation of each treatment component across and within sessions. Reporting of treatment integrity in terms of overall integrity, component integrity, and session integrity is recommended (Gresham, 1997; Schlosser, 2002). *Overall treatment integrity* addresses the degree to which all components were implemented across sessions. *Component integrity* refers to the integrity of implementing each treatment component across sessions. *Session integrity* refers to integrity of all treatment components within one session. Overall treatment integrity and session treatment integrity can be calculated by summing the components that were correctly implemented (across sessions or within one session, respectively) and dividing this number by the total number of components, expressing integrity as a percentage. Component integrity can be calculated by summing the number of sessions during which a component was correctly implemented and dividing this number by the total number of sessions, expressing integrity as a percentage. This approach is regarded as the most relevant because it evaluates the accuracy of treatment implementation (Schlosser, 2002). However, taking into account perspectives of only one observer can weaken this approach; thus, obtaining adequate levels of

interrater agreement is also necessary. The current state of the literature on treatment integrity in psychotherapy research indicates that a high integrity level may be demonstrated by 80% to 100% integrity, moderate integrity by a 60% to 80% range, and low integrity by less than 60% integrity (Gansle & McMahon, 1997; Gresham, Gansle, Noell, et al., 1993; Holcombe, Wolery, & Snyder, 1994; Noell et al., 2002).

Calculating all three estimates of treatment integrity is essential, because even though overall treatment integrity may be high, component integrity and/or session integrity may be low. For example, therapist performance may vary as a function of client difficulty (e.g., Foley et al., 1987), and such variability may result in inconsistent treatment delivery within sessions. Although component integrity may be high across sessions, session integrity may be low for a particular session. Failure to measure session integrity may hinder the evaluation of results, especially when treatment failed to produce significant change on dependent measures while overall treatment integrity was high. Treatment components may be more or less critical for successful treatment implementation (e.g., providing a rationale for treatment may be less crucial than contingent delivery of positive reinforcement; Gresham, 1997). Monitoring within-session integrity may supply important information on the degree of competency and consistency in administering each treatment component. Such fine-grained analysis of treatment integrity permits better evaluation of treatment outcome and enhances the credibility and replicability of results. Further, such analysis helps distinguish among treatment components that are essential for change, sufficient, and just facilitative, thus aiding in the identification of the mechanisms of change.

Quantitative adherence and competence data must be presented in a publication on study outcomes, and these data should be informative of the treatment integrity levels. Reporting that utilization of treatment components between treatments was significantly higher for one treatment than for the other does not adequately inform the research consumer about the adherence levels. There may be a statistically significant difference without either treatment having high treatment integrity levels. That is, 50% treatment integrity may be significantly higher than 20% treatment integrity; however, neither represents adequate integrity levels. Further, treatment integrity is sometimes evaluated by asking raters to classify videotapes of therapy sessions by the employed treatment modality (e.g., which tape belongs to cognitive versus interpersonal therapy). This method does not indicate the degree to which therapists were adhering to a manual or were competent in treatment delivery. A tape may be correctly classified because the number of components within a session was higher for one treatment than for the other. However, this classification does not demonstrate that all of the prescribed

components were utilized during a session or that proscribed strategies and procedures were avoided.

Only absolute values (not relative to each other) may be informative of treatment integrity levels. To be considered as informative, data may be presented as a percentage (e.g., 85% integrity of treatment delivery), a specific score within a clearly defined range (e.g., the median adherence score on the 5-point Likert scale was 3.8), or a number that can be easily converted into a percentage (e.g., the proportion of strategies consistent with prescribed treatment modality was .82).

CONCLUSION

The question of what represents a satisfactory assurance that treatment was implemented as designed may not have a straightforward answer. Multiple considerations may affect how treatment integrity is addressed, including available funding, study design, setting, level of risk of treatment inaccuracies, and nature of a treatment. Gresham (2009) discussed treatment integrity flexibility, where the required adherence levels depend on the type of the research study, with higher levels necessitated for efficacy studies and less stringent adherence accepted for effectiveness research. This discussion is based on a notion that treatment integrity is related to outcome and that treatments with "drifts" from a protocol can still produce positive effects, while allowing for flexibility under less controlled conditions. However, the main objective of establishing treatment integrity is not to increase the strength of the treatment but to inform the degree to which treatment was implemented as intended. The higher the treatment integrity level, the more closely the implemented treatment approximates the intended treatment. Low treatment integrity levels do not indicate that a treatment is weak but rather that a treatment is different from that originally intended.

A flexible approach to treatment integrity levels can lead to loose operational definition of a treatment and thus increase ambiguity in the interpretation of the findings. Yet, the call for a flexible approach highlights the challenges of consistent treatment delivery in different settings and under different conditions. The flexibility of treatment implementation can be built into the treatment protocol and accounted for in the treatment integrity measures, thus allowing for accurate estimations of integrity levels under variable conditions without compromising the interpretability of the results. Flexibility can be built into the treatment protocol by specifying (a) procedures for handling difficult cases, (b) additional treatment components for addressing comorbid problems, and (c) approaches for working with various populations (e.g., cultural issues, language barriers) and in different settings

(e.g., outpatient care, residential facility, inpatient units). Additionally, intervention ingredients are not equally important (e.g., Sanetti & Kratochwill, 2009), and differentiating required components from auxiliary strategies can further augment flexible treatment delivery. These procedures would allow therapists to adapt treatment delivery to the needs of a client and the limitations of a particular setting, while delivering all key elements.

The call for accountability in mental health care necessitates establishing a scientific basis for interventions. Requiring manipulation checks on treatment delivery will improve the quality of research and clinical practice. Although it may be laborious and costly, this adjustment is necessary to advance the psychotherapy field. Redefinition of criteria for clearing the "evidence-based" threshold, standardization and enforcement of treatment integrity procedures, increased funding for research, and provision of incentives for implementing integrity procedures may aid in this quest.

REFERENCES

Armstrong, K. J., Ehrhardt, K. E., Cool, R. T., & Poling, A. (1997). Social validity and treatment integrity data: Reporting in articles published in the *Journal of Developmental and Physical Disabilities*, 1991–1995. *Journal of Developmental and Physical Disabilities, 9,* 359–367. doi:10.1023/A:1024982112859

Barber, J., & Crits-Christoph, P. (1996). Development of a therapist adherence/competence scale for supportive-expressive dynamic psychotherapy: A preliminary report. *Psychotherapy Research, 6,* 81–94. doi:10.1080/10503309612331331608

Barber, J. P., Liese, B., & Abrams, M. J. (2003). Development of the Cognitive Therapy Adherence and Competence Scale. *Psychotherapy Research, 13,* 205–221. doi:10.1093/ptr/kpg019

Beidas, R. S., & Kendall, P. C. (2010). Training therapists in evidence-based practice: A critical review of studies from a systems-contextual perspective. *Clinical Psychology: Science and Practice, 17,* 1–30. doi:10.1111/j.1468-2850.2009.01187.x

Bergan, J. R., & Kratochwill, T. R. (1990). *Behavioral consultation and therapy.* New York, NY: Plenum Press.

Bohart, A. C., O'Hara, M., & Leitner, L. M. (1998). Empirically violated treatments: Disenfranchisement of humanistic and other psychotherapies. *Psychotherapy Research, 8,* 141–157. doi:10.1080/10503309812331332277

Bond, G. R., Evans, L., Salyers, M. P., Williams, J., & Kim, H. W. (2000). Measurement of fidelity in psychiatric rehabilitation. *Mental Health Services Research, 2,* 75–87. doi:10.1023/A:1010153020697

Borrelli, B., Sepinwall, D., Ernst, D., Bellg, A. J., Czajkowski, S., Breger, R., . . . Orwig, D. (2005). A new tool to assess treatment fidelity and evaluation of treatment

fidelity across 10 years of health behavior research. *Journal of Consulting and Clinical Psychology, 73,* 852–860. doi:10.1037/0022-006X.73.5.852

Breitenstein, S. M., Fogg, L., Garvey, C., Hill, C., Resnick, B., & Gross, D. (2010). Measuring implementation fidelity in a community-based parenting intervention. *Nursing Research, 59,* 158–165. doi:10.1097/NNR.0b013e3181dbb2e2

Carroll, K. M., Nich, C., & Rounsaville, B. J. (1998). Utility of therapist session checklists to monitor delivery of coping skills treatment for cocaine abusers. *Psychotherapy Research, 8,* 307–320. doi:10.1080/10503309812331332407

Carroll, K. M., Nich, C., Sifry, R. L., Nuro, K. F., Frankforter, T. L., Ball, S. A., . . . Rounsaville, B. J. (2000). A general system for evaluating therapist adherence and competence in psychotherapy research in the addictions. *Drug and Alcohol Dependence, 57,* 225–238. doi:10.1016/S0376-8716(99)00049-6

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement, 20,* 37–46. doi:10.1177/001316446002000104

Dane, A. V., & Schneider, B. H. (1998). Program integrity in primary and early secondary prevention: Are implementation effects out of control? *Clinical Psychology Review, 18,* 23–45. doi:10.1016/S0272-7358(97)00043-3

Detrich, R. (1999). Increasing treatment fidelity by matching interventions to contextual variables within the educational setting. *School Psychology Review, 28,* 608–620.

DiGennaro-Reed, F. D., Codding, R., Catania, C. N., & Maguire, H. (2010). Effects of video modeling on treatment integrity of behavioral interventions. *Journal of Applied Behavior Analysis, 43,* 291–295. doi:10.1901/jaba.2010.43-291

Dobson, K. S., & Shaw, B. F. (1988). The use of treatment manuals in cognitive therapy: Experience and issues. *Journal of Consulting and Clinical Psychology, 56,* 673–680. doi:10.1037/0022-006X.56.5.673

Docherty, J. P. (1984). Implications of the technological model of psychotherapy. In J. B. W. Williams & R. L. Spitzer (Eds.), *Psychotherapy research: Where are we and where should we go?* (pp. 139–147). New York, NY: Guilford Press.

Drozd, J. F., & Goldfried, M. R. (1996). A critical evaluation on the state-of-the-art in psychotherapy outcome research. *Psychotherapy: Theory, Research, Practice, Training, 33,* 171–180. doi:10.1037/0033-3204.33.2.171

Dusenbury, L., Brannigan, R., Falco, M., & Hansen, W. B. (2003). A review of research on fidelity of implementation: Implications for drug abuse prevention in school settings. *Health Education Research, 18,* 237–256. doi:10.1093/her/18.2.237

Dusenbury, L., Brannigan, R., Hansen, W. B., Walsh, J., & Falco, M. (2005). Quality of implementation: Developing measures crucial to understanding the diffusion of preventive interventions. *Health Education Research, 20,* 308–313. doi:10.1093/her/cyg134

Ehrhardt, K. E., Barnett, D. W., Lentz, F. E., Jr., Stollar, S. A., & Reifin, L. H. (1996). Innovative methodology in ecological consultation: Use of scripts to promote

treatment acceptability and integrity. *School Psychology Quarterly, 11,* 149–168. doi:10.1037/h0088926

Elkin, I., Pilkonis, P. A., & Sotsky, S. M. (1988). Conceptual and methodological issues in comparative studies of psychotherapy and pharmacotherapy, I: Active ingredients and mechanisms of change. *American Journal of Psychiatry, 145,* 909–917.

Fairburn, C. G., & Cooper, Z. (2011). Therapist competence, therapy quality, and therapist training. *Behaviour Research and Therapy, 49,* 373–378. doi:10.1016/j.brat.2011.03.005

Foley, S. H., O'Malley, S., Rounsaville, B. J., Prusoff, B. A., & Weissman, M. M. (1987). The relationship of patient difficulty to therapist performance in interpersonal psychotherapy of depression. *Journal of Affective Disorders, 12,* 207–217. doi:10.1016/0165-0327(87)90029-2

Frances, A., Sweeney, J., & Clarkin, J. (1985). Do psychotherapies have specific effects? *American Journal of Psychotherapy, 39,* 159–174.

Gansle, K. A., & McMahon, C. M. (1997). Component integrity of teacher intervention management behavior using a student self-monitoring treatment: An experimental analysis. *Journal of Behavioral Education, 7,* 405–419. doi:10.1023/A:1022851117439

Gresham, F. M. (1989). Assessment of treatment integrity in school consultation and prereferral interventions. *School Psychology Review, 18,* 37–50.

Gresham, F. M. (1997). Treatment integrity in single-subject research. In R. D. Franklin, D. B. Allison, & B. S. Gorman (Eds.), *Design and analysis of single-case research* (pp. 93–117). Mahwah, NJ: Erlbaum.

Gresham, F. M. (2009). Evolution of the treatment integrity concept: Current status and future directions. *School Psychology Review, 38,* 533–540.

Gresham, F. M., Gansle, K. A., & Noell, G. H. (1993). Treatment integrity in applied behavior analysis with children. *Journal of Applied Behavior Analysis, 26,* 257–263. doi:10.1901/jaba.1993.26-257

Gresham, F. M., Gansle, K. A., Noell, G. H., Cohen, S., & Rosenblum, S. (1993). Treatment integrity of school-based behavioral intervention studies: 1980–1990. *School Psychology Review, 22,* 254–272.

Gresham, F. M., MacMillan, D. L., Beebe-Frankenberger, M. E., & Bocian, K. M. (2000). Treatment integrity in learning disabilities intervention research: Do we really know how treatments are implemented? *Learning Disabilities Research & Practice, 15,* 198–205. doi:10.1207/SLDRP1504_4

Heimberg, R. G., & Becker, R. E. (1984). Comparative outcome research. In M. Hersen, L. Michelson, & A. S. Bellack (Eds.), *Issues in psychotherapy research* (pp. 251–283). New York, NY: Plenum Press.

Hogue, A., Liddle, H. A., & Rowe, C. (1996). Treatment adherence process research in family therapy: A rationale and some practical guidelines. *Psychotherapy: Theory, Research, Practice, Training, 33,* 332–345. doi:10.1037/0033-3204.33.2.332

Hohmann, A. A., & Shear, M. K. (2002). Community-based intervention research: Coping with the "noise" of real life in study design. *American Journal of Psychiatry, 159,* 201–207. doi:10.1176/appi.ajp.159.2.201

Holcombe, A., Wolery, M., & Snyder, E. (1994). Effects of two levels of procedural fidelity with constant time delay on children's learning. *Journal of Behavioral Education, 4,* 49–73. doi:10.1007/BF01560509

Hoyt, W. T. (2000). Rater bias in psychological research: When is it a problem and what can we do about it? *Psychological Methods, 5,* 64–86. doi:10.1037/1082-989X.5.1.64

Jensen, P. S., Weersing, R., Hoagwood, K. E., & Goldman, E. (2005). What is the evidence for evidence-based treatments? A hard look at our soft underbelly. *Mental Health Services Research, 7,* 53–74. doi:10.1007/s11020-005-1965-3

Jones, H. A., Clarke, A. T., & Power, T. J. (2008). Expanding the concept of intervention integrity: A multidimensional model of participant engagement. *In Balance, 23*(1), 4–5.

Jones, K., Wickstrom, K., & Friman, P. (1997). The effects of observational feedback on treatment integrity in school-based behavioral consultation. *School Psychology Quarterly, 12,* 316–326. doi:10.1037/h0088965

Kazdin, A. E. (1977). Artifact, bias, and complexity of assessment: The ABCs of reliability. *Journal of Applied Behavior Analysis, 10,* 141–150. doi:10.1901/jaba.1977.10-141

Kazdin, A. E. (2003). *Research design in clinical psychology* (4th ed.). Boston, MA: Allyn & Bacon.

Kazdin, A. E. (2007). Mediators and mechanisms of change in psychotherapy research. *Annual Review of Clinical Psychology, 3,* 1–27. doi:10.1146/annurev.clinpsy.3.022806.091432

Kratochwill, T. R., Sheridan, S. M., Rotto, P. C., & Salmon, D. (1991). Preparation of school psychologists to serve as consultants for teachers of emotionally disturbed children. *School Psychology Review, 20,* 530–550.

Krumholz, H. M., Radford, M. J., Wang, Y., Chen, J., Heiat, A., & Marciniak, T. A. N. (1998). National use and effectiveness of beta-blockers for the treatment of elderly patients after acute myocardial infarction: National Cooperative Cardiovascular Project. *JAMA: Journal of the American Medical Association, 280,* 623–629. doi:10.1001/jama.280.7.623

Lauritsen, J. (1992, March 30). FDA documents show fraud in AZT trials. *New York Native*. Retrieved from http://www.virusmyth.com/aids/hiv/jlfraud.htm

Marcus, S. M., Gorman, J. M., Tu, X., Gibbons, R. D., Barlow, D. H., Woods, S. W., & Shear, M. K. (2006). Rater bias in a blinded randomized placebo-controlled psychiatry trial. *Statistics in Medicine, 25,* 2762–2770. doi:10.1002/sim.2405

Margison, F. R., McGrath, G., Barkham, M., Clark, J. M., Audit, K., & Connell, J. (2000). Measurement and psychotherapy: Evidence-based practice and practice-based evidence. *British Journal of Psychiatry, 177,* 123–130. doi:10.1192/bjp.177.2.123

Marks, S. E., & Tolsma, R. J. (1986). Empathy research: Some methodological considerations. *Psychotherapy: Theory, Research, Practice, Training, 23,* 4–20. doi:10.1037/h0085591

McGrew, J. H., Bond, G. R., Dietzen, L., & Salyers, M. (1994). Measuring the fidelity of implementation of a mental health program model. *Journal of Consulting and Clinical Psychology, 62,* 670–678. doi:10.1037/0022-006X.62.4.670

McLeod, B. D., & Islam, N. Y. (2011). Using treatment integrity methods to study the implementation process. *Clinical Psychology: Science and Practice, 18,* 36–40. doi:10.1111/j.1468-2850.2010.01232.x

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. doi:10.1080/15374411003691750

McMahon, P. M. (1987). Shifts in intervention procedures: A problem in evaluating human service interventions. *Social Work Research & Abstracts, 23,* 13–16. doi:10.1093/swra/23.4.13

Miller, S. J., & Binder, J. L. (2002). The effects of manual-based training on treatment fidelity and outcome: A review of the literature on adult individual psychotherapy. *Psychotherapy: Theory, Research, Practice, Training, 39,* 184–198. doi:10.1037/0033-3204.39.2.184

Moncher, F. J., & Prinz, R. J. (1991). Treatment fidelity in outcome studies. *Clinical Psychology Review, 11,* 247–266. doi:10.1016/0272-7358(91)90103-2

Moore, J. W., & Fisher, W. W. (2007). The effects of videotape modeling on staff acquisition of functional analysis methodology. *Journal of Applied Behavior Analysis, 40,* 197–202. doi:10.1901/jaba.2007.24-06

Moos, R. H., & Finney, J. W. (1983). The expanding scope of alcohol treatment evaluation. *American Psychologist, 38,* 1036–1044. doi:10.1037/0003-066X.38.10.1036

Morris, S. B., Turner, C. W., & Szykula, S. A. (1988). Psychotherapy outcome research: An application of new methods for evaluation research methodology. *Psychotherapy: Theory, Research, Practice, Training, 25,* 18–26. doi:10.1037/h0085319

Mowbray, C. T., Holter, M. C., Teague, G. B., & Bybee, D. (2003). Fidelity criteria: Development, measurement, and validation. *American Journal of Evaluation, 24,* 315–340. doi:10.1177/109821400302400303

Nelsen, J. C. (1985). Verifying the independent variable in single-subject research. *Social Work Research & Abstracts, 21,* 3–8. doi:10.1093/swra/21.2.3

Nezu, A. M., & Nezu, C. M. (2008). *Evidence-based outcome research: A practical guide to conducting randomized controlled trials for psychosocial interventions*. New York, NY: Oxford University Press.

Noell, G. H., Gresham, F. M., & Gansle, K. A. (2002). Does treatment integrity matter? A preliminary investigation of instructional implementation and mathematical performance. *Journal of Behavioral Education, 11,* 51–67. doi:10.1023/A:1014385321849

O'Leary, K. D., & Kent, R. N. (1973). Behavior modification for social action: Research tactics and problems. In L. A. Hamerlynck, P. O. Davidson, & L. E. Acker (Eds.), *Critical issues in research and practice* (pp. 69–96). Champaign, IL: Research Press.

Perepletchikova, F. (2009). *Treatment integrity in treatment outcome research: Adequacy of procedures, associated factors, implications for research and practice, guidelines and recommendations*. Saarbrüken, Germany: Lambert Academic Publishing.

Perepletchikova, F., Hilt, L. M., Chereji, E., & Kazdin, A. E. (2009). Barriers to implementing treatment integrity procedures: Survey of treatment outcome researchers. *Journal of Consulting and Clinical Psychology, 77,* 212–218. doi:10.1037/a0015232

Perepletchikova, F., & Kazdin, A. E. (2005). Treatment integrity and therapeutic change: Issues and research recommendations. *Clinical Psychology: Science and Practice, 12,* 365–383. doi:10.1093/clipsy.bpi045

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. doi:10.1037/0022-006X.75.6.829

Peterson, L., Homer, A. L., & Wonderlich, S. A. (1982). The integrity of independent variables in behavior analysis. *Journal of Applied Behavior Analysis, 15,* 477–492. doi:10.1901/jaba.1982.15-477

Plumb, J. C., & Vilardaga, R. (2010). Assessing treatment integrity in acceptance and commitment therapy: Strategies and suggestions. *International Journal of Behavioral Consultation and Therapy, 6,* 263–295.

Resnicow, K., Davis, M., Smith, M., Lazarus-Yaroch, A., Baranowski, T., Baranowski, J., . . . Wang, D. T. (1998). How best to measure implementation of school health curricula: A comparison of three measures. *Health Education Research, 13,* 239–250. doi:10.1093/her/13.2.239

Robbins, J. R., & Gutkin, T. B. (1994). Consultee and client remedial and preventive outcomes following consultation: Some mixed empirical results and directions for future researchers. *Journal of Educational & Psychological Consultation, 5,* 149–167. doi:10.1207/s1532768xjepc0502_5

Rounsaville, B. J., Chevron, E. S., & Weissman, M. M. (1984). Specification of technique in interpersonal therapy. In J. B. W. Williams & R. L. Spitzer (Eds.), *Psychotherapy research: Where are we and where should we go?* (pp. 160–171). New York, NY: Guilford Press.

Sanetti, L. M. H., & Kratochwill, T. R. (2009). Toward developing a science of treatment integrity: Introduction to the special series. *School Psychology Review, 38,* 445–459.

Schinke, S. P., Gilchrest, L. D., & Snow, W. H. (1985). Skills intervention to prevent cigarette smoking among adolescents. *American Journal of Public Health, 75,* 665–667. doi:10.2105/AJPH.75.6.665

Schlosser, R. W. (2002). On the importance of being earnest about treatment integrity. *Augmentative and Alternative Communication, 18,* 36–44.

Sechrest, L., White, S. O., & Brown, E. D. (Eds.). (1979). *The rehabilitation of criminal offenders: Problems and prospects.* Washington, DC: National Academy of Sciences.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference.* New York, NY: Houghton Mifflin.

Sharpless, B. A., & Barber, J. P. (2009). A conceptual and empirical review of the meaning, measurement, development, and teaching of intervention competence in clinical psychology. *Clinical Psychology Review, 29,* 47–56. doi:10.1016/j.cpr.2008.09.008

Shaw, B. F., Elkin, I., Yamaguchi, J., Olmsted, M., Vallis, T. M., Dobson, K. S., . . . Imber, S. D. (1999). Therapist competence ratings in relation to clinical outcome in cognitive therapy of depression. *Journal of Consulting and Clinical Psychology, 67,* 837–846. doi:10.1037/0022-006X.67.6.837

Sheridan, S. M., Swanger-Gagne, M., Welch, G. W., Kwon, K., & Garbacz, S. A. (2009). Fidelity measurement in consultation: Psychometric issues and preliminary examination. *School Psychology Review, 38,* 476–495.

Sholomskas, D. E., Syracuse-Siewert, G., Rounsaville, B. J., Ball, S. A., Nuro, K. F., & Carroll, K. M. (2005). We don't train in vain: A dissemination trial of three strategies of training clinicians in cognitive-behavioral therapy. *Journal of Consulting and Clinical Psychology, 73,* 106–115. doi:10.1037/0022-006X.73.1.106

Stein, K. F., Sargent, J. T., & Rafaels, N. (2007). Intervention research: Establishing fidelity of the independent variable in nursing clinical trials. *Nursing Research, 56,* 54–62. doi:10.1097/00006199-200701000-00007

Sterling-Turner, H. E., Watson, T. S., Wildmon, M., Watkins, C., & Little, E. (2001). Investigating the relationship between training type and treatment integrity. *School Psychology Quarterly, 16,* 56–67. doi:10.1521/scpq.16.1.56.19157

Tobin, D. L., Banker, J. D., Weisberg, L., & Bowers, W. (2007). I know what you did last summer (and it was not CBT): A factor analytic model of international psychotherapeutic practice in the eating disorders. *International Journal of Eating Disorders, 40,* 754–757. doi:10.1002/eat.20426

Vermilyea, B. B., Barlow, D. H., & O'Brien, G. T. (1984). The importance of assessing treatment integrity: An example in the anxiety disorders. *Journal of Behavioral Assessment, 6,* 1–11. doi:10.1007/BF01321456

Waltz, J., Addis, M. E., Koerner, K., & Jacobson, N. S. (1993). Testing the integrity of a psychotherapy protocol: Assessment of adherence and competence. *Journal of Consulting and Clinical Psychology, 61,* 620–630. doi:10.1037/0022-006X.61.4.620

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 & Adolescent Psychology, 31,* 168–180. doi:10.1207/S15374424JCCP3102_03

Weingardt, K. R., Cucciare, M. A., Bellotti, C., & Lai, W. P. (2009). A randomized trial comparing two models of web-based training in cognitive-behavioral therapy for substance abuse counselors. *Journal of Substance Abuse Treatment, 37,* 219–227. doi:10.1016/j.jsat.2009.01.002

Weissman, M. M., Rounsaville, B. J., & Chevron, E. (1982). Training psychotherapists to participate in psychotherapy outcome studies. *American Journal of Psychiatry, 139,* 1442–1446.

Weisz, J. R., Weiss, B., & Donenberg, G. R. (1992). The lab versus the clinic: Effects of child and adolescent psychotherapy. *American Psychologist, 47,* 1578–1585. doi:10.1037/0003-066X.47.12.1578

Wickstrom, K. F., Jones, K. M., LaFleur, L. H., & Witt, J. C. (1998). An analysis of treatment integrity in school-based behavioral consultation. *School Psychology Quarterly, 13,* 141–154. doi:10.1037/h0088978

Wiese, M. R. R. (1992). A critical review of parent training research. *Psychology in the Schools, 29,* 229–236. doi:10.1002/1520-6807(199207)29:3<229::AID-PITS2310290305>3.0.CO;2-Q

Wu, S. M., Whiteside, U., & Neighbors, C. (2007). Differences in inter-rater reliability and accuracy for a treatment adherence scale. *Cognitive Behaviour Therapy, 36,* 230–239. doi:10.1080/16506070701584367