

# Treatment Integrity and Therapeutic Change: Issues and Research Recommendations

Francheska Perepletchikova, Yale University  
Alan E. Kazdin, Yale University School of Medicine

**One of the most important aspects of treatment outcome research is establishing treatment integrity. Integrity of the treatment refers to the degree to which treatment is implemented as intended. Research examining the relationship between treatment integrity and therapeutic change has produced conflicting results. However, assessment, design strategies, and the possible confound of integrity with other variables may explain the inconsistency in findings. This paper elaborates the limitations of existing strategies for evaluating the relationship between treatment integrity and outcome. Recommendations for future research include controlling possible confounding variables, experimentally manipulating treatment integrity, and using novel assessment and evaluation strategies.**

**Key words:** treatment integrity, adherence, competence, therapeutic change, treatment outcome. [*Clin Psychol Sci Prac* 12: 365–383, 2005]

Treatment integrity refers to the extent to which the intervention was implemented as intended (Vermilyea, Barlow, & O'Brien, 1984; Yeaton & Sechrest, 1981). Interpretations of treatment effects or lack of treatment effects require some assurance that the treatment was carried out as it was designed (e.g., Marks & Tolsma, 1986; Morris, Turner, & Szykula, 1988; Quay, 1977).

---

Address correspondence to Alan E. Kazdin, Child Study Center, Yale University School of Medicine, 230 S. Frontage Road, New Haven, CT 06520–7900. E-mail alan.kazdin@yale.edu.

Treatment integrity has also been implicated as a key ingredient of intervention success (e.g., Gresham, 1989; Peterson, Homer, & Wonderlich, 1982). A high level of treatment integrity has been associated with increased probability of changes on treatment outcome measures. However, several studies have shown that level of treatment integrity is irrelevant for successful outcome (e.g., Bein et al., 2000; Burke, 1996; Patton, 1998; Toffalo, 2000). Inconsistency in the relationship between treatment integrity and therapeutic change may stem from inadequate methodology. The purposes of this paper are (a) to discuss characteristics of the treatment, therapist, and client that are associated with treatment integrity and treatment outcome; (b) to elaborate the need for empirical examination of the relationship between treatment integrity and treatment outcome; and (c) to provide recommendations for the changes that are needed in the measurement and evaluation of integrity.

## **TREATMENT INTEGRITY AND TREATMENT OUTCOME**

Treatment integrity includes three components: treatment adherence, therapist competence, and treatment differentiation (e.g., Margison et al., 2000; Waltz, Addis, Koerner, & Jacobson, 1993). Adherence refers to the degree of utilization of specified procedures by the therapist. Competence refers to the level of skill and judgment shown by the therapist in delivering the treatment. Differentiation refers to whether treatments under investigation differ from each other along critical dimensions. For example, in the evaluation of the effects of cognitive and interpersonal therapies on depression, adherence would mean following the manual provided

for each therapy by performing all prescribed tasks and activities; competence would refer to how sensitively strategies are applied to a particular problem, timing of the interventions, and ability to focus the session on topics appropriate for the specific therapy; and differentiation would mean implementing procedures prescribed by the manual for cognitive therapy and avoiding procedures prescribed for interpersonal therapy, 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 in fact different (Waltz et al., 1993). However, adherence cannot be substituted for or combined with competence because they are not highly related (Miller & Binder, 2002).

The breakdown in any of these aspects may compromise treatment integrity. For example, a therapist may be very skilled in delivering the treatment, but he or she may utilize techniques that are not prescribed by the manual. This may lead to a low adherence rating. Diffusion of treatment may also occur if the utilized techniques were unique or essential for one of the treatments and not to be used for another treatment condition. On the other hand, a therapist may faithfully follow the treatment protocol but may exhibit inflexibility in the use of techniques, have critical tendencies, and lack of respect for client (e.g., failure to listen, interruptions, disapproving or critical comments, low reinforcement schedule). Such negative indicators of competence may compromise treatment progress and result in an unfair test of the intervention.

Treatment integrity plays an essential role in establishing what the treatment is and in evaluating its effectiveness. Failure to ensure treatment integrity compromises the experimental validity of the study (Gresham, Donald, MacMillan, Beebe-Frankenberger, & Bocian, 2000; Kazdin, 2003; Moncher & Prinz, 1991). Without ensuring treatment integrity, inferences about the obtained results would be ambiguous.

#### **Key Findings**

Ensuring treatment integrity is not merely a methodological nicety or end in its own right. An underlying assumption is that if an effective treatment has been identified, it will be important to ensure that the treatment is carried out faithfully to achieve the desired

end. Stated more simply, treatment integrity ought to relate to therapeutic change. However, the literature disagrees as to whether treatment integrity is related to treatment outcome. Treatment integrity is studied in the context of adult psychotherapy, school intervention program, wraparound services for children, and prevention, all areas to which we refer.

Several studies indicate that decreased treatment integrity is associated with decreased likelihood of therapeutic change (Erhardt, Barnett, Lentz, Stollar, & Raifin, 1996; Frank, Kupfer, Wagner, McEachran, & Cornes, 1991; Greenwood, Terry, Arreaga-Mayer, & Finney, 1992; Gresham, Gansle, Noell, Cohen, & Rosenblum, 1993b; Henggeler, Melton, Brondino, Scherer, & Hanley, 1997; Huey, Henggeler, Brondino, & Pickrel, 2000). For example, in the examination of the efficacy of interpersonal psychotherapy (IPT) as maintenance treatment of recurrent depression, the levels of treatment integrity (defined in terms of specificity and purity of treatment) were related to the length of patients' survival time without new episodes of major depression (Frank et al., 1991). High ratings on the specificity and purity of IPT were associated with increased survival time, median ratings were associated with median survival time, and low ratings were associated with low survival time.

Some studies find only partial support for the association between treatment integrity and outcome, demonstrating relationship for some but not all conditions or participants (Jones, K., Wickstrom, & Friman, 1997; McEvoy, Shores, Wehby, Johnson, & Fox, 1990; Sterling-Turner, Watson, & Moore, 2002). For example, evaluation of the impact of social skills training on student performance in special education classes produced ambiguous results (McEvoy et al., 1990). The social skills training consisted of two elements: the environmental arrangement phase, which included provision of activities that promote social interactions (e.g., board games) and verbal encouragements for interactions with peers, and the direct instruction phase, which included specific teacher prompts for interaction and contingent teacher praise. Teachers who implemented social skills training with high integrity achieved superior outcomes across dependent measures as compared to teachers who implemented the intervention with low integrity. However, the significant group difference was obtained

only for the direct instruction phase. Groups did not differ for the environmental arrangement phase.

Studies that do not support the relationship between integrity and outcome suggest that although treatment resulted in a significant change on the dependent measures, the level of treatment integrity was irrelevant for the outcome (e.g., Bein et al., 2000; Burke, 1996; Noell, Witt, Gilbertson, Ranier, & Freeland, 1997; Ogrodniczuk, 1998; Patton, 1998; Toffalo, 2000; Weisman et al., 2002). For example, evaluation of the data from the National Institute of Mental Health Treatment of Depression Collaborative Research Program indicated that although patients showed significant reduction in depressive symptoms in both cognitive-behavioral and interpersonal treatment conditions, treatment adherence was not related to outcome of either intervention (see Elkin, 1999).

The relationship between integrity and outcome has been evaluated empirically by varying the number of components implemented between various levels of integrity (Gansle & McMahon, 1997). Three levels of component implementation of a self-monitoring treatment were compared on their impact on student's self-report of positive and negative classroom behaviors. The results were ambiguous as they found support for the relationship for some conditions but not others (only self-recording of positive target behaviors was affected by component integrity levels).

Several studies examined the impact of treatment integrity using single-subject experimental designs. For example, continuous delivery of the delayed prompt was compared to intermittent delivery, while other procedural components were held constant (Holcombe, Wolery, & Snyder, 1994). The results were equivocal, because for some participants the level of integrity did not affect achievement. Similarly, evaluation of the impact of varying levels of prompt implementation produced ambiguous results (Noell, Gresham, & Gansle, 2002). The outcome across some cases was similar regardless of condition.

#### **Interpretations of the Discrepant Findings**

The disparity in findings on the relationship between treatment integrity and treatment outcome can stem from several factors. Studies that rely on indirect measures of treatment integrity may have difficulty in

interpreting results because such data can overrepresent or underrepresent treatment integrity levels. For example, therapist self-report is one of the indirect methods of assessment that may inflate integrity rating due to a demand characteristic and a need for social approval. Therapists may wish to portray themselves as adhering to the manual more closely than they actually do.

The breakdown in treatment integrity can both reduce and enhance the effectiveness of the treatment (Gresham et al., 2000). Deviations can augment procedures, add more effective techniques, and alter the protocol to better suit the treated population. For example, in the study of the factors associated with therapeutic effectiveness, therapists who were successful with more disturbed patients with stress-response disorders deviated from the treatment protocol and modified the prescribed model to better address the specific difficulties of their clients (Jones, E. E., Cumming, & Horowitz, 1988). Thus, low treatment integrity levels may not necessarily attenuate treatment strength. Low integrity does not mean that the treatment is weak, just that it is different from that which was originally intended.

Studies that evaluate the effect of treatment integrity on treatment outcome primarily rely on posttreatment integrity data. These studies were mentioned previously (e.g., Frank et al., 1991; Huey et al., 2000; McEvoy et al., 1990; Toffalo, 2000). Evaluation at the end of the treatment can only provide preliminary evidence on the association because other possible influences are not controlled. Treatment integrity may be just a proxy variable for other influences that may impact and even obscure the relationship between integrity and outcome.

Empirical demonstrations of the relationship between integrity and outcome are surprisingly scarce and have produced equivocal result. Ambiguous results may stem from employing problematic strategies, such as varying number of treatment components between integrity levels, exemplified in the study by Gansle and McMahon (1997, see above). When several but not all treatment components are utilized for the low integrity condition, then treatment may have a lower impact on behavior. The integrity of treatment implementation may be irrelevant for the outcome if treatment itself is not very effective. Further, it would be difficult to establish the construct underlying the abridged treatment, because

it is usually difficult to isolate components that are primarily responsible for behavior change.

Results of the single-subject studies may have been inconclusive because the design methodology may be inappropriate for the test. Single-subject experimental designs usually rely on only a handful of participants. As integrity decreases, responses to the intervention may become less predictable and more variable (Greenwood, Terry, Arreaga-Mayer, & Finney 1992; Holcombe, Wolery, & Snyder, 1994). Conclusions about the relationship between integrity and outcome may be difficult to reach without tests of statistical significance, comparisons of various subgroups, and analyses of possible moderating variables. Group designs may be more useful when performance is expected to be highly variable because of the options they provide for data analyses.

Overall, research examining the relationship between treatment integrity and treatment outcome has produced conflicting results. The methods may have been inadequate for evaluating the association. The methods have lacked the necessary controls, employed problematic strategies, or included specific features that would weaken the test.

#### **ASSOCIATED VARIABLES**

Treatment integrity may explain treatment outcome effects. Yet, investigators rarely consider other variables associated with treatment integrity and outcome. Treatment integrity may just be a proxy variable for other influences that account for therapeutic change. Characteristics of the treatment, therapists, and clients may affect or be confounded with integrity in treatment outcome studies.

#### **Treatment Characteristics**

Several treatment characteristics may influence treatment integrity, including complexity of the treatment, required multiple resources and materials, number of treatment agents, time needed for treatment implementation, rate of behavioral change, and acceptability of treatment by therapists and clients (Gresham, 1989; Gresham et al., 2000; Noell, Gresham, & Gansle, 2002; Reimers, Wacker, & Koeppel, 1987). The complexity of the treatment refers to the number of treatment components and may be inversely related to the level

of treatment integrity (Yeaton & Sechrest, 1981). Treatments that are complex might be specifically at a risk for procedural degradation because of an increased difficulty in establishing and maintaining integrity. For example, in the evaluation of the direct instruction model for comprehensive educational intervention with the disadvantaged, the high number of treatment components was interfering with maintaining high integrity levels (Becker & Carnine, 1981).

Treatments that require multiple materials and resources are likely to be implemented with lower integrity. Multiple materials and special resources (e.g., expensive supplies, technical equipment) might not always be readily available, acceptable, or cost effective. Therefore, their implementation might vary within and across therapists and sessions, and as a function of time. The utilization of the resources may have adequate short-term integrity, but integrity may not be sustained in the long term due to difficulties in the maintenance of the procedures. For example, within one year of implementing video-disk curriculum designed to teach fractions to students, only about 30% of teachers in the special education classes were still using the program, despite its demonstrated effectiveness (Woodward, 1993).

Interventions that require more than one treatment agent per client may be less likely to be conducted with a high integrity level than treatments with one therapist. The higher the number of additional treatment agents, the higher the probability of the failure on the part of the treatment agents to follow the specified protocol. For example, poor integrity can result when treatment involves cooperation of parents, teachers, spouses, and relatives. It is not always possible or feasible to monitor treatment integrity when treatment is implemented by a third party, such as by relatives at home or teachers at the classroom, as is sometimes the case in treating children. When manipulation checks are not provided, it may only be tenuously assumed that the third party is implementing treatment with integrity (Mortenson & Witt, 1998). Interventions requiring multiple therapists are also usually more complex and time consuming.

Time required for delivering the intervention (e.g., limited number of sessions) is another factor that may be related to integrity. Lack of time is one of the most commonly specified reasons for failing to implement the treatment as planned (Happe, 1982). For example,

a client may require support, feedback, or clarifications of presented information and techniques that are surplus to the specified protocol. Such additional procedures may impede the therapist's ability to deliver all the tasks, instructions, and activities specified for the session. The more complex the treatment, the greater amount of time may be required for its implementation, and the more likely are the lapses in treatment integrity (Noell & Gresham, 1993).

Rate of behavior change may also influence treatment implementation. Treatments that result in early symptomatic improvements may be utilized with greater integrity than the slower-acting treatments. For example, in the study of supportive expressive dynamic therapy, higher therapists' adherence ratings were associated with rapid behavior change (Barber, Crits-Christoph, & Luborsky, 1996).

Treatment acceptability may play an important role in how faithfully a therapist will follow the intended procedures. Acceptability refers to a judgment of whether treatment procedures are effective, reasonable, fair, or appropriate for a given problem or client. Therapists who perceive treatment as acceptable may implement prescribed procedures with greater integrity than when treatment is perceived as unacceptable (Gresham, 1989). Perceived effectiveness of the treatment may influence treatment acceptability. Therapies that are perceived as more effective by treatment agents may be implemented with higher level of integrity than therapies that are viewed as less effective. Positive interventions (e.g., positive reinforcement, token economy) are rated more acceptable than negative interventions (e.g., time-out, response cost; e.g., Blampied & Kahan, 1992; Cross Calvert & Johnston, 1990; Kalfus & Burk, 1989; Kazdin, 1984; Miller & Kelley, 1992).

Acceptance of treatment by the client may also impact treatment integrity. Clients may be more amenable to a specific treatment because this particular approach is more congruent with their conceptualization of their own problem. For those clients who do not find treatment suitable, resistance may be increased, while involvement and compliance may be reduced (see Cross Calvert & Johnston, 1990). For example, in the NIMH Treatment of Depression Collaborative Research Program, clients who were more amenable

to treatment were more likely to stay in treatment and to develop a positive therapeutic relationship (Elkin et al., 1999). Clients' willingness to participate fully and cooperate may have an impact on the treatment process and, therefore, on the degree to which procedures are implemented as intended (Witt & Elliott, 1985). For example, a client may be less likely to follow a therapist's suggestions during session, implement learned techniques, or complete homework assigned by a therapist. Such resistance may hinder treatment progress and result in a need to revisit past material or provide additional sessions that are not included in the original treatment protocol. Clients' acceptance of the intervention can be significantly increased by enhancing their knowledge about the treatment, therapeutic goals, and potential side effects (e.g., Singh & Katz, 1985).

#### **Therapist Characteristics**

Treatment integrity can be affected by several therapist characteristics, including experience and motivation to work with a particular client (Gresham, 1989; Miller & Binder, 2002; Weissman, Rounsaville, & Chevron, 1982). Experience may negatively impact the adherence to protocol. Experience may have solidified therapists' working styles and, therefore, can hinder new learning. Further, therapists who treated a greater variety of clients may begin to include techniques from other treatments over time, which may result in treatment drift from the prescribed procedures (Clarke, 1995). Highly experienced therapists tend to integrate elements from different treatments and are more likely to deviate from rules and guidelines and use novel problem-solving techniques (see Margison et al., 2000).

The motivation of a therapist may also affect treatment integrity. Motivation refers to the extent to which a therapist is inclined to work with a particular client. When a client is difficult to manage (e.g., a disruptive child in the regular classroom) or has characteristics that result in negative outcome early in treatment (e.g., low-income patients), the treatment agent may be more eager to refer this client elsewhere (e.g., special education), rather than attempt to remediate (Lorion, 1974; Ysseldyke, Christenson, Pianta, & Algozzine, 1983). In such cases, interventions are less likely to be implemented with high integrity (Gresham, 1985).

### **Client Characteristics**

Several client characteristics are known to affect treatment integrity, including client difficulty, anger and hostility, problem severity, duration, and comorbidity (e.g., Detrich, 1999; Elliott, 1986; Foley, O'Malley, Rounsaville, Prusoff, & Weissman, 1987; Waltz et al., 1993). Therapist adherence to a prescribed technique may be higher for easy and uncomplicated patients, while more impaired and complicated cases are likely to be associated with poorer adherence. 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. Client difficulty has been defined in terms of client's resistance, defensiveness, anger, and hostility (Foley et al., 1987). Therapists may be more emotionally distant with difficult patients and may refrain from training techniques or engaging a client into a therapeutic process (e.g., Patterson & Chamberlain, 1994; Patterson & Forgatch, 1985). Thus, patient difficulty may negatively affect administration of the treatment. For example, supervisor's ratings of overall therapist skills and therapist self-report of performance decreased as client difficulty increased in the study of the interpersonal psychotherapy for depression (Foley et al., 1987).

An angry or hostile client can prevent a therapist from successfully implementing prescribed strategies, as the accuracy and consistency of treatment delivery may be impacted. Interactions with hostile clients may significantly differ from the overall pattern. The therapeutic relationship with a hostile client may be less reinforcing for the therapist and may require greater effort. Greater effort in the face of little success may discourage the faithful rendition of the plan. For example, in the study of psychotherapy process in the interpersonal psychotherapy of depression, lower supervisor ratings of therapist competence were related to higher patient hostility (Rounsaville et al., 1987). Further, specific procedures for addressing a client's negativity may not be outlined in the manual. Therapists' resilience and willingness to address and resolve issues of client anger or hostility may be variable (Teyber & McClure, 2000).

The severity and duration of the problem may negatively impact the delivery of the intervention as originally intended. A therapist may find strict compli-

ance with guidelines to be challenging because more work may be necessary with more disturbed clients. In such cases the therapist may incorporate additional techniques. Severe cases may require more direction and coaching. Comorbidity may be especially problematic for treatment integrity. To address comorbid problems, treatment agents may utilize additional interventions, not originally specified in the treatment plan. For example, the severity of antisocial behavior and the comorbidity of criminal and substance abuse problems were inversely related to therapist adherence in the study of multisystemic therapy in community-service settings (Schoenwald, Halliday-Boykins, & Henggeler, 2003).

### **Associated Variables and Treatment Outcome**

The previous discussion outlined several treatment, therapist, and client characteristics that may influence the integrity of treatment implementation. Several of these characteristics are also associated with treatment outcome. It is important to address this relationship in our discussion because of the possibility that treatment integrity is just a proxy variable for other influences. That is, the association between treatment integrity and treatment outcome may be mediated by other variables that are related to both constructs (predictors of treatment outcome will be addressed only briefly here because this complex topic is beyond the scope of this paper).

Treatment outcome may be influenced by the acceptability of the procedures by the therapist (Reimers et al., 1987). Treatments that are not acceptable to the treatment agent are less likely to be attempted and successfully implemented. Therapist's motivation for treatment is also known to affect the effectiveness of the intervention. Higher motivation is related to increased therapist efficacy, reduced client attrition, and increased behavior change (e.g., Thomas, 2002; Wild, Cunningham, & Hobdon, 1998). The training level of therapists is also associated with increased therapeutic efficacy (e.g., Beutler, 1997; Driscoll et al., 2003; Stein & Lambert, 1995).

Further, severity of psychological disturbance, duration of the disorder, and comorbidity are among the best predictors of treatment response (Brent et al., 1998; Hamilton & Dobson, 2002; Mynors-Wallis & Gath, 1997; Petry, Tennen, & Affleck, 2000; Scheibe & Albus,

1997). Cognitive deficits are associated with poorer treatment outcome (Whisman, Miller, Norman, & Keitner, 1995). A client's negativity and hostility are also related to decreased chances of treatment success (Rounsaville et al., 1987; Teyber & McClure, 2000). Further, treatment acceptability by a client has been linked to more favorable outcomes (e.g., Collins & Hyer, 1986; Crane, Criffin, & Hill, 1986; Hamilton & Dobson, 2002; Keijsers, Schaap, & Hoogduin, 2000).

#### **General Comments**

Treatment integrity may account for or contribute to therapeutic change. Indeed, the statement seems obvious in many ways. Quite surprisingly, there is very little evidence that provides a strong link between how well or faithfully treatment is carried out and the extent to which clients improve. There are studies correlating integrity with therapeutic change as well as studies showing no such connection. Consider only those studies that do show a relation. In these studies, showing a correlation between integrity and therapeutic change is a bit different from showing that integrity accounts for or is the ingredient responsible in whole or in part for therapeutic change.

Among the key problems, many other variables associated with treatment integrity may account for or affect the relationship between integrity and outcome. Characteristics of the therapy, therapist, and client are among the major contenders. We have discussed these variables in isolation for purposes of presentation. It is not difficult to conceive of these in more complex ways in which some of the variables previously discussed interact as confounding influences. Associated characteristics may function as moderating or mediating variables in the relationship, or may just be the confounding influences that may serve as alternative explanations of the obtained outcome.

#### **RECOMMENDATIONS**

Many recommendations address how to establish, assess, evaluate, and report integrity (e.g., Carroll & Nuro, 2002; Gresham, 1997; Gresham et al., 2000; Schlosser, 2002; Waltz et al., 1993). Our recommendations build on and extend these by focusing specifically on the evaluation of the relationship between treatment integrity and treatment outcome.

#### **Experimental Tests of Integrity**

There is no substitute for a randomized controlled trial in the evaluation of the relationship between treatment integrity and therapeutic change. In such a study integrity would be manipulated experimentally, and conditions (i.e., levels of integrity) would be assigned randomly to subjects and therapists to control for the influences that affect both integrity and treatment outcome. Random assignment would make implausible that treatment integrity is just a proxy variable for other influences, including characteristics of the treatment, participants, and therapists. Needless to say, such design would not be feasible in clinical research, that is, in a clinical setting where the priorities for patient care are essential. However, analogue studies or research in more controlled and contrived settings would focus on measurement development and validation and direct tests of critical hypotheses.

Much of psychotherapy research has been criticized because it is conducted in controlled settings. Some of that research, with college students and with graduate student therapists, would permit experimental manipulation of integrity and go a long way toward elaborating the consequences of lapses in integrity or lapses in one of the components (e.g., adherence). It would allow researchers to determine whether integrity lapses can be separate from confounds that normally occur in the context of clinical work and clinical trials. The translation from laboratory to clinic and back is one that ought to rely more on experimental tests that cannot be conducted in clinical settings.

*Treatment Integrity Levels.* Treatment integrity can be established by explicitly specifying a protocol of treatment implementation, providing careful training of therapists, and monitoring therapists' adherence to prescribed procedures and competence of intervention delivery. Treatment integrity can be manipulated by varying the extent to which these procedures are implemented between levels.<sup>1</sup> In order to increase the expected differences between groups, only two integrity levels may be compared on their influence on treatment outcome. The groups that are more extreme are more likely to show a stronger effect. Therefore, to maximize the dissimilarity of experimental conditions only two

levels of integrity may be compared, low and high, versus three levels, which would include the medium integrity condition.

A high level of integrity may include (a) a treatment manual that specifies treatment rationale and procedures, spells out statements to be made by the therapist, and outlines characteristics of treatment process, sequencing of techniques, and procedures for handling deviations; (b) careful training of treatment agents by direct methods and involving opportunities for practice such as role-playing, modeling, feedback, rehearsal, and periodic booster sessions; (c) explicit, continued supervision to enhance accuracy and consistency of treatment delivery; and d) therapists who are made explicitly aware that the purpose of observations or videotaping is to assess treatment integrity.

A low level of integrity may include (a) the same manual utilized for the high integrity condition, and therapists should be directed to follow the manual to ensure the uniformity in the quantity of treatment components between levels, and to decrease variability within and between treatment agents; (b) indirect training of therapists, including didactic instructions about the intervention but no opportunities for rehearsal of tasks and procedures; and (c) monitoring treatment delivery only for the purposes of integrity assessment and ensuring that levels are distinct but without direct supervision and feedback, but therapists should not be made aware of the purpose of observations or videotaping.

The reason for keeping constant the extent to which treatment is manualized between levels is to assure that the same treatment components are delivered in both integrity conditions. The same number of treatment components has to be used for each integrity condition. Abridged treatment may have lower impact on the therapeutic change, regardless of how faithfully components are implemented. Lower effectiveness of an abridged treatment may have contributed to the ambiguity of the findings in the evaluation of the relationship between integrity and outcome, where the number of implemented components varied between levels (Gansle & McMahon, 1997). Because it may not be known which treatment components are responsible for change, implementation of all treatment components for each integrity condition may be necessary to control

for treatment effectiveness. Further, it would be difficult to evaluate what is the treatment and why it is expected to produce an effect. There needs to be a strong argument that the same treatment was utilized for each level in order for the interpretations of the results to be unambiguous.

Now, we recognize that not all treatments or facets of treatment can be manualized. The task may be challenging, as it is not always possible or feasible to include all potential scenarios of intervention delivery or to consider various comorbid diagnoses, patient's difficulty, resistance, or hostility. However, the empirical investigation we are proposing may utilize treatments that are already manualized and can be clearly specified (e.g., behavioral interventions for phobias).

The initial demonstration of the effect may also yield some insights into the nature of the relationship. In the low-integrity condition, treatment acceptability may diminish over time for both therapists and clients. Low integrity may adversely influence the satisfaction with intervention delivery. However, when acceptability is held constant at pretreatment by increasing therapists' and clients' understanding of the intervention in both groups, diminished acceptability in the low-integrity condition over the course of treatment may serve as a moderating variable in the relationship between integrity and outcome. Testing for the moderating effect of treatment acceptability may supply important clues pertaining to the nature of the relationship of integrity and therapeutic change.

*Procedures for Establishing Integrity.* The above recommendations are based on the evidence that treatment integrity is influenced by the extent of protocol specification, training of therapists, and supervision of treatment delivery (e.g., Erhardt, Barnett, Lentz, Stollar, & Raifin, 1996; Moos & Finney, 1983; Schinke, Gilchrest, & Snow, 1985; Yeaton & Sechrest, 1981). In this section we will address each of these procedures in greater detail to elaborate important points pertaining to their implementation.

Treatment procedures can be detailed in a manual form. The purpose of a psychotherapy manual is to specify the treatment and strategies for its acceptable implementation. Manuals guide and standardize intervention delivery by discussing a theoretical basis of



the intervention, specifying rationales for adherence, spelling verbatim statements to be made by treatment agents, describing characteristics of the treatment process, sequencing the techniques, providing examples of treatment operations, and procedures for handling deviations (e.g., Dobson & Shaw, 1988; McMahon, 1987, Nelson, 1985). Manuals reduce the variability in treatment implementation (Drozd & Goldfried, 1996; Rounsaville, Chevron, & Weissman, 1984) and enhance treatment integrity (Erhardt et al., 1996; Schinke et al., 1985). However, manuals have been frequently accused of limiting therapist flexibility in addressing issues that are relevant to treatment but beyond the scope of the protocol (e.g., Weissman et al., 1982). Further, tailoring treatment to the individual needs of each client is thought to enhance treatment gains and maintenance (e.g., Garfield, 1998; Goldfried & Wolfe, 1998; Persons, 1991). However, recent evidence suggests that the degree of therapist flexibility and treatment tailoring do not predict favorable outcomes (Kendall & Chu, 2000). Further, evaluation of the data from the NIMH Treatment of Depression Collaborative Research Program indicated that therapists could adhere to the manual but still incorporate flexibility in therapeutic technique and style by adjusting information presentation, directional statements, questions and clarifications according to patient's individual needs (Connolly Gibbons et al., 2002).

Clear and unambiguous specification of the independent variable, however, does not insure that the manipulation will be implemented as planned. Even the most detailed, comprehensive, and user-friendly script will not be sufficient for treatment integrity without a careful training of therapists (Golberg, 1984). Well-trained therapists are less susceptible to deviation from specified treatment protocol (Kazdin, 2003; Yeaton & Sechrest, 1981). Training is also associated with an increase in therapist competence (Milner, Baker, Blackburn, James, & Reichelt, 1999). Thus, training of therapists influences treatment adherence and therapist competence, the two major aspects of treatment integrity.

Training procedures can be roughly divided into indirect and direct categories. Indirect training includes didactic instructions about the intervention and written materials describing rationale, script, tasks, and activities. Direct training provides opportunities for practice, such

as role-playing, modeling, feedback, rehearsal, and periodic booster sessions. Faithful rendition of the treatment is more likely with direct training procedures (e.g., Kratochwill, Elliott, & Busse, 1995; Kratochwill, Sheridan, Rotto, & Salmon, 1991; Sterling-Turner, Watson, & Moore, 2002).

Careful training of therapists is obviously important. However, training has to be supplemented with continued supervision to ensure accuracy and consistency. Monitoring of treatment delivery can help reduce therapeutic drift, which refers to the deviation from the protocol and gradual alteration of treatment plan (Kazdin, 2003; Moos & Finney, 1983). Adherence and competence can be facilitated by ongoing supervision, viewing therapy tapes, providing regular feedback, role playing on how to approach difficult situations, and having regular meetings with staff. These procedures enhance homogeneity across therapists and can adjust ongoing deviations from the prescribed treatment plan.

#### **Nonexperimental Tests of Treatment Integrity**

Most of the research conducted to date involved non-experimental evaluation of the effects of integrity on treatment outcome (e.g., correlational analyses of the posttreatment integrity data). As discussed above, such examinations may have been inconclusive because possible confounding variables were not controlled. However, for practical reasons (e.g., limited funding), incorporating treatment integrity evaluations into other outcome research may be more compelling than conducting separate controlled trials. The posthoc regression analyses of integrity data may indeed provide initial demonstrations of the relationship between integrity and outcome when possible confounding variables, such as treatment, client, and therapist characteristics, are ruled out.

In the nonexperimental investigation, treatments that are complex, require extensive time and materials, employ multiple treatment agents, are slow acting, and may be perceived as unacceptable, should be avoided. Because these variables may negatively affect treatment integrity, it may be difficult to achieve high levels of integrity. Effects of treatment integrity on treatment outcome should be examined with an established, evidence-based treatment in order to control for the effectiveness of the treatment. If the utilized

therapy is not established as effective, then levels of integrity may be irrelevant for treatment outcome. With ineffective therapy, even a high integrity level may result in treatment failure.

Employing fairly inexperienced therapists, such as graduate students, may help control for the therapist characteristics. Graduate students are more likely to have similar levels of experience, training, and supervision, and such uniformity may reduce between-therapist variability in treatment implementation. Training on new manuals is more likely to be successful with novice rather than seasoned therapists. Novice therapists are also more likely to find treatment acceptable than seasoned clinicians (Elliott, 1988, Reimers et al., 1987). Acceptability of treatment is related to greater integrity and should be assessed before the intervention. Assessment of treatment acceptability should continue throughout treatment to control for the possible changes in perception.

Therapists should be advised on treatment effectiveness, rationales, and procedures before initial assessment to increase their knowledge and understanding of the selected intervention. A therapist may be less inclined to adhere to prescribed procedures if he or she does not fully understand the treatment (see Reimers et al., 1987). Further, increasing a therapist's knowledge about treatment and its effectiveness, and enhancing therapist awareness of an array of other possible interventions, may increase ratings of treatment acceptability (Elliott, 1988, Reimers et al., 1987).

To control for client characteristics that are known to negatively affect treatment integrity and treatment outcome, clients should be carefully screened. The selected participants should exhibit low severity and shorter duration of the problem, perceive treatment as acceptable, and should not have comorbid disorders. Furthermore, clients' acceptability of the selected treatment should be assessed before treatment implementation to understand clients' attitudes and beliefs. Monitoring of acceptability should be continued during treatment to control for possible changes in perception. Enhancing clients' knowledge and understanding of treatment rationales and procedures may be attempted to increase pretreatment acceptability ratings.

Homogeneity of the sample characteristics, low levels of severity, duration and number of identified problems,

and restricted range in therapy and therapists' characteristics may threaten the generality of the findings. The results of the study may be questioned on their applicability to the more diverse clinical population and with more complex treatments. However, in the evaluation of the effects of treatment integrity on treatment outcome, especially early in a program of studies, internal validity has priority over external validity. The study is concerned with the demonstration of the relationship, where other possible influences are ruled out. In order to examine the generality of findings, one must first have unambiguous findings about the relationship. Once the relationship between treatment integrity and outcome is established, further investigations may include other variables (e.g., how different levels of integrity would affect treatment outcome with a more heterogeneous sample, wider range in therapist and client characteristics, and more complex treatments). Further, studies examining the generality of the findings may provide insights into the possible mechanisms in the relationship between integrity and outcome by evaluating which variables mediated or moderated the association.

#### **Assessment Strategies**

The accuracy of the treatment integrity measurement depends on the degree to which components are specified, the competence of raters, the level of methodological rigor, and the properties of the selected measures. Various components of the treatment can be specified in a form of a manual. Such specification would minimize the amount of inference required in coding and may simplify rater training.

Rater competence requires rigorous training on all of the major and minor treatment components, subtle aspects of the treatment and treatment manual. Raters that are themselves skilled in the treatment delivery, if available, may be especially suitable for integrity ratings. However, such raters may be directly involved in the project and have an obvious investment in the demonstration of integrity. When raters are affiliated with the project, their ratings may be biased. Sophistication of the integrity measure can also have a significant impact on the accuracy of rating, with less sophisticated measures, such as indirect assessment (e.g., therapist self-report), contributing to higher integrity rates

(Miller, S. J., & Binder, 2002; Robbins & Gutkin, 1994; Wickstrom, Jones, LaFleur, & Witt, 1998). Demand characteristics and a need for social approval can affect the accuracy of the self-reported adherence to treatment regimen. Therapists may exaggerate the degree of their adherence because they know that they are being assessed. Such awareness may result in endorsement of responses that corroborate experimenter's expectations and are more socially desirable. Direct assessment methods, such as observations or videotaping, may yield more accurate results because they may be less susceptible to bias and distortions in self-interest.

Assessment of treatment integrity may involve direct observations or videotaping. Because treatment integrity is an independent variable, multiple sessions, randomly selected from each phase of treatment, should be observed and analyzed. Accuracy of the representation of the obtained data depends upon the number and length of observations, and collection of data across therapists, situations, cases, and sessions (e.g., Moncher & Prinz, 1991; Peterson et al., 1982). However, observations can alter performance of the therapist and may result in higher adherence to specified procedures during the observed sessions (Jones, K., et al, 1997; McMahon, 1987). Performance may change when individuals are aware that they are being evaluated. Differential adherence may artificially inflate estimates of treatment integrity and compromise the accuracy of integrity ratings. When multiple sessions are observed or videotaped on a variable-time schedule, differential adherence may be less likely. A more uniform performance may be achieved when an individual knows that he will be observed but is unaware of when.

Indirect assessment strategies should also be used to allow closer examination of intervention delivery. Indirect assessment strategies include therapist self-reports, subjects' report of what has been done during the treatment session, and permanent products of treatment implementation (e.g., written homework assignments, data collection sheets). Although there is a low agreement between direct and indirect methods (Carroll, Nich, & Rounsaville, 1998; Gresham, 1997; Wickstrom et al., 1998), self-reports can supplement data from direct assessment. Indirect measures of integrity can be compared to observational data and can be used to clarify implementation issues (Bergan &

Kratochwill, 1990; Gresham, 1989). Indirect measures may also offer an immediate access to the integrity data. Self-reports can be obtained and evaluated right after session administration, while observational data from independent raters may take a substantial amount of time to obtain. Immediacy may be useful for the ongoing monitoring of performance and possible adjustments in intervention delivery. Performance feedback may increase integrity when low levels are detected during treatment sessions (Jones, K., et al., 1997). Further, requiring self-reports may cue therapists to implement treatments with integrity (Gresham, 1997). Since only a fraction of sessions is usually utilized for coding of observational data, therapist reports of session content can be employed for a more detailed assessment of treatment delivery.

Assessment of treatment integrity should encompass all three aspects involved in its specification: treatment adherence, therapist competence, and treatment differentiation. Adherence measures can include items that are also useful in addressing treatment distinctiveness and can be sufficient to measure both aspects. Adherence measures should include items pertaining to four types of therapist behaviors: (a) those unique and essential to the specific treatment (e.g., assigning homework in behavior therapy); (b) those essential but not unique to the treatment (e.g., setting treatment goals); (c) those compatible with the treatment, 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; Waltz et al., 1993).

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, rating 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).

Because characteristics of the treatments may differ in treatment components and requirements for competent implementation, integrity measures may be developed

specifically for each treatment. Such measures should correspond to the operational definition of the intervention or the devised manual. Constructing measures that focus on one therapy, rather than interventions from multiple therapies, may have an advantage of parsimony, which may significantly reduce the length of rater training and degree of rater burden, and may simplify formulation of rater instructions (McGlinchey & Dobson, 2003).

#### **Validation of the Treatment Integrity Measures**

Efforts to establish validity of the integrity measures are largely lacking in the literature. However, when investigators assume that the devised scales include the characteristics of interest without providing supporting evidence, credibility of the results is jeopardized. If a measure is not validated, it is not possible to determine whether it actually assesses integrity.

Recommendations for treatment integrity assessment have primarily focused on the conceptualization of integrity and its measures (e.g., McGlinchey & Dobson, 2003). Additional attention is needed to provide concrete strategies for integrity assessment. In this section we are detailing concrete ways for the validation of treatment integrity measures. Validation of the integrity measure represents a particular challenge because the measure should encompass two unrelated constructs: adherence to specified protocol and therapist competence. Low associations between adherence and competence indicate that they should not be combined into one variable (Miller & Binder, 2002).

*Validating Adherence Measures.* Construct validity of the adherence measure may be evaluated by examining the association of treatment characteristics with adherence levels. Construct validity refers to the relation of a measure to other measures or domains of functioning. Treatment that is complex and time consuming and that requires multiple materials, resources, and therapists may yield lower adherence ratings than the less complicated treatment. Two treatments that differ on the characteristics that negatively affect integrity may be compared on their effects on adherence levels. Construct validity of the adherence measure may be supported by demonstrating that the characteristics of

the interventions are differentially associated with levels of integrity.

Construct validity can also be examined by evaluating the effect of training type on integrity levels. Faithful rendition of treatment protocol is more likely with direct training of therapists that offers opportunities for practice and feedback (e.g., role playing, modeling, feedback, rehearsal, and periodic booster sessions) than with indirect therapist training (e.g., didactic instructions about the intervention and written materials describing rationale, script, tasks, and activities). Direct training may be expected to be associated with higher integrity levels than indirect training. Differential training effect may serve to support construct validity of the adherence measure.

Other types of validity, such as discriminant and concurrent validity, may also be used to support construct validity of a measure. Discriminant validity refers to the correlation between measures that are expected not to relate to each other. Discriminant validity is suggested if the measure shows little or no correlation with measures with which it is not expected to correlate. Discriminant validity of the adherence measure may be supported when the measure of adherence to treatment protocol is associated with the intervention for which the protocol was originally devised (e.g., cognitive-behavioral treatment) and is not associated with the intervention for which treatment 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 the therapy type in the expected direction may be indicative of its ability to discriminate between interventions. This approach has been used in the study of the psychometric properties of the Yale Adherence and Competence Scale (YACS; Carroll et al., 2000). Three treatment subscales of the YACS, namely Clinical Management (CM), Twelve Step Facilitation (TSF), and Cognitive-Behavioral Treatment (CBT), have been evaluated on their ability to discriminate between the three corresponding interventions. Ratings on items composing the treatment subscale were significantly higher for the given treatment than for the two comparison treatments. That is, CM items were significantly higher for CM treatment, TSF items were significantly higher for TSF treatment,

and CBT items were significantly higher for CBT treatment.

Concurrent validity refers to the association of a measure with performance on another measure at the same point in time. Associations between adherence ratings, client and therapist acceptability of treatment, and therapist motivation for treatment delivery may be used to support concurrent validity of the adherence measure. Because these variables may affect treatment integrity, differential integrity may be expected. That is, higher ratings of acceptability and motivation may predict higher levels of adherence, while lower rating of acceptability and motivation may predict lower adherence.

An example of a validation effort of a measure of therapist adherence included examination of the relationship between therapist's adherence and treatment outcome, and a link between supervisor adherence and therapist adherence in the study of Multisystemic Therapy (MST, Henggeler & Schoenwald, 1999; Henggeler, Schoenwald, Liao, Letourneau, & Edwards, 2002). Associations of high therapist adherence with significant changes on dependent measures (e.g., arrest and incarceration rates, association with deviant peers) have been used to empirically support the validity of the measures of therapist adherence. Although validation efforts are commendable, relying on the relationship between adherence rates and therapeutic outcome seems premature because literature does not agree on whether the relationship is indeed present. Even through clinical trials have supported the effectiveness of MST in treating juvenile offenders, the association between the integrity of MST and treatment outcome cannot be assumed because other variables may influence the relationship. On the other hand, the association between supervisor adherence to the MST supervisory protocol and therapist adherence to the MST principals seem appropriate for validation of the adherence measures. As already discussed, supervision and monitoring of treatment delivery can help reduce therapeutic drift and may facilitate adherence to the specified treatment protocol. Thus, concurrent validity of integrity measure may be supported via its association with supervisory adherence.

*Validating Competence Measures.* Supervision and ongoing monitoring may also enhance the competence of intervention delivery. Thus, the relationship between

the measures of the quality of the provided supervisions with the measure of therapist competence may serve to support concurrent validity of the competence measure. Concurrent validity of the competence measure can also be examined by evaluating the association of therapist competence with the measures of client characteristics. Therapist performance may vary as a function of client difficulty (e.g., Foley et al., 1987) and competence may be expected to be lower with clients who have higher severity and duration of the presented problems, comorbid diagnoses, anger and hostility, and low levels of involvement or cooperation. Overall, a separate evaluation of the validity of adherence and competence measures may be needed before examining the relationship between treatment integrity and treatment outcome. A pilot study may be conducted with more diverse characteristics of the therapies, therapists, and clinical population to allow for the examination of the validity of the integrity measures.

#### **Strategies for Evaluation and Report of Integrity**

Guidelines for acceptable integrity within each level have to be developed before treatment is implemented in order to facilitate interpretation of the results. Definition of what constitutes adequate integrity may vary, depending on the treatment, its complexity, and relative weight of each component. Components can be more or less critical to treatment success and the relative importance of each component must be considered when guidelines for evaluating integrity are developed (Gresham et al., 2000). For example, providing rationale for treatment may be less crucial than contingent delivery of positive reinforcement. Decision rules are arbitrary because a conventional criterion for the adequate level of integrity has not yet been established. High integrity level may be represented by 80–100% integrity, whereas low integrity condition may be represented by 50% integrity or less. These guidelines reflect the current state of literature on integrity in treatment outcome studies (Burke, 1996; Galloway & Sheridan, 1994; Gansle & McMahon, 1997; Gresham, Gansle, & Noell, 1993; Holcombe et al., 1994; Noell et al., 2002).

In order to facilitate the accurate evaluation of relationship between integrity and treatment outcome, treatment integrity data should be reported in terms of overall integrity, component integrity, and session

integrity (Gresham, 1997; Schlosser, 2002). *Overall integrity* reflects integrity of treatment components across sessions. Overall integrity addresses the degree to which all components were implemented competently and according to the manual throughout treatment. *Component integrity* refers to the integrity of implementing each treatment component across sessions. *Session integrity* refers to the integrity of all treatment components within one session. Although overall integrity may be high, a treatment may fail to produce a significant outcome because of poor component integrity and/or session integrity. 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 all treatment components were implemented across sessions, session integrity may be low. Failure to measure session integrity may hinder the evaluation of results, especially when the treatment failed to produce significant change on dependent measures while overall integrity was high. Monitoring within session integrity may supply important information on the degree of competency and consistency in administering each treatment component. Such evaluation allows fine-grained analysis of the data, necessary for the unambiguous interpretation of the results.

## CONCLUSIONS

Treatment integrity has important implications for drawing inferences about the impact of the intervention. Treatment integrity has been associated with treatment outcome. However, investigations of this relationship produced conflicting results and either relied on posttreatment data, utilized questionable strategies, or employed experimental designs problematic for this test. Further, variables that are known to be associated with treatment integrity and treatment outcome are usually overlooked in the literature examining this relationship. Several client, therapist and treatment characteristics have been implicated. In order to control for these influences, the effects of treatment integrity on therapeutic change has to be studied empirically.

An obstacle to research on treatment integrity and treatment outcome that is beyond the scope of this review is the need to understand how treatment works and the mechanisms of therapeutic change (Kazdin, in

press). Presumably integrity is critically important for those facets of treatment that are responsible for change, and less important or indeed not important at all for ancillary factors that do not influence outcome. For current therapies for children, adolescents, and adults we do not know what the mechanisms of action are. Consequently, the study of integrity does not address the critical question, "Integrity of what?"

Research on treatment integrity and treatment outcome can actually contribute to our understanding of the mechanisms of therapy. If carefully developed measures of integrity include subscales or facets of different components of treatment, some of these components more than others may relate to therapeutic change. Such research by itself would not show what caused the change or what the mechanisms are, but the findings could alert us to likely leads for direct evaluation.

Integrity of treatment and therapeutic change remain a critically important topic. The purpose of this paper was to convey current findings and the difficulty in interpreting them. We provided several recommendations to shed light on the relationship between how treatment is carried out and improvements in clients. There are broad implications for work in this area including extending treatments to clinical practice and providing experiences to mental health professionals during training.

## NOTE

1. For the purposes of presentation we are discussing levels of integrity as high versus low. There might well be a continuum of treatment integrity levels that is more suitable depending on the treatment. The key concept here is an empirical manipulation of integrity rather than a commitment to a continuous or ordinal conceptualization of treatment integrity.

## ACKNOWLEDGMENT

This work was supported in part by grants from the Leon Lowenstein Foundation, the William T. Grant Foundation (98-1872-98), and the National Institute of Mental Health (MH59029). We are very grateful to Teresa Treat and Jerome J. Singer for their help on previous versions of this paper.

## REFERENCES

Barber, J. P., Crits-Christoph, P., & Luborsky, L. (1996). Effects of therapist adherence and competence on patient

- outcome in brief dynamic therapy. *Journal of Consulting and Clinical Psychology*, 64, 619–622.
- Becker, W. C., & Carnine, D. W. (1981). Direct instruction: A behavioral therapy model for comprehensive educational intervention with disadvantaged. In S. W. Bijou & R. Ruiz (Eds.), *Behavior modification: Contribution to education* (pp. 145–210). Hillsdale, NJ: Erlbaum.
- Bein, E., Anderson, T., Strupp, H. H., Henry, W. P., Schacht, T. E., Binder, J. L., et al. (2000). The effects of training in time-limited dynamic psychotherapy: Changes in therapeutic outcome. *Psychotherapy Research*, 10, 199–132.
- Bergan, J., & Kratochwill, T. (1990). *Behavioral consultation and therapy*. New York: Plenum Press.
- Beutler, L. E. (1997). The psychotherapist as a neglected variable in psychotherapy: An illustration by reference to the role of therapist experience and training. *Clinical Psychology: Science and Practice*, 4, 44–52.
- Blampied, N. M., & Kahan, E. (1992). Acceptability of alternative punishments: A community survey. *Behavior Modification*, 16, 400–413.
- Brent, D. A., Kolko, D. J., Birmaher, B., Baugher, M., Bridge, J., Roth, C., et al. (1998). Predictors of treatment efficacy in a clinical trial of three psychosocial treatments for adolescent depression. *Journal of the American Academy of Child and Adolescent Psychiatry*, 37, 906–914.
- Burke, R. V. (1996). Effects of treatment integrity and risk for children abuse on parent training outcomes. *Dissertation Abstracts International*. Vol. 56(7-A), 2884, US: University Microfilms International.
- 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.
- Carroll, K. M., Nich, C., Sifry, R. L., Nuro, K. F., Frankforter, T. L., Ball, S. A., et al. (2000). A general system for evaluating therapist adherence and competence in psychotherapy research in the addictions. *Drug and Alcohol Dependence*, 57, 225–238.
- Carroll, K. M., & Nuro, K. F. (2002). One size cannot fit all: A stage model for psychotherapy manual development. *Clinical Psychology: Science and Practice*, 9, 396–406.
- Clarke, G. N. (1995). Improving the transition from basic efficacy research to effectiveness studies: Methodological issues and procedures. *Journal of Consulting and Clinical Psychology*, 63, 718–725.
- Collins, J., & Hyer, L. (1986). Treatment expectancy among psychiatric inpatients. *Journal of Clinical Psychology*, 42, 562–596.
- Connolly Gibbons, M. B., Crits-Christoph, P., Levinson, J., Gladis, M., Siqueland, L., Barber, J. P., et al. (2002). Therapist interventions in the interpersonal and cognitive therapy sessions of the Treatment of Depression Collaborative Research Program. *American Journal of Psychotherapy*, 56, 3–26.
- Crane, D. R., Griffin, W., & Hill, R. D. (1986). Influence of therapist skills on client perceptions of marriage and family therapy outcome: Implications for supervision. *Journal of Marital and Family Therapy*, 12, 91–96.
- Cross Calvert, S., & Johnston, C. (1990). Acceptability of treatment for child behavior problems: Issues and implications for future research. *Journal of Clinical Child Psychology*, 19, 61–74.
- Detrich, R. (1999). Increasing treatment fidelity by matching interventions to contextual variables within the educational setting. *School Psychology Review*, 28, 608–620.
- 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.
- Driscoll, K. A., Cukrowicz, K. C., Reitzel, L. R., Hernandez, A., Petty, S. C., & Joiner, T. E. (2003). The effect of trainee experience in psychotherapy on client treatment outcome. *Behavior Therapy*, 34, 165–177.
- Droz, J. F., & Goldfried, M. R. (1996). A critical evaluation on the state-of-the-art in psychotherapy outcome research. *Psychotherapy*, 33, 171–180.
- Elkin, I. (1999). A major dilemma in psychotherapy outcome research: Disentangling therapists from therapies. *Clinical Psychology: Science and Practice*, 6, 10–32.
- Elkin, I., Yamaguchi, J. L., Arnkoff, D. B., Glass, C. R., Sotsky, S. M., & Krupnick, J. L. (1999). “Patient-treatment fit” and early engagement in therapy. *Psychotherapy Research*, 9, 437–451.
- Elliott, S. N. (1986). Children’s ratings of the acceptability of classroom interventions for misbehavior: Findings and methodological considerations. *Journal of School Psychology*, 24, 23–35.
- Elliott, S. N. (1988). Acceptability of behavioral treatments: Review of variables that influence treatment selection. *Professional Psychology: Research and Practice*, 19, 68–80.
- Erhardt, K. E., Barnett, D. W., Lentz, F. E., Stollar, S. A., & Raifin, L. H. (1996). Innovative methodology in ecological consultation: Use of scripts to promote treatment acceptability and integrity. *School Psychology Quarterly*, 11, 149–168.
- Foley, S. H., O’Malley, S, Rounsaville, B. J., Prusoff, B. A., & Weissman, M. M. (1987). The relationship between patient difficulty and therapist performance in interpersonal

- psychotherapy of depression. *Journal of Affective Disorders*, 12, 207–217.
- Frank, E., Kupfer, D. J., Wagner, E. F., McEachran, A. B., & Cornes, C. (1991). Efficacy of interpersonal psychotherapy as a maintenance treatment of recurrent depression. *Archives of General Psychiatry*, 48, 1053–1059.
- Galloway, J., & Sheridan, S. M. (1994). Implementing scientific practices through case studies: Examples using home-school interventions and consultation. *Journal of School Psychology*, 32, 385–413.
- 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.
- Garfield, S. L. (1998). Some comments on the empirically supported treatments. *Journal of Consulting and Clinical Psychology*, 66, 121–125.
- Golberg, S. C. (1984). “Nuts and bolts”: The technology of psychotherapy. In J. B. W. Williams & R. L. Spitzer (Eds.), *Psychotherapy research: Where are we and where should we go?* (pp. 190–192). New York: Guilford Press.
- Goldfried, M. R., & Wolfe, B. E. (1998). Towards a more clinically valid approach to therapy research. *Journal of Consulting and Clinical Psychology*, 66, 143–150.
- Greenwood, C. R., Terry, B., Arreaga-Mayer, C., & Finney, R. (1992). The classwide peer tutoring program: Implementation factors moderating students’ achievement. *Journal of Applied Behavior Analysis*, 25, 101–116.
- Gresham, F. M. (1985). Behavior disorder assessment: Conceptual, definitional, and practical considerations. *School Psychology Review*, 14, 495–509.
- Gresham, F. M. (1989). Assessment or 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., Donald, L., 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 and Practice*, 15, 198–205.
- 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.
- Gresham, F. M., Gansle, K. A., Noell, G. H., Cohen, S., & Rosenblum, S. (1993). Treatment integrity in school-based intervention studies: 1980–1990. *School Psychology Review*, 22, 254–272.
- Hamilton, K. E., & Dobson, K. S. (2002). Cognitive therapy for depression: Pretreatment patient predictors of outcome. *Clinical Psychology Review*, 22, 875–893.
- Happe, D. (1982). Behavioral intervention: It doesn’t do any good in your briefcase. In J. Grimes (Ed.), *Psychological approaches to problems of children and adolescents* (pp. 15–41). Des Moines, IA: Iowa Department of Public Instruction.
- Henggeler, S. W., Melton, G. B., Brondino, M. J., Scherer, D. G., & Hanley, J. H. (1997). *Journal of Consulting and Clinical Psychology*, 65, 821–833.
- Henggeler, S. W., & Schoenwald, S. K. (1999). The role of quality assurance in achieving outcomes in MST programs. *Journal of Juvenile Justice and Detention Services*, 14, 1–17.
- Henggeler, S. W., Schoenwald, S. K., Liao, J. G., Letourneau, E. J., & Edwards, D. L. (2002). Transporting efficacious treatments to field settings: The link between supervisory practices and therapist fidelity in MST programs. *Journal of Clinical Child Psychology*, 31, 155–167.
- 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.
- Huey, S. J., Henggeler, S. W., Brondino, M. J., & Pickrel, S. G. (2000). Mechanisms of change in multisystemic therapy: Reducing delinquent behavior through therapist adherence and improved family and peer functioning. *Journal of Consulting and Clinical Psychology*, 68, 451–467.
- Jones, E. E., Cumming, J. D., & Horowitz, M. (1988). Another look at the nonspecific hypothesis of therapeutic effectiveness. *Journal of Consulting and Clinical Psychology*, 56, 48–55.
- 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.
- Kalfus, G. R., & Burk, B. K. (1989). The influence of case history and treatment mediation on ratings of acceptability of child treatment. *Child and Family Behavior Therapy*, 11, 45–55.
- Kazdin, A. E. (1984). Acceptability of aversive procedures and mediation as treatment alternatives for deviant child behavior. *Journal of Abnormal Child Psychology*, 12, 289–302.
- Kazdin, A. E. (2003). *Research design in clinical psychology* (4th ed.). Boston, MA: Allyn & Bacon.
- Kazdin, A. E. (in press). Mechanisms of change in psychotherapy: Advances, breakthroughs, and cutting edge research (do not exist). In R. R. Bootzin (Ed.), *Festschrift in honor of Lee Sechrest*. Washington, DC: American Psychological Association.



- Keijsers, G. P. J., Schaap, C. P. D., & Hoogduin, C. A. L. (2000). The impact of interpersonal patient and therapist behavior on outcome in cognitive-behavioral therapy: A review of empirical studies. *Behavior Modification, 24*, 264–297.
- Kendall, P. C., & Chu, B. C. (2000). Retrospective self-reports of therapist flexibility in a manual-based treatment for youths with anxiety disorders. *Journal of Clinical Child Psychology, 29*, 209–220.
- Kratochwill, T. R., Elliott, S. N., & Busse, R. T. (1995). Behavioral consultation: A five-year evaluation of consultant and client outcomes. *School Psychology Quarterly, 10*, 87–117.
- 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, 4*, 530–550.
- Lorion, R. P. (1974). Patient and therapist variables in the treatment of low-income patients. *Psychological Bulletin, 81*, 344–354.
- Margison, F. R., Barkham, M., Evans, C., McGrath, G., Clark, J. M., Audit, K., et al. (2000). Measurement and psychotherapy: Evidence-based practice and practice-based evidence. *The British Journal of Psychiatry, 177*, 123–130.
- Marks, S. E., & Tolsma, R. J. (1986). Empathy research: Some methodological considerations. *Psychotherapy, 23*, 4–20.
- McEvoy, M. A., Shores, R. E., Wehby, J. H., Johnson, S. M., & Fox, J. J. (1990). Special education teachers' implementation of procedures to promote social interaction among children in integrated settings. *Education and Training in Mental Retardation, 25*, 267–276.
- McGlinchey, J. B., & Dobson, K. S. (2003). Treatment integrity concerns in cognitive therapy for depression. *Journal of Cognitive Psychotherapy: An International Quarterly, 17*, 299–318.
- McMahon, P. M. (1987). Shifts in intervention procedures: A problem in evaluating human service interventions. *Social Work Research and Abstracts, 23*, 13–16.
- Miller, D. L., & Kelley, M. L. (1992). Treatment acceptability: The effects of parent gender, marital adjustment, and child behavior. *Child and Family Behavior Therapy, 14*, 11–23.
- 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, and Training, 39*, 184–198.
- Milner, D. L., Baker, C., Blackburn, I. M., James, I., & Reichelt, K. (1999). Effectiveness of cognitive therapy training. *Journal of Behavior Therapy and Experimental Psychiatry, 30*, 81–92.
- Moncher, F. J., & Prinz, R. J. (1991). Treatment fidelity in outcome studies. *Clinical Psychology Review, 11*, 247–266.
- Moos, R. H., & Finney, J. W. (1983). The expanding scope of alcohol treatment evaluation. *American Psychologist, 38*, 1036–1044.
- Morris, S. B., Turner, C. W., & Szykula, S. A. (1988). Psychotherapy outcome research: An application of new methods for evaluation research methodology. *Psychotherapy, 25*, 18–26.
- Mortenson, B., & Witt, J. (1998). The use of weekly performance feedback to increase teacher implementation of a prereferral academic intervention. *School Psychology Review, 27*, 613–627.
- Mynors-Wallis, L., & Gath, D. (1997). Predictors of treatment outcome for major depression in primary care. *Psychological Medicine, 27*, 731–736.
- Nelson, J. C. (1985). Verifying the independent variable in single-subject research. *Social Work Research and Abstracts, 21*, 3–8.
- Noell, G. H., & Gresham, F. M. (1993). Functional outcome analysis: Do the benefits of consultation and prereferral intervention justify the costs? *School Psychology Quarterly, 27*, 200–226.
- 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.
- Noell, G. H., Witt, J. C., Gilbertson, D. N., Ranier, D. D., & Freeland, J. T. (1997). Increasing teacher intervention implementation in general education settings through consultation and performance feedback. *School Psychology Quarterly, 12*, 77–88.
- Ogrodniczuk, J. S. (1998). Therapist adherence to treatment manuals and its relation to the therapeutic alliance and therapy outcome: Scale development and validation. *Dissertation Abstracts International*. Vol. 58(10-B), 5652, US: University Microfilms International.
- Patterson, G. R., & Chamberlain, P. (1994). A functional analysis of resistance during parent training therapy. *Clinical Psychology: Science and Practice, 1*, 53–70.
- Patterson, G. R., & Forgatch, M. S. (1985). Therapist behavior as a determinant for client noncompliance: A paradox for the behavior modifier. *Journal of Consulting and Clinical Psychology, 53*, 846–851.
- Patton, J. R. (1998). A study in reliability and validation: What is the nature of the unique and joint contributions of therapist adherence and therapeutic alliance to treatment outcome? *Dissertation Abstracts International*. Vol. 58(9-B), 5136, US: University Microfilms International.

- Persons, J. B. (1991). Psychotherapy outcome studies do not accurately represent current models of psychotherapy: A proposed remedy. *American Psychologist, 46*, 99–106.
- Peterson, L., Homer, A., & Wonderlich, S. (1982). The integrity of independent variables in behavior analysis. *Journal of Applied Behavior Analysis, 15*, 477–492.
- Petry, N. M., Tennen, H., & Affleck, G. (2000). Stalking the elusive client variable in psychotherapy research. In C. R. Snyder & R. E. Ingram (Eds.), *Handbook of psychological change: Psychotherapy processes and practices for 21<sup>st</sup> century* (pp. 88–108). Hoboken, NJ: John Wiley & Sons.
- Quay, H. C. (1977). The three faces of evaluation. *Criminal Justice and Behavior, 4*, 341–354.
- Reimers, T. M., Wacker, D. P., & Koepl, G. (1987). Acceptability of behavioral interventions: A review of the literature. *School Psychology Review, 16*, 212–227.
- 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 and Psychological Consultation, 5*, 149–167.
- Rounsaville, B. J., Chevron, E. S., Prusoff, B. A., Elkin, I., Imber, S., Sotsky, S., et al. (1987). The relationship between specific and general dimensions of the psychotherapy process in interpersonal psychotherapy for depression. *Journal of Consulting and Clinical Psychology, 55*, 379–384.
- 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: Guilford Press.
- Scheibe, G., & Albus, M. (1997). Predictors and outcome in panic disorder: A 2-year prospective follow-up study. *Psychopathology, 30*, 177–184.
- Schinke, S. P., Gilchrist, L. D., & Snow, W. H. (1985). Skills intervention to prevent cigarette smoking among adolescents. *American Journal of Public Health, 75*, 665–667.
- Schlosser, R. W. (2002). On the importance of being earnest about treatment integrity. *Augmentative and Alternative Communication, 18*, 36–44.
- Schoenwald, S. K., Halliday-Boykins, C. A., & Henggeler, S. W. (2003). Client-level predictors of adherence to MST in community service settings. *Family Process, 42*, 345–359.
- Singh, N. N., & Katz, R. C. (1985). On the modification of acceptability ratings for alternative child treatments. *Behavior Modification, 9*, 375–386.
- Sterling-Turner, H. E., Watson, T. S., & Moore, J. W. (2002). The effects of direct training and treatment integrity on treatment outcomes in school consultation. *School Psychology Quarterly, 17*, 47–77.
- Stein, D. M., & Lambert, M. J. (1995). Graduate training in psychotherapy: Are therapy outcomes enhanced? *Journal of Consulting and Clinical Psychology, 63*, 182–196.
- Teyber, E., & McClure, F. (2000). Therapist variables. In C. R. Snyder & R. E. Ingram (Eds.), *Handbook of psychological change: Psychotherapy processes and practices for 21<sup>st</sup> century* (pp. 62–87). Hoboken, NJ: John Wiley & Sons.
- Thomas, V. (2002). Conducting research with community agencies: Meeting recruitment and collaboration challenges. *Journal of Marital and Family Therapy, 28*, 9–14.
- Toffalo, D. A. D. (2000). In investigation of treatment integrity and outcomes in wraparound services. *Journal of Child and Family Studies, 9*, 351–361.
- 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.
- 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.
- Weisman, A., Tompson, M. C., Okazaki, S., Gregory, J., Goldstein, M. J., Rea, M., et al. (2002). Clinician's fidelity to a manual-based family treatment as a predictor of the one-year course of bipolar disorder. *Family Processes, 41*, 123–131.
- Weissman, M. M., Rounsaville, B. J., & Chevron, E. (1982). Training psychotherapists to participate in psychotherapy outcome studies. *American Journal of Psychiatry, 139*, 1442–1446.
- Wickstrom, K., Jones, K., LaFleur, L., & Witt, J. (1998). An analysis of treatment integrity in school-based behavioral consultation. *School Psychology Quarterly, 13*, 141–154.
- Wild, T. C., Cunningham, J., & Hobdon, K. (1998). When do people believe that alcohol treatment is effective? The importance of perceived client and therapist motivation. *Psychology of Addictive Behaviors, 12*, 93–100.
- Whisman, M. A., Miller, I. W., Norman, W. H., & Keitner, G. I. (1995). Hopelessness depression in depressed inpatients: Symptomatology, patient characteristics, and outcome. *Cognitive Therapy and Research, 19*, 377–398.
- Witt, J. C., & Elliott, S. N. (1985). Acceptability of classroom intervention strategies. In T. R. Kratochwill (Ed.), *Advances in school psychology* (Vol. 4, pp. 251–288). Mahwah, NJ: Erlbaum.

- Woodward, J. (1993). The technology of technology-based instructions: Comments on the research, development, and dissemination approach to innovation. *Education and Treatment of Children, 16*, 345–360.
- Yeaton, W., & Sechrest, L. (1981). Critical dimensions in the choice and maintenance of successful treatments: Strength, integrity, and effectiveness. *Journal of Consulting and Clinical Psychology, 49*, 156–167.
- Ysseldyke, J. E., Christenson, S., Pianta, B., & Algozzine, B. (1983). An analysis of teachers' reasons and desired outcome for students referred for psychoeducational assessment. *Journal of Psychoeducational Assessment, 1*, 73–83.

Received April 24, 2004; revised July 21, 2004; accepted November 23, 2004.